Institute of Exercise Training and Sport Informatics
German Sport University Cologne

From Atomic Events to Sequences: Analyzing Soccer Data on Various Time Scales

Doctoral thesis accepted for the degree
Dr. rer. nat.

by
Henrik Bernhard Rolf Biermann
from
Witten

Cologne, 2025

First reviewer: Prof. Dr. Daniel Memmert,
Institute of Exercise Training and Sport Informatics,
German Sport University Cologne, Germany

Second reviewer: Prof. Dr. Ricardo da Silva Torres,
Professor in Data Science and Artificial Intelligence,
Wageningen University & Research, Netherlands

Chair of the doctoral committee: Prof. Dr. Mario Thevis,
Institute of Biochemistry,
German Sport University Cologne, Germany

Thesis defended on: June 11th 2025

Affidavits following §7 section 2 No. 9 of the doctoral regulations from the German Sport University Cologne, July 9th 2024:

Hereby I declare:
The work presented in this thesis is the original work of the author except where acknowledged in the text. This material has not been submitted either in whole or in part for a degree at this or any other institution. Those parts or single sentences, which have been taken verbatim from other sources, are identified as citations.
I further declare that I complied with the actual "guidelines of qualified scientific work" of the German Sport University Cologne.

H. B

06.12.2024, Henrik Biermann

Acknowledgements

I would like to express my gratitude to my supervisor, Daniel, for the trust and freedom he has granted me.

Thank you to everyone who has supported me along the way.
It was a truly special opportunity to pursue my PhD at the Sport University Cologne.

Dedicated to Prof. Dr. Lothar Georg Gerritzen. 25.08.1941 - 13.3.2024.

Zusammenfassung

In den letzten Jahren hat die angewandte Wissenschaft der Sportanalytik die Strategien in US-amerikanischen Sportarten wie Baseball, Basketball und American Football maßgeblich beeinflusst. Im Fußball hingegen sind trotz Fortschritten in Technologien und Datenerhebung derartige Veränderungen bisher kaum sichtbar.

Die vergleichsweise geringere Adaption datengetriebener Analysen im Fußball lässt sich auf verschiedene Faktoren wie die torarme Natur des Spiels, die taktische Komplexität und die Fluidität durch wenig strukturellen Komponenten zurückführen. Um die Lücke zu schließen, zielt die vorliegende Dissertation darauf ab, die grundlegenden Voraussetzungen für ein tieferes Verständnis des Fußballs zu verbessern. Insbesondere wird der Mangel an verbindenden Ansätzen zwischen domänenspezifischen Begriffen und Konzepten sowie deren Überführung in konkrete, regelbasierte Definitionen adressiert.

In der Sportwissenschaft existieren verschiedene Taxonomien und Annotationsschemata für Angriffs- und Verteidigungsprozesse im Fußball. Allerdings ermöglicht die oft eher narrative als deskriptive Natur dieser Schemata keine granulare Annotation, wie in der Informatik benötigt. Darüber hinaus beinhalten die Schemata mitunter semantisch komplexe, subjektive Beschreibungen, sind nicht ausreichend validiert oder beinhalten keine klaren Definitionen, die regelbasierte Entscheidungen ermöglichen. Dadurch sind sie weder leicht durch Menschen nachvollziehbar noch durch automatische Modelle replizierbar.

Die vorliegende Dissertation adressiert diese Lücke auf verschiedenen Zeitskalen. Bezüglich eines kurzen Zeithorizonts wurde eine einheitliche, hierarchische Taxonomie für die Annotation atomarer Ereignisse vorgeschlagen, welche durch eine Annotationsstudie mit Experten validiert wurde. Außerdem wurden die Ereignisannotationen durch einen Ansatz der zeitreihenbasierten Ereigniserkennung verfeinert, um die atomaren Ereignisse mit den zugrunde liegenden Positionsdaten zu synchronisieren. Bezüglich längerer Zeithorizonte wurde, unter Beibehaltung der Hierarchie zwischen den Zeitskalen, ein hierarchisches Annotationsschema für Sequenzannotationen zum Konzept der taktischen Periodisierung vorgeschlagen und ebenfalls in einer Annotationsstudie mit Experten validiert. Schließlich wurde eine grundlegende Struktur zur sequenzspezifischen Analyse im Fußball vorgestellt und auf das Konzept von Konterangriffen angewandt.

Ein zentrales Ergebnis der Annotationsexperimente war die Wichtigkeit einer hierarchischen Struktur in Annotationsschemata, welche die semantische Komplexität effektiv abbildet und es ermöglicht, unterschiedliche Annotationen durch Experten in gemeinsame übergeordnete Klassen zu integrieren.

Ein wichtiger Befund beim Training automatischer Modelle war, dass die Nachbildung manueller Sequenzannotationen zwar gelang, jedoch Uneinigkeit zwischen Experten die Genauigkeit der automatischen Modelle erheblich beeinträchtigte. Dies zeigte, dass automatische Ansätze zwar für domänenspezifische Konzepte mit hoher Übereinstimmung zwischen Experten zielführend sind, manuelle Annotationen trotz ihres Arbeitsaufwands für semantisch komplexe Konzepte aber weiterhin notwendig bleiben.

Die Ergebnisse der sequenzspezifischen Analyse in dieser Dissertation zeigten, dass

manuell erstellte, nachvollziehbare Merkmale wertvoll bleiben, da sie es im Vergleich zu maschinenbasierten Merkmalen ermöglichen, Forschungsergebnisse in umsetzbare Richtlinien zu übersetzen.

Abschließend wurden negative Ergebnisse in der Vorhersage verschiedener Sequenzresultate mit unterschiedlichen Kombinationen von Modelltypen unter Verwendung von Merkmalen wie Teamposition, Kompaktheit, oder Anzahl von Spielern zwischen Ball und Tor erreicht. Dies ließ keine Aussage zu der konkreten Verbindung der untersuchten Merkmale mit Sequenzresultaten zu.

Zusammenfassend sind die gewonnenen Erkenntnisse für Fußballpraktiker und Forscher sowohl in der Sportwissenschaft als auch in der Informatik von Bedeutung. Wenn Annotationen in einem dieser Bereiche vorgenommen werden, sollte die Bedeutung klarer, regelbasierter Definitionen und hierarchischer Annotationsschemata beachtet werden, um die semantische Komplexität der untersuchten Konzepte zu erfassen. Für das Training automatisierter Studien muss die Übereinstimmung zwischen Experten bei den untersuchten Konzepten gewährleistet sein, während die Interpretierbarkeit der verwendeten Merkmale und Modelle ein zentraler Punkt der verwendeten Architektur sein sollte. Diese Leitlinien können dazu beitragen, die Lücke zwischen theoretischer Forschung und praktischer Anwendung in Fußballvereinen zu schließen. Letztlich legt diese Arbeit das Fundament für weitere Fortschritte in der Verbindung zwischen Informatik und Sportwissenschaft, bei denen ähnliche experimentelle Studien auf anderen Zeitskalen zu einem ganzheitlichen Verständnis des Spiels beitragen können.

Abstract

Over the past years, the applied science of *sports analytics* has significantly influenced strategies in the US sports baseball, basketball, and American football. In soccer, despite advancements in technology and data collection exist, such significant changes are yet to be seen.

Reasons for the comparably lower adaptation of such data-driven analysis in soccer can be found in soccer's low-scoring nature, tactical complexity, and fluidity due to a low amount of structural components. To bridge this gap, this dissertation aims to improve on the foundational requirements of understanding soccer. Specifically, the lack of connecting approaches between domain-specific terms and concepts into concrete, rule-based definitions is addressed.

In the sports science community, various taxonomies and annotation schemes for the attacking and defending process in soccer have been defined. However, due to the oftentimes more narrative than descriptive nature of these schemes they do not enable the degree granular annotation required in computer science. Moreover, they sometimes include *semantically complex* subjective definitions, are not well validated, or lack a clear rule-based decision system and can not easily be understood by humans or recreated by automatic models.

In the present dissertation, this problem is addressed on various time scales. On a short time scale, a unified, hierarchic taxonomy for *atomic* event annotation was proposed and validated by a multi-expert agreement study. Moreover, the annotated events were refined by a time-series event detection approach to synchronize the *atomic* events with the underlying position data. On an intermediate time scale, maintaining the hierarchy when moving between time scales, a hierarchic annotation scheme for the *sequence*-based concept of *tactical periodization* was proposed and also validated in a multi-expert annotation experiment. Finally, a framework for *sequence*-specific analyses in soccer was proposed and applied to the concept of *counterattacks*.

A central result of the annotation experiments was the importance of a hierarchical structure in annotation schemes, effectively representing *semantic complexity* and integrating differing expert annotations into shared parent classes.

An important finding of the automatic model training was that while the recreation of manual *sequence* annotations was possible, expert disagreement generally impeded model accuracy. This showed that while automatic approaches are interesting for domain-specific concepts with high expert agreement, manual annotations, despite their labor intensity, remain necessary for *semantically complex* concepts.

The results of the *sequence*-specific analysis suggested that manually crafted, comprehensible features remain valuable as they enable the translation of research findings into actionable guidelines.

Finally, negative results regarding the prediction of various *sequence* outcomes were encountered, despite different types models and features such as team position, compactness, or players between the ball and the goal were used. This left the interconnection of the investigated features with regard to concrete *sequence* outcomes uncertain.

To conclude, the insights of this disseration are valuable for soccer practitioners and researchers, both in the sports science and computer science communities. When

annotations are made in either of those domains, the importance of clear, rule-based definitions and hierarchical annotation schemes should be respected to capture the *semantic complexity* of the investigated concepts. For training automatic studies, expert agreement of the examined concepts should be ensured while the interpretability of the used features and models should be a focal point of the model architecture. These guidelines can help towards integrating more theoretical research into practical applications in soccer clubs. Ultimately, this work lays a foundation for further advances in the connection of sports science and computer science, where similar experimental studies on other time scales might contribute to a more holistic understanding of the game.

Overview of the Articles

Ten articles have been published over the course of this doctoral program. Out of these ten, six include a first authorship with one shared first authorship. Five of these six articles are part of the synopsis of this cumulative dissertation and will be discussed in detail. The remaining articles are also part of this synopsis, but will not be discussed in detail. An overview of the scientific output and naming scheme is provided below.

Article #		Status
Articles that are part of this synopsis and will be discussed in detail		
I	**Biermann, H.**, Theiner, J., Bassek, M., Raabe, D., Memmert, D., & Ewerth, R. (2021). A unified taxonomy and multimodal dataset for events in invasion games. *Proceedings of the 4th International Workshop on Multimedia Content Analysis in Sports*	Published
II	**Biermann, H.**, Komitova, R., Raabe, D., Müller-Budack, E., Ewerth, R., & Memmert, D. (2023). Synchronization of passes in event and spatiotemporal soccer data. *Scientific Reports*, 13(1), 15878.	Published
III	**Biermann, H.**, Memmert, D., Petersen, N., Raabe, D. (2025). Contextualization of soccer analysis with tactical periodization and machine learning. *Data Mining and Knowledge Discovery*, 39, 23.	Published
IV	**Biermann, H.**, Wieland, F. G., Timmer, J., Memmert, D., & Phatak, A. (2022). Towards Expected Counter - Using Comprehensible Features to Predict Counterattacks. *Machine Learning and Data Mining for Sports Analytics. MLSA 2022. Communications in Computer and Information Science (1783)*, 3-13.	Published
V	**Biermann, H.**, Yang, W., Wieland, F. G., Timmer, J., & Memmert, D. (2023). Quantification of Turnover Danger with xCounter. *Machine Learning and Data Mining for Sports Analytics. MLSA 2023. Communications in Computer and Information Science (2035)*, 36–51.	Published
Articles that are part of this synopsis but will not be discussed in detail		
VI	Wunderlich, F., **Biermann, H.**, Yang, W., Bassek, M., Raabe, D., Elbert, N., Memmert, D., & Garnica-Caparrós, M. (2025). Assessing Machine Learning and Data Imputation Approaches to Handle the Issue of Data Sparsity in Sports Forecasting. *Machine Learning*. 114, 48 (2025).	Published
VII	Stival, L., Pinto, A., Andrade, F. D. S. P. D., Santiago, P. R. P., **Biermann, H.**, Torres, R. D. S., & Dias, U. (2023). Using machine learning pipeline to predict entry into the attack zone in football. *PloS one*, 18(1).	Published
VIII	Raabe, D., **Biermann, H.**, Bassek, M., Memmert, D., & Rein, R. (2024). The dual problem of space: Relative player positioning determines attacking success in elite men's football. Journal of Sports Sciences, 1-10.	Published
IX	**Biermann, H.**, Memmert, D., Romeike, C., Knäbel, P., & Furley, P. (2024). Relative age effect inverts when looking at career performance in elite youth academy soccer. Journal of Sports Sciences, 1–6.	Published
X	Raabe, D., **Biermann, H.**, Bassek, M., Wohlan, M., Komitova, R., Rein, R., Kuppens Groot, T. & Memmert, D. floodlight - A high-level, data-driven sports analytics framework. *Journal of Open Source Software*, 7(76), 4588.	Published

Contents

Chapter 1

Introduction

One significant cultural difference between the United States and Europe is the variation of public interest in sports. Although soccer is the dominant sport in Europe (Nielsen Sports, 2018), it has not achieved the same status in the US, where American football, baseball, and basketball (Norman, 2018) are the top favored sports. In addition to differing sports preferences, there are also discrepancies in the adoption and success of sport analysis techniques, often referred to as *sports analytics* in the US (Tuyls et al., 2021). Notably, for baseball, the data-driven recruiting approach *Moneyball* (Lewis, 2004) stands out as one of the most prominent examples of how a major sport in the US is influenced by *sports analytics*. Examples of successful applications of *sports analytics* also exist in other popular US sports. For instance, in basketball, shot selection analyses revealed the inefficiency of long two-point shots compared to close two-point or three-point shots (Goldsberry, 2019). Consequently, the frequency of long two-point shots has significantly decreased compared to previous seasons (Goldsberry, 2019). Similarly, statistical analyses of play conversion rates in American football have influenced strategies for comebacks when trailing late in games (Fox, 2021).

However, in soccer, a similarly strong influence of *sports analytics* on gameplay has yet to emerge. This can be attributed to several factors, including limited participation in systematic data collection (Tuyls et al., 2021) and specific characteristics of soccer that make its analysis more challenging compared to American football, baseball, and basketball. These characteristics include the scarcity of scoring events (Anzer & Bauer, 2021), the comparably large number of players (G. Liu et al., 2020), the complex outdoor playing environment (Tuyls et al., 2021), the significant role of randomness (Wunderlich et al., 2021), the limited scope for structural interventions (Fernández et al., 2021), as well as the intricate complexity of events and player interactions during a match (McHale et al., 2012).

1.1 Data Analysis in Soccer

Soccer data has been recorded and analyzed for decades, with Reep and Benjamin (1968) often cited as the pioneers in this field. While at that time the authors recorded passes and goals manually in notebooks, recent technological advances have facilitated the collection and analysis of various types of soccer data (Pino-Ortega et al., 2022; Pappalardo et al., 2019).

To adequately capture the complexity of soccer, three types of soccer data have gained popularity within the community. (i) *Video data*, which includes both single-camera tactical feeds and multi-camera broadcast streams. (ii) *Event data*, which consists of a stream of manually or semi-automatically captured player, team, and match events, defined by varying event catalogs from *event data providers* such as Impect (2024), StatsPerform (2024), StatsBomb (2024), and Wyscout Spa (2024). (iii) *Position* or *tracking data*, which records player movements on the pitch as two- or three-dimensional coordinates at a specified sampling frequency (Pino-Ortega et al., 2022; Pappalardo et al., 2019).

To date, *video data* remains the predominant source of information (Mehta et al., 2024), primarily due to its accessibility and comprehensiveness. In recent years, however, *position* and *event data* have emerged as valuable additions to complement *video data* in analysis (Mehta et al., 2024). Nevertheless, concerns regarding the accuracy of the data sources as well as difficulties in synchronizing *position* and *event data* have been reported (Anzer & Bauer, 2021; H. Liu et al., 2013; Pino-Ortega et al., 2022). These challenges are further examined in the next chapter, which outlines issues when working with soccer data and discusses factors limiting a seamless integration of automated, data-driven processes into the soccer analysis workflow.

1.2 Research Gaps & Research Questions

As previously highlighted, the application of *sports analytics* (in practical settings) is more advanced in US sports compared to soccer (Tuyls et al., 2021). However, this discrepancy is not solely due to technological limitations, as access to diverse data sources and advanced analytical methods is readily available. Yet, the conversion of these resources into practicable, actionable insights remains notably limited (Mehta et al., 2024). This shortcoming points to other potential factors that may have hindered more influential research in this field.

One possible factor for this development is the lack of universally recognized soccer-specific concepts and truths to anchor technological applications (Goes et al., 2021). Although automated models have successfully predicted various in-match developments (Fassmeyer et al., 2021; Sanford et al., 2020), these studies largely focus on concepts in enclosed environments and on very short time scales (such as passes or shots), referred to as *atomic events*. Yet, as these predictive approaches are of limited relevance to practitioners, an integration of predictive models into practical applications (e.g. to suggest tactical shofts during a game), remains largely unexplored.

A potential way to address this discrepancy is to define fundamental tactical patterns, substantiated by traditional sports science, validated across multiple data sources, and agreed upon by domain experts (Goes et al., 2021; Tuyls et al., 2021). To date, practitioners commonly use colloquial terms such as *possession play*, *counterpress*, or *counterattack* to describe recurring patterns in soccer matches. However, these terms have yet to be translated into formalized, rule-based concept definitions.

The present dissertation focuses on performing this translation step for concepts on various time scales, from *atomic events* on the shortest time scale up to the analysis of entire *sequences*. Following Low et al. (2020), for each time scale, widely accepted concepts are used as a foundation and is gradually advanced. Therefore, the term *semantic complexity* is introduced to encapsulate the uncertainty surrounding sport-specific definitions, where clear guidelines for the rule-based identification of the underlying concepts are lacking.

Thus, four research gaps that have potentially hindered a more advanced adaptation of theoretical research on soccer into practice are identified. By filling these research gaps, the integration of soccer-specific research (in the sports and computer science community) into practical applications is tackled.

RG 1 **Lack of globally valid atomic event annotations**: There is a shortage of studies that investigate the validity of *atomic events*. Due to differing event catalogs within the community, there are numerous definitions that lack unified guidelines and standards. Although efforts have been made to integrate event definitions into a unified language (Decroos et al., 2019; FIFA, 2022), neither the validity of these integrative classes nor the potential semantic overlap between definitions have been thoroughly explored. Thus, if meaningful insights are retained using *event data*, the lack of scientific validation for the source event definitions limits the reliability of outcomes derived.

One potential approach to address this gap is through automatic computer vision techniques that detect events directly from *video data*. However, as the community primarily focuses on high-level video understanding, the fine-grained and reliable detection of events remains a highly complex problem that has yet to be fully solved (Deliege et al., 2021; Naik et al., 2022).

RG 2 **Issues with synchroniz wation between event and position data**: Challenges exist in achieving temporal alignment between *position* and *event data*. As a consequence of this misalignment, analysis techniques that integrate *position* and *event data* without synchronization may produce (unrecognized) biased results.

To address this problem, Anzer and Bauer (2021) proposed a synchronization technique to align shot events in *position* and *event data*. Although their algorithm revealed promising results, they have not extended their proposed solution to more frequently occurring events. More recently, Van Roy et al. (2023) built on the previous approach by Anzer and Bauer (2021) and proposed the rule-based ETSY algorithm that synchronizes various events in *event data* with *position data*. However, while their algorithm provides a substantial first step towards synchronizing *event* and *position data*, the hand-crafted rules are not validated by comparison against possible automatic solutions. Furthermore, the algorithm requires event location information, which is not always present in all *event data* sources.

RG 3 **Lack of globally valid sequence annotations**: There is a scarcity of validated annotations for *semantically complex* team behavior during *sequences*. As a result, sports practitioners must either perform labor-intensive manual annotations or rely on non-validated rule-based algorithms when analyzing specific situations of interest within a game.

A foundational step towards establishing a validated automatic model for generating *sequence*-based annotations was proposed by Bauer and Anzer (2021). The authors introduced an automated detection of *counterpressing* situations, defined as the immediate attempt to regain *ball possession* after a *turnover*. This solution is highly valuable to the community as it facilitates the accurate detection of *counterpressing* in relevant situations. Yet, their algorithm is limited to these specific situations such that a holistic annotation of a soccer match cannot be performed. Thus, there remains a lack of general automatic approaches that can provide *sequence* annotations for soccer matches.

RG 4 **Lack of sequence-specific analysis**: The previously described lack of validated *sequence* annotations for *semantically complex* concepts has impeded a robust automatic analysis of specific situations within a match. Although a team behavior analysis has been performed using aggregated match statistics (Fernandez-Navarro et al., 2016), this approach does not offer fine-grained insights into specific situations. In contrast, individual *ball possession* intervals have been analyzed using manually crafted *sequence* annotations (Kempe et al., 2014); however, this method is labor-intensive and potentially subjective.

To tackle these challenges, fine-grained automatic *sequence*-specific analysis techniques are highly valuable, as they significantly reduce the workload of human analysts while improving the objectivity and granularity of the insights gained. Yet, to be fully accepted by the practitioners community, a critical requirement of these analysis techniques is a sufficient degree of interpretability.

In this dissertation, these four research gaps are addressed on time scales spanning from *atomic events* to *sequences*, with the goal of integrating automatic, data-driven applications into practitioners' workflows.

Chapter 2 provides background information and describes the soccer-specific and algorithmic methodologies employed throughout this work.

Chapter 3 focuses on RG 1 and RG 2, both relating to the manual annotation of *event data*. First, a unified taxonomy for *atomic events* is introduced (Chapter 3.2). Second, a synchronization algorithm for *position* and event data is proposed (Chapter 3.3).

Chapter 4 addresses RG 3 through a novel method to manually annotate and to automatically detect tactical periodization labels in soccer.

Chapter 5 tackles RG 4 by isolating and analyzing *counterattacks* in a soccer match.

Finally, Chapter 6 summarizes the results of this dissertation, discusses limitations and the application to practice, and provides potential directions for future work.

A summary of the discussed research gaps, resulting research questions, and connected articles in this dissertation is provided in Table 1.

Time Scale	Research Gap		Research Questions		Article
Atomic	**RG 1**	Globally valid atomic event annotation	**RQ 1**	How can reliable, validated, and semantically rich manual annotations for *atomic events* be created?	**I**
Atomic	**RG 2**	Synchronization between event and position data	**RQ 2**	To what extent is it possible to synchronize manual *event data* with automatically tracked *position data*?	**II**
Sequence	**RG 3**	Globally valid sequence annotation	**RQ 3A**	How can reliable, validated, and contextually rich manual *sequence* annotations for team behavior be created?	**III**
			RQ 3B	Can automatic models accurately replicate manual *sequence* annotations?	
			RQ 3C	How can automatic *sequence* annotations enhance common soccer analysis tasks?	
Sequence	**RG 4**	Sequence analysis	**RQ 4A**	To what extent is it possible to create isolated ratings for individual *sequences*?	**IV & V**
			RQ 4B	How can interpretable, data-driven insights for optimal *sequence*-specific team behavior be generated?	

Table 1: Overview of research questions tackled in this dissertation.

Chapter 2

Background

This chapter first provides an overview of *soccer analytics* (Chapter 2.1). Then, the methodological foundations for the used sport-specific concepts (Chapter 2.2) and algorithmic methods (Chapter 2.3) are described.

2.1 Soccer Analytics

The upcoming chapter introduces *soccer analytics* by outlining typical workflows in soccer analyses (Chapter 2.1.1) and discussing common soccer data sources and their applications (Chapter 2.1.2).

2.1.1 Workflows

The sports science literature commonly classifies sport games into three distinct families, based on shared properties, so-called family resemblances (Wittgenstein & Anscombe, 2003):

1. *Net and wall games*, which have no set time limits and are determined by scores (e.g., tennis, squash, volleyball).

2. Striking/fielding games, structured in innings (e.g., cricket, baseball).

3. *Invasion games* (e.g., handball, basketball, rugby), which are played within a set time and involve directing an object (e.g., ball, frisbee, puck) towards a target (e.g., goal or basket) to score points.

Soccer is classified as an *invasion game*, as teams aim to move a ball within defined boundaries into an opponent's goal within a set time (IFAB, 2024). The intrinsic complexity of soccer arises from several factors such as the ball's properties, the physical demands of maneuvering the ball primarily with the feet (Davids et al., 2000), the high running speeds of players (Pino-Ortega et al., 2022), the fluidity of the game and low degree of structural intervention (G. Liu et al., 2020), the strategic coordination required among the eleven players on each team.

Due to this complexity, the importance of strategy in soccer has evolved over the past decades. This development has led teams to increasingly rely on specialized staff, such as coaches, managers, scouts, and analysts, to optimize the underlying tasks in a soccer team including tactical preparation, strategic decisions, and player acquisitions (Relvas et al., 2010).

These domain experts need to adapt to the rapidly changing environment present at soccer clubs (e.g., due to a tight schedule with many matches). Still, the detailed and comprehensive nature of qualitative analyses remain highly valued by practitioners (Mehta et al., 2024). Yet, the manual labor involved in this process is often characterized as cumbersome and time-consuming as analysis results - covering many players, teams, and competitions - must be regularly updated (Decroos et al., 2018; Martin et al., 2021; McKenna et al., 2018). This results in a large workload that are often demanded by practitioners in soccer clubs.

A potential way to lighten this workload is the integration of automated processes in soccer clubs, with the aim of providing objective assessments and complementing human judgment (allowing the human to focus on more nuanced tasks).

2.1.2 Soccer Data & Usages

The integration of automatic models in soccer analysis depends on the availability of input data for training, validation, and evaluation. There exist three primary data sources in the soccer analysis community:

(i) **Video data** includes various formats of camera recordings of soccer matches, such as single-camera tactical views and multi-camera broadcast feeds. Each frame, an RGB pixel image, varies in resolution depending on the camera setup, with typical resolutions such as 1280 x 720 pixels for HD. Frames are captured multiple times per second, determined by the camera configuration. Although *video data* generates vast quantities of information, which can pose challenges for automatic processing (Verma & Agrawal, 2016), it remains a preferred source due to the extensive insights it provides. Professional match analysts, in particular, favor *video data* (Mehta et al., 2024).

(ii) **Event data** refers to the manual (or semi-automated) notational descriptions of a soccer match (H. Liu et al., 2013; Wright et al., 2012). A single instance of *event data* typically includes the characterization of an event type (from a predefined catalog), the location on the pitch, and the timestamp of the event. Different event types range from individual player actions like passes to team-tactical behavior such as a *counterpress*. However, individual definitions are not unified and vary between *event data providers* (Impect, 2024; StatsBomb, 2024; StatsPerform, 2024; Wyscout Spa, 2024).

    Due to its specificity, *event data* is used in various applications, such as generating match statistics and aiding video analysis (HUDL Sportscode, 2023; Shih, 2018). However, while the latter application can significantly reduce analysts' workloads, the focus of *event data* on on-ball actions limits its ability to analyze off-ball movements and broader tactical contexts (Herold et al., 2022).

(iii) **Position (Tracking) Data** captures the xy-coordinates of all 22 players and the ball on the pitch, with an occasional third dimension (for the ball). Like *video data*, *position data* frames are recorded multiple times per second at a fixed frequency, which varies between *event data providers* (Kinexon, 2023; Tracab, 2024; Wyscout Spa, 2024). Optical tracking systems are commonly used due to their non-invasiveness (Pino-Ortega et al., 2022), but GPS systems can offer higher accuracy and provide additional metrics like heart rate (Shih, 2018).

In practice, *position data* is used to assess physical performance metrics, such as speed or distance covered (SkillCorner, 2024), while applications for analyzing tactical aspects are currently being developed (Goes et al., 2021).

While research has traditionally focused on individual data sources, integrating multiple sources offers substantial potential for enhanced insights. For instance, as demonstrated by Bauer and Anzer (2021), combining *position data* with *event data* can provide a more comprehensive understanding of the match. However, despite the exploration of this concept, issues with temporal alignment (see RG 2) have complicated research in this direction.

2.2 Soccer-Specific Methods

This chapter presents the theoretical background of the soccer-specific methods used to address the domain-specific challenges in this dissertation. First, the notational description of soccer (Chapter 2.2.1) is discussed. Second, a framework for analyzing individual *sequences* in soccer (Chapter 2.2.2) is introduced.

2.2.1 Soccer Annotation Schemes

A common problem in *soccer analytics* is the analysis of individual soccer-specific concepts, that are deemed relevant to experts or scientists. Yet, the previously discussed lack of extensive public databases of expert annotations (Pappalardo et al., 2019) and automatic algorithms for annotations (see RG 3) necessitates the conduction of manual expert annotation studies (Decroos et al., 2018; Goes et al., 2021; Pappalardo et al., 2019).

However, it is problematic that expert annotations in scientific studies are often not sufficiently validated (see RG 1 and RG 3), as expert annotations often rely on individual judgment and encompass various factors, such as ball position, player movement, and team formation. Thus, there exists a clear need for concise, unambiguous, and objective annotation schemes that offer detailed, granular definitions of annotated concepts and serve as guidelines to ensure reproducible results.

To facilitate the creation of annotation schemes, several key criteria that promote validity and comparability are identified. In Chapters 3 and 4 of this dissertation, these criteria are followed to ensure a transparent design process of the used annotation schemes.

Given the complexity and dynamic nature of soccer (see Chapter 2.1.1), developing comprehensive annotation schemes is inherently challenging. To accommodate specific use-cases, annotation schemes should be *modular and expandable*.

To address varying levels of *semantic complexity* (see Chapter 1.2), annotation schemes can benefit from a *hierarchical* structure. This design enables a clear and concise framework with varying different level of detail, where events can default to their common properties at a higher hierarchical (parent) level if needed.

Annotation schemes should aim to be *minimal and unambiguous*, reducing redundancy and promoting clarity. Each class within the scheme should be clearly defined and distinct, ensuring that every real-time development on the pitch corresponds uniquely to one annotation class.

Annotation schemes should also strive for a *holistic* description of a match, ensuring a comprehensive coverage of all match situations and avoiding blind spots. Yet, this characteristic must be balanced with the previously discussed criteria of a *minimal and unambiguous* annotation scheme.

The key criteria for annotation schemes are summarized below:

C1 *Modular and expendable*: Schemes should begin with a foundational structure and enable the integration of detailed, sport-specific modules as needed.

C2 *Hierarchical*: Schemes should employ a hierarchical structure to manage varying levels of *semantic complexity* effectively.

C3 *Minimal and unambiguous*: Schemes should contain a minimal number of precise, non-overlapping classes.

C4 *Holistic*: Schemes should comprehensively describe match situations, ensuring that no critical developments are overlooked.

2.2.2 Framework for Analyzing Complex Sequences in Soccer

Given the complexity and fluidity of soccer, as outlined in Chapter 2.1.1, segmenting matches and performing isolated analyses of individual concepts is a promising approach in sports science (Memmert et al., 2017). While various studies have explored this strategy, they often focus on either narrow or broad time scales, while intermediate time scales are less commonly explored.

On shorter time scales, approaches like the *expected goals* (Anzer & Bauer, 2021) or *expected passes* (Dick et al., 2022) predict the outcomes of individual events within a confined context, leveraging a limited set of possible outcomes. Conversely, larger-scale analyses predict outcomes across entire matches (Wunderlich & Memmert, 2021) or even seasons (Groll et al., 2019), relying on aggregated data to simplify complex individual processes.

However, intermediate time scales, referred to as *sequences* in this work, present unique challenges. *Sequence*-specific analyses on these time scales struggle with defining a confined analytical environment while also lacking the aggregation benefits of broader time scales. To address these challenges, a framework that streamlines the identification and analysis of *sequences* is introduced and applied in Chapter 5 of this dissertation.

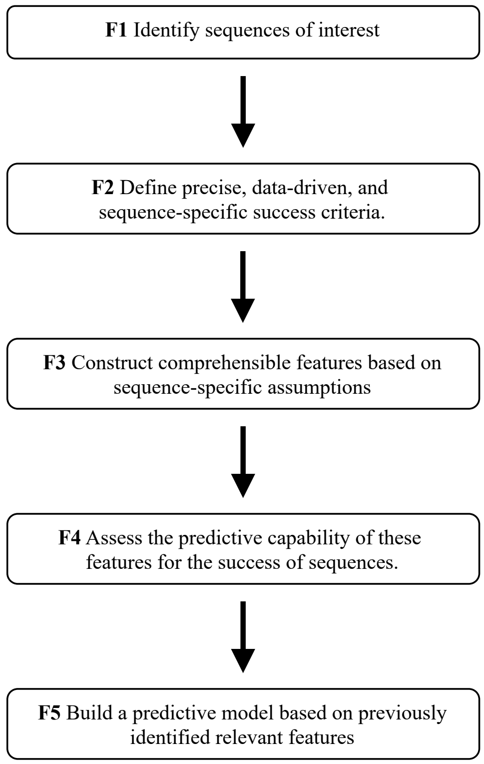

Figure 2.1: Framework for Analyzing Complex Sequences in Soccer, displayed as a flowchart.

The first step of the framework is the precise, rule-based definition of *sequences* of interest. As highlighted in Chapter 1.2, converting colloquial terms into rule-based definitions is a critical, yet often overlooked, component for reliable analyses

Secondly, quantifying the impact of the identified *sequences* of interest on match outcomes is challenging, due to the variability of success strategies in soccer (as discussed in Chapter 2.1.1). The framework addresses this problem by establishing precise, data-driven, *sequence*-specific success criteria that provide a comprehensive description of *sequence* outcomes, enabling clear assessments of success.

Next, even with well-defined *sequences* and success criteria, the inherent complexity of individual actions within *sequences* necessitates the creation of comprehensible features. These features should be specifically tailored to describing the *sequence* developments and should be evaluated for their predictive power in relation to *sequence* success.

The critical evaluation is essential before finally advancing to the development of predictive models based on the identified relevant features. The framework can be represented as a successive list of actions, summarized in Figure 2.1.

2.3 Algorithmic Methods

This chapter provides the theoretical background for two key machine learning problems and algorithms relevant to this dissertation. An overview of time series event detection methods is given in Chapter 2.3.1 and different models for the binary classification problem are introduced in Chapter 2.3.2.

2.3.1 Time Series Event Detection

In time series analysis, an event is defined as a recurring pattern within the time series that holds particular significance for the user (Guralnik & Srivastava, 1999). The characteristics of such events can vary, including peaks, level shifts, or changes in spectral patterns, depending on the application (Gensler & Sick, 2018).

Event detection in this context refers to identifying the specific time points at which events occur, also known as change point detection (Guralnik & Srivastava, 1999). Applications of time series event detection include traffic monitoring, web access logging, and medical diagnostics (Ihler et al., 2006).

Numerous algorithms, including popular machine learning methods, have been developed to address this problem. However, Makridakis et al. (2018) found that classical statistical approaches often outperform machine learning algorithms while also being computationally less expensive. One example of a statistical approach is the *SwiftEvent* algorithm for time series event detection (Gensler & Sick, 2018), which is applied to a soccer-specific event detection problem in Chapter 3.3.

SwiftEvent Algorithm *SwiftEvent* is an efficient algorithm developed for detecting events in time series data with a single pass over the dataset.

Therefore, *SwiftEvent* first uses a sliding window approach to segment time series data into windows that either do or do not contain events of interest. The algorithm then transforms the segmented data into points within a low-dimensional feature space, consisting of a concise set of descriptive features that effectively capture the characteristics of the events of interest. This transformation results in a distinct separation within the feature space between windows containing events of interest and those without. This separation is leveraged in a training step.

During training, data points in the low-dimensional feature space are assigned an event label if that event occurred at the center of the corresponding time series window. This labeling process is repeated for each data point, resulting in the formation of clusters containing points with differing labels. To complete the training, these clusters are used to parameterize multivariate Gaussian distributions, encapsulating the statistical characteristics of each cluster.

Finally, the Gaussian distributions are used to perform event detection in unseen time series. The time series is segmented and transformed into a low-dimensional feature space. For each data point generated in this manner, Mahalanobis distances are computed to each of the multivariate Gaussian distributions. An event is detected for a data point if the Mahalanobis distance to a multivariate Gaussian distribution falls below a previously determined threshold and is lower than the distance to all other distributions.

A significant advantage of this process is its ability to detect events that reoccur with slight variations (such as changes in time or magnitude) even when normally distributed noise is present. Despite these variations, events can still be accurately detected and assigned to the correct event class. Moreover, the efficient and relatively simple training procedure, which does not require a large number of labeled samples, makes *SwiftEvent* well-suited for event detection tasks related to soccer. For more detail on *SwiftEvent* refer to Gensler and Sick (2018).

2.3.2 Binary Classification

Binary classification is a fundamental task in machine learning, where the objective is to categorize data points into one of two distinct categories based on a set of input features. This process involves using an input vector that represents the characteristics of a data point to determine its class. Each data point must be assigned to one of the two categories, with no possibility for a third option such as undecided.

Binary classification problems are typically approached within the context of supervised learning, where each data point is labeled with its true class. These labels are crucial for guiding the training of the machine learning model and evaluating its performance on unseen data.

Numerous supervised machine learning algorithms are available for binary classification, each varying in complexity. In the upcoming chapter, two of those architectures are presented and an overview of popular metrics used for binary classification is given. For a more detailed discussion of the architectures and the binary classification problem, refer to Bishop and Nasrabadi (2006).

Metrics for Binary Classification

A common metric for evaluating model performance in binary classification tasks is class accuracy. This metric is easily interpretable. For a given class, it represents the ratio of correctly classified data points to the total number of data points classified to be of that class. However, in highly imbalanced binary classification problems, class accuracy can be misleading. For example, if a classifier always predicts the majority class and never predicts the minority class, its accuracy will correspond to the frequency of the majority class. This can result in high accuracy values, even though the minority class is never correctly classified. To address this issue, commonly used metrics such as precision and recall are often preferred.

For a given class, precision describes the ratio of correctly classified data points to the total number of data points predicted to belong to that class. It answers the question: 'If the output class was predicted, how often was the prediction correct?'. Recall is the ratio of correctly classified data points of that class to the total number of data points of that class. It answers the question: 'How many of the total data points of the given class were correctly predicted?'. Finally, the F1-score is calculated as the harmonic mean of precision and recall, combining both metrics into a single value.

Logistic Regression

Logistic regression is often used as a baseline for comparison with more complex models. In this dissertation, *logistic regression* is applied in Chapters 4 and 5.

Logistic regression belongs to the family of probabilistic discriminative models, which focus on maximizing the likelihood function based on conditional class distributions—a technique known as discriminative training. Before applying *logistic regression*, the original input vector is transformed into a feature space, often using a fixed nonlinear function. This approach ensures that while the decision boundaries in the feature space are linear, they correspond to nonlinear decision boundaries in the original input space, enabling *logistic regression* to model more complex relationships between variables.

At the core of *logistic regression* is the formulation of the posterior probability of class membership using a sigmoid function. This function maps the transformed feature space to a probability between 0 and 1, representing the likelihood that a data point belongs to a particular class. In an M-dimensional feature space, the model will have M adjustable parameters to optimize. Fitting the adjustable model parameters involves maximizing the likelihood of the observed data. Thus, an error function is defined as the negative logarithm of the likelihood. The parameters are iteratively updated to minimize this errorx. This process continues until the model converges to a set of parameters deemed optimal.

Neural Networks and Backpropagation

A *neural network* is an example of a more complex machine learning architecture. In this dissertation, *neural networks* are applied in Chapter 4.

Neural networks are a class of supervised machine learning architectures that perform a series of matrix-vector multiplications to transform an input vector into an output. Due to their flexible structure, they are a popular choice for binary classification problems.

Neuron *Neural networks* are designed to emulate the functioning of the human brain (Rosenblatt, 1958), with their basic building block known as a *neuron*. Similar to biological neurons, which process and transmit information, the artificial *neurons* in a *neural network* transform multi-dimensional input data into a one-dimensional scalar output. This transformation involves two fundamental components: the parameters of the *neuron* and the *activation function*.

The parameter vector of a *neuron* consists of weights, each multiplied by corresponding elements of the *neuron's* input. Additionally, it contains a unique weight, known as the bias, which is not directly multiplied with an input element but is added to the weighted sum of the inputs. While the parameter vector is adaptable and modified during training, the *activation function* is predefined and remains fixed. It operates on the weighted sum of the inputs, including the bias, to produce the *neuron's* output. Consequently, the weights and the bias are referred to as the trainable parameters of the *neuron*.

Feed-Forward Neural Network Similar to the processes in the human brain, a multitude of interconnected *neurons* is required to work together to create meaningful outputs. In artificial *neural networks*, the parallel connection of multiple *neurons* is known as a perceptron (Rosenblatt, 1958). Each individual neuron within the perceptron processes the same input simultaneously. Thus, the *neurons* collectively determine the perceptron's output, where the number of *neurons* directly affects the output size.

Due to the simple linear operations executed by a single *neuron*, the computational power of a single perceptron is inherently limited. To overcome this limitation, it is common practice to concatenate multiple layers of perceptrons, forming what is known as a multilayer perceptron (Rosenblatt, 1958). To build this type of structure, a second perceptron is added to the first, with each *neuron* in the second perceptron receiving the outputs from the *neurons* of the first perceptron. This configuration, known as a two-layer feed-forward neural network, can be further expanded by adding more layers; in practice, so-called *deep neural networks* often consist of several layers.

Network Training During network training, the trainable parameters of a *neural network* are adjusted to minimize an error function that depends on these parameters. The objective of network training is to minimize this error function, which is often visualized as navigating a geometric landscape to find the point where the error function reaches its lowest value, representing the optimal parameters.

To guide this search, the error gradient is computed. The gradient indicates the direction of the steepest increase in the error function, based on the current parameter values. The training process involves searching for a point in the weight space where this gradient is zero. At this point, no further adjustments to the parameters can reduce the error function value, indicating that the optimal parameters have been found.

Given the complexity of analytically identifying where the gradient is zero, an iterative numerical search technique is typically employed. This process begins with a randomly initialized parameter value, which is updated successively. Updates are made using the gradient information at the current parameters to take small steps in the direction of the negative gradient, with the step size being determined by the learning rate. Due to the geometric interpretation of the error function, this method is known as *gradient descent*.

To evaluate the gradient effectively at the current weight vector, the backpropagation method has been developed (LeCun et al., 1988; Rumelhart et al., 1986). This technique involves propagating partial derivatives backward through the network, from the output layer towards the input layer, using the chain rule of calculus. By first calculating the derivatives in the last layer and storing these values, the algorithm can recursively compute the derivatives for preceding layers. This enables efficient gradient evaluation independent of the network's specific topology.

Chapter 3

Atomic Event Annotation of Soccer Matches

This chapter addresses the challenge of creating reliable and synchronized *event data* for soccer matches and examines associated issues, previously summarized in RG 1 and RG 2.

Chapter 3.1 outlines the state-of-the-art annotation principles used by *event data providers* and explains how these principles may lead to issues with validity and temporal precision.

Chapter 3.2 proposes an innovative *event data* annotation framework in the form of a unified, hierarchical taxonomy of events in *invasion games*, addressing RQ 1. It is based on the content of Article I.

Chapter 3.3 presents time-series event detection techniques and introduces an algorithm for refining *event data* temporally, contributing to the answer for RQ 2. It presents the content of Article II.

3.1 Principles of Annotation by Event Data Providers

A significant portion of the *event data* used by soccer clubs is provided by companies such as Impect (2024); StatsBomb (2024); StatsPerform (2024); Wyscout Spa (2024), referred to as *event data providers* in this work. Although scientific institutes have released open *event data* sets (Pappalardo et al., 2019), these datasets have limited coverage as they contain only selected competitions within a certain time span. This is insufficient for clubs that, as discussed in Chapter 2.1.1, require continuously updated *event data* across various competitions worldwide.

The data from *event data providers* relies on human annotators who analyze video footage to assign predefined event classes (H. Liu et al., 2013). This process has been reported to be highly time-consuming, requiring up to two hours and three expert annotators per match, depending on the desired level of granularity and validity (Pappalardo et al., 2019).

The range of event classes available for annotation is based on specific definition catalogs, which generally include basic events like shots or passes. However, these catalogs differ in their definitions and incorporation of more nuanced events. For instance, StatsBomb's catalog includes the event type *fifty-fifty* to denote a challenge between two players for the ball (StatsBomb, 2024). Yet, this event type is absent from StatsPerform's catalog, which instead includes various other event types such as *tackle*, *take-on/dribble*, or *smother* (StatsPerform, 2024). Although the overall inter-annotator validity of data from *event data providers* has been confirmed, significant discrepancies have been observed regarding agreement on more specific event classes, as well as temporal misalignment between events annotated by different annotators (Anzer & Bauer, 2021; H. Liu et al., 2013; McKinley, 2019).

To standardize data from various *event data providers*, the universal event language SPADL has been introduced by Decroos et al. (2019). This language supports post-hoc unification of events from different *event data providers*, integrating them to joined event classes. While their approach is an important step toward unifying *event data* in soccer, it inevitably overlooks specific information unique to a single *event data provider*'s annotations. Additionally, it cannot resolve differences in event definitions across different catalogs.

In this regard, two research gaps were previously identified. RG 1 discusses how comparing event annotations from different *event data providers* can be challenging, due to unclear inter-annotator agreement and validity of different event classes. Furthermore, RG 2 explains the previously reported temporal misalignment of *event data* with *position* and *video data* (Anzer & Bauer, 2021). To address these research gaps, the upcoming chapter validates a unified, hierarchical taxonomy for events in *invasion games* using a multi-expert set of *atomic event* annotations (Chapter 3.2). Furthermore, a temporal refinement algorithm for *event data* is developed and evaluated using a set of precisely annotated pass events (Chapter 3.3).

3.2 Unified Taxonomy for Events in Invasion Games

This chapter outlines the design of a unified taxonomy for event annotation. First, prior research that underscores the lack of comprehensive definitions for player-ball interactions is reviewed (Chapter 3.2.1). Then, insights from the related work are used to develop a hierarchical taxonomy for events applicable to *invasion games* (Chapter 3.2.2). Finally RQ 1 is addressed (Chapter 3.2.3).

3.2.1 Previous Research

While Dodge et al. (2008) made initial efforts to define overarching concepts of player movement patterns, specific developments in *invasion games*, such as player-ball interactions, remain less explored. For instance, although the concept of a rebound has been thoroughly analyzed in sports like handball and basketball (Burger et al., 2013; Evangelos & Nikolaos, 2004), a similar analytical approach for soccer was only recently proposed by Litwitz et al. (2024).

The computer vision community has explored various types of events in soccer (Naik et al., 2022). Deliege et al. (2021) introduced soccer event categories with definitions encompassing meta, on-ball, and referee-based events. While they achieved high accuracy in automatically detecting these events from *video data*, their focus on fundamental characteristics of soccer broadcast videos (such as substitutions, ball out-of-bounds, yellow cards) limits the number of event classes related to player-ball interactions. In contrast, the openly available dataset of manually collected soccer events published by Pappalardo et al. (2019) places a greater emphasis on such. However, despite the inclusion of these event classes, the authors did not provide detailed event definitions and did not report inter-annotator agreement.

Decroos et al. (2019) have introduced a streamlined language for describing *event data* from different *event data providers* in a unified format. The format integrates a multitude of definitions from the underlying event catalogs of different *event data providers* into shared classes. However, to our knowledge, no validation study on inter-annotator agreement for these event definitions has been conducted. Similarly, the FIFA has proposed a unified language with guidelines on annotating different sections of a soccer match (FIFA, 2022), but we are not aware of an annotation study validating these definitions.

Recognizing the importance of standardizing domain-specific concepts for soccer annotation, Fernandes et al. (2019) introduced an observational instrument for defensive *sequences*, achieving sufficient inter-annotator agreement between experts. Yet, the annotation scheme is limited towards the frame-wise description of *atomic events*. J. Kim et al. (2019) proposed a taxonomy for analyzing offensive *sequences* but did not conduct inter-annotator validation. To our knowledge, no unified annotation scheme for *atomic events* in soccer has been proposed.

3.2.2 Design of Taxonomy for Atomic Events

To address the shortcomings in practical applications and theoretical literature (see RG 1) a unified, hierarchical taxonomy for *atomic events* was proposed in Article I. This taxonomy was designed to facilitate event annotations for *invasion games*, with a specific focus on soccer. In the design process of the taxonomy, the criteria for annotation schemes (see Chapter 2.2.1) were followed and insights from discussions with domain experts were incorporated. Three categories of taxonomy classes were defined, all based on fundamental properties common to *invasion games*.

First, the concept of *activeness* was addressed. *Activeness* is defined by the rules (IFAB, 2024) and refers to actions governed by the referee, such as interruptions, resumptions of play, ball repositioning, and sanctions for players. Events that result in changes to *activeness* were summarized in the highest level of the taxonomy, with additional contextual information added at lower hierarchical levels.

Second, the concept of *ball possession* was examined. This concept refers to the assignment of the ball to a player or team. Traditionally, *ball possession* has been attributed to one of the two teams at all times; however, a third category, undefined *possession* (under contest), has recently been introduced (FIFA, 2022). Yet, as this assignment can be challenging, it was found that common convention is to maintain the *ball possession* assignment of the last player until a new player can be clearly identified.

Thus, all events that trigger a *turnover* (change in *ball possession*), were encapsulated under the second category of the taxonomy.

Third, the individual interactions of players with the ball were studied. It was was found that these interactions are strongly intertwined with the concept of *ball possession*, as they typically initiate or terminate an *individual ball possession* phase (Link & Hoernig, 2017). Therefore, the technical skill at the beginning of an individual *ball possession* was defined as a *ball reception* while the end of an individual *ball possession* was defined as a *ball release*. At the lower levels of the hierarchy, it was further differentiated between the underlying cause (intentional or unintentional) and specific movements and intentions (e.g., shot or pass). All of these events were compiled in the third category of the taxonomy.

For a visualization of the taxonomy structure, refer to Figure 2 in Article I. Detailed definitions of the individual taxonomy classes can be found in the Appendix of Article I.

3.2.3 Answer to Research Question 1 (RQ 1)

Article I performed a validation study involving the annotation of taxonomy events for handball and soccer by nine annotators, where experts and inexperienced annotators were consulted. For each event class in the taxonomy, both, the agreement of the specific event type and the temporal alignment were measured.

For the event classes concerning *activeness* and *ball possession*, a byproduct of the *atomic event* annotation were signals with a duration (e.g., *atomic events* for *activeness* resulted in phases of active and inactive match). For these signals, annotator agreement was measured as the ratio of temporal intersection of two annotators over the total temporal union of their annotations. The results confirmed the validity of the proposed taxonomy for game status event. For *activeness*, a high average of 0.92 was achieved. For *ball possession*, the intersection over union decreased to 0.78, confirming a general validity, but leaving some room for future improvement of the *ball possession* definition. One possibility for such an extension would be adding an undefined *ball possession* option, as proposed by the FIFA (2022). Notably, the strong agreement values held for the comparison of experienced and inexperienced annotators, demonstrating the flexibility and clarity of the involved definitions.

For *atomic events* annotation, inter-annotator agreement was measured using the number of cases where two annotators annotated the same event type within 0.44 seconds. Here, for the *ball release* event, situated at the top of the hierarchy, a strong annotator agreement of 93% was found. However, for more *semantically complex* events at lower hierarchy levels, the annotator agreement declined. For example, the agreement for the *intercepted pass* drastically decreased to 45%. This finding emphasized the importance of using a hierarchical annotation scheme and supported the placement of events with varying *semantic complexity* at different levels within the hierarchy.

Another finding of Article I was that successful events generally showed stronger agreement compared to unsuccessful defensive events. For example, for the *unintentional release* event, even though the tolerance window was increased to 2.04 seconds, the annotator agreement was only at 18%. This highlighted the need for clearer definitions of unsuccessful defensive components in soccer event annotation schemes.

In response to RQ 1, the proposed unified annotation scheme, which generalizes across *invasion games*, offers a methodological way to create rich event annotations in soccer. The multi-expert validation of event classes and the use of a hierarchical structure to capture *semantic complexity* (see Chapter 1.2) were particularly important components of the proposed annotation scheme. Essentially, the latter provided the possibility to integrate diverse event annotations into a shared parent class which allows users to account for expert disagreement in individual events. Thus, when adhering to the guidelines for annotation schemes in soccer (Chapter 2.2.1), it is generally possible to create *atomic event* annotations that are reliable, valid, contextually rich, and granular.

3.3 Algorithm for Synchronizing Event and Position Data

This chapter introduces a synchronization algorithm for *event* and *position data*, designed to address the temporal misalignment identified in RG 2. First, existing literature on synchronization and event detection with a specific focus on soccer is reviewed (Chapter 3.3.1). Next, a time series event detection algorithm is refined for synchronizing pass events in soccer (Chapter 3.3.2). Finally, it is discussed how the proposed algorithm improves the synchronization between *event* and *position data* and RQ 2 is addressed (Chapter 3.3.3).

3.3.1 Previous Research

Time-series event detection involves identifying distinct patterns within time-series data. Techniques for detecting these events range from classical statistical methods to modern machine learning approaches. Notably, Makridakis et al. (2018) have suggested that classical statistical methods often outperform machine learning algorithms for this task. One example of a classical approach is the *SwiftEvent* algorithm, proposed by Gensler and Sick (2018) and introduced in Chapter 2.3.1.

In soccer research, time-series event detection was frequently employed, although most studies focused on *video data* (Deliege et al., 2021; Fakhar et al., 2019), audio data (Sanabria et al., 2019), or social network data (Wunderlich & Memmert, 2020). Thus, these approaches were primarily proposed for automatic summarizations of soccer matches, such as creating highlight reels, rather than for fine-grained event detection that could assist in data synchronization.

Sanford et al. (2020) applied a method for time-series event detection to player-ball interactions. The authors focused on the fine-grained detection of passes, shots, and receptions with *deep neural networks* using *position* and *video data*. Their results showed high F1-scores of 0.7 to 0.8 for the examined event types, however, their study included several systematic issues. First, the annotations from *event data providers* used for training and testing were not validated for temporal accuracy or semantic validity. As a result, neither inter-annotator agreement nor consistency in event definitions was addressed. This limitation affected the reliability of their findings, as inaccurate event annotations possibly hindered consistent and accurate assessments of fine-grained event detection. Second, the authors used an evaluation metric that groups detected passes with their nearest ground truth counterpart, which may have led to a biased accuracy since the potential impact of many-to-one mappings was not accounted for.

A more tailored approach to event synchronization was presented by Anzer and Bauer (2021), who proposed a refinement step that uses player-ball distances and ball acceleration to adjust the timestamps of shots the *event data*. Their refinement significantly improved the temporal alignment of *position* and *event data*. However, since this synchronization step was part of a broader analysis, the authors did not provide detailed insights into the underlying design choices and limitations. Additionally, an extension of their algorithm to event types (such as passes) was not performed.

Recently, Van Roy et al. (2023) have proposed a rule-based approach, called ETSY, for synchronizing event and tracking data. They extended the previous work by Anzer and Bauer (2021) to enable the synchronization of various event types. Therefore, they proposed to ingest both *position data* and locations contained in the *event data* to compute a likelihood function around the timestamp of the event. Thus, they assigned the annotated event to the timestamp of the maximum likelihood value. This publicly available approach has been highly valuable to the community, as it was empirically shown to decrease the misalignment of the *event* to the *position data*. Yet, the hand-crafted rules of ETSY for computing the likelihood function have not been compared to algorithmic approaches. Furthermore, it requires the location of the event, which may not always be available in the *event data*.

3.3.2 Pass Event Refinement with SwiftEvent

Article II focused on the temporal misalignment between *event data* and *position data*, summarized in RG 2. Unlike previous efforts (Anzer & Bauer, 2021), this study specifically targeted pass events in soccer, chosen for their high frequency and relatively clear definition (Aquino et al., 2017).

As previously reported by Anzer and Bauer (2021), the independent collection of *position* and *event data*, with each system operating on its own clock, leads to a systematic offset between the two data sources. In addition to this offset, a human error in the *event data* was found to further contribute to misalignment, causing discrepancies between the recorded event (e.g., a pass) and the corresponding event in the *position data* (e.g., the frame where the ball leaves the player's foot).

Since the systematic offset can be easily accounted for (Anzer & Bauer, 2021), Article II focused on correcting the human error in the data by applying a refinement algorithm for passes. The algorithm modified *SwiftEvent* by Gensler and Sick (2018). Thus, player-ball distances and ball acceleration were computed from the *position data*, transformed into a low-dimensional feature space, and used to train the pass event detection. For ground truth, a manual set of precisely annotated pass events was used.

To refine pass events in a frame-wise manner, the probabilities from *SwiftEvent* were combined with a search window around the event timestamp. Thus, the pass event was assigned to the timestamp with the maximum pass probability in the search window, similar to the approach of Van Roy et al. (2023).

3.3.3 Answer to Research Question 2 (RQ 2)

The results of Article II demonstrated the benefits of refining pass annotations from *event data providers* using *SwiftEvent*. In the examined dataset, the temporal misalignment of passes between *event* and *position data* was lowered from 2.34 seconds to 1.4 seconds. Furthermore, the ratio of passes with a relatively small misalignment of less half a second increased from 14% to 70%.

In a comparative analysis *SwiftEvent* outperformed rule-based and *logistic regression*-based refinement algorithms. The overall validity of the results was reinforced by a matching logic that ensures one-to-one mappings between ground truth and refined passes.

However, despite the significant reduction of provider-based temporal errors, *SwiftEvent* was unable to reduce the average temporal error to below one second. This limitation suggested that perfect synchronization between *event* and *position data* cannot be achieved using *SwiftEvent*. Beneath shortcoming of the algorithm, this can be attributed to noise in the data, where passes were sometimes completely absent from the *event data* or where the *position data* tended to show unrealistic spikes in some occasions.

Yet, *SwiftEvent* proved valuable for cases where a pass in the *event data* did not align with a corresponding behavior in the *position data*. This unrealistic behavior was successfully detected using *SwiftEvent* with a threshold for outcome probability of the refined pass. Results of this outlier detection showed that for certain thresholds, the average player-ball distance of the excluded events at the moment of the pass was as high as 25 meters.

In response to RQ 2, synchronization of *event* and *position data* can be significantly enhanced using an automatic event detection algorithm. However, challenges remain especially for less recent datasets (e.g. from the 2014/15 season) where the included noise and the non-available event location complicate an accurate synchonization. Thus, a fair comparison of *SwiftEvent* with related algorithms using this event location (Anzer & Bauer, 2021; Van Roy et al., 2023) is not possible. In that sense, applying *SwiftEvent* to newer datasets is an interesting perspective for future work.

To conclude, for cases where precise, frame-exact synchronization are required, manual inspection of individual cases is still necessary. When using automatic event detection methods, it is crucial to thoroughly evaluate their performance and eliminate any potential many-to-one mappings that could bias the evaluation metrics.

Chapter 4

Sequence Annotation of Soccer Matches

This chapter turns from the previously evaluated *atomic events* (Chapter 3) to the challenge of generating *semantically complex* annotations for *sequences* in soccer matches, as highlighted in RG 3. Thus, it summarizes the content of Article III.

First, Chapter 4.1 presents the concept of *tactical periodization* and discusses its potential as a blueprint for *sequence* annotations.

Next, Chapter 4.2 outlines the process of translating the practical concept of *tactical periodization* to a quantifiable, frame-by-frame scheme, aimed at addressing RQ 3A.

Chapter 4.3 explores how this annotation scheme enables the creation of an automatic annotation model for *sequences*, providing insights into RQ 3B.

Finally, Chapter 4.4 presents a use-case for the automatic annotation model and addresses RQ 3C.

4.1 Tactical Periodization in Practice

One of the previously discussed differences between soccer and US sports is soccer's general lack of structured play. In US sports like basketball and American football, clear *ball possessions*, distinct offensive and defensive phases, and enforced dynamics through time and space restrictions are present, whereas soccer lacks similar structuring rules.

This absence of clear divisions in play creates challenges for *sports analytics* in soccer, as it complicates the evaluation and comparison of recurring sub-components of a match. Moreover, tactical concepts in soccer are often tied to situation-specific player behavior, requiring a reliable method for accurately identifying these situations.

Practitioners have addressed the lack of structure in soccer through the concept of *tactical periodization*, a method that segments matches into distinct *moments of play*. This approach aids in contextualizing match analyses and formulating clear, situation-specific guidelines for players (Hewitt et al., 2016). Therefore, the fundamental concept of *ball possession* has been used to structure a match into five distinct *moments of* *play*: *established offense*, *established defense*, offensive *transition*, defensive *transition* (which both occur shortly after a change of *ball possession*), and *set piece*. These *moments of play* logically reflect the flow of the match, as changes from *established offense* to *established defense* and vice versa must always pass through the respective *transition* phases.

While the first four *moments of play* are collectively referred to as *open play*, the fifth moment, *set piece*, originates from game restarts such as free kicks, corners, or throw-ins (Sousa & Garganta, 2001). This category is unique because it can precede and follow any of the other four *moments of play* and lead back to them. Moreover, due to their potential for creating scoring opportunities and distinct positional arrangements, *set pieces* enjoy a strong strategical importance (Sousa & Garganta, 2001).

However, based on previous findings of decreasing expert agreement with increasing *semantic complexity* (discussed in Chapter 3.2), it was expected to encounter similar challenges with the *semantically complex* concept of periodization. Thus, reviewed existing research was carefully reviewed with the aim of translating *tactical periodization* into a reliable, quantified annotation scheme.

4.2 Formalization of Tactical Periodization

Building on the previously established annotation scheme for *atomic events* (Chapter 3.2), this chapter follows a similar approach to develop an annotation model for *tactical periodization*. First, related work is reviewed (Chapter 4.2.1). Then, a hierarchical annotation scheme is derived (Chapter 4.2.2) and RQ 3A is addressed (Chapter 4.2.3).

4.2.1 Previous Research

Various studies have analyzed the tactical implications of play across the five *moments of play* using manual annotations or correlations with match statistics (Hewitt et al., 2016). Notably, Kempe et al. (2014) provided a dichotomous description of offensive play, introducing a metric that categorizes an offensive *ball possession sequence* into two contrasting styles: *direct play* and *possession play*. The authors characterized *direct play* as a quick, vertical playing style with few passes per *sequence*, and *possession play* as slower, longer periods of play with more passes and greater ball retention. While the proposed metric was successfully validated through multi-expert inter-annotator agreements, the original study only assigned one label per *sequence*, thus limiting the possibility for more fine-grained analyses.

Regarding defensive *moments of play*, the concept of *team pressure* (varying levels of urgency with which a team attempts to regain ball possession) has been examined by various studies. Andrienko et al. (2017) developed a computational approach to quantify individual *team pressure* based on player velocities and inter-player distances. However, the authors did not present an expert validation study to assess the algorithm's suitability, particularly concerning team-wide group dynamics and behavioral patterns associated with *team pressure* (Forcher et al., 2022). Conversely, Bauer and Anzer (2021) introduced the notion of *team pressure* specifically during *counterpressing*. They formalized their concept as an effort by a team to regain *ball possession* immediately after losing the ball. The used *counterpressing* definition was validated based manual annotations with high pairwise accuracy between two domain experts, confirming its reliability. Yet, they did not extend their approach towards a more holistic description of defensive behavior.

In contrast, a systematic framework for the holistic annotation of defensive team behavior was proposed by Fernandes et al. (2019). The authors defined a variety of fine-grained observational classes for defensive behavior and distinguished between defenders controlling space, delaying the attack, forcing the opponent direction, and reducing space. However, while their approach was tailored towards the notational analysis of individual players, the required degree of detail interferes with the broader notion annotating team behavior, leaving the applicability of the framework for *tactical periodization* uncertain.

4.2.2 Design of Annotation Scheme for Match Phases

To address RG 3 regarding the absence of frame-wise *sequence* annotations for *semantically complex* contextual concepts, Article III proposed a hierarchical, frame-wise annotation scheme for *tactical periodization*. Similar to the previously developed taxonomy for *atomic events* (see Chapter 3.2) this scheme followed the criteria, outlined in Chapter 2.2.1. The scheme used a hierarchical structure, progressively refining class definitions through step-wise extensions in *semantic complexity*. The individual classes of the annotation scheme for *tactical periodization* were referred to as *match phases*.

The annotation scheme contained six distinct *match phases*, defined for active parts of the match. Furthermore, it was designed as a succession of pair-wise hierarchical binary decisions. The binary decisions included *open play* vs. *set pieces*, as well as the assignment of *ball possession* (offense vs. defense) for both of these *match phases*. Finally, for *open play* it was further distinguished between *possession play* vs. *direct play* for *offensive open play*, and, respectively, *team pressure* vs. *no team pressure* for *defensive open play*. For further details on the individual definitions used within the annotation scheme, refer to Article III, Chapter 4.3. For a visualization of the annotation scheme, see Article III, Figure 2.

4.2.3 Answer to Research Question 3A (RQ 3A)

The results of the three-expert annotation study in Article III were measured for each of the step-wise binary decisions using Cohen's kappa scores (Cohen, 1960).

The lowest agreement of 0.66 was measured for *offensive set piece* vs. *defensive set piece*, which still corresponded to a substantial agreement according to Cohen (1960). Further analysis of this case showed that the low expert agreement was caused due to disagreement about the precise endpoint of a *set piece*, where two out of three annotators frequently shifted to *open play* while one remained with *set piece*.

For the remaining decisions, higher Cohen's kappa scores were obtained. The expert's annotated *open play* vs. *set piece* with an almost perfect agreement of 0.85. Similarly, almost perfect agreement of 0.89 was achieved for the distinction of *offensive open play* vs. *defensive open play*. For the ultimate decisions of *possession* vs. *direct* *play* and *team pressure* vs. *no team pressure*, the results showed substantial agreement with kappa scores of 0.78.

Overall, the inter-annotator agreement confirmed the general validity of the annotation scheme and the *match phase* definitions. The decreasing expert agreement when moving down the hierarchy (e.g., from *open play* vs. *set piece* to the concepts of *possession play*, *direct play*, and *team pressure*), validated the step-wise hierarchical structure used to manage *semantic complexity*. Moreover, the variance in the granular expert agreement supported the approach of disentangling each match phase annotation as a series of sequential binary decisions.

Interestingly, the findings from Article I regarding decreased expert agreement with increasing *semantic complexity* were replicated in Article III, further highlighting the benefits of a general hierarchical structure for soccer annotation schemes.

To answer RQ 3A, the implemented annotation scheme for *tactical periodization* facilitates the creation of *semantically complex sequence* annotations with substantial to perfect inter-expert agreement. This demonstrates the feasibility of applying the general criteria for soccer annotation schemes, outlined in Chapter 2.2.1, for developing *semantically complex* conceptualizations.

On the contrary, the expert disagreement about *set pieces* highlighted the need to refine their temporal end point definitions to improve clarity and reliability. Moreover, the annotation scheme did not provide a *holistic* description (see Chapter 2.2.1) of team-tactical behavior. Commonly used concepts like *chance creation*, *build-up play*, *counterattacks*, and *transitions* (Seidenschwarz et al., 2020) may have not been adequately represented within the scheme. Yet, given the already present decrease in agreement, adding such concepts to the annotation scheme could potentially reduce its validity. Therefore, developing *sequence* annotation schemes for *semantically complex* concepts require methodical formalization and careful assessment of the introduced complexity. This process should include conducting multi-expert validation experiments to ensure the robustness and reliability of the annotation scheme.

4.3 Automatic Detection of Match Phases

The upcoming chapter explores the implementation of automatic *sequence* annotation models. First, previous research on detecting individual *match phases* is introduced (Chapter 4.3.1). Next, automatic models for *match phase* prediction are proposed (Chapter 4.3.2). Finally, the performances of these models are compared and RQ 3B is addressed (Chapter 4.3.3).

4.3.1 Previous Research

Although many studies have focused on the broad detection of specific team behavior, the frame-wise automatic detection of *tactical periodization* in soccer has less frequently been explored.

The concept of *activeness* was successfully automated by Lang et al. (2022), who trained an AdaBoost model (Freund & Schapire, 1995) using manually annotated *activeness* annotations and *position data*. The authors reported high frame-wise accuracy, effectively automating the prediction of game status using *position data*.

Regarding the concept of *ball possession*, Link and Hoernig (2017) developed an unsupervised method to detect individual player *ball possession* based on proximity to the ball and ball acceleration. Although the authors reported high agreement of the algorithm's output with single annotator assessments, it did not undergo a multi-expert validation, leaving its practical performance uncertain.

H. Kim et al. (2022) leveraged the drastic formation changes of teams during *set pieces* to automatically detect them. They identified change points in role-adjacency matrices to detect situations where players' spatial arrangements and movements significantly deviated from typical play. Yet, the authors only qualitatively evaluated their approach while a detailed multi-expert frame-wise validation was not reported.

Bauer and Anzer (2021) detected *counterpressing* with a XGBoost model (Chen & Guestrin, 2016) using on *position data* and validated manual annotations. They reported a high model accuracy, confirming the model's effectiveness in accurately detecting *counterpressing*.

However, despite these advancements in detecting individual components of *tactical periodization*, no comprehensive model combining these approaches into a unified system for periodization has yet been proposed.

4.3.2 Models for Automatic Detection

Article III used multi-expert *sequence* annotations from the previously introduced annotation scheme for *tactical periodization* to explore the automatic, frame-wise detection of *match phases*. As noted by Raabe et al. (2023), when using *position data*, comparing various data representations is essential, particularly in fields where prior research is limited. In Article III, three data representations were compared:

1. *Position representation*: Composed by stacking the normalized xy-coordinates of the player- and ball-positions into vectors.

2. *Feature representation*: Contained common features from the literature associated with *match phases*. (i) Defensive components metrics such as player-player distance, player-ball distance, and player velocity, to capture *team pressure* and assist with *ball possession* allocation (Andrienko et al., 2017; Link & Hoernig, 2017). (ii) Offensive component metrics such as the number of players positioned ahead of the ball to catpture *direct play* (Kempe et al., 2014).

3. *Combined representation*: Integrated the features from the *feature representation* with a condensed description of team centroids and variances (in x- and y-direction). Used to assist the detection of positional changes during *set pieces* (H. Kim et al., 2022).

The step-wise binary structure of the periodization annotation scheme (Chapter 4.2) guided the design of a multi-level architecture for each model, composed of multiple binary classifiers. This layered approach broke down the complex decision-making process for detecting *match phases* into individual binary decisions, supporting both the model's interpretability and flexibility.

For all binary decisions, different automatic models with various complexity were compared to determine which aspects of periodization require complex algorithms and which can be addressed with simpler methods. A *logistic regression*, a random forest classifier (Ho, 1995), and a rule-based model were employed as baselines. A state-of-the-art graph *neural network* (Raabe et al., 2023) and a custom recurrent *neural network* architecture based on a Gated Recurrent Unit (GRU) (Cho et al., 2014) represented more complex models. For more detailed information on the architecture and implementation of these models, as well as for more details on the used data representations, refer to Article III.

4.3.3 Answer to Research Question 3B (RQ 3B)

To validate automatic models for *match phase detection* in Article III, the alignment of model predictions with the merged expert annotation was computed. Therefore, the three expert annotations from the validation study (see Chapter 4.2.3) were combined into a single merged expert annotation using a frame-wise democratic decision, selecting the *match phase* chosen by the majority of experts. Thus, accuracy, precision, and recall (see Chapter 2.3.2) were computed using the merged annotation as ground truth.

The comparison of multiple automatic detection models and data representations revealed that the highest F1-score for all binary decisions was obtained by the GRU model, using the data in the *combined representation* as input. This aligned with the expectation that the *position data* contains valuable contextual information (e.g. for *set pieces*), while the feature calculation provided valuable granular, domain-specific information.

Due to the binary, hierarchical structure of the annotation scheme, variations in the performance of the GRU model between different binary decisions could be observed. The results from this granular comparison revealed that the highest F1-scores, around 0.8, were achieved for *open play*, *direct play*, and *team pressure*, while lower F1-scores, around 0.75, were recorded for *no team pressure* and *offensive set piece*. The lowest F1-score was encountered for *defensive set piece*. Interestingly, the low automatic performance for *set pieces* aligned with the previously identified lack of expert agreement for this particular *match phase*. This finding pointed to a fundamental lack of expert agreement hindering the successful implementation of automatic annotation models.

The individual binary decisions of the GRU model were combined to create a holistic frame-wise *match phase* annotation. This annotation achieved 80% accuracy (on unseen data), which showed that the model generally succeeds in replicating domain experts' annotations with high accuracy.

To address RQ 3B, the results indicate that the automatic model can replicate expert opinions on the *semantically complex* concept of *tactical periodization* accurately. This suggests that the time-intensive and potentially subjective manual annotation process in soccer can be complemented by deploying fast and objective automatic models. However, the merging of expert annotations inherently introduces some ambiguity, making perfect alignment of the automatic annotations with the combined expert annotations challenging. This highlights the complexity and subjectivity involved in manually annotating *semantically complex* concepts, making expert subjectivity an interesting topic for future research.

4.4 Application of Automatic Match Phase Detection

In this chapter, the previously created automatic annotations for *tactical periodization* are applied in a real-world scenario. This scenario demonstrates how automatic models can support match analysts in practical use-cases.

4.4.1 Team Average Positions

Understanding team performance is a critical aspect of soccer analysis, as discussed in Chapter 2.1.1. One common approach to analyzing team dynamics is through the use of average team formations (DFL, 2021).

Average team formations have typically been calculated in combination with the post-hoc role-assignment approach proposed by Bialkowski et al. (2014). This algorithm tracks role changes among players during the game. Thus, the average formation of a team is independent of player identities as it is computed based on the roles that players fulfill during a match.

However, while averaging formations over an entire match provides a generalized view, it might obscure important tactical details. For example, teams often adjust their tactical behavior depending on whether they have *ball possession*. Consequently, these generalized average formations may not capture the nuanced shifts in strategy that potentially occur during different phases of the game.

4.4.2 Answer to Research Question 3C (RQ 3C)

The schematized use-case, presented in Article III, Chapter 6, compared different average team formations. One formation was computed for an entire half of a soccer match, while a respective formation was computed for every of the four *match phases*, based on automatic frame-wise annotations of the GRU (see Chapter 4.3.3).

The results demonstrated how automatic *match phase* annotations can be used to generate formations for specific *match phases*, effectively visualizing a team's positional shifts during the match. This application significantly enhanced the traditional computation of average formations by offering a higher level of contextualization. Moreover it allowed users to obtain a more detailed and dynamic understanding of team strategies and adjustments throughout the match.

In response to RQ 3C, automatic models for *semantically complex* concepts can complement traditional tasks performed by soccer analysts. The contextualization provided by these models has the potential to reduce both the workload and the level of subjectivity typically associated with manual video annotations. As described in Chapter 2.1.1, this automation enables analysts to focus on more detailed strategic aspects. However, it is important to note that the automatic *match phase annotations* cannot replicate domain experts' opinions with frame-exact precision, as a certain degree of uncertainty remains. Therefore, these automatic models are not suitable for *sequence*-based analysis tasks that require precise, frame-wise annotations.

Chapter 5

Sequence Analysis of Soccer Matches

After the implementation of *sequence* annotations (Chapter 4), this chapter focuses on *sequence*-specific analyses, specifically on *transitions* between *offensive open play* and *defensive open play*. As a whole, this chapter presents the content of Articles IV and V.

Chapter 5.1 first explores the tactical significance of *transitions*, showing how they impact the dynamics and strategy of a match.

Chapter 5.2 discusses methodologies for identifying and evaluating *transitions*, addressing RQ 4A.

Finally, Chapter 5.3 presents a statistical analysis of *transitions* and their practical application, offering insights to RQ 4B.

5.1 Tactical Importance of Transitions

The *tactical periodization* model proposed by Hewitt et al. (2016), as discussed in Chapter 4.2, divides a soccer match into five distinct *moments of play*. Two critical moments are the *transitions* between offense and defense. According to Hewitt et al. (2016), a *turnover* (one team loses *ball possession* to the other team) always triggers a *transition*, which must occur before a team can establish *ball possession*. As a result, *transitions* are a crucial aspect of the match and are often a key focus in practitioners' strategic preparations (Bauer & Anzer, 2021).

Transitions in soccer are typically categorized into two types: defensive *transitions*, which occur after losing the *ball possession* after a *turnover*, and offensive *transitions*, which happen after gaining *ball possession* after a *turnover*.

For defensive *transitions*, the concept of *counterpressing* is well-documented (Bauer & Anzer, 2021). It involves an immediate attempt to regain *ball possession* after committing a *turnover*, with the dual goal of preventing an opponent's attack and potentially creating new attacking opportunities. Effective *counterpressing* behavior requires defensive players to quickly converge on the ball carrier at high speed (Andrienko et al., 2017; Bauer & Anzer, 2021), while remaining players are strategically positioned to cut off passing options (Bauer & Anzer, 2021; Forcher et al., 2022).

In contrast, offensive *transitions* are often characterized by *counterattacking*. This strategy is defined as a rapid, forward-moving attack aimed at the opposing goal (Tenga et al., 2009). Usually, this attacking style involves few direct vertical passes (Lago-Ballesteros et al., 2012), and targets areas near the opposing goal (Cooper, 2023). *Counterattacks* benefit from the opposing team's temporary disorganization immediately after losing *ball possession*, which creates space for the countering team (Tenga et al., 2009).

However, a key challenge in analyzing *transitions* is the lack of a clear endpoint. While the start of a *transition* is clearly marked by a *turnover*, there is no established end for these phases. This ambiguity makes it difficult to determine which subsequent events should fall under team-tactical concepts like *counterpressing* or *counterattacking*. As a result, analyzing and classifying these situations can become inconsistent and complex. This issue and its implications for accurate analyses are further examined in Chapter 5.2.

5.2 Isolation & Rating of Transitions

To perform situation-specific analyses effectively, it is essential to have a clear definition of the examined situations. This chapter focuses on isolating and rating *transitions* in soccer. Existing literature on the analysis of *transitions* is reviewed in Chapter 5.2.1. A novel metric to locate and rate *transitions* is designed in Chapter 5.2.2. The insights gained from this process are then employed to address RQ 3A (Chapter 5.2.3).

5.2.1 Previous Research

While *transitions* are a widely discussed topic in soccer, concrete studies that isolate *transitions* in a match are rather rare.

To classify defensive *transitions* as *counterpressing*, Bauer and Anzer (2021) used a five-second rule, commonly referenced by practitioners. According to this rule, a team successfully performs *counterpressing* if they regain *ball possession* within five seconds after committing a *turnover*. Additionally, the authors also included shots and goals occurring within 20 seconds of a *turnover* as part of the preceding *counterpressing* effort.

Regarding offensive *transitions*, Raudonius and Allmendinger (2021) developed an automatic mechanism to detect *counterattacks* based on specific events that occur after a *turnover*. Similarly, Sahasrabudhe and Bekkers (2023) identified *counterattacks* by analyzing *sequences* in which the ball advanced towards the opposing goal at a minimum speed of five meters per second. Building on these approaches, the upcoming Chapter introduces a novel rule-based metric to identify and evaluate *turnovers* based on their potential to lead to a *counterattack*.

5.2.2 Definition of Rule-based Criteria

From a tactical perspective, *turnovers* vary significantly in their potential for a successful *counterattack*, depending on factors such as pitch location and team positioning.

To quantify this variability, Articles IV and V proposed the *expected counter* (*xCounter*). This metric captures the expected potential of a *counterattack* given the situation at the time of a *turnover*. Similar to the concept of an *expected goals*, which quantifies the probability of scoring at the moment of a shot (Anzer & Bauer, 2021), *xCounter* was designed to assess and rate the potential effectiveness of *counterattacks* based on specific game situations. As a prerequisite for the definition of *xCounter*, the first two steps of the previously introduced framework for analyzing complex *sequences* (see Chapter 2.2.2) were followed.

F1. Identify Sequences of Interest

Since *counterattacks* occur during offensive *transitions*, they begin at the time of a *turnover*. However, defining the end point of a *counterattack* was more challenging. Lago-Ballesteros et al. (2012) proposed that the goal of a *counterattack* is for the ball-winning team to advance into zones near the opponent's goal. According to this definition, a location-dependent temporal endpoint of a *counterattack* was proposed. Based on a ball progression velocity of five meters per second, previously proposed by Sahasrabudhe and Bekkers (2023), this endpoint was chosen as the time span it takes (at this constant speed) to reach the opponent's goal from the *turnover* location.

F2. Define Success Criteria

After defining the temporal boundaries for *counterattacks*, it was found that the relatively low frequency of scoring in soccer (discussed in Chapter 2.1) necessitates alternative success criteria beyond goals. Various metrics, such as *expected possession value* (Fernández et al., 2021), *expected goals* (Anzer & Bauer, 2021), and *expected threat* (Merhej et al., 2021), have been developed to value space in soccer. As a common feature of these metrics, an increased value of space with decreasing distance to the goal was identified. Therefore, a linear relationship between goal distance and value was proposed as a simple rating model for *sequence*-specific success criteria. For future experiments, this rating metric is easily extendable. Refer to Article V for more detail on the identification and rating step for *xCounter*.

5.2.3 Answer to Research Question 4A (RQ 4A)

The rule-based definitions for identification and rating were applied in Articles IV and V to identify *counterattacks* and to rate situations using *xCounter*.

By clearly defining start and end points, relevant *sequences* were isolated and assigned to a rating score based on the specific context. Additionally, *match phases* were incorporated as a filtering step to only inspect *open play turnovers*. This improved the comparability of individual situations, ensuring that *turnovers* were analyzed within a consistent strategic framework, thus making comparisons more meaningful.

Article IV also explored the use of a post-hoc filter to exclude chaotic situations characterized by frequent *turnovers*. However, this approach was revised in Article V, as it was found to interfere with the primary goal of analyzing the situation at the time *turnover* (in advance of a potential *counterattack*). This revision highlights the dynamic nature of research in this field and the need to carefully evaluate and adjust methodologies, especially regarding how certain analysis steps can introduce unintended dependencies.

However, since Articles IV and V did not include a validation step with domain experts, it remains uncertain if the identified situations align with expert interpretations of potential *counterattacks*. Although qualitative results from both articles indicated that *counterattacks* are among the situations with high *xCounter* values, it is possible that the dataset contains examples that experts might not agree with. As such, RQ 4A cannot be conclusively answered without a multi-expert validation study, similar to those conducted in Articles I and III.

5.3 Analysis & Prediction of Transitions

Building on the isolation and rating criteria for *counterattacks*, this chapter aims to generate data-driven insights into effective behaviors during *turnovers*. First, prior research on *counterattacks* is reviewed (Chapter 5.3.1). Next, a predictive *xCounter* model is developed using comprehensible features, assessed for their predictive capability (Chapter 5.3.2). Finally, it is demonstrated how the obtained insights can be applied practically to improve strategic decision-making in soccer, addressing RQ 4B (Chapter 5.3.3).

5.3.1 Previous Research

In their analysis of *counterpressing*, Bauer and Anzer (2021) developed an automatic model to identify situations where the defensive team employs *counterpressing*. They used manually crafted, validated annotations to train a XGBoost prediction model (Chen & Guestrin, 2016), which achieved a high accuracy in predicting defensive behavior during *transitions*. However, even though the authors report high accuracy, translating the predictive results into concrete, sport-specific terms or actionable guidelines remained challenging.

In offensive situations, the effectiveness of *counterattacks* were evaluated using manual annotations (Lago-Ballesteros et al., 2012) and descriptive statistics (Fernandez-Navarro et al., 2016), with varying results across studies. More recently, automated methods were used to analyze offensive *transitions* more dynamically. For example, Raudonius and Allmendinger (2021) developed an automated system to quantify each player's contribution during a *counterattack* using four different performance indicators. However, the applicability of their findings was limited by their relatively small dataset.

In a more recent approach, Sahasrabudhe and Bekkers (2023) used a graph *neural network* to analyze success factors during *counterattacks*. In their results, the authors reported that the vertical speed of the attack, along with the angles to the goal and the ball, are key determinants of success.

To this date, most automatic approaches have focused on post-hoc analysis of behavior during the *transition* phase, rather than assessing the situation at the time of the *turnover*. Addressing this gap could greatly improve tactical analyses by concentrating on the strategic opportunities available immediately after a *turnover*, offering more immediate and actionable insights for practitioners.

5.3.2 Feature Construction, Assessment & Prediction

As previously discussed in RG 4, interpretability is an important trait of automatic models for soccer analysis. To achieve this, an interpretable model was designed and a choice of relevant, comprehensible features was selected in Article V. Thus, the steps F3 to F5 for understanding complex *sequences* (see Chapter 2.2.2) were followed.

F3: Construct Comprehensible Features

An essential step of the framework involved selecting features that are relevant to understanding the dynamics of a *turnover*. This step was found to be critical, as it determines the investigated factors in the *counterattack* analysis.

Bauer and Anzer (2021) previously used systematic combination approach as an effective strategy for selecting features. This involved segmenting a team into logical subgroups based on their tactical roles or positions during a match (for example players near the ball). For every subgroup, different features (such as the number of players, player position, and compactness were computed. This approach was adopted in Articles IV and V.

For more detailed rationales and explanations on the choice of subgroups and metrics, refer to Articles IV and V, which discuss the methodologies and results of these feature selection processes in greater depth.

F4. Assess Predictive Capability of Features

The predictive capability of the previously selected features was evaluated with respect to the outcome of a *transitions* (following *turnovers*).

As essential finding of Articles IV and V was that in the evaluation of predictive capabilities for *counterattacks*, the correlation between goal-distance and *xCounter* value needs to be incorporated. Given the *xCounter* definition (see Chapter 5.2.2), *turnovers* close to the opponent's goal already possessed a high rating score. Thus, features linked to *turnover* position were found to accordingly have inherent predictive capabilities, regardless of their specific attributes.

To address this, Article V introduced a novel *position-agnostic* method for feature assessment, using a full-pitch comparison. The assessment procedure involved setting a threshold value for a specific feature, which categorizes all *turnovers* into two distinct groups. In a discretized version of the pitch (divided into one-meter by one-meter squares), *turnovers* falling below and above this threshold were assigned to specific squares, and the average evaluation score for each square was calculated.

This *position-agnostic* approach enabled a systematic comparison across the entire pitch, quantifying the predictive capability of each feature based on a set threshold, independent of specific field locations. This limited possible spatial biases in the resulting assessment results. For further details on the assessment process and strategies for mitigating small sample size issues, refer to Article V.

F5. Build a predictive model

Once the most predictive features were identified, they were fed to a predictive model aimed at predicting the outcomes of *transitions*. Given the exploratory nature of the *xCounter* analysis, employing a broad set of features was important to ensure generalizable insights and to capture a wide range of tactical scenarios and player interactions during *turnovers*.

In Articles IV and V, the insights gained from the assessment process were leveraged to develop an automatic model that predicted the *xCounter* score at the moment of the *turnover*. This process involved compiling the most predictive features into a long vector, which is then fed into a XGBoost model (Chen & Guestrin, 2016), successfully applied in previous soccer studies (Anzer & Bauer, 2021; Bauer & Anzer, 2021; Lang et al., 2022).

The model's performance was compared against feature- and location-based baselines, which treated their inputs independently and overlooked interaction effects. Thus, it was evaluated whether the XGBoost model integrates and analyzes feature interactions. For a detailed discussion on the model's performance, including metrics and comparative results, refer to Article V.

5.3.3 Answer to Research Question 4B (RQ 4B)

The feature assessment and model building steps were applied in Articles IV and V to learn about impactful factors for predicting *counterattacks* at the time of the *turnover*. Moreover, based on the feature assessment, a ranking of comprehensible features was created and concrete values for optimal values in the ranked features were identified.

Concerning defensive features (for the team losing *possession* after committing the *turnover*), the overall most predictive feature for preventing *counterattacks* was a small vertical compactness. Optimal values were 12 meters for vertical compactness of the whole team, and 10 meters for players around the ball. Additionally, if the ball was lost in the opposing third, it was important for players behind the ball to maintain a distance less than 20 meters to the ball. Finally, having less than four players positioned in front of the ball was beneficial for preventing *counterattacks*.

The results of a beneficial influence of small compactness aligned with the tactical notion of *counterpressing*, where a compact team unit aims to quickly regain *ball possession* after losing the ball (Bauer & Anzer, 2021). Yet, the indication that no more than three should participate in the attack in front of the ball introduced an additional component in the risk-reward evaluation of attacks.

Concerning offensive features (for the team gaining *possession* after the turnover), a large horizontal compactness was a crucial factor. Thus, optimal values of horizontal compactness were at least 15 meters for all players, 14 meters for players around the ball, and 16 meters for players in front of the ball. Furthermore, for winning *turnovers* close to the own goal or at the sidelines, it was beneficial to have at least three open passing options (players without opponents within a five-meter radius).

The findings for offensive features suggested that, in addition to their defensive responsibilities, players should also focus on anticipating ball wins and positioning themselves horizontally as open passing options, in case the team gains *possession*. This result could provide interesting directions and motivation for the future research regarding player anticipation in soccer (Gonçalves et al., 2015).

Overall, the results from the feature assessment in Articles IV and V demonstrate how the inspection of features, in combination with rating criteria, can be used to derive clear, actionable guidelines for optimal behavior during *turnovers*. Using this approach it is possible to rank features, provide concrete thresholds where the feature impact is the largest, and report pitch positions where the features are most influential. These novel insights present interesting directions for practical applications, such as optimizing team formations for build-up play, developing strategies for and against *team pressure*, or enabling favorable team behavior for initiating successful *counterattacks*.

However, the results of the predictive models for *xCounter* were less clear. Different configurations of XGBoost were able to outperform the baselines in each individual metric. However, no single model performed superior in all evaluation metrics. These results suggest a non-linear influence of the features on *xCounter* that was not sufficiently represented by the XGBoost model.

The best-performing XGBoost models (in each evaluation metric) used a relatively small number of features. This could indicate that the models struggle to effectively incorporate the additional information provided by the features beyond the already strong predictive capability of the *turnover* location. Thus, a more refined algorithm would be needed to effectively leverage the predictive potential of the features in conjunction with the location data.

To answer RQ 4B, the feature assessment step in the *xCounter* analysis offers several practical applications. It provided clear guidelines for player behavior in both offensive and defensive *turnover* scenarios, identifying optimal feature values, as well as the impact of these features and their most influential zones on the pitch. Additionally, the proposed predictive model could be used to assess player and team performance in *transition* situations served as an automatic search tool for finding exceptional positive or negative *transition* instances. These application examples show how situation-specific analyses lead to practical insights that can possibly inform (post-)match analyses, player scouting, or the design of training sessions.

Chapter 6

Conclusion

The concluding chapter of this dissertation summarizes the key findings (Chapter 6.1), discusses the gained insights with a focus on limitations and practical applications (Chapter 6.2), and offers perspectives for future research (Chapter 6.3).

6.1 Summary

The upcoming chapter recaps the content of Chapters 3, 4, and 5. Therefore, it presents different fields in which progress towards filling the previously presented research gaps (Chapter 1.2) was made.

6.1.1 Atmomic Event Annotations

Analyzing soccer *event data* from various sources is a common practice among clubs and researchers. However, in Chapter 3 of this dissertation, the critical need to validate both the semantic and temporal accuracy of such data was highlighted.

On a semantic level, the definitions used by *event data providers* were found to be diverse and complex, yet they often fail to account for the inherent differences in the *semantic complexity* of events (Chapter 3.1). To address this issue outlined in RG 1, a highly general, unified taxonomy was developed (Chapter 3.2). The taxonomy was based on the similarities across *invasion games* and introduced varying degrees of *semantic complexity* through a structured, multi-level hierarchy. Using this hierarchy, it was demonstrated how expert agreement generally tends to decrease as *semantic complexity* increases.

On a temporal level, it was shown that data from *event data providers* often exhibits inaccuracies, primarily due to human errors that are inherent in manual annotation. Moreover, it was found that discrepancy leads to a lack of synchronization between *event* and *position data*, summarized in RG 2. To address this issue, a synchronization algorithm for *event* and *position data* (Chapter 3.3) was proposed. In contrast to existing approaches (Anzer & Bauer, 2021; Van Roy et al., 2023), the proposed algorithm was based on time-series event detection, did not require hand-crafted rules, and did not require event locations in the data. However, for a perfect alignment, manual intervention was still required.

6.1.2 Sequence Annotations

Chapter 4 of this dissertation moved from *atomic events* to *sequences*, addressing the challenge of creating *sequence* annotations (see RG 3).

Therefore, the widely discussed concept of *tactical periodization* (Chapter 4.1) was formalized into a multi-level hierarchical annotation scheme (Chapter 4.2). The annotation scheme was successfully validated by a three-expert annotation study, where general agreement between three expert annotators was reached.

Subsequently, a merging step was implemented to consolidate the multiple expert opinions into a unified annotation. Using the merged annotation, the automatic detection of *tactical periodization* was approached (Chapter 4.3). Results showed that the automatic annotations aligned with the merged expert annotation with a high accuracy. Thus, a broad-grained detection of events was generally possible, however, a validation by human experts did remain necessary to ensure accuracy on a granular frame-by-frame level.

Finally, the practical applicability of the automatic detection model was demonstrated through the automation of a common soccer analysis task (Chapter 4.4). In this analysis, team formations were contextualized to extract valuable insights into the offensive and defensive tactics of the examined team. These results underlined the significant value of automatic models for *tactical periodization*, offering a fast, objective alternative to hand-crafted annotations.

6.1.3 Sequence Analysis

After examining the automation of *sequence* annotations, Chapter 5 of this dissertation focused on *sequence*-specific analyses, addressed in RG 4.

Further drawing from the concept of *tactical periodization*, the dynamics of *transitions* were investigated. Specifically, the widely discussed concept of a *counterattacks* after a *turnover* (see Chapter 5.1) was discussed.

In Chapter 5.2, *counterattacks* were isolated and rated by measuring forward progress of the team gaining *possession* after a *turnover*. Using this metric, the concept of *expected counter* (*xCounter*) was proposed in analogy to the well-established concept of *expected goals* (Anzer & Bauer, 2021).

Given a dataset of *turnovers* with *xCounter* ratings, an analysis of various features describing team tactical behavior was conducted (Chapter 5.3). Introducing a *position-agnostic* systematic assessment procedure, various hand-crafted features were ranked according to their predictive capability for *counterattacks*. Moreover, optimal feature values and feature-relevant pitch positions were identified. In this way, concrete insights into beneficial characteristics of a team's positioning during *turnovers* could be provided. However, the translation of these insights into a predictive *xCounter* model did not reach a high prediction performance.

6.2 Discussion

In this dissertation, various soccer-specific concepts were successfully formalized and analyzed. Starting from short time scales in *atomic event* annotation (Chapter 3), this dissertation progressed to *sequence* annotation (Chapter 4) and *sequence*-specific analyses (Chapter 5).

An important concept defined in this dissertation is *semantic complexity* (Chapter 1.2), which was introduced to describe the fact that sport-specific terms and concepts often lack a clear, rule-based definition. The results of the expert annotation studies in this work showed that *semantic complexity* coincidentally increases with larger time scales, from *atomic events* to *sequences*. Moreover, the results revealed that higher *semantic complexity* generally causes expert disagreement on the annotated concepts. Notably, even on short time scales, increasing *semantic complexity* was already found to lead to expert disagreement. Adding to this relation, the experiments concerning automatic models indicated that expert disagreement impedes the classification performance of automatic models. Therefore, a critical takeaway is the necessity of detailed validation for any expert annotation, including discussions, pilot annotations, and comparisons between experts.

Among the proposed criteria for annotation schemes (see Chapter 2.2.1), the hierarchical structure proves particularly beneficial. Thus, that manual annotations should be performed using a hierarchical scheme. In that sense, this dissertation adds to the argument that studies should start with simple concepts that experts can agree on before progressing to more sophisticated analyses (Low et al., 2022). This provides a valuable extension to the field, where comparable state-of-the-art studies often rely on flat annotation schemes lacking a similar structure.

More broadly, the results of this dissertation also raised concerns about the validity of data from *event data providers*, particularly concerning event definitions involving high *semantic complexity*. This supports prior findings (H. Liu et al., 2013; McKinley, 2019) suggesting possible semantic overlap between event definitions. Due to these concerns, data from *event data providers* should be rigorously validated. In this sense, a valuable future step is to integrate a validation step with the previously presented SPADL method of unification for *event data* from different *event data providers* (Decroos et al., 2019).

Finally, this dissertation also addressed the problem of finding an optimal *position data* representation, previously raised by Raabe et al. (2023). The results highlighted the relevance of manually crafted features, as the employed models performed superior when using features compared to raw *position data*. While this may be attributed to the simplicity of the designed *neural networks*, another advantage of hand-crafted features was their ability to provide actionable insights that can be directly communicated to practitioners. Thus, a trade-off between predictive capability and interpretability should be carefully considered in future research, as practical applications may favor concise insights over higher predictive power.

In conclusion, the results of this dissertation contribute to bridging the gap between computer science and sports science, emphasized as an important step for both of those communities to support the integration of more theoretical research into practice (Goes et al., 2021; Low et al., 2020).

6.2.1 Limitations

As previously discussed, this dissertation focuses on a subset of possible time scales examined in soccer research. Yet, apart from short and intermediate time scales, there are other time scales on which soccer analyses are highly valuable. For example, an analysis on a large time scale was conducted in Article VI. This study focused on predicting match outcomes on a dataset of over 300.000 matches around the world. Regarding this approach, problems with data quality and data availability beneath the topflight leagues were encountered. Due to this shortcoming, the integration of meta information (such as rest days, market values, betting odds, etc.) was more applicable than using the previously presented *event data*, *position data*, or *video data* sources. In that sense, the algorithms presented in this dissertation are subject to data availability.

Moreover, while the developed automatic models for *sequence* annotations performed their respective task with relatively high accuracy, a certain degree of error remained. For a given application, it needs to be carefully evaluated if this error can be tolerated or not. For example, the automatic *match phase* annotation with 80% accuracy was able to provide a general impression about a team's average formation in certain *match phases* (see Chapter 4.4). Yet, an analogous frame-wise analysis of these *match phase sequence* could not be performed.

Moreover, the magnitude of errors was even more severe for models predicting the future of *sequences*. For example, the predictive model for the outcome of *counterattacks* was not able to outperform its respective baselines. Similar difficulties are also encountered in Articles VII and VIII, where the resulting forward progress of a team during *sequences* of *open play* was predicted. In this regard, examining the limiting factor of randomness in soccer, which has already been examined for goals by Wunderlich and Memmert (2021), still needs to be investigated for other components such as *turnovers*, passes, or *set pieces*.

Finally, the algorithms in this dissertation were proposed with the aim of integrating these structures into the workflow of practitioners working in soccer clubs. Yet, tailoring studies towards exact requirements of clubs is difficult, as little is known about the concrete working principles and requirements of soccer clubs. As the clubs are often careful to publish their insights due to the concern of losing a competitive advantage, it remains challenging to assess whether publicly available research is adopted by soccer clubs. Nevertheless, the upcoming chapter discusses how the content of this dissertation could possibly be applied to practice.

6.2.2 Application to Practice

An important achievement of this dissertation is the automatic contextualization of soccer data. Contextualized applications have recently been encouraged by Low et al. (2022), to enable the integration of more theoretical research into practice (Goes et al., 2021). In that sense, this dissertation provided multiple ways to segment a match of soccer into *sequences*, using automatic *match phase* annotations. After completing the segmentation of the match, there are multiple available use-cases.

Directly applied to qualitative video analyses, contextualization is useful to facilitate the search for relevant situations such as *transitions*. By outsourcing this task to automatic models, the video analyst can save a significant amount of time, as the video can be analyzed more effectively, without the need of manually searching for these situations.

On a more advanced level, contextualization might offer an interesting perspective into already existing algorithms present at soccer clubs. As discussed in Chapter 2.1.1, clubs commonly rate players based on certain evaluation metrics. For example, if the pass conversion rate is used to quantify the passing abilities of players, it might be valuable to also compute the conversion rate while facing *team pressure*. Thus, the automatic annotation of *match phases* offers fast, fine-grained insights into player performance.

Another achievement of this dissertation is the identification of valuable team behavior for *counterattacks*. Using the *xCounter* approach, an optimal vertical compactness that facilitates a *counterattack* when gaining *possession* after a *turnover* can be identified. Such a threshold value could be applied in training procedures, to improve the team behavior for potential ball wins, but also to detect that characteristic in the opponent team and prevent a ball loss in these situations.

Finally, the developed annotation scheme structure can be extended for any annotation study performed within a club. During the scouting process of potential new players, many clubs rely on qualitative video analyses (Mehta et al., 2024). Therefore, scouting reports contain the subjective estimation of the scout that assessed the player. To extend this process in terms of objectivity, and re-usability of the scouting reports, defining a custom annotation scheme offers a sound way to evaluate player performance. In that sense, adopting such a scheme would enable a granular validation (of a scouts subjective report) as well as a quick, detailed summary of player performances.

In conclusion, the insights from this dissertation can be applied with various degrees of complexity, based on the circumstances at a given soccer club. However, as this situation is not always publicly known, it remains important for research in soccer analysis to offer different levels of complexity to be applicable for the majority of clubs despite of differences in their technological progress.

6.3 Future Work

While dissertation addressed research gaps (see Chapter 1.2), several challenges remain that provide directions for future research.

Data Advancements With the ongoing development of local positioning systems and the increasing acceptance of wearable sensors among players and practitioners, *position data* becomes similarly more available and precise. In general, this development is beneficial for algorithms that require precise *position data*. Moreover, when local positioning systems are worn by players during the match, the *position data* is available in real-time. This advancement is especially interesting for real-time analysis approaches, such as load management, injury prevention, and in-game tactical analyses, which recommend concrete actions (such as a substitution, or tactical shifts) during the game.

A recently emerging technology is the tracking of three-dimensional player and ball movements, including limb-tracking (FIFA, 2022). On the one hand, this data type presents tremendous opportunities for improving current metrics, e.g., by refining existing *expected goals* models (Anzer & Bauer, 2021) using the body position of the shooting and the defending players. On the other hand, the processing this complex data source is challenging for scientists and practitioners working with the data.

Besides advancements in *position data*, there also exists an ongoing surge of other data sources. Given the long history of human interaction with soccer, there exists a large source of historical data in soccer. This data might be accessed through recordings of radio commentaries, documentation of video games (in some form), or social media posts (Wunderlich & Memmert, 2020). From that perspective, multimodal analyses involving databases from diverse sources could provide a more comprehensive understanding of players and teams. Thus, analyzing and combining data from different sources is a promising avenue to obtain more holistic insights.

Integrating Insights from Different Time Scales and Research Fields In this dissertation, soccer data is analyzed on various time scales. However, the connected analysis of these insights and the translation from one time scale to another is limited. For future work, it would be interesting to observe how *atomic events* correlate to different *match phases*, or how aggregated *match phases* influence the win probability of a team. Thus, filling the gaps between time scales and connecting the individual results could improve prediction results and generate for more nuanced insights.

Similarly, while this dissertation largely focuses on the tactical analysis, there are other research fields (e.g., physiology and psychology) that are highly relevant for the performance of players and teams. As an example for such an alternative research field, the relation between relative age of youth players and their career length was examined Article IX. Thus, an interesting direction for future research is the connection of methods from different research fields.

To facilitate research in these directions we have developed an open-source software package in Article X. This package is designed to allow users to quickly process and analyze soccer *position* and *event data*.

Alternative Experimental Settings Finally, this dissertation analyzed existing data of soccer matches that have already been played. Yet, there are alternative possibilities for gaining insights into the underlying principles of soccer. For instance, the evaluation of different team formations in a laboratory-like setting, proposed by Low et al. (2022), offers an interesting perspective and presents an intriguing direction for future soccer research.

Bibliography

Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., Von Landesberger, T., & Weber, H. (2017). Visual analysis of pressure in football. *Data Mining and Knowledge Discovery*, *31*(6), 1793–1839. https://doi.org/10.1007/s10618-017-0513-2

Anzer, G., & Bauer, P. (2021). A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.624475

Aquino, R., Puggina, E. F., Alves, I. S., & Garganta, J. (2017). Skill-related performance in soccer: a systematic review. *Human Movement*, *18*(5), 3–24. https://doi.org/10.1515/humo-2017-0042

Bauer, P., & Anzer, G. (2021). Data-driven detection of counterpressing in professional football: A supervised machine learning task based on synchronized positional and event data with expert-based feature extraction. *Data Mining and Knowledge Discovery*, *35*(5), 2009–2049. https://doi.org/10.1007/s10618-021-00763-7

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2014). Identifying Team Style in Soccer Using Formations Learned from Spatiotemporal Tracking Data. In *2014 IEEE International Conference on Data Mining Workshop* (pp. 9–14). Shenzhen, China: IEEE. https://doi.org/10.1109/ICDMW.2014.167

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). New York, NY: Springer.

Burger, A., Rogulj, N., Foretić, N., & Čavala, M. (2013). Analysis of Rebounded Balls in a Team Handball Match. *Sportlogia*, *9*(1), 53–58. https://doi.org/10.5550/sgia.130901.en.007B

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *ArXiv preprint ArXiv:1406.1078*.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

https://doi.org/10.1177/001316446002000104
Cooper, A. (2023). *StatsPerform playing styles - an introduction.* Retrieved 2024-05-01, from https://www.statsperform.com/resource/stats-playing-styles-introduction
Davids, K., Lees, A., & Burwitz, L. (2000). Understanding and measuring coordination and control in kicking skills in soccer: Implications for talent identification and skill acquisition. *Journal of Sports Sciences*, *18*(9), 703–714. https://doi.org/10.1080/02640410050120087
Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions Speak Louder than Goals: Valuing Player Actions in Soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1851–1861). Anchorage AK USA: ACM. https://doi.org/10.1145/3292500.3330758
Decroos, T., Van Haaren, J., & Davis, J. (2018). Automatic Discovery of Tactics in Spatio-Temporal Soccer Match Data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 223–232). London United Kingdom: ACM. https://doi.org/10.1145/3219819.3219832
Deliege, A., Cioppa, A., Giancola, S., Seikavandi, M. J., Dueholm, J. V., Nasrollahi, K., . . . Van Droogenbroeck, M. (2021). SoccerNet-v2: A Dataset and Benchmarks for Holistic Understanding of Broadcast Soccer Videos. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 4503–4514). Nashville, TN, USA: IEEE. https://doi.org/10.1109/CVPRW53098.2021.00508
DFL. (2021). *Bundesliga Match Facts: DFL and AWS revamp 'Average Positions'.* Retrieved 2024-09-07, from https://www.bundesliga.com/en/bundesliga/news/match-facts-dfl-aws-revamp-average-positions-trends-14706
Dick, U., Link, D., & Brefeld, U. (2022). Who can receive the pass? A computational model for quantifying availability in soccer. *Data Mining and Knowledge Discovery*, *36*(3), 987–1014. https://doi.org/10.1007/s10618-022-00827-2
Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, *7*(3-4), 240–252. https://doi.org/10.1057/PALGRAVE.IVS.9500182
Evangelos, T., & Nikolaos, A. (2004). Registration of rebound possession zones in basketball. *International Journal of Performance Analysis in Sport*, *4*(1), 34–39. https://doi.org/10.1080/24748668.2004.11868289
Fakhar, B., Rashidy Kanan, H., & Behrad, A. (2019). Event detection in soccer videos using unsupervised learning of Spatio-temporal features based on pooled spatial pyramid model. *Multimedia Tools and Applications*, *78*(12), 16995–17025. https://doi.org/10.1007/s11042-018-7083-1
Fassmeyer, D., Anzer, G., Bauer, P., & Brefeld, U. (2021). Toward Automatically Labeling Situations in Soccer. *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.725431
Fernandes, T., Camerino, O., Garganta, J., Pereira, R., & Barreira, D. (2019). Design

and validation of an observational instrument for defence in soccer based on the Dynamical Systems Theory. *International Journal of Sports Science & Coaching*, *14*(2), 138–152.
https://doi.org/10.1177/1747954119827283

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

Fernández, J., Bornn, L., & Cervone, D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*, *110*(6), 1389–1427.
https://doi.org/10.1007/s10994-021-05989-6

FIFA. (2022). *The FIFA Football Language.* Retrieved 2024-09-07, from https://www.fifatrainingcentre.com/en/game/performance-analysis/football-language-analysis/the-fifa-football-language.php

FIFA. (2022). *Limb-tracking technology offers new array of possibilities.* Retrieved 2024-10-11, from https://inside.fifa.com/technical/football-technology/news/limb-tracking-technology-offers-new-array-of-possibilities

FIFA. (2022). *Possession control.* Retrieved 2024-05-01, from https://www.fifatrainingcentre.com/en/fwc2022/efi-metrics/efi-metric--possession-control.php

Forcher, L., Altmann, S., Forcher, L., Jekauc, D., & Kempe, M. (2022). The use of player tracking data to analyze defensive play in professional soccer - A scoping review. *International Journal of Sports Science & Coaching*, *17*(6), 1567–1592.
https://doi.org/10.1177/17479541221075734

Fox, L. (2021). *How the NFL uses analytics, according to the lead analyst of A super bowl champion.* Retrieved 2024-05-01, from https://www.forbes.com/sites/liamfox/2021/08/12/how-the-nfl-uses-analytics-according-to-the-lead-analyst-of-a-super-bowl-champion/?sh=53458e38424e

Freund, Y., & Schapire, R. E. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory* (pp. 23–37).
https://doi.org/10.1006/jcss.1997.1504

Gensler, A., & Sick, B. (2018). Performing event detection in time series with SwiftEvent: an algorithm with supervised learning of detection criteria. *Pattern Analysis and Applications*, *21*(2), 543–562.
https://doi.org/10.1007/s10044-017-0657-0

Goes, F., Meerhoff L. Bueno M. Rodrigues D. Moura F. Brink, M., Elferink-Gemser, M., Knobbe, A., Cunha, S., Torres, R., & Lemmink, K. (2021). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science*, *21*(4), 481–496.
https://doi.org/10.1080/17461391.2020.1747552

Goldsberry, K. P. (2019). *Sprawlball: a visual tour of the new era of the NBA*. Boston: Houghton Mifflin Harcourt.

Gonçalves, E., Gonzaga, A. d. S., Cardoso, F. d. S. L., & Teoldo, I. (2015). Anticipation in soccer: a systematic review. *Human Movement*, *16*(2), 91–101. https://doi.org/10.1515/humo-2015-0032

Groll, A., Ley, C., Schauberger, G., & Van Eetvelde, H. (2019). A hybrid random forest to predict soccer matches in international tournaments. *Journal of Quantitative Analysis in Sports*, *15*(4), 271–287. https://doi.org/10.1515/jqas-2018-0060

Guralnik, V., & Srivastava, J. (1999). Event detection from time series data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 33–42). https://doi.org/10.1145/312129.312190

Herold, M., Hecksteden, A., Radke, D., Goes, F., Nopp, S., Meyer, T., & Kempe, M. (2022). Off-ball behavior in association football: A data-driven model to measure changes in individual defensive pressure. *Journal of Sports Sciences*, *40*(12), 1412–1425. https://doi.org/10.1080/02640414.2022.2081405

Hewitt, A., Greenham, G., & Norton, K. (2016). Game style in soccer: what is it and can we quantify it? *International Journal of Performance Analysis in Sport*, *16*(1), 355–372. https://doi.org/10.1080/24748668.2016.11868892

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278–282). https://doi.org/10.1109/ICDAR.1995.598994

HUDL Sportscode. (2023). *Premiership Power Leicester Tigers Leverage Video to Prepare.* Retrieved 2024-05-01, from https://www.hudl.com/blog/leicester-tigers

IFAB. (2024). *IFAB Laws of the Game.* Retrieved 2024-05-01, from https://www.theifab.com

Ihler, A., Hutchins, J., & Smyth, P. (2006). Adaptive event detection with time-varying poisson processes. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 207–216). Philadelphia PA USA: ACM. https://doi.org/10.1145/1150402.1150428

Impect. (2024). *Impect.* Retrieved 2024-09-07, from https://www.impect.com/en/products/

Kempe, M., Vogelbein, M., Memmert, D., & Nopp, S. (2014). Possession vs. direct play: evaluating tactical behavior in elite soccer. *International Journal of Sports Science*, *4*(6A), 35–41.

Kim, H., Kim, B., Chung, D., Yoon, J., & Ko, S.-K. (2022). SoccerCPD: Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3146–3156). Washington DC USA: ACM. https://doi.org/10.1145/3534678.3539150

Kim, J., James, N., Parmar, N., Ali, B., & Vučković, G. (2019). The Attacking Process in Football: A Taxonomy for Classifying How Teams Create Goal Scoring Opportunities Using a Case Study of Crystal Palace FC. *Frontiers in Psychology*, *10*.
https://doi.org/10.3389/fpsyg.2019.02202

Kinexon. (2023). *KINEXON becomes world's first "FIFA Preferred Provider Live Player & Ball Tracking"*. Retrieved 2024-05-01, from https://kinexon.com/blog/fifa-preferred-provider-live-player-and-ball-tracking/

Lago-Ballesteros, J., Lago-Peñas, C., & Rey, E. (2012). The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *Journal of Sports Sciences*, *30*(14), 1455–1461.
https://doi.org/10.1080/02640414.2012.712715

Lang, S., Wild, R., Isenko, A., & Link, D. (2022). Predicting the in-game status in soccer with machine learning using spatiotemporal player tracking data. *Scientific Reports*, *12*(1).
https://doi.org/10.1038/s41598-022-19948-1

LeCun, Y., Touresky, D., Hinton, G., & Sejnowski, T. (1988). A theoretical framework for back-propagation. In *Proceedings of the 1988 connectionist models summer school* (Vol. 1, pp. 21–28). San Mateo, CA, USA.

Lewis, M. (2004). *Moneyball: the art of winning an unfair game*. New York, NY: Norton.

Link, D., & Hoernig, M. (2017). Individual ball possession in soccer. *PLOS ONE*, *12*(7), 1-15.
https://doi.org/10.1371/journal.pone.0179953

Litwitz, K., Memmert, D., & Wunderlich, F. (2024). Rebounds in football: A systematic investigation of characteristics of goals scored after rebounded balls in english premier league seasons 2012/2013 to 2018/2019. *International Journal of Sports Science & Coaching*.
https://doi.org/10.1177/17479541241269007

Liu, G., Luo, Y., Schulte, O., & Kharrat, T. (2020). Deep soccer analytics: learning an action-value function for evaluating soccer players. *Data Mining and Knowledge Discovery*, *34*(5), 1531–1559.
https://doi.org/10.1007/s10618-020-00705-9

Liu, H., Hopkins, W., Gómez, A. M., & Molinuevo, S. J. (2013). Inter-operator reliability of live football match statistics from OPTA Sportsdata. *International Journal of Performance Analysis in Sport*, *13*(3), 803–821.
https://doi.org/10.1080/24748668.2013.11868690

Low, B., Coutinho, D., Gonçalves, B., Rein, R., Memmert, D., & Sampaio, J. (2020). A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. *Sports Medicine*, *50*(2), 343–385.
https://doi.org/10.1007/s40279-019-01194-7

Low, B., Rein, R., Schwab, S., & Memmert, D. (2022). Defending in 4-4-2 or 5-3-2 formation? small differences in footballers' collective tactical behaviours. *Journal of Sports Sciences*, *40*(3), 351–363.
https://doi.org/10.1080/02640414.2021.1993655

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, *13*(3), e0194889.
https://doi.org/10.1371/journal.pone.0194889

Martin, D., O Donoghue, P. G., Bradley, J., & McGrath, D. (2021). Developing a framework for professional practice in applied performance analysis. *International Journal of Performance Analysis in Sport*, *21*(6), 845–888.
https://doi.org/10.1080/24748668.2021.1951490

McHale, I. G., Scarf, P. A., & Folker, D. E. (2012). On the Development of a Soccer Player Performance Rating System for the English Premier League. *Interfaces*, *42*(4), 339–351.
https://doi.org/10.1287/inte.1110.0589

McKenna, M., Cowan, D. T., Stevenson, D., & Baker, J. S. (2018). Neophyte experiences of football (soccer) match analysis: a multiple case study approach. *Research in Sports Medicine*, *26*(3), 306–322.
https://doi.org/10.1080/15438627.2018.1447473

McKinley, E. (2019). *Shots in the dark: How data providers tell us different versions of what happened.* Retrieved 2024-09-07, from https://www.americansocceranalysis.com/home/2019/9/30/shots-in-the-dark-how-data-providers-tell-us-different-versions-of-what-happened

Mehta, S., Furley, P., Raabe, D., & Memmert, D. (2024). Examining how data becomes information for an upcoming opponent in football. *International Journal of Sports Science & Coaching*, *19*(3), 978–987.
https://doi.org/10.1177/17479541231187871

Memmert, D., Lemmink, K. A. P. M., & Sampaio, J. (2017). Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine*, *47*(1), 1–10.
https://doi.org/10.1007/s40279-016-0562-5

Merhej, C., Beal, R. J., Matthews, T., & Ramchurn, S. (2021). What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 3394–3403). Virtual Event Singapore: ACM.
https://doi.org/10.1145/3447548.3467090

Naik, B. T., Hashmi, M. F., & Bokde, N. D. (2022). A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. *Applied Sciences*, *12*(9).
https://doi.org/10.3390/app12094429

Nielsen Sports. (2018). *Nielsen Sports World Football Report.* Retrieved 2024-05-01, from https://www.nielsen.com/wp-content/uploads/sites/2/2019/04/world-football-report-2018.pdf

Norman, J. (2018). *Football still americans' favorite sport to watch.* Retrieved 2024-05-01, from https://news.gallup.com/poll/224864/football-americans-favorite-sport-watch.aspx

Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A public data set of spatio-temporal match events in soccer

competitions. *Scientific Data*, *6*(1).
https://doi.org/10.1038/s41597-019-0247-7

Pino-Ortega, J., Oliva-Lozano, J. M., Gantois, P., Nakamura, F. Y., & Rico-González, M. (2022). Comparison of the validity and reliability of local positioning systems against other tracking technologies in team sport: A systematic review. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, *236*(2), 73–82.
https://doi.org/10.1177/1754337120988236

Raabe, D., Nabben, R., & Memmert, D. (2023). Graph representations for the analysis of multi-agent spatiotemporal sports data. *Applied Intelligence*, *53*(4), 3783–3803.
https://doi.org/10.1007/s10489-022-03631-z

Raudonius, L., & Allmendinger, R. (2021). Evaluating Football Player Actions During Counterattacks. In *Intelligent Data Engineering and Automated Learning – IDEAL 2021* (Vol. 13113, pp. 367–377). Cham: Springer International Publishing.
https://doi.org/10.1007/978-3-030-91608-4_36

Reep, C., & Benjamin, B. (1968). Skill and Chance in Association Football. *Journal of the Royal Statistical Society. Series A (General)*, *131*(4), 581–585.
https://doi.org/10.2307/2343726

Relvas, H., Littlewood, M., Nesti, M., Gilbourne, D., & Richardson, D. (2010). Organizational structures and working practices in elite european professional football clubs: Understanding the relationship between youth and professional domains. *European Sport Management Quarterly*, *10*(2), 165–187.
https://doi.org/10.1080/16184740903559891

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.
https://doi.org/10.1037/h0042519

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.
https://doi.org/10.1038/323533a0

Sahasrabudhe, A., & Bekkers, J. (2023). A Graph Neural Network deep-dive into successful counterattacks. In *MIT Sloan Sports Analytics Conference* (Vol. 17).

Sanabria, M., Sherly, Precioso, F., & Menguy, T. (2019). A Deep Architecture for Multimodal Summarization of Soccer Games. In *Proceedings Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports* (pp. 16–24). Nice France: ACM.
https://doi.org/10.1145/3347318.3355524

Sanford, R., Gorji, S., Hafemann, L. G., Pourbabaee, B., & Javan, M. (2020). Group Activity Detection from Trajectory and Video Data in Soccer. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 3932–3940). Seattle, WA, USA: IEEE.
https://doi.org/10.1109/CVPRW50498.2020.00457

Seidenschwarz, P., Rumo, M., Probst, L., & Schuldt, H. (2020). High-Level Tactical Performance Analysis with SportSense. In *Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports* (pp. 45–52).

https://doi.org/10.1145/3422844.342305
Shih, H.-C. (2018). A Survey of Content-Aware Video Analysis for Sports. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(5), 1212–1231. https://doi.org/10.1109/TCSVT.2017.2655624
SkillCorner. (2024). *SkillCorner.* Retrieved 2024-05-01, from https://skillcorner.com/football
Sousa, T., & Garganta, J. (2001). The Importance of Set-plays in Soccer. In *Proceedings of the IV Congress of Notational Analysis of Sport* (pp. 53–57).
StatsBomb. (2024). *StatsBomb.* Retrieved 2024-05-01, from https://statsbomb.com/what-we-do/soccer-data/
StatsPerform. (2024). *StatsPerform.* Retrieved 2024-05-01, from https://www.statsperform.com
Tenga, A., Kanstad, D., Ronglan, L. T., & Bahr, R. (2009). Developing a New Method for Team Match Performance Analysis in Professional Soccer and Testing its Reliability. *International Journal of Performance Analysis in Sport*, *9*(1), 8–25. https://doi.org/10.1080/24748668.2009.11868461
Tracab. (2024). *Tracab technologies.* Retrieved 2024-05-01, from https://tracab.com/products/tracab-technologies
Tuyls, K., Omidshafiei, S., Muller, P., Wang, Z., Connor, J., Hennes, D., . . . Hassabis, D. (2021). Game Plan: What AI can do for Football, and What Football can do for AI. *Journal of Artificial Intelligence Research*, *71*, 41–88. https://doi.org/10.1613/jair.1.12505
Van Roy, M., Cascioli, L., & Davis, J. (2023). Etsy: A rule-based approach to event and tracking data synchronization. In *International workshop on machine learning and data mining for sports analytics* (pp. 11–23). https://doi.org/10.1007/978-3-031-53833-9_2
Verma, J. P., & Agrawal, S. (2016). Big Data Analytics: Challenges And Applications For Text, Audio, Video, And Social Media Data. *International Journal on Soft Computing, Artificial Intelligence and Applications*, *5*(1), 41–51. https://doi.org/10.5121/ijscai.2016.5105
Wittgenstein, L., & Anscombe, G. E. M. (2003). *Philosophical investigations: the German text, with a revised English translation* (3rd ed.). Malden, MA: Blackwell Pub.
Wright, C., Atkins, S., & Jones, B. (2012). An analysis of elite coaches' engagement with performance analysis services (match, notational analysis and technique analysis). *International Journal of Performance Analysis in Sport*, *12*(2), 436–451. https://doi.org/10.1080/24748668.2012.11868609
Wunderlich, F., & Memmert, D. (2020). Innovative Approaches in Sports Science—Lexicon-Based Sentiment Analysis as a Tool to Analyze Sports-Related Twitter Communication. *Applied Sciences*, *10*(2), 431. https://doi.org/10.3390/app10020431
Wunderlich, F., & Memmert, D. (2021). Forecasting the outcomes of sports events: A review. *European Journal of Sport Science*, *21*(7), 944–957. https://doi.org/10.1080/17461391.2020.1793002

Wunderlich, F., Seck, A., & Memmert, D. (2021). The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. *Journal of Sports Sciences*, *39*(20), 2322–2337.
https://doi.org/10.1080/02640414.2021.1930685

Wyscout Spa. (2024). *Make the most of Wyscout Data.* Retrieved 2024-05-01, from https://footballdata.wyscout.com

Appendix

Abstracts of the ten articles that have been published over the course of this doctoral program are listed on the following pages.

I. Biermann, H., Theiner, J., Bassek, M., Raabe, D., Memmert, D., & Ewerth, R. (2021). A unified taxonomy and multimodal dataset for events in invasion games. *Proceedings of the 4th International Workshop on Multimedia Content Analysis in Sports*

The automatic detection of events in complex sports games like soccer and handball using positional or video data is of large interest in research and industry. One requirement is a fundamental understanding of underlying concepts, i.e., events that occur on the pitch. Previous work often deals only with so-called lowlevel events based on well-defined rules such as free kicks, free throws, or goals. High-level events, such as passes, are less frequently approached due to a lack of consistent definitions. This introduces a level of ambiguity that necessities careful validation when regarding event annotations. Yet, this validation step is usually neglected as the majority of studies adopt annotations from commercial providers on private datasets of unknown quality and focuses on soccer only. To address these issues, we present (1) a universal taxonomy that covers a wide range of low and high-level events for invasion games and is exemplarily refined to soccer and handball, and (2) release two multi-modal datasets comprising video and positional data with gold-standard annotations to foster research in fine-grained and ball-centered event spotting. Experiments on human performance demonstrate the robustness of the proposed taxonomy, and that disagreements and ambiguities in the annotation increase with the complexity of the event. Datasets are available at https://github.com/mm4spa/eigd.

II. Biermann, H., Komitova, R., Raabe, D., Müller-Budack, E., Ewerth, R., & Memmert, D. (2023). Synchronization of passes in event and spatiotemporal soccer data. *Scientific Reports*, 13(1), 15878.

The majority of soccer analysis studies investigates specifc scenarios through the implementation of computational techniques, which involve the examination of either spatiotemporal position data (movement of players and the ball on the pitch) or event data (relating to signifcant situations during a match). Yet, only a few applications perform a joint analysis of both data sources despite the various involved advantages emerging from such an approach. One possible reason for this is a non-systematic error in the event data, causing a temporal misalignment of the two data sources. To address this problem, we propose a solution that combines the SwiftEvent online algorithm (Gensler and Sick in Pattern Anal Appl 21:543–562, 2018) with a subsequent refnement step that corrects pass timestamps by exploiting the statistical properties of passes in the position data. We evaluate our proposed algorithm on ground-truth pass labels of four top-fight soccer matches from the 2014/15 season. Results show that the percentage of passes within half a second to ground truth increases from 14 to 70data. A comparison with other models shows that our algorithm is superior to baseline models and comparable to a deep learning pass detection method (while requiring signifcantly less data). Hence, our proposed lightweight framework ofers a viable solution that enables groups facing limited access to (recent) data sources to efectively synchronize passes in the event and position data.

III. Biermann, H., Memmert, D., Petersen, N., Raabe, D. (2025). Contextualization of soccer analysis with tactical periodization and machine learning. *Data Mining and Knowledge Discovery*, 39, 23.

It has become common practice in topflight leagues to track position data of soccer players and the ball. Analyzing sports performance based on this high-resolution data is a non-trivial task due to the great complexity and simultaneous lack of structure of the game. Sports practitioners tackle this problem through tactical periodization, i.e., mapping the course of the game onto different states, so-called match phases. However, creating manual tactical periodizations is a timeconsuming task prone to subjective biases. Automatic approaches are thus preferred, but validated and open match phase models are currently lacking. The present study addresses this issue by (i) formalizing a domain-specific, qualitative match phase annotation scheme from related work, (ii) creating and validating a multi-annotator set of annotations, (iii) training several supervised machine learning architectures to fully automate the task of annotation, and (iv) demonstrating the usefulness by conducting a contextualized detection of playing formations with the best model, referred to as FeatGRU. Steps (ii) through (iv) were performed on a set of real-world soccer data and the bestperforming model is made available. FeatGRU is of value to the soccer community as it provides a fully automatic, frame-by-frame match phase annotation that matches domain experts' opinions with 80accuracy while being modular extendable for future work. Moreover, we found a strong relation between semantic complexity of matchphases, expert agreements, and classification performance, highlighting the importance of valid label generation. Thus, our approach presents an interesting benchmark to domains where automatic approaches are required while ambiguity between human expert opinions exists.

IV. Biermann, H., Wieland, F. G., Timmer, J., Memmert, D., & Phatak, A. (2022). Towards Expected Counter - Using Comprehensible Features to Predict Counterattacks. ***Machine Learning and Data Mining for Sports Analytics. MLSA 2022. Communications in Computer and Information Science (1783)*****, 3-13.**

Soccer is a low-scoring game where one goal can make the difference. Thus, counterattacks have been recognized by modern strategy as an effective way to create scoring opportunities from a position of stable defense. This coincidentally requires teams on offense to be mindful of taking risks, i.e. losing the ball. To assess these risks, it is crucial to understand the involved mechanisms that turn ball losses into counterattacks. However, while the soccer analytics community has made progress predicting outcomes of single actions (shots or passes) [1,2] up to entire matches [15], individual sequences like counterattacks have not been predicted with comparable success. In this paper, we give reasons for this and create a framework that allows understanding complex sequences through comprehensible features. We apply this framework to predict counterattacks before they happen. Therefore, we find turnovers in soccer matches and create transparent counterattack labels from spatiotemporal data. Subsequently, we construct comprehensible features from sportspecific assumptions and assess their influence on counterattacks. Finally, we use these features to create a simple binary logistic regression model that predicts counterattacks. Our results show that players behind the ball are the most important predictive factors. We find that if a team loses the ball in the center and more than two players are not behind the ball, they concede a counterattack in almost 30% of cases. This stresses the importance to avoid ball losses in build-up play. In the future, we plan to extend this approach to generate more differentiated insights.

V. Biermann, H., Yang, W., Wieland, F. G., Timmer, J., & Memmert, D. (2023). Quantification of Turnover Danger with xCounter. ***Machine Learning and Data Mining for Sports Analytics. MLSA 2023. Communications in Computer and Information Science (2035)*****, 36–51.**

Counterattacks in soccer are an important strategical component for goal scoring. Previous work in the literature has described their impact and has formulated descriptive advice on successful actions during a counterattack. In contrast, in this work, we propose the notion of expected counter, i.e., quantifying forward progress by the ball-winning team at the moment of the turnover. Therefore, we apply a previously proposed framework for understanding complex sequences in soccer. Using this framework, we perform a novel feature-specific assessment that yields (a) critical feature values, (b) relevant feature pitch zones, and (c) feature prediction capabilities. The insights from this assessment step allow for creating concrete guidelines for optimal behavior in and out of possession. Thus, we find that preparing horizontally spaced pass options facilitates an own counterattack in case of a ball win while moving as a compact unit prevents an opposing counterattack in case of a ball loss. As a final step, we generalize our results by creating a predictive XGBoost model that outperforms a location-based baseline but still shows room for improvement.

VI. Wunderlich, F., Biermann, H., Yang, W., Bassek, M., Raabe, D., Elbert, N., Memmert, D., & Garnica-Caparrós, M. (2025). Assessing Machine Learning and Data Imputation Approaches to Handle the Issue of Data Sparsity in Sports Forecasting. *Machine Learning*. 114, 48 (2025).

Sparsity is a common characteristic for datasets used in the domain of sports forecasting, mainly derived from inconsistencies in data coverage. Typically, this issue is circumvented by cutting the number of features (depth-focused) or the sample size (breadth-focused) for analysis. The present study uses an experimental approach to analyse the effects of depth- or breadth-focused analyses and data imputation to enable usage of the full sample size and feature wealth. Two forecasting models following a hybrid (i.e., a combination of classical statistical and machine learning) and a full deep learning approach are introduced to perform experiments on a dataset of more than 300,000 soccer matches. In contrast to typical soccer forecasting studies, the analysis was not restricted to one-matchahead forecasts but used a longer forecasting horizon of up to two months ahead. Systematic differences between the two types of models were identified. The hybrid model based on classical statistical rating models, performs strongly on depth-focused approaches while not or only marginally improving for approaches with high data breadth. The deep learning model, however, performs weakly in a depth-focused approach but profits strongly from data breadth. The improved predicting performance in cases of high data breadth suggests that a rich feature set offers better training opportunities than a less detailed set with a larger sample size. Additionally, we showcase that data imputation can be used to address data sparsity by enabling full data depth and breadth. The presented findings are relevant for advancing predictive accuracy and sports forecasting methodologies, emphasizing the viability of imputation techniques to increase data coverage in different analytical approaches.

VII. Stival, L., Pinto, A., Andrade, F. D. S. P. D., Santiago, P. R. P., Biermann, H., Torres, R. D. S., & Dias, U. (2023). Using machine learning pipeline to predict entry into the attack zone in football. *PloS one*, 18(1).

Sports sciences are increasingly data-intensive nowadays since computational tools can extract information from large amounts of data and derive insights from athlete performances during the competition. This paper addresses a performance prediction problem in soccer, a popular collective sport modality played by two teams competing against each other in the same field. In a soccer game, teams score points by placing the ball into the opponent's goal and the winner is the team with the highest count of goals. Retaining possession of the ball is one key to success, but it is not enough since a team needs to score to achieve victory, which requires an offensive toward the opponent's goal. The focus of this work is to determine if analyzing the first five seconds after the control of the ball is taken by one of the teams provides enough information to determine whether the ball will reach the final quarter of the soccer field, therefore creating a goal-scoring chance. By doing so, we can further investigate which conditions increase strategic leverage. Our approach comprises modeling players' interactions as graph structures and extracting metrics from these structures. These metrics, when combined, form time series that we encode in two-dimensional representations of visual rhythms, allowing feature extraction through deep convolutional networks, coupled with a classifier to predict the outcome (whether the final quarter of the field is reached). The results indicate that offensive play near the adversary penalty area can be predicted by looking at the first five seconds. Finally, the explainability of our models reveals the main metrics along with its contributions for the final inference result, which corroborates other studies found in the literature for soccer match analysis.

VIII. Raabe, D., Biermann, H., Bassek, M., Memmert, D., & Rein, R. (2024). The dual problem of space: Relative player positioning determines attacking success in elite men's football. Journal of Sports Sciences, 1-10.

The concept of space has been successfully modelled in football using spatiotemporal player data and Voronoi diagrams. Current approaches, however, are narrow in scope by focusing on an inter-team allocation of space to measure space control. The present work extends this widespread perspective with an intra-team application of the Voronoi diagram and its dual Delaunay triangulation to measure space management. Both models are leveraged to derive novel performance metrics, which assess how teams use triangular positioning and how players tie up defenders during attacks. The outcome of N = 128,187 attacking sequences from 306 elite men's football matches is analysed using linear mixed-effects models to validate the proposed performance metrics. Results show that attacking success is characterized by player positioning which promotes forming of large triangles especially in ball proximity, whereas the overall number of triangles is of no relevance. Furthermore, players tie up more defenders and thus create free teammates more often during successful attacks. The results demonstrate that a new perspective on space is helpful to better quantify and understand the effect of space management and player positioning on attacking performance in football.

IX. Biermann, H., Memmert, D., Romeike, C., Knäbel, P., & Furley, P. (2024). Relative age effect inverts when looking at career performance in elite youth academy soccer. Journal of Sports Sciences, 1–6.

The aim of this study was to investigate the well-known selection bias favouring players born earlier in the year commonly referred to as the Relative Age Effect (RAE). Therefore, we analysed a group of top-tier youth academy players. The players joined and left the youth academy at different stages, where the first stage was at 8 years (U9) and the last stage was at 18 years of age (U19). Subsequently, we followed the career paths of all players in terms of minutes played in professional competitions. For that purpose, we collected competition information from transfermarkt.de. We label a competition as professional if the included teams obtain an average market value of 100.000€ (aggregated value of players playing in the team) and recorded the professional career minutes (PCM) for all players. Results show that in the youth academy, there are fewer players that are born late in the year compared to players that are born early in the year, confirming previous findings of RAE. However, we also find that players that are born late in the year achieve more PCM on average. This indicates that players that survive the RAE selection bias are exceptionally good at achieving long, successful careers.

X. Raabe, D., Biermann, H., Bassek, M., Wohlan, M., Komitova, R., Rein, R., Kuppens Groot, T. & Memmert, D. floodlight - A high-level, data-driven sports analytics framework. ***Journal of Open Source Software*****, 7(76), 4588.**

The present work introduces floodlight, an open source Python package built to support and automate team sport data analysis. It is specifically designed for the scientific analysis of spatiotemporal tracking data, event data, and game codes in disciplines such as match and performance analysis, exercise physiology, training science, and collective movement behavior analysis. It is completely providerand sports-independent and includes a high-level interface suitable for programming beginners. The package includes routines for most aspects of the data analysis process, including dedicated data classes, file parsing functionality, public dataset APIs, pre-processing routines, common data models and several standard analysis algorithms previously used in the literature, as well as basic visualization functionality. The package is intended to make team sport data analysis more accessible to sport scientists, foster collaborations between sport and computer scientists, and strengthen the community's culture of open science and inclusion of previous works in future works.