Institute of Exercise Training and Sport Informatics
German Sport University Cologne
Head of Institute: Prof. Dr. Daniel Memmert

Deep Learning with Spatiotemporal Data for Team Sports Analysis

Doctoral thesis accepted for the degree

Doktor der Naturwissenschaft (Dr. rer. nat.)

by

Dominik Raabe

from Neukirchen

Cologne, 2024

First reviewer: Prof. Dr. Daniel Memmert
Second reviewer: Prof. Dr. Ralph Ewerth

Chair of the doctoral committee: Prof. Dr. Mario Thevis
Thesis defended on: 12.11.2024

Affidavits following §7 section 2 No. 9 of the doctoral regulations from the German Sport University Cologne, March 30th 2020:

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.

Cologne, 12.11.2024

Abstract

Team sports analysis is crucial for enhancing performance in the global sports market. A recent game changer in this respect has been the widespread acquisition of data, such as spatiotemporal data from player tracking. However, extracting meaningful information on collective movements from this data is a difficult task. Previous studies have thus begun to utilize deep learning algorithms due to their high predictive power. Although these studies propose isolated applications, a systematic investigation on the usage of deep learning for analyzing spatiotemporal data in team sports is lacking.

The present thesis addresses this gap through a series of peer-reviewed publications. First, theoretical considerations on spatiotemporal data in team sports analysis are discussed, leading to the proposal and evaluation of a neural network architecture based on graph representations of spatiotemporal sports data. Numerical results reveal that graph representations and corresponding deep learning models can achieve state-of-the-art performance in sports-related prediction tasks by leveraging the characteristics of spatiotemporal data, while being computationally more efficient than comparable architectures.

Subsequently, several team sports analyses are conducted using the proposed network. These analyses focus on the relationship between machine and human performance in label prediction as well as on task and model designs that promote robust applications and analyses previously deemed infeasible. Results show that deep learning models are well suited for automation tasks such as supervised label generation. Yet, as the ambiguity in ground-truth labels increases, sound operationalization and careful interpretations of findings are required. If these conditions are met, deep learning is capable of producing results that open new frontiers for team sports analyses.

The empirical work of this thesis is complemented by contributions addressing technological challenges identified along the way. For example, profound programming experience is often required to conduct advanced team sports analysis. To meet these challenges, extensive code bases and sample datasets are released under open source licensing to significantly simplify such analyses.

It can be concluded that deep learning is a powerful tool for team sports analysis with spatiotemporal data which outperforms previous approaches in the investigated cases. The presented findings offer valuable insights on the effective use of deep learning with respect to task and model design as well as result interpretation. Furthermore, the released code bases facilitate fast and reproducible implementations in the area of team sports analysis as demonstrated.

Zusammenfassung

Auf der globalen Bühne des Teamsports sind Analysen entscheidend für die Leistungsoptimierung. In letzter Zeit hat die umfassende Datenerhebung stark an Relevanz für solche Teamsportanalysen gewonnen, wie beispielsweise die Erhebung von raum-zeitlichen Bewegungsdaten aus modernen Trackingsystemen. Die Extraktion aussagekräftiger Informationen über kollektive Bewegungen aus diesen Daten ist jedoch schwierig. Studien haben daher mit der Nutzung von Deep-Learning-Algorithmen, bekannt für ihre hohe Vorhersagekraft, begonnen. Obwohl diese Studien vereinzelte Anwendungen entwickeln, fehlt eine systematische Erforschung über die Verwendung von Deep Learning zur Analyse von raum-zeitlichen Daten in Teamsportanalysen.

Die vorliegende Arbeit bearbeitet diese Lücke in der Forschung durch eine Reihe von begutachteten Veröffentlichungen. Dazu werden zunächst theoretische Überlegungen zu raum-zeitlichen Daten in der Teamsportanalyse erörtert, welche anschließend zur Entwicklung und Auswertung einer neuronalen Netzwerkarchitektur auf der Grundlage von Graphrepräsentationen raum-zeitlicher Daten führen. Numerische Ergebnisse zeigen, dass Graphrepräsentationen und entsprechende Deep-Learning-Modelle durch Ausnutzung der Eigenschaften raum-zeitlicher Daten in sportbezogenen Vorhersageaufgaben exzellent abschneiden und dabei effizienter sind als vergleichbare Architekturen.

Im Anschluss werden mehrere Teamsportanalysen unter Verwendung der entwickelten Netzwerkarchitektur durchgeführt. Diese Analysen konzentrieren sich auf die Beziehung zwischen maschineller und menschlicher Leistung bei der Label-Generierung, sowie auf das Design von Aufgaben und Modellen, welche robuste Anwendungen und zuvor nicht realisierbare Analysen ermöglichen. Die Ergebnisse zeigen, dass Deep-Learning-Modelle gut für Automatisierungsaufgaben wie die überwachte Label-Generierung geeignet sind. Doch mit steigender Ambiguität der verwendeten ground-truth-Labels sind eine solide Operationalisierung und eine sorgfältige Interpretation der Ergebnisse erforderlich. Sind diese Bedingungen erfüllt, kann die Nutzung von Deep Learning die Grenzen des Machbaren in der Teamsportanalyse erweitern.

Die empirischen Ergebnisse dieser Dissertation werden durch Beiträge ergänzt, welche die im Laufe der Zeit erkannten technologischen Hürden reduzieren. So erfordern beispielsweise fortgeschrittene Teamsportanalysen oft umfassende Programmierkenntnisse. Um diesen Herausforderungen zu begegnen, werden umfangreiche Programmcodes und Beispieldatensätze unter Open-Source-Lizenzen veröffentlicht, um derartige Analysen signifikant zu vereinfachen.

Abschließend lässt sich sagen, dass Deep Learning ein leistungsfähiger Ansatz für die Teamsportanalyse mit raum-zeitlichen Daten ist, welches frühere Ansätze in den Untersuchungen übertrifft. Die präsentierten Ergebnisse bieten wertvolle Einblicke in die effektive Nutzung von Deep Learning für das Design von Aufgaben und Modellen, sowie die Interpretation von Ergebnissen. Darüber hinaus erleichtert der veröffentlichte Programmcode nachweisbar schnelle und reproduzierbare Implementierungen von Teamsportanalysen.

Acknowledgements

This dissertation is the result of an academic journey made possible by many different people. I would like to take this chance to express my deepest gratitude towards everyone who has taught, mentored, challenged, encouraged, and supported me along the way.

I am grateful to my supervisor Prof. Dr. Daniel Memmert for giving me the opportunity to pursue a PhD, for trusting me to take on responsibilities, and for providing me with the freedom to develop new ideas. Special thanks are also due to Prof. Dr. Jürgen Perl and to Dr. Robert Rein for their generous academic guidance and mentorship over the last decade.

I would also like to thank Prof. Dr. Ralph Ewerth for reviewing this thesis, the examination board and the doctoral committee for evaluating my doctorate, as well as the Bundesministerium für Bildung und Forschung (BMBF) for providing the funding for large portions of my work.

Over the course of my PhD, I got the chance to work with a lot of wonderful people, many of whom have become close friends. I am grateful to all my colleagues, fellow PhDs, and students for their endless help and inspiration. And finally, what cannot be put in words, I would like to thank my partner and family.

Articles Overview

The following articles were published in peer-reviewed journals or are currently under revision. They are discussed in detail in the present work.

Study I **Raabe, D.**, Nabben, R., & Memmert, D., (2023). Graph representations for the analysis of multi-agent spatiotemporal sports data. *Applied Intelligence, 53*, 3783–3803. https://doi.org/10.1007/s10489-022-03631-z [Impact Factor 7.04]{Q2 Artifical Intelligence}

Study II Bassek, M.*, **Raabe, D.***, Banning, A., Memmert, D., & Rein, R. (2023). Analysis of contextualized physical performance in elite Men's handball using graph-based deep learning. *Journal of Sports Sciences, 41*(13), 1299-1308. https://doi.org/10.1080/02640414.2023.2268366 [IF 3.87]{Q1 Sports Science}
** both authors contributed equally to this work*

Study III **Raabe, D.**, Biermann, H., Bassek, M., Wohlan, M., Komitova, R., Rein, R., Kuppens Groot, T., & Memmert, D. (2022). floodlight - A high-level, data-driven sports analytics framework. *Journal of Open Source Software, 7*(76), 4588. https://doi.org/10.21105/joss.04588 [IF 5.2]{Q1 Software Engineering}

Study IV **Raabe, D.**, Biermann, H., Bassek, M., Memmert, D., & Rein, R. (2024). The dual problem of space: player positioning determines attacking success in elite men's football. *Journal of Sports Sciences, 42*(19), 1821-1830. https://doi.org/10.1080/02640414.2024.2414363[IF 3.87]{Q1 Sports Science}

Study V Bassek, M., **Raabe, D.**, Memmert, D., & Rein, R., (2023). Analysis of Motion Characteristics and Metabolic Power in Elite Male Handball Players. *Journal of Sports Science and Medicine, 22*(2), 310-316. https://doi.org/10.52082/jssm.2023.310 [IF 3.3]{Q1 Orthopedics and Sports Medicine}

Study VI 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*, 1-10.

https://doi.org/10.1145/3475722.3482792 {A* - Computer vision and multimedia computation}

Study VII 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.
https://doi.org/10.1038/s41598-023-39616-2 [IF 4.44]{Q1 Multidisciplinary}

Study VIII Biermann, H., Memmert, D., Petersen, N., & **Raabe, D.** (2024, under review). Contextualization of Soccer Analysis with Tactical Periodization and Machine Learning. *Data Mining and Knowledge Discovery*.

Study IX Furley, P., Mehta, S., **Raabe, D.**, & Memmert, D. (2024). Objectivity of match analysis in football: Testing the level of agreement between coaches' interpretations of video data. *International Journal of Sports Science & Coaching*.
https://doi.org/10.1177/17479541241278603 [IF 2.58]{Q1 Social Sciences (miscellaneous)}

Study X Komitova, R., **Raabe, D.**, Rein, R., & Memmert, D. (2022). Time Series Data Mining for Sport Data: a Review. *International Journal of Computer Science in Sport, 21*(2), 17-31.
https://doi.org/10.2478/ijcss-2022-0008 [IF 1.42]{Q3 Computer Science (miscellaneous)}

Study XI Mehta, S., Furley, P., **Raabe, D.**, & Memmert, D. (2023). 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 [IF 2.58]{Q1 Social Sciences (miscellaneous)}

Study XII Schlenger, J., Wunderlich, F., **Raabe, D.**, & Memmert, D. (2023). Systematic Analysis of Position-Data-based Key Performance Indicators. *International Journal of Computer Science in Sport, 22*(1), 80-101.
https://doi.org/10.2478/ijcss-2023-0006 [IF 1.42]{Q3 Computer Science (miscellaneous)}

Study XIII Wunderlich, F., Biermann, H., Yang, W., Bassek, M., **Raabe, D.**, Elbert, N., Memmert, D., & Garnica Caparrós, M. (2024, accepted). Machine learning and data imputation approaches to handle the issue of data sparsity in sports forecasting. *Machine Learning*.

Contents

List of Tables

List of Acronyms

AI Artificial Intelligence. 2, 12

ANN Artificial Neural Networks. 10, 11, 15, 18

CNN Convolutional Neural Networks. 11, 19, 20, 27

DOTS Dynamic Optical Tracking Systems. 9, 18

EIGD-H Events in Invasion Games Dataset for Handball. 28

EIGD-S Events in Invasion Games Dataset for Soccer. 28

EPTS Electronic Performance and Tracking Systems. 8, 9, 10

FIFA Fédération Internationale de Football Association. 8, 28

GNN Graph Neural Networks. 11, 18, 19

GNSS Global Navigational Satellite Systems. 9

GPS Global Positioning System. 9

HBL German Men's Handball Bundesliga. 27, 28

IL Imitation Learning. 19

kNN k-Nearest Neighbors. 15, 18, 19

LPS Local Positioning systems. 9

LSTM Long Short Term Memory Networks. 11, 18, 19, 20, 29

MDP Markov Decision Processes. 19

RL Reinforcement Learning. 19

RNN Recurrent Neural Networks. 11, 18, 19

SOM Self-Organizing Maps. 10, 15

SOTS Stationary Optical Tracking Systems. 9

STSD Spatiotemporal Sports Data. 1, 2, 3, 5, 6, 8, 9, 10, 14, 15, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 34, 35

SVM Support Vector Machines. 15

TGNet Tactical Graph Networks. 26, 27, 33

VAE Variational Autoencoders. 18

Notation

The following notation is used throughout the main text of this thesis.

Scalars, Tuples and Sets

a, α, or A	A scalar (integer or real)
(a, b)	A tuple with (two) scalar entries
$\mathcal{A}$ or $\{...\}$	A set
$\{1, 2, ..., N\}$	A set containing all elements from 1 to N
$\mathcal{A} \subseteq \mathcal{B}$	$\mathcal{A}$ is a subset of $\mathcal{B}$

Indexing

a_i	A scalar for a given index i
$a_{t,i}$	A scalar at a given time point t and for index i
$\mathcal{A}_t$	A set with elements from a given time point t
$\mathcal{A}_{t_0,...,t_1}$	A set with elements from multiple given time points from t_0 to t_1

1. Introduction

Success in professional sports is of immense economical, social, political, and personal relevance for a heterogeneous group of stakeholders. More than five million people within the European Union worked in the sports sector in 2012, contributing to 2.12% of the total gross domestic product - with rising trend (European Commission, 2018). Athletes, clubs, fans, companies, and policy makers may differ in their motivations, but they share the goal of winning sports competitions. In the resulting race for competitiveness, research constitutes an integral part. In fact, empirical evidence (Lippi et al., 2008) as well as widespread opinion (Frevel et al., 2022) suggest that technological and scientific advancements are the major driving force in pushing the limits of human athletic performance. Research concerned with performance and competitiveness in sports can thus be safely regarded as an important field of study, with undeniably high economic and social impact.

Team sports analysis, i.e., match and performance analysis in team invasion games, is the primary tool for the assessment, monitoring, and improvement of competitiveness and success in professional sports. In the recent decade, a cornerstone of the technological and scientific advancement in professional sports has been the widespread collection of data (Jayal et al., 2018; Memmert & Raabe, 2023; Morgulev et al., 2018; Rein & Memmert, 2016). This development is in line with a general increase in the relevance of data science (Cao, 2018), and today team sports analyses are firmly integrated in professional sports. As research shows, such data analyses have even had direct influence on tactical and strategic decision-making, e.g., when it comes to scoring strategies in basketball (Morgulev et al., 2018). The processing of data thus offers great opportunities to provide a novel and more powerful form of quantified performance analysis (Patel et al., 2020). More precisely, data may render performance analysis more objective, comprehensive and profound - at faster speeds and lower costs compared to conventional standards set by qualitative and manually conducted analyses (Herold et al., 2019; Rein & Memmert, 2016; Stein et al., 2017).

Spatiotemporal Sports Data (STSD), i.e., athlete positions tracked over time, are a novel high-resolution data source recorded by optical or sensor-based technology. STSD provide an exhaustive and precise description of athlete movement, which enables detailed physical and tactical analyses (Rein & Memmert, 2016; Torres-Ronda et al., 2022). This separates STSD from related data sources, such as or event data logs or raw video footage. Yet, it also significantly increases the requirements on the used analysis methods. The complexity of STSD demand advanced algorithms to bridge a large semantic gap between the data generating process and sophisticated movement behaviors, including several layers of abstraction (Dutt-Mazumder et al., 2011; Kovalchik, 2023; Lord et al., 2020; Perl, 2011; Raabe et al., 2023; Stein et al., 2017). This is especially true given that (multi-agent) movement patterns themselves are regarded as highly complex phenomena, exceeding the capabilities of many established algorithms (Dutt-Mazumder et al., 2011; Fujii, 2021; Kovalchik, 2023; Lord et al., 2020; Morgulev et al., 2018; Stein et al., 2017). Conventional methods of model building and analysis thus often fall short in detecting meaningful patterns and anomalies or even linking the data to performance outcomes (Rein & Memmert, 2016). From a technological, methodological, and conceptual perspective, new analysis techniques are required to unleash the full potential of STSD for team sports analysis applications.

Deep learning, i.e., types of machine learning approaches such as neural networks characterized by high model complexity, is a promising form of data analysis within the domain of Artificial Intelligence (AI). Deep learning models are strong function approximators and have demonstrated unprecedented performance on classification and prediction tasks with complex data sources across a broad range of domains (Janiesch et al., 2021; LeCun et al., 2015; Schulz & Behnke, 2012). As a consequence, this class of algorithms has been singled out as a promising route to handle the challenges arising when analyzing STSD in team sports analysis (Dutt-Mazumder et al., 2011; Herold et al., 2019; Rein & Memmert, 2016; Tuyls et al., 2021). However, the increased complexity of deep learning models offers predictive power at the cost of several limitations. These limitations include low interpretability of models (they are commonly referred to as black box approaches), high requirements on data volume and variety, or high computational costs (Janiesch et al., 2021).

1.1 Current State of Research

Recent research in team sports analysis has started to incorporate deep learning for the processing of STSD. Deep learning has been successfully used to acquire and pre-process raw data (Manafifard et al., 2017; Omidshafiei et al., 2022), detect player actions on the pitch (Richly et al., 2017; Rodrigues et al., 2020), predict favorable outcomes of plays (Wagenaar et al., 2017; Yurko et al., 2020), simulate characteristic behavioral patterns (Le, Yue, et al., 2017; Schmid et al., 2021), or quantify goal-scoring likelihoods (Fernández et al., 2021; Sicilia et al., 2019). These works seem to justify the promise, and their comparison of deep learning models with less complex baselines reveal moderate to drastic performance increases in favor of the former.

The existing body of work shows promising results and indicates that deep learning can be a good fit for analyzing STSD. However, the research landscape only offers a collection of isolated applications at this point, which often do not address the general challenges of team sports analysis, and particular challenges involved with analyzing STSD. As a consequence, the proposed works often lack generalizability across sports (Ghosh et al., 2023), reproducibility (Bullock et al., 2022; Herold et al., 2019), or consistency across disciplines (Goes et al., 2021). Furthermore, only few works have directly addressed the limitations of deep learning models in their applications, including interpretability concerns (Bullock et al., 2022) or data sparsity (Ghosh et al., 2023; Müller et al., 2022).

In summary, a thorough analysis of the technological, methodological, and conceptual foundations for applying deep learning to STSD in team sports analysis is currently lacking (Rein & Memmert, 2016). There still remain many unanswered questions regarding optimal data representations for effective learning (Kovalchik, 2023), reasonable applications of black box approaches (Bullock et al., 2022), as well as the necessary technological requirements for the advancement of the field (Rein & Memmert, 2016).

1.2 Thesis Outline

The present cumulative dissertation addresses this research gap by thoroughly investigating *the usage of deep learning and spatiotemporal sports data for team sports analysis*. With regard to the identified limitations discussed above, five research questions are raised and answered through a series of peer-reviewed articles. Keeping in mind the complexity of the data source, the methods and the investigated processes, this thesis provides a systematic analysis of the advantages and disadvantages, opportunities and challenges, as well as prerequisites and pitfalls of embracing novel analysis techniques in order to push the limits of human performance in team sports. To this end, the following steps are undertaken:

- (i) The current scientific literature on the characteristics of team sports analysis, STSD and deep learning applications is reviewed,
- (ii) different data representations are tested empirically,
- (iii) multiple applications are created and evaluated with an emphasis on their usability for practical performance analysis,

- (iv) and extensive code and data bases are released to enable sustainable progression in the field.

The present work is structured as follows. Chapter 2 summarizes relevant background information on team sports analysis, spatiotemporal sports data, and deep learning, with a special emphasis on the challenges that need to be addressed. Subsequently, related work on the application of deep learning to spatiotemporal sports data for team sports analysis is reviewed in Chapter 3. Following this analysis, open research questions are identified in Chapter 4. The research contributions from 13 peer-reviewed articles are summarized in 5. All contributions are thoroughly discussed in Chapter 6 with respect to the research questions of this thesis. The corresponding articles can be found in the Appendix.

2. Background

To investigate team sports analysis by means of STSD and deep learning, its three components need to be established first. This chapter serves as an introduction to *team sports analysis* (Chapter 2.1), *spatiotemporal sports data* (Chapter 2.2), and *deep learning* (Chapter 2.3), by providing definitions, background information and associated research challenges. All identified challenges are summarized by Table 2.1.

2.1 Team Sports Analysis

Team sports analysis is an umbrella term that includes all analyses which attempt to summarize, model, explain, or predict performance-related aspects of *team invasion games*. The focus of the present work is on *quantitative analyses* that involve some form of data usage (e.g. STSD or event data, cf. Chapter 2.2). *Qualitative analyses*, such as conventional video analyses conducted by domain experts, are excluded from the scope of this work. Team invasion games include all time-based sports games where a team tries to send an object (ball, puck, flying disc) to a location (goal, basket, zone) as often as possible within a given time limit, while preventing an opponent from doing the same (Biermann et al., 2021; Lord et al., 2020). This includes games such as football (also known as association football or soccer), handball, basketball, American football, or hockey.

2.1.1 Applications

The volume and variety of stakeholders that may have an interest in data-informed team sports analysis is large (Morgulev et al., 2018; Patel et al., 2020). However, depending on primary stakeholder interests, a high-level overview of use cases becomes apparent. *Sports practitioners* includes all personnel of teams that directly participate in competitions, e.g., athletes, coaches, and staff members. Their main goal is to increase the competitiveness of a team by preparing for upcoming matches, improving the team roster, managing training and development, or monitoring performance. Potential use cases for data-informed analysis comprise match and performance analysis including pre-match and post-match analysis

(Lord et al., 2020; Low et al., 2021; Memmert et al., 2017; Rein & Memmert, 2016; Sarmento et al., 2014), design and management of training sessions (Fister et al., 2015), or talent identification, recruitment and development (Sarmento et al., 2018). Closely related are scientific applications in the area of *sports science and medicine*, which follow the general aim of understanding human performance and increasing athlete well-being. Corresponding applications are similar to those of sports practitioners and include assessment of training and competition demands including physiological and metabolic responses (Dellaserra et al., 2014; Torres-Ronda et al., 2022), predicting injury likelihood due to excessive intensities (Nassis et al., 2023), activity recognition (Ghosh et al., 2023), or movement recognition (Cust et al., 2019).

Furthermore, *industry* companies create technological solutions for practitioners and sports scientists. They also serve a wide range of markets that emerge around team sports, including sports broadcasting, fan engagement, or fantasy sports (Beal et al., 2019; Morgulev et al., 2018; Patel et al., 2020). For example, forecasting match outcomes (Wunderlich & Memmert, 2020) is of high relevance for the sports betting market (Horvat & Job, 2020). Other stakeholders include *public health institutions* or *policy makers* where data may aid management decisions (Morgulev et al., 2018; Patel et al., 2020). Team sports data also provide a valuable testbed and are used frequently by other scientific disciplines as model systems (Tuyls et al., 2021). For example, domains such as human trajectory modeling (Rudenko et al., 2020) or the study of multi-agent systems (Stone & Veloso, 2000), including humanoid robot development (Asada & Von Stryk, 2020), employ STSD for transdisciplinary research.

2.1.2 Challenges in Team Sports Analysis

Across the wide range of involved disciplines, use cases and methodologies, a series of overarching challenges can be identified. Some of these challenges are innate to the field, whereas others are more general data science challenges (Cao, 2018). A major concern raised continuously and in unison is the need for contextualized team sports analyses that integrate the multi-faceted information available, instead of evaluating performance metrics in isolation (Gudmundsson & Horton, 2017; Kovalchik, 2023; Low et al., 2021; Morgulev et al., 2018; Rein & Memmert, 2016; Sarmento et al., 2014; Stein et al., 2017). Given the many factors that influence human performance, this endeavor may be singled out as the major challenge of team sports analysis within the big data era. However, integrating and contextualizing analyses requires underlying conceptual frameworks (Glazier, 2010, 2017; Mackenzie & Cushion, 2013), to which the community has yet to agree (Rein & Memmert, 2016; Rein et al., 2017). This comes as no surprise, given that team sports are regarded as highly complex phenomena (Morgulev et al., 2018) and behavioral aspects that influence performance, such as joint group movement patterns, are notoriously difficult to understand and analyze (Dutt-Mazumder et al., 2011; Fujii, 2021; Lord et al., 2020; Stein et al., 2017). Proposed solutions to this challenge emphasize the importance of athlete interactions and advocate the use of dynamical systems theory (Lames & McGarry, 2007; Sarmento et al., 2014).

Besides these conceptual challenges, other issues are of a methodological nature. The lack of sound and widely adopted operational definitions has led to reduced comparability of studies in the domain of notational analysis (Mackenzie & Cushion, 2013; Sarmento et al., 2014) and beyond (Morgulev et al., 2018). Furthermore, efforts of studies to attempt and ensure replicability has been claimed to be subpar (Bullock et al., 2022; Herold et al., 2019; Sarmento et al., 2014) and several concerns regarding the validity and reliability of team sports analyses have been raised (Herold et al., 2019; Lames & McGarry, 2007; Torres-Ronda et al., 2022). In summary, these challenges may play a vital role in the limited applicability of research in practice (Herold et al., 2019; Low et al., 2021; Mackenzie & Cushion, 2013).

On more practical terms, it has been found that researchers conducting team sports analyses often come from different domains, e.g., sports science and computer science, with fundamentally different questions and methodologies (Goes et al., 2021). As a consequence, a range of proposed methods lack knowledge in the application domain (Herold et al., 2019; Rein & Memmert, 2016; Stein et al., 2017) and interdisciplinary research is highly favored (Goes et al., 2021). This is due to the high technological requirements which scale with the complexity of advanced team sports analysis algorithms. The demands on technological infrastructure for data acquisition, storage, processing and visualization are increasing (Goes et al., 2021; Mackenzie & Cushion, 2013; Rein & Memmert, 2016) and typically exceed the technical skill set of sports scientists (Rein & Memmert, 2016). Examples of such requirements are the many different sources and formats of data (Patel et al., 2020) that need to be integrated for (potentially multi-modal) analyses (Morgulev et al., 2018) as well as advanced programming skills (Goes et al., 2021; Rein & Memmert, 2016). Another prevalent issue is data availability (Cao, 2018). Limited sample sizes have traditionally been a concern, especially for notational analyses (Dellaserra et al., 2014; Mackenzie & Cushion, 2013). Although in theory, the big data era has led to an increase in data acquisition (Rein & Memmert, 2016), limited data is openly available for researchers as propriety interests often prevent its distribution (Fister et al., 2015; Kovalchik, 2023). Beyond data, other open science principles crucial for the advancement of the field, e.g., transparent reporting of methods, code sharing, or standardized data processes, are highly underdeveloped (Bullock et al., 2022; Herold et al., 2019; Rein & Memmert, 2016).

2.2 Spatiotemporal Sports Data

Spatiotemporal sports data (STSD) are data records that contain coordinate positions of athletes and potentially a playing device (e.g. ball, puck, or frisbee) which evolve over time. Other common terms used for this kind of data are *position data* or *tracking data*. STSD are collected by different *Electronic Performance and Tracking Systems (EPTS)* during training or competition. In contrast to related sports data sources such as event data or videos (see Stein et al., 2017, for an overview), STSD provide a highly granular and time-continuous description of sports games.

Following Raabe et al. (2023), STSD can be formally defined as a collection of trajectories given as two-dimensional Cartesian coordinates for a fixed period of time. More precisely, for a set $\mathcal{P} = \{1, 2, ..., N\}$ of N players and a fixed time point $t_0 \leq t \leq t_1$, player coordinates (x, y) in the Cartesian plane form a set of tuples

$$\mathcal{C}_t = \{(x_{t,1}, y_{t,1}), (x_{t,2}, y_{t,2}), ..., (x_{t,N}, y_{t,N})\}.$$

Three-dimensional descriptions can be constructed similarly, however, are less common in practice. Equivalently, the playing device can be denoted as $(b_{t,1}, b_{t,2})$. The union of these sets for multiple (discretized) time points results in a set $\mathcal{C}_{t_0,...,t_1}$ of trajectories for the given time sequence. This set includes the trajectory of an individual player i, i.e.,

$$\{(x_{t_0,i}, y_{t_0,i}), (x_{t_0+1,i}, y_{t_0+1,i}), ..., (x_{t_1,i}, y_{t_1,i})\} \subseteq \mathcal{C}_{t_0,...,t_1}.$$

2.2.1 Acquisition

Four major types of EPTS capable of acquiring STSD at varying quality (Lutz et al., 2020; Rico-González et al., 2020) can currently be identified. As some EPTS require athletes to wear sensors during competition, their usage is typically regulated by the federations of each sport. For example, the Fédération Internationale de Football Association (FIFA) has principally allowed the usage of EPTS in football (FIFA, 2015) and maintains a list of approved EPTS that passed a standardized quality test (FIFA, 2024). EPTS vary by their technology, temporal resolution, spatial accuracy, usability, ball tracking capability, and cost. Thus, the overall usage of different types of EPTS varies in practice (Rico-González et al., 2020). The selection of an EPTS does not only depend on the requirements of the respective use case, but also has a strong impact on the resulting STSD in terms of data quality and accuracy. It therefore remains important to acknowledge the specific peculiarities and systematic errors of the source EPTS when processing STSD.

Stationary Optical Tracking Systems (SOTS) use a set of cameras which are mounted to the respective venue architecture such as stadium roof constructions (Linke et al., 2020). The cameras cover fixed and overlapping portions of the playing surface and the respective images can be combined to full multi-views. Athlete and ball positions are calculated with the help of advanced computer vision algorithms (see Manafifard et al., 2017, for an overview) with good overall validity (Di Salvo et al., 2006; Linke et al., 2020; Redwood-Brown et al., 2012).

Dynamic Optical Tracking Systems (DOTS) also generate athlete and ball positions from video. Instead of using multiple stationary cameras, however, only a single movable camera such as a common television broadcast feed are used (Theiner et al., 2022). This technique is much simpler and does not require access to the venue. Instead, any available video feed can be used. This advantage renders DOTS extremely valuable for many stakeholders, as the data coverage of professional matches can be widely expanded. However, some players are regularly outside the camera's field of view and the limited depth perception of a single camera renders the exact determination of player positions much more difficult than with multiple perspectives (Omidshafiei et al., 2022; Theiner et al., 2022). As DOTS are a relatively new phenomenon, conclusive validation studies are currently lacking.

EPTS based on *Global Navigational Satellite Systems (GNSS)* such as the Global Positioning System (GPS) technology require players to wear sensors and determine their positions via satellite communication (Lutz et al., 2020). These systems are comparatively flexible and easy in their usage, and often contain additional measurement sensors. However, sampling frequency and validity vary strongly, especially for high intensity activities (Hoppe et al., 2018; Lutz et al., 2020; Scott et al., 2016). Data quality also depends on external factors such as satellite coverage or atmospheric conditions, and indoor usage is limited (Dellaserra et al., 2014).

Local Positioning systems (LPS) are also sensor-based, however, communicate via radio frequency signals with base stations mounted locally to the respective venue architecture (Lutz et al., 2020). Thus, infrastructural overhead is similarly high as for DOTS and the systems need to be carefully calibrated before usage. On the other hand, this allows for high sampling frequencies and good validity of LPS (Blauberger et al., 2021; Frencken et al., 2010; Hoppe et al., 2018; Seidl et al., 2016).

2.2.2 Challenges in Processing STSD

Sound algorithmic processing of STSD needs to consider substantial challenges involved, irrespective of the methods used for analysis. STSD vary in their spatial and temporal resolution and contain non-systematic measurement errors that can be directly linked to the source EPTS (Schmid & Lames, 2023). As a result, derived performance metrics may also vary strongly between systems (Hoppe et al., 2018; Linke et al., 2018; Randers et al., 2010). Correction can be performed by transforming derived metrics obtained from one EPTS to another via linear (Buchheit et al., 2014; Taberner et al., 2020) or non-linear regression (Schmid & Lames, 2023). However, this approach requires many transformations and a lot of ground truth data (Taberner et al., 2020), which increases the urgency to improve raw data quality. One approach to remove noise from raw STSD involves filtering, however, these approaches are currently understudied and of low prevalence (Ellens et al., 2022). Furthermore, limited evidence exists on how the temporal resolution of STSD may affect the precise calculation of resulting performance metrics (Corsie & Swinton, 2023).

The high spatial and temporal dimensionality of the data also poses problems by itself. Early works using machine learning and STSD identified the need for dimensionality reduction during processing in order to cope with data volumes and resulting information redundancies (Perl, 2011). Following this approach, researchers have used neural network techniques such as Self-Organizing Maps (SOM) to reduce the complexity of the data in football (Grunz et al., 2012), basketball (Kempe et al., 2015), and handball (Schrapf et al., 2017) applications. This observation has also been confirmed systematically, finding that a majority of published studies use spatial and temporal aggregation methods during analysis (Goes et al., 2021). In summary, algorithmic processing of STSD need to account for several data-related challenges regarding data quality, temporal and spatial resolution and data complexity.

2.3 Deep Learning

Deep learning describes a subset of machine learning methods, particularly Artificial Neural Networks (ANN), that distinguish themselves by their increased complexity and number of trainable parameters (Janiesch et al., 2021; LeCun et al., 2015). Corresponding *deep learning models* are *parameterized* models that can be *trained* to perform a specific *learning task*. Formally speaking, such a task can be defined as finding a *target function* mapping data from an *input space* (such as, e.g., images) to an *output space* (such as, e.g., image descriptions). In this scenario, learning means altering the model parameters in order to improve the approximation of such a target function. Depending on the usage of labelled information during the learning process, deep learning models can be broadly categorized into *supervised*, *semi-supervised*, and *unsupervised* methods. For example, annotated training data is used in supervised learning to calculate the difference between a predicted output and the true output (via so-called *loss functions*) for a given input, and use this information to iteratively improve the prediction (Goodfellow et al., 2016).

Whereas "classical" or "shallow" machine learning models typically consist of only a few processing layers (Jordan & Mitchell, 2015), deep learning models are typically constructed by concatenating many layers or smaller computational blocks (Janiesch et al., 2021; LeCun et al., 2015). This results in large architectures, with potentially billions of parameters. The general idea guiding the construction of such large models is to assume a hierarchically structured processing of data, where subsequent blocks learn data features of increasing complexity (LeCun et al., 2015; Schulz & Behnke, 2012). This idea is based on the reductionist assumption that the learned concepts can be disassembled hierarchically, where high-level concepts are composed of lower-level concepts.

One major breakthrough of deep neural networks can be dated to the year 2012, when deep models outperformed shallow models in several image and speech recognition benchmark tasks (Ciresan et al., 2012; Hinton et al., 2012; Krizhevsky et al., 2017). Since then, the combination of architecture design, increased model complexity, and better computational resources has proven unmatched effectiveness for many applications (LeCun et al., 2015; Schulz & Behnke, 2012), including some team sports analysis tasks (Cust et al., 2019). This development has also led to an uncountable number of proposed methods (Schmidhuber, 2015), most prominently ranging from Convolutional Neural Networks (CNN) for image-processing (LeCun et al., 1998; LeCun et al., 2015), Recurrent Neural Networks (RNN) such as Long Short Term Memory Networks (LSTM) (Hochreiter & Schmidhuber, 1997) for sequential data processing, Graph Neural Networks (GNN) (Battaglia et al., 2018; Bronstein et al., 2017) for non-Euclidean data processing, or transformers (Vaswani et al., 2017) for versatile representation learning.

2.3.1 Strenghts and Limitations

The primary advantage of deep learning models, to which they also owe their popularity, is their strong predictive power. In theory, methods such as feed-forward ANN can approximate any target function arbitrarily close given enough trainable parameter, including complex non-linear relationships (Goodfellow et al., 2016). Furthermore, deep learning can handle large amounts of data and is very flexible adapting to complex input and output spaces. Deep learning models can process structured and unstructured data, sequential data, or handle missing data. In combination, this makes deep learning strong end-to-end learners, i.e., algorithms that can learn target functions from a wide range of input spaces with minimal prior assumptions or human intervention (Janiesch et al., 2021).

The complexity of deep learning models with respect to the high amount of trainable parameters leads to strong predictive power, but this also comes at a cost. Most notably, deep learning models are inherently intransparent, as the approximated target functions are difficult to comprehend for humans (Murdoch et al., 2019; Zednik, 2019). As a consequence, they are often regarded as *opaque* or *black box approaches*. Furthermore, deep learning methods are primarily predictive (McCall et al., 2017), i.e., they do not provide associative results or explanations for causal relationships between variables out of the box. How to effectively use these predictive models for scientific advancement remains a major question in current AI research (Wang et al., 2023).

Furthermore, the demands on training data to train deep learning models can be enormous, both in terms of quality and volume. Parameter tuning happens during training, and the number of parameters scales with the amount of training data needed for successful adaptation. Furthermore, to ensure generalizability on learning tasks, training data need to be representative of the input space. Closely related is the fact that model training requires vast amounts of computational power, reflected by high hardware costs and potentially long training times.

The end-to-end learning capabilities of deep learning models is a strong advantage, but also leads to a series of problems. Little prior assumptions on input spaces and target functions lend deep learning its flexibility and convenience in applications, but lack of human intervention also means a lack of control. This causes severe ethical and safety concerns, as training data characteristics scale to the trained model, potentially without notice, causing unwanted side-effects (Ntoutsi et al., 2020; Rudin, 2019).

2.3.2 Challenges in Deep Learning Applications

The discussed advantages and disadvantages of deep learning pose a series of challenges that need to be addressed when using deep learning for team sports analysis. A major current research questions is finding appropriate minimal assumptions on input and target function spaces in form of inductive biases (Mitchell, 1980) that shape the learned approximations (Battaglia et al., 2018; Bronstein et al., 2021). A major lever in this regard is the choice of appropriate data representation, which is considered the most important questions in applying deep learning to team sports analysis by some researchers (Kovalchik, 2023). Furthermore, black box approaches with low interpretability are not ideal candidates to explain relationships (Herold et al., 2019). Especially if not reported properly, these models cannot be easily validated, and caution should be exerted, e.g., when designing interventions (Bullock et al., 2022). Researchers have also raised concerns about the generalizability and transferability across sports in team sports analysis (Ghosh et al., 2023; Gudmundsson & Horton, 2017). Additionally, more general team sports analysis challenges coincide with deep learning limitations, including data availability and technological requirements.

Table 2.1: Summary of identified challenges with respect to their domain.

Abbr.	Challenge	Discussed in
Team sports analysis		
C1	High complexity in team sports analysis and movement patterns	Dutt-Mazumder et al., 2011; Fujii, 2021; Lord et al., 2020; Morgulev et al., 2018; Stein et al., 2017
C2	Lack of theoretical framework and conceptual unclarity	Glazier, 2010, 2017; Mackenzie and Cushion, 2013
C3	Necessity of contextualization	Gudmundsson and Horton, 2017; Kovalchik, 2023; Low et al., 2021; Morgulev et al., 2018; Rein and Memmert, 2016; Sarmento et al., 2014; Stein et al., 2017
C4	Need for emphasis on interactions	Lames and McGarry, 2007; Sarmento et al., 2014
C5	Lack of sound operational definitions	Mackenzie and Cushion, 2013; Morgulev et al., 2018; Sarmento et al., 2014
C6	Possibility of and efforts for replicability	Herold et al., 2019; Sarmento et al., 2014
C7	Integration of domain knowledge needed	Goes et al., 2021; Herold et al., 2019; Rein and Memmert, 2016; Stein et al., 2017
C8	Technological requirements	Goes et al., 2021; Mackenzie and Cushion, 2013; Morgulev et al., 2018; Patel et al., 2020; Rein and Memmert, 2016
C9	Small sample sizes and lack of (openly available) data	Fister et al., 2015; Kovalchik, 2023; Mackenzie and Cushion, 2013
C10	Validity and reliability concerns	Herold et al., 2019; Lames and McGarry, 2007; Torres-Ronda et al., 2022
C11	Research-practice gap	Herold et al., 2019; Low et al., 2021; Mackenzie and Cushion, 2013
STSD analysis		
C12	Data quality concerns	Hoppe et al., 2018; Linke et al., 2018; Randers et al., 2010; Schmid and Lames, 2023; Taberner et al., 2020
C13	Need for filtering	Ellens et al., 2022
C14	Problems with low temporal resolution	Corsie and Swinton, 2023
C15	Need for abstraction and aggregation	Goes et al., 2021; Perl, 2011; Stein et al., 2017
C16	Need of methodological and theoretical guidelines for modeling	Rein and Memmert, 2016
Deep Learning		
C17	Selecting appropriate data representation	Kovalchik, 2023
C18	Interpretability	Bullock et al., 2022; Herold et al., 2019
C19	Generalizability	Ghosh et al., 2023; Gudmundsson and Horton, 2017

3. Deep Learning with Sports Data

This chapter discusses related work where deep learning techniques have been applied to analyze STSD for team sports analysis. An overview of reviewed articles can be found in Table 3.1. The corresponding scope is limited to the definitions established in Chapter 2. Works are included only if *(i)* the application is directly involved with team sports analysis, *(ii)* STSD are used as the primary data source, and *(iii)* deep learning methodologies are adapted. The first point excludes deep learning applications primarily involved with acquisition of STSD (cf. Manafifard et al., 2017; Theiner et al., 2022). The last point particularly excludes machine learning methods that are "shallow", such as smaller ANN, SOM, k-Nearest Neighbors (kNN), Support Vector Machines (SVM), or tree-based methods (cf. Herold et al., 2019).

Table 3.1: Summary of related work with respect to applications.

Work	**Sample**	**Task**	**Methods**
Action Recognition			
Richly et al. (2017)	Football (4 matches)	Detection of passes, receptions clearances, and shots on target	ANN
Rodrigues et al. (2020)	Futsal (4 matches)	Detection of running, passing, and shooting activity	ANN, LSTM, DBMM
Trajectory Prediction			

Continued on next page

Table 3.1 – continued from previous page

Work	Sample	Task	Methods
Felsen et al. (2018)	Basketball (95,002 sequences)	Predicting trajectories of players	cVAE
Yeh et al. (2019)	Football (7500 sequences), Basketball (120.991) sequences)	Predicting trajectories of players and ball	GNN, vRNN
Alcorn and Nguyen (2021)	Basketball (632 matches)	Predicting trajectories of players and ball	Transformer
Hauri et al. (2021)	Basketball (632 matches)	Predicting trajectories of players	LSTM
Data Imputation			
Omidshafiei et al. (2022)	Football (105 matches)	Impute partially observed trajectories	GNN, vRNN
Imitation Learning			
Le, Yue, et al. (2017)	Football (45 matches)	Learn policies for defensive behavior	LSTM
Le, Carr, et al. (2017)	Football (100 matches)	Learn characteristic defensive behaviors	LSTM
Seidl et al. (2018)	Basketball (30.764 sequences)	Learn characteristic defensive behaviors	LSTM

Continued on next page

Table 3.1 – continued from previous page

Work	Sample	Task	Methods
Lindström et al. (2020)	Football (79 matches)	Learn characteristic offensive behaviors	LSTM
Schmid et al. (2021)	Am. Football (unspecified)	Learn characteristic defensive movements	LSTM
Prediction			
Wagenaar et al. (2017)	Football (29 matches)	Predict goal scoring opportunities	CNN
Yurko et al. (2020)	Am. Football (14.167 plays)	Predict yard gainage of plays	LSTM, ANN
Stival et al. (2023)	Football (10 matches)	Predict entries to attacking zone	CNN
Probability estimation			
Sicilia et al. (2019)	Basketball (134.000 sequences)	Estimate expected point outcome of offensive sequences	LSTM
Fernández and Bornn (2021)	Football (740 matches)	Estimate probability of pass locations	CNN
Fernández et al. (2021)	Football (633 matches)	Estimate probability of scoring or conceding goal	CNN
Müller et al. (2022)	Handball (15 matches)	Estimate probability of scoring	LSTM, Transformer

3.1 Action Recognition

Action recognition refers to the task of detecting individual player actions (such as passes, shots, or tackles) from a given data source. This application is of high relevance as it attempts to generate event data automatically from other data sources, whereas the manual acquisition of event data is both time consuming and potentially subjective (Bassek, Raabe, Banning, et al., 2023; Cust et al., 2019). Although a large majority of proposed approaches has been conducted on video data (Cuevas et al., 2020; Mahaseni et al., 2021), some works have suggested to use STSD as a resource for action recognition. For example, Richly and colleagues (Richly et al., 2017) used neural networks to recognize frequent actions in football matches (passes, receptions, clearances and shots on target). STSD were enriched with hand-engineered features (e.g. velocities, accelerations, changes of direction) and compared to manually collected event data. F1 scores were reported to vary between 0.89 for matches recorded at 10 Hz and 0.49 for matches recorded at 25 Hz. In Futsal, Rodrigues and colleagues (Rodrigues et al., 2020) compared a standard ANN, an LSTM and a dynamic Bayesian mixture model (DBMM, an ensemble classifier composed of naive Bayes, kNN, ANN) to recognize player activity and actions (walking, running, passing, shooting) enriched with physiological data (muscle activity). As STSD and video data vary in their information content, multi-modal approaches could be a promising route for future research in action recognition.

3.2 Trajectory Prediction and Data Imputation

An important field of research that extends well beyond the sports domain is *trajectory prediction* of (potentially multiple) objects, especially humans (Rudenko et al., 2020). There is a large body of work that has used sports as a model system to evaluate predicted trajectories generated by deep learning architectures. This body exceeds the scope of this work, as this application is not primarily concerned with team sports analysis tasks. However, it should be noted that the range of used deep learning methods is quite diverse, including conventional LSTMs (Hauri et al., 2021), GNNs and variational RNN (Yeh et al., 2019), conditional Variational Autoencoders (VAE)s (Felsen et al., 2018), and transformer architectures (Alcorn & Nguyen, 2021).

Trajectory prediction furthermore plays a crucial role for STSD pre-processing in terms of *data imputation*. As STSD may contain longer sequences of missing player positions, particularly partial recordings coming from DOTS, imputation methods are required to create continuous data streams. Compared to classical and simple machine learning methods for data imputation (Kontos & Karlis, 2023; Long, 2016), recent works have demonstrated superior performance achieved by deep learning models. For example, Omidshafiei and colleagues (Omidshafiei et al., 2022) proposed a Graph Imputer architecture which combines LSTMs with GNNs and variational RNNs (Chung et al., 2015). The Graph Imputer outperformed classical imputation models as well as less complex deep learning models with respect to the distances between true and predicted trajectories in football. On a side note, similar results have also been found when predicting player trajectories based off event data (Everett et al., 2023).

3.3 Imitation Learning

Imitation Learning (IL) describes a group of algorithms which are trained to imitate a specific expert labelling behavior (Hussein et al., 2018). In IL settings, the learning task is typically designed as a Markov Decision Processes (MDP), where a *policy* π is derived that can optimally predict the best *action* for a given *state*. In contrast to Reinforcement Learning (RL), optimization is not performed via reward functions, but by evaluating the proximity of suggested policy actions to expert demonstration, i.e., actual data. The first work that extended IL to STSD goes back to Le and colleagues (Le, Yue, et al., 2017), who proposed an algorithm based on LSTMs to imitate defensive team behaviors in football derived from defensive sequences of 45 actual matches. This approach was later applied by deriving and evaluating characteristic behavioral responses to given match situations in football (Le, Carr, et al., 2017; Lindström et al., 2020), basketball (Seidl et al., 2018), and American football (Schmid et al., 2021).

3.4 Prediction and Forecasting

The *prediction* of in-game situations has also received considerable attention. These approaches are typically defined as classification problems, where a short sequence of match play is used to predict the future event that will most likely happen. For example, Wagenaar and colleagues (Wagenaar et al., 2017) used several CNNs to predict whether sequences of ten seconds will result in a goal scoring opportunity or loss of ball possession in football, outperforming kNN baselines. Similar works have used CNNs to predict whether a football team will reach the attacking quarter of the pitch given the first five seconds of each ball possession phase (Stival et al., 2023) or time-continuously predict the yards gained by American football plays at different moments in time (Yurko et al., 2020).

In terms of match outcome *forecasting*, other data sources with higher availability are usually preferred over STSD. However, STSD based analyses may benefit the performance in outcome prediction. This depends on the availability as these predictions typically rely on large datasets with less deep information available. This leads to a depth versus breath problem, which is explored in detail by Wunderlich and colleagues (Wunderlich et al., 2024).

3.5 Probability Estimation

Similar to the prediction of certain in-game events occurring during team invasion games is the task of *probability estimation*. This approach has the advantage that it allows to calculate expected outcomes and what-if analyses of sports plays. Originally, Cervone and colleagues have proposed a stochastic framework for estimating the expected outcome of play sequences in basketball (Cervone et al., 2016). Based on this work, other researchers have used deep learning techniques to derive (parts of the) probability estimations. For example, Fernandez and Bornn (Fernández & Bornn, 2021) used CNNs to derive two-dimensional probability surfaces that map the likelihood of pass locations in football. This approach was subsequently extended to estimate the more general probability of scoring or conceding goals of a team (Fernández et al., 2021). Similar approaches have used LSTMs and transformers to estimate the scoring probability of offensive sequences in basketball (Sicilia et al., 2019) and handball (Müller et al., 2022).

3.6 Benchmark Datasets

The evaluation of computationally expensive deep learning models often incorporates a trade-off between performance and complexity. As a consequence, proposed models are usually tested on characteristic tasks performed on public benchmark datasets to demonstrate the power of a novel method. For example, the SoccerNet-v2 dataset (Deliege et al., 2021) contains a large corpus of annotated video material from football. The accompanying SoccerNet challenge includes tasks such as action spotting or shot boundary detection (Giancola et al., 2022). Similar datasets and challenges exist for basic match meta information (Berrar et al., 2019; Dubitzky et al., 2019) or event data (Pappalardo et al., 2019; StatsBomb, 2023).

However, this common practice in deep learning research has not yet been fully established with respect to STSD. Only few datasets are available and their permanent availability is not always secured (NFL, 2023; Pettersen et al., 2014; SkillCorner, 2023). A potential reason for this is the restricted access of proprietary data collected during elite team sports competitions (Fister et al., 2015; Kovalchik, 2023). Data owners such as clubs or organizations might be hesitant to share data publicly as their acquisition is still costly and sharing could give away information they perceive as increasing their competitiveness. Nevertheless, the lack of publicly available STSD and respective team sports analysis tasks poses a strong limiting factor on the development of the field (Cao, 2018; Rein & Memmert, 2016). This circumstance could also serve an explanation as to why current works focus on isolated applications which are rarely compared to existing approaches (Herold et al., 2019).

4. Thesis Synopsis

Based on the findings in Chapter 2 and 3, the present chapter identifies important research gaps and derives the thesis synopsis (see Table 4.1 for an overview). Recent works discussing deep learning and STSD for team sport analysis have raised a series of difficulties and paths for future work that provide a clear outline of the current research frontier. Additionally, challenges connected to team sports analysis (Chapter 2.1.2), STSD (Chapter 2.2.2), and deep learning (Chapter 2.3.2) need to be considered for a successful application of deep learning algorithms (cf. Table 2.1).

4.1 Overall Research Questions

The *overall research question (ORQ)* of this thesis can be stated as follows:

> *ORQ: How can novel deep learning methods be effectively used for analyzing spatiotemporal sports data for team sports analysis?*

To provide an answer to the *ORQ*, the following, more detailed research questions (*RQ*) are addressed by the published works. Each research question covers a certain facet of the *ORQ*, such as methodological and conceptual (*RQ1, RQ2, RQ3*), practical (*RQ4*), and technological aspects (*RQ5*) of deep learning with STSD in team sport analysis.

4.2 Research Question 1

> *How can novel deep learning methods be effectively used for analyzing spatiotemporal sports data for team sports analysis?*

A fundamental question concerns the general design choices involved in the construction of large deep learning architectures. Researchers have mourned a general lack of modeling guidelines (*C16*), and this circumstance is likely amplified by an increase of model complexity. It is known that team sport behaviors such as elaborate movement patterns are innately complex (*C1*) and dependent on interactions (*C4*), but also that abstraction and aggregations are needed to draw inferences from STSD (*C15*). Principally, deep learning appears like a promising candidate to connect complex data with complex phenomena, due to its end-to-end learning capabilities on unstructured domains. However, classical team sport analysis has repeatedly stressed the importance of contextualized analyses (*C3*) and inclusion of domain knowledge (*C7*). Thus, it remains unclear how much additional prior knowledge needs to be introduced to the learning algorithm to achieve adequate performance. These considerations are subsumed in *RQ1* and discussed in Study I and II.

4.3 Research Question 2

> *What are suitable data representations for effective deep learning with STSD, especially with respect to computational cost and training data requirements?*

The practical implications of *RQ1* are addressed in *RQ2* and empirically evaluated in Studies I, II, and VIII. More specifically, *RQ2* addresses the choice of optimal data representations for STSD in deep learning applications. Unsuitable choices may induce unfavorable biases that impact generalizability (*C19*) and may drastically increase training data requirements and computational costs (Battaglia et al., 2018; Bronstein et al., 2017). This is problematic, as sample sizes are traditionally limited within team sports analysis (*C9*). Therefore, deriving and testing suitable data representations remains an important question and has even been stated as the main challenge for deep learning in a sports context (*C17*) by Kovalchik (2023).

4.4 Research Question 3

> *What is the relation between human and machine performance in the detection and interpretation of complex movement patterns?*

Whereas *RQ1* and *RQ2* address methodological aspects of the data and methods, the third research question (*RQ3*) is aimed at the team sport analysis concepts to be learned. As even "simple" individual movements such as sprints or changes of directions sometimes lack sound operationalization (*C5*), similar challenges may be anticipated for complex group movements (*C1*), such as sophisticated attacking strategies. Poor operationalization, in turn, may seriously affect the validity and reliability of the trained models (*C10*). *RQ3* addresses this operationalization issue by investigating the quality of ground truth labels and their relation to model performance in Studies II and VIII.

4.5 Research Question 4

> *How can non-interpretable deep learning methods be optimally used in team sports analysis?*

Answering the methodological aspects of deep learning raised by the first three research questions may lead to more effective deep learning models. However, it does not guarantee successful practical application. There still remains a noticeable research-practice gap (*C11*), and narrowing this gap constitutes the ultimate measure of deep learning's success in the sports domain. In this regard, the often criticized low interpretability of opaque approaches (*C18*) can be regarded as highly problematic. *RQ4* elaborates on optimal applications of deep learning that can provide value for sports scientists and practitioners. The question is answered with respect to the applications designed in Studies II and VIII.

4.6 Research Question 5

> *What are eminent technological challenges that need to be addressed and how can open science principles be used to foster accessible and reproducible research?*

A series of technological challenges faced in team sports analysis can be observed as a prerequisite for successful deep learning applications. These challenges include high technological requirements for the implementation of processing pipelines (*C8*), lack of openly available data (*C9*) and replication efforts due to limited code sharing (*C6*). In summary, the general conditions for successful implementations of deep learning models are less than ideal. *RQ5* collectively addresses these circumstances by calling on open science practices and their implementation in Studies III, IV, V, VI, VII, VIII, and X.

Table 4.1: Summary of the research questions discussed in this thesis.

Abbr.	**Question**	**Discussed in**
ORQ	*How can novel deep learning methods be effectively used for analyzing spatiotemporal sports data for team sports analysis?*	
RQ1	What are the characteristics of team sports analysis and STSD that may have an impact on the design of deep learning applications?	Study I & II
RQ2	What are suitable data representations for effective deep learning with STSD, especially with respect to computational cost and training data requirements?	Study I, II, & VIII
RQ3	What is the relation between human and machine performance in the detection and interpretation of complex movement patterns?	Study II & VIII
RQ4	How can non-interpretable deep learning methods be optimally used in team sports analysis?	Study II & VIII
RQ5	What are eminent technological challenges that need to be addressed and how can open science principles be used to foster accessible and reproducible research?	Study III-VIII & X

5. Publications

A number of articles were published or are currently under review to answer the outlined research questions. Their contributions are summarized below. Article abstracts and references can be found in the Appendix.

5.1 Summary of Findings

The overall research project presented in this thesis was originally initiated with an investigation on the usage of different STSD representations for deep learning applications (Raabe et al., 2023). To this end, three major data representations used for team sport analysis were identified from the literature: (i) hand-engineered features, (ii) state vectors, and (iii) images. Next, three important characteristics of STSD processing for team sport analysis were derived and discussed. First, group movement patterns that constitute tactical behaviors in team games may likely be *invariant* under isometric transformations such as translations, reflections, rotations, or permutations. Second, these patterns can predominantly be characterized by player *relations* rather than absolute positions. Third, their hierarchical organization suggests some sort of *compositionality*.

Following these theoretical considerations, a fourth data representation based on graphs was proposed in the same article (Raabe et al., 2023). This representation was motivated by the extracted data characteristics, the importance of relational inductive biases (Battaglia et al., 2018) and the success of non-Euclidean data representations (Bronstein et al., 2017). Furthermore, a corresponding deep learning approach termed *Tactical Graph Networks (TGNet)* was proposed to process these graph representations. TGNet is a lightweight and hybrid neural network that can ingest STSD as well as domain knowledge for team sport related classification and prediction tasks.

Empirical testing comprised an extensive ablation study, where TGNet and its components were optimized. Furthermore, a state-of-the-art comparison tested all four data representations and six corresponding learning algorithms on a generic classification task based on $N = 34$ football matches. Results showed that some data representations (hand engineered features and state vectors) and corresponding methods fall short in performance. State-of-the-art performance was achieved by TGNet as well as large CNN working with image representations. However, these CNN contain around 100 times more parameters than TGNet and showed considerably longer training and inference times. In summary, the study demonstrated that exploiting characteristics of STSD and team sport analysis is crucial in deep learning model design and provided strong evidence that graphs are an optimal STSD representation.

Based on the promising results from Study I, the aim of Study II was to apply TGNet in a performance analysis task with practical impact (Bassek, Raabe, Banning, et al., 2023). Manual label generation in team sport analysis is extremely costly, but needed for contextualized performance analyses (Cust et al., 2019). Deep learning, in turn, may be an appropriate method to mimic expert labeling behavior to automatically generate labels.

To test this empirically, handball possession and activity phases (inactivity, counter attacks, position attacks) were manually annotated for a set of 10 matches. Annotations were compared with two additional raters confirming very high inter-rater agreement ($\kappa = .93$). A TGNet variant was subsequently trained and optimized to predict manual annotations from STSD. Evaluation on a test match revealed 86.2% balanced accuracy, which was deemed sufficient for further processing. After successful training, the deep learning model was used to automatically generate labels for $N = 539$ matches of the German Men's Handball Bundesliga (HBL). Match intensity metrics were calculated with respect to different activity phases, and statistical testing confirmed the initial hypothesis.

Within the scope of this thesis, the actual performance analysis results are of less relevance. More interesting, instead, is the confirmation that deep learning can be used effectively to generate labels for contextualized performance analysis fully automatically. Compared to related work, this approach was able to increase the used sample sizes of annotated data from 15-90 matches to over 500 matches. In this regard, it was demonstrated that deep learning can enable analyses that are infeasible with conventional methodology.

In reminiscence to Study II, a similar task was addressed in Study VIII, i.e., to learn automatic match phase detection based on hand annotations in football (Biermann et al., 2024, under review). The derived annotation scheme was extended from basic match phases to include playing styles, and multiple deep learning models were optimized and compared, confirming the results discussed above. The best performing model was furthermore evaluated on an exemplary performance analysis task, i.e., to detect formations, and released to the public.

During the completion of the empirical studies discussed above, it became evident what had previously been reported abundantly by other researchers: the technological requirements on team sport analysis are extensive and manifold. As a first and fundamental step to meet these requirements, an open source software package written in Python was created and released in Study III (Raabe et al., 2022). The package itself contains relevant algorithms for all parts of a processing pipeline, including data loading for more than ten different input data formats, direct access to publicly available data sets, essential data processing techniques (e.g., transformations, sequencing, filtering), basic visualization methods, and a selection of widely used analysis algorithms. Additionally, extensive documentation including multiple tutorials was created to provide easy access and multiple entry points.

The overall aim of Study III was to enable sport scientists with limited programming experience to implement advanced data analysis procedures, and provide a platform as well as standardized recipes for code sharing among scientists. The package has been widely adopted, including several professional sports clubs, companies, and research groups worldwide, and recently received funding by the Deutsche Forschungsgemeinschaft (DFG, ME 2678/43 1) for further development.

The effectiveness of floodlight with regard to simplifying team sports analysis projects was also demonstrated in later work. Two subsequent studies, Study IV and Study V, made extensive use of the package to load, pre-process, analyze, and visualize data in football and handball. More precisely, new performance metrics were proposed in Study IV to study spatial aspects of relative player positioning during attacks (Raabe et al., 2024). In Study V, the energy expenditure of elite male handball players in the HBL was examined (Bassek, Raabe, Memmert, & Rein, 2023)

In addition to Study III, two further technological requirements of team sports analysis have been addressed. First, a taxonomy of individual player actions in team invasion games was proposed in Study VI (Biermann et al., 2021). This taxonomy provides a hierarchical, minimal, exact and modularly expandable categorization of such actions. It can be used for creating well-defined ground truth annotations needed for deep learning tasks such as action recognition in video data or STSD. Additionally, the work publishes annotated football and handball datasets containing 125 minutes of match play separated in five sequences of five minutes from five different games, respectively. The *Events in Invasion Games Dataset for Soccer (EIGD-S)* contains video from FIFA World Cup competitions and hand annotations from multiple annotators following the presented taxonomy. The Events in Invasion Games Dataset for Handball (EIGD-H) contains video from the HBL, synchronized STSD, as well as hand annotations analogously to the football dataset. Both datasets serve as valuable ground truth information for deep learning tasks such as action spotting. Furthermore, they provide insights into rating agreement between multiple experts and a non-expert for seemingly basic annotation tasks.

Second, a synchronization algorithm was developed in Study VII to temporally align event data and STSD for joint processing (Biermann et al., 2023). Although data integration is highly relevant and a pre-requisite for multi-modal processing, time-scale differences and annotation inaccuracies prevent an easy fusion of STSD with event data. The proposed algorithm performs action recognition via time series motif discovery in STSD and subsequent candidate selection given the event data. Results showed significant improvements in event data timestamps, as the fraction of passes within a half-second error margin increased from 14% to 70%. During the development of this algorithm, it became evident that time series data mining algorithms also constitute an interesting class of algorithms for team sport analysis, as most sports data contain a temporal dimension. To identify the potential in this regard, a literature review was conducted in Study X (Komitova et al., 2022).

The works discussed above address the main questions raised in Chapter 4. Besides these works, other projects have been conducted that do not directly fit the scope of this thesis. These include an empirical evaluation on how analysts use sports data to conduct opponent analyses in football (Mehta et al., 2024), and how coaches perceive such reports (Furley et al., 2024). Furthermore, the relation between performance metrics based on STSD and betting odds was investigated (Schlenger et al., 2023). Lastly, data imputation techniques and LSTM models were used for improved match outcome prediction in the case of missing data (Wunderlich et al., 2024).

6. Discussion

The present chapter discusses the overall research project with respect to the thesis synopsis derived in Chapter 4. To this end, the findings of the research articles summarized in Chapter 5 are related to the stated research questions (cf. Table 4.1) in Chapter 6.1. Furthermore, important limitations (6.2) and ethical considerations (6.3) of the research project are outlined. The practical applicability of the work is evaluated (6.4) and possibilities for future studies are highlighted (6.5).

6.1 Reply to Research Questions

The first research question is concerned with characteristics of team sports analysis and STSD relevant for the design of deep learning applications. The theoretical part of Study I has identified multiple characteristics of STSD and team sports analysis, and their importance for effective learning were demonstrated by empirical results. Furthemore, the inclusion of domain knowledge in hybrid network architectures has shown improved performance during model ablation. These results underpin the following reply to *RQ1*:

> *Reply to RQ1*: Important characteristics of STSD and team sports analysis (specifically: invariances of movement patterns, relationality of player interactions, compositional structure of tactical organization, and a priori inclusion of domain knowledge) have an impact on the design of deep learning architectures and should be considered as guidelines during model implementation.

The second research question asks about effective STSD representations for processing in deep learning models. The state of the art comparison presented in Study I as well as derived applications in Studies II and VIII provide strong evidence on the choice of graphs as an optimal data representation. Furthermore, whereas only a single reviewed study (Table 2.1) applied their model to more than a single team sport, the present work explicitly tested proposed data representations and models on handball and football datasets to ensure generalizability (*C19*).

> *Reply to RQ2*: Graph representations are a suitable and numerically effective data representation for STSD in deep learning applications. The corresponding models based on geometric deep learning show high performance compared to similar approaches and results indicate that they require smaller sample sizes and less computational complexity in team sports analysis settings.

The third research question aims at the relation between human and machine performance during assessment of complex movement patterns. With respect to the designed applications and conducted experiments in Studies II and VIII, evidence was collected on human interpretation of complex team sports analysis phenomena and their influence on learning performance. As supervised deep learning models base their inference on the structures derived from the training data, sound operationalization of concepts remains crucial. The results confirm that inter-rater agreement can be used as a measure to assess clarity in ground truth labels (Study II). However, they also demonstrate that increasing subjectivity in team sports analysis concepts correlates with prediction performance (Study VIII). Taken together, the following reply can be stated:

> *Reply to RQ3*: (Supervised) Deep learning models may be leveraged to learn high-level team sports analysis concepts from data, however, inference quality can be linked to the (un)ambiguity of ground truth labels used during training. As the models don't provide an assessment of ground truth label quality per se, sound operationalization of concepts and external validation measures such as inter-rater agreements are vital to ensure the usability of derived models.

The fourth research question looks at the challenge of interpretability in deep learning applications. Studies II and VIII both use deep learning to automatically derive match phases in handball and football that can be used for contextualized performance analysis on large datasets. Transferring the results on *RQ1* through *RQ3* into practical settings, the results demonstrate how deep learning can be leveraged to design applications with a strong benefit for sports scientific research and practical sports settings. Both models work fully automatic by delivering insights based on raw data ingestion with minimal human intervention needed. Such automation tasks draw on the strengths of deep learning, while at the same time its limitations are of less relevance. More precisely, the predictive power provides performance superior to classical or shallow approaches, whereas lowered explanatory power and low interpretability are easily neglected by task design. To sum up, this leads to the following reply to *RQ4*:

> *Reply to RQ4*: Deep learning may be used most effectively in tasks that require

models with high predictive power, such as automation, label generation for large-scale contextualized team sports analyses, or forecasting. On the contrary, tasks that require high explanatory power or interpretability, such as discovery of causal relations or intervention design, may be less ideal without additional steps undertaken.

The fifth research question is concerned with the technological challenges connected to team sports analysis, and how open science principles can help to increase the accessibility and reproducibility of related research. With respect to the technological challenges faced by the field, Studies III-VIII, and X contained important contributions. Study III released an extensive team sports analysis framework for easier processing of STSD and related data sources. Its quick adoption within the field confirms the necessity of such a common-ground coding platform.

Furthermore, its contribution to reproducible research and lowered technical skill requirements for team sports analyses became evident during Study IV and V. Study VI provides a flexible and exhaustive taxonomy of individual actions performed in team sports. Additionally, important datasets for action recognition and other team sports analysis tasks were released that were previously missing (cf. Chapter 3.6). Study VII proposed and evaluated methodology for synchronization of multiple data sources which is an important pre-processing step and pre-requisite for sound multi-modal analyses. Last but not least, Study VIII releases the best trained deep learning models for automatic recognition of football match phases, which are an important contextualization feature and can be used by other researchers in future works.

> *Reply to RQ5*: Important technological challenges in the field (no general open source coding platform available, lack of openly available datasets, lack of essential pre-processing methodologies, little transparent publication of trained models) were identified and tackled by following open science principles. Following these open science principles allowed to address important challenges and provide the conditions needed for an advancement of the field.

The insights on the five derived research questions provide the basis for a summarizing answer to the overall research question, which asks how deep learning can be effectively used to analyze STSD for team sports analysis:

> *Reply to ORQ*: Novel deep learning methods provide unprecedented value for STSD analysis in team sports. To use deep learning methods effectively, they need to be designed appropriately by respecting domain and data characteristics that directly influence the quality and cost of deep learning models. In

this regard, graphs and (hybrid) geometric deep learning provide a promising class of algorithms. Sound operationalization and validation of team sports analysis concepts, as well as learning task designs that optimize the predictive versus explanatory power trade-off associated with deep learning are equally important for effective application. Last but not least, boundary conditions for deep learning in the team sports analysis domain are currently subpar compared to other application domains, and open science practices are needed to advance the field.

6.2 Limitations

A series of overarching limitations concerning the presented studies need to be discussed. One major limitation concerns the volume of training data used to train neural network architectures (Studies I, II, & VIII). It should be noted that TGNet models were specifically designed to make inferences based on minimal amounts of training data, as one of their main purpose was to tackle the challenge of limited data availability. However, increasing the volume and variety of annotated ground truth data for training could potentially lead to better generalization capabilities of the trained models beyond the training set. The selection of training datasets thus remains a trade-off between costs of dataset creation and performance of the trained model. Given the complexity and individuality of learning tasks and learning models, there is little information available on how to choose the right amount of training data a priori. In this sense, it can be argued that the datasets used for TGNet training might be too small to fully solve the addressed problems.

Another important issue that has not been fully resolved concerns the errors introduced by deep learning predictions to subsequent analyses. For example, Study II found high overall label classification accuracies, but precision and recall values of underrepresented activity classes (counter attacks) still remained comparatively low. The problem was addressed by deriving theory-guided class definitions, assessing inter-rater agreement, and providing data samples for qualitative assessment. Still, performance analyses based on automatically generated labels (such as the one presented in Study II) contain a source of errors caused by misclassifications. As the learning algorithms inherits characteristics and biases of the generated training data to its predictions, the resulting errors may stem from quality issues within the training data set, insufficient operationalization or inadequate model design. Although the overall classification accuracy remains strong, the precise impact of these errors on subsequent statistical analysis also remains to be investigated at this point.

It should also be noted that the present thesis did not address all of the numerous challenges summarized by Table 2.1. For example, it was found that deep learning can handle STSD quite well, however, not all challenges originating from the data were fully addressed (*C12-14*). Similarly, the problem of opaqueness (*C18*) was not addressed by exploring explainable or interpretable models (Murdoch et al., 2019), but instead circumvented by exploring tasks where interpretability is of less importance.

6.3 Ethical Considerations

The application of deep learning algorithms, especially considering some of their limitations such as low model interpretability, raises a series of ethical considerations that need to be considered. First, researchers need to be aware of biases of trained models that originate from the characteristics of the training data (Ntoutsi et al., 2020). In this regard, two primary biases in team sport analysis data sets can be identified. On the one hand, a strong imbalance in study prevalence favors the sports of football and basketball, as shown by Table 3.1 and recent reviews (Ellens et al., 2022; Rico-González et al., 2020). On the other hand, female athlete data is strongly underrepresented in sport scientific research (Cowley et al., 2021). It can be assumed that this effect also extends to STSD collection and no reviewed study (cf. Table 3.1) or conducted study within this dissertation used data from competitions with female athletes. As a consequence, caution must be exerted when drawing conclusions from deep learning applications in terms of their applicability and usage.

6.4 Practical Applicability

The ultimate goal of team sports analysis remains to transfer scientific and technological advancement to practical settings. This has previously not always been the case, leading to a noticeable research practice gap (*C11*). In this respect, it remains important to highlight the presented findings with regards to their practical applicability.

First and foremost, it can be summarized that the present thesis (e.g. with respect to *RQ4*) advocates and empirically supports the fact that deep learning should preferably be used as a powerful tool for team sports analysis, rather than a theoretical framework (*C2*) that promises to explain team sports performance. In this regard, applications that can be translated to predictive learning tasks were deliberately chosen for empirical testing.

At the same time, most pipelines were not constructed as demo cases that operate under idealized conditions, but as robust and validated end-to-end pipelines. For example, the architectures designed in Study II and VIII generate complex labels for all games (irrespective of boundary conditions such as player dismissals) right from the raw data. In this fashion, the trained models can be used in the field out-of-the-box with minimal programming experience. Sports scientists and practitioners can use these models (one of which is also being released) for label generation and contextualization of analyses with high usability and transparency.

Furthermore, the present thesis has addressed a series of technological challenges prevalent in the field. Most importantly, the open source software package floodlight can be used under permissive licensing by anybody to process data and replicate many state-of-the-art applications. Thus, the present work has significantly lowered the coding skill requirements and entry barriers to team sports analysis for fellow researchers and practitioners.

6.5 Future Work

The presented findings provide novel insights to pressing questions on the role of deep learning in team sport analysis research. At the same time, usable and effective deep learning methods themselves have only been around for little more than a decade. Much more work on all aspects of deep learning with STSD for team sports analysis is needed to unfold the full potential of this approach. Especially with respect to the highlighted challenges and the rapid development within the deep learning community, novel algorithms need to be designed, evaluated, and tested empirically on team sports analysis tasks. A particularly interesting path constitute advancements in interpretable and explainable deep learning to address the problems of black-box-approaches.

Within this endeavor, the present thesis has identified some pressing issues that should be addressed by future work. With respect to sample sizes, it appears as if the problem of small populations and limited sample sizes (*C9*) shifts to the challenge of selecting appropriate training datasets in terms of volumen and variety. More research is needed to provide insights on how training datasets should be constructed accordingly and with respect to domain-specific characteristics. In a similar fashion, the biases introduced by training data to model predictions need further exploration. Statistical analyses based on automatically generated information can only be valid if potential error sources are better understood and bounded.

To further improve the effectiveness of deep learning in team sports analyses, several measures can be taken in future work. The traditional problem of operational definitions (*C5*) is even more critical in deep learning applications. Abstract team sports analysis concepts such as complex movement patterns need to be operationalized and validated to be used in learning tasks. Valid methods and measures are needed to provide robust ground truth data and high-quality models. Last but not least, the continued implementation of open science practices such as code and data sharing needs to be stressed as a necessary condition to provide value and ultimately narrow the research practice gap.

References

Alcorn, M. A., & Nguyen, A. (2021). Baller2vec: A Multi-Entity Transformer For Multi-Agent Spatiotemporal Modeling. *arXiv:2102.03291 [cs].*

Asada, M., & Von Stryk, O. (2020). Scientific and Technological Challenges in RoboCup. *Annual Review of Control, Robotics, and Autonomous Systems*, *3*(1), 441–471. https://doi.org/10.1146/annurev-control-100719-064806

Bassek, M., Raabe, D., Banning, A., Memmert, D., & Rein, R. (2023). Analysis of contextualized intensity in Men's elite handball using graph-based deep learning. *Journal of Sports Sciences*, 1–10. https://doi.org/10.1080/02640414.2023.2268366

Bassek, M., Raabe, D., Memmert, D., & Rein, R. (2023). Analysis of Motion Characteristics and Metabolic Power in Elite Male Handball Players. *Journal of Sports Science and Medicine*, 310–316. https://doi.org/10.52082/jssm.2023.310

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., … Pascanu, R. (2018). Relational inductive biases, deep learning, and graph networks. *arXiv:1806.01261*.

Beal, R., Norman, T. J., & Ramchurn, S. D. (2019). Artificial intelligence for team sports: A survey. *The Knowledge Engineering Review*, *34*, e28. https://doi.org/10.1017/S0269888919000225

Berrar, D., Lopes, P., Davis, J., & Dubitzky, W. (2019). Guest editorial: Special issue on machine learning for soccer. *Machine Learning*, *108*(1), 1–7. https://doi.org/10.1007/s10994-018-5763-8

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. https://doi.org/10.1038/s41598-023-39616-2

Biermann, H., Petersen, N., Memmert, D., & Raabe, D. (2024, under review). Contextualization of Soccer Analysis with Tactical Periodization and Machine Learning. *Data Mining and Knowledge Discovery*.

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*, 1–10. https://doi.org/10.1145/3475722.3482792

Blauberger, P., Marzilger, R., & Lames, M. (2021). Validation of Player and Ball Tracking with a Local Positioning System. *Sensors*, *21*(4), 1465. https://doi.org/10.3390/s21041465

Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. https://doi.org/10.48550/ARXIV.2104.13478

Bronstein, M. M., Bruna, J., Lecun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, *34*(4), 18–42. https://doi.org/10.1109/MSP.2017.2693418

Buchheit, M., Allen, A., Poon, T. K., Modonutti, M., Gregson, W., & Di Salvo, V. (2014). Integrating different tracking systems in football: Multiple camera semi-automatic system, local position measurement and GPS technologies. *Journal of Sports Sciences*, *32*(20), 1844–1857. https://doi.org/10.1080/02640414.2014.942687

Bullock, G. S., Hughes, T., Arundale, A. H., Ward, P., Collins, G. S., & Kluzek, S. (2022). Black Box Prediction Methods in Sports Medicine Deserve a Red Card for Reckless Practice: A Change of Tactics is Needed to Advance Athlete Care. *Sports Medicine*, *52*(8), 1729–1735. https://doi.org/10.1007/s40279-022-01655-6

Cao, L. (2018). Data Science: A Comprehensive Overview. *ACM Computing Surveys*, *50*(3), 1–42. https://doi.org/10.1145/3076253

Cervone, D., D'Amour, A., Bornn, L., & Goldsberry, K. (2016). A Multiresolution Stochastic Process Model for Predicting Basketball Possession Outcomes. *Journal of the American Statistical Association*, *111*(514), 585–599. https://doi.org/10.1080/01621459.2016.1141685

Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., & Bengio, Y. (2015). A recurrent latent variable model for sequential data. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 28). Curran Associates, Inc.

Ciresan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3642–3649. https://doi.org/10.1109/CVPR.2012.6248110

Corsie, M., & Swinton, P. A. (2023). Reliability of spatial-temporal metrics used to assess collective behaviours in football: An in-silico experiment. *Science and Medicine in Football*, *7*(3), 297–305. https://doi.org/10.1080/24733938.2022.2100460

Cowley, E. S., Olenick, A. A., McNulty, K. L., & Ross, E. Z. (2021). "Invisible Sportswomen": The Sex Data Gap in Sport and Exercise Science Research. *Women in Sport and Physical Activity Journal*, *29*(2), 146–151. https://doi.org/10.1123/wspaj.2021-0028

Cuevas, C., Quilón, D., & García, N. (2020). Techniques and applications for soccer video analysis: A survey. *Multimedia Tools and Applications*, *79*(39-40), 29685–29721. https://doi.org/10.1007/s11042-020-09409-0

Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2019). Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance. *Journal of Sports Sciences*, *37*(5), 568–600. https://doi.org/10.1080/02640414.2018.1521769

Deliege, A., Cioppa, A., Giancola, S., Seikavandi, M. J., Dueholm, J. V., Nasrollahi, K., Ghanem, B., Moeslund, T. B., & Van Droogenbroeck, M. (2021). SoccerNet-v2: A Dataset and Benchmarks for Holistic Understanding of Broadcast Soccer Videos. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 4508–4519.

Dellaserra, C. L., Gao, Y., & Ransdell, L. (2014). Use of Integrated Technology in Team Sports. *Journal of Strength and Conditioning Research*, *28*(2), 556–573. https://doi.org/10.1519/JSC.0b013e3182a952fb

Di Salvo, V., Collins, A., McNeill, B., & Cardinale, M. (2006). Validation of Prozone: A new video-based performance analysis system. *International Journal of Performance Analysis in Sport*, *6*(1), 108–119. https://doi.org/10.1080/24748668.2006.11868359

Dubitzky, W., Lopes, P., Davis, J., & Berrar, D. (2019). The Open International Soccer Database for machine learning. *Machine Learning*, *108*(1), 9–28. https://doi.org/10.1007/s10994-018-5726-0

Dutt-Mazumder, A., Button, C., Robins, A., & Bartlett, R. (2011). Neural network modelling and dynamical system theory: Are they relevant to study the governing dynamics of association football players? *Sports Medicine*, *41*(12), 1003–1017. https://doi.org/10.2165/11593950-000000000-00000

Ellens, S., Middleton, K., Gastin, P. B., & Varley, M. C. (2022). Techniques to derive and clean acceleration and deceleration data of athlete tracking technologies in team sports: A scoping review. *Journal of Sports Sciences*, *40*(16), 1772–1800. https://doi.org/10.1080/02640414.2022.2054535

European Commission. (2018). *Study on the economic impact of sport through sport satellite accounts* (tech. rep.). Publications Office of the European Union. Luxembourg. https://doi.org/10.2766/156532

Everett, G., Beal, R. J., Matthews, T., Early, J., Norman, T. J., & Ramchurn, S. D. (2023). Inferring Player Location in Sports Matches: Multi-Agent Spatial Imputation from Limited Observations. https://doi.org/10.48550/ARXIV.2302.06569

Felsen, P., Lucey, P., & Ganguly, S. (2018). Where will they go? predicting fine-grained adversarial multi-agent motion using conditional variational autoencoders. *Proceedings of the European Conference on Computer Vision (ECCV)*, 732–747.

Fernández, J., & Bornn, L. (2021). SoccerMap: A Deep Learning Architecture for Visually-Interpretable Analysis in Soccer. In Y. Dong, G. Ifrim, D. Mladenić, C. Saunders, & S. Van Hoecke (Eds.), *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track* (pp. 491–506, Vol. 12461). Springer International Publishing. https://doi.org/10.1007/978-3-030-67670-4_30

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. (2015). Approval of electronic performance and tracking system (EPTS) devices. *Federation Internationale de Football Association.*, *Circular*(1494).

FIFA. (2024). Electronic Performance and Tracking Systems (EPTS).

Fister, I., Ljubič, K., Suganthan, P. N., Perc, M., & Fister, I. (2015). Computational intelligence in sports: Challenges and opportunities within a new research domain. *Applied Mathematics and Computation*, *262*, 178–186. https://doi.org/10.1016/j.amc.2015.04.004

Frencken, W. G., Lemmink, K. A., & Delleman, N. J. (2010). Soccer-specific accuracy and validity of the local position measurement (LPM) system. *Journal of Science and Medicine in Sport*, *13*(6), 641–645. https://doi.org/10.1016/j.jsams.2010.04.003

Frevel, N., Beiderbeck, D., & Schmidt, S. L. (2022). The impact of technology on sports – A prospective study. *Technological Forecasting and Social Change*, *182*, 121838. https://doi.org/10.1016/j.techfore.2022.121838

Fujii, K. (2021). Data-Driven Analysis for Understanding Team Sports Behaviors. *Journal of Robotics and Mechatronics*, *33*(3), 505–514. https://doi.org/10.20965/jrm.2021.p0505

Furley, P., Mehta, S., Raabe, D., & Memmert, D. (2024). Objectivity of match analysis in football: Testing the level of agreement between coaches' interpretations of video data. *International Journal of Sports Science & Coaching*. https://doi.org/10.1177/17479541241278603

Ghosh, I., Ramasamy Ramamurthy, S., Chakma, A., & Roy, N. (2023). Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective. *WIREs Data Mining and Knowledge Discovery*, *13*(5), e1496. https://doi.org/10.1002/widm.1496

Giancola, S., Cioppa, A., Deliège, A., Magera, F., Somers, V., Kang, L., Zhou, X., Barnich, O., De Vleeschouwer, C., Alahi, A., Ghanem, B., Van Droogenbroeck, M., Darwish, A., Maglo, A., Clapés, A., Luyts, A., Boiarov, A., Xarles, A., Orcesi, A., … Li, Z. (2022). SoccerNet 2022 Challenges Results. *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*, 75–86. https://doi.org/10.1145/3552437.3558545

Glazier, P. S. (2010). Game, set and match? Substantive issues and future directions in performance analysis. *Sports Medicine*, *40*(8), 625–634. https://doi.org/10.2165/11534970-000000000-00000

Glazier, P. S. (2017). Towards a Grand Unified Theory of sports performance. *Human Movement Science*, *56*, 139–156. https://doi.org/10.1016/j.humov.2015.08.001

Goes, F. R., Meerhoff, L. A., Bueno, M. J., Rodrigues, D. M., Moura, F. A., Brink, M. S., Elferink-Gemser, M. T., Knobbe, A. J., Cunha, S. A., Torres, R. S., & Lemmink, K. A. (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

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Grunz, A., Memmert, D., & Perl, J. (2012). Tactical pattern recognition in soccer games by means of special self-organizing maps. *Human Movement Science*, *31*(2), 334–343. https://doi.org/10.1016/j.humov.2011.02.008

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

Hauri, S., Djuric, N., Radosavljevic, V., & Vucetic, S. (2021). Multi-modal trajectory prediction of NBA players. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1640–1649.

Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., & Meyer, T. (2019). Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science and Coaching*, *14*(6), 798–817. https://doi.org/10.1177/1747954119879350

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., & Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97. https://doi.org/10.1109/MSP.2012.2205597

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hoppe, M. W., Baumgart, C., Polglaze, T., & Freiwald, J. (2018). Validity and reliability of GPS and LPS for measuring distances covered and sprint mechanical properties in team sports. *PLoS ONE*, *13*(2), 1–21. https://doi.org/10.1371/journal.pone.0192708

Horvat, T., & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. *WIREs Data Mining and Knowledge Discovery*, *10*(5), e1380. https://doi.org/10.1002/widm.1380

Hussein, A., Gaber, M. M., Elyan, E., & Jayne, C. (2018). Imitation Learning: A Survey of Learning Methods. *ACM Computing Surveys*, *50*(2), 1–35. https://doi.org/10.1145/3054912

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jayal, A., McRobert, A., Oatley, G., & O'Donoghue, P. (2018, June). *Sports Analytics: Analysis, Visualisation and Decision Making in Sports Performance* (1st ed.). Routledge. https://doi.org/10.4324/9781315222783

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kempe, M., Grunz, A., & Memmert, D. (2015). Detecting tactical patterns in basketball: Comparison of merge self-organising maps and dynamic controlled neural networks. *European Journal of Sport Science*, *15*(4), 249–255. https://doi.org/10.1080/17461391.2014.933882

Komitova, R., Raabe, D., Rein, R., & Memmert, D. (2022). Time Series Data Mining for Sport Data: A Review. *International Journal of Computer Science in Sport*, *21*(2), 17–31. https://doi.org/10.2478/ijcss-2022-0008

Kontos, C., & Karlis, D. (2023). Football analytics based on player tracking data using interpolation techniques for the prediction of missing coordinates. *Statistica Applicata - Italian Journal of Applied Statistics*, (2). https://doi.org/10.26398/IJAS.0035-010

Kovalchik, S. A. (2023). Player Tracking Data in Sports. *Annual Review of Statistics and Its Application*, *10*(1), 677–697. https://doi.org/10.1146/annurev-statistics-033021-110117

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Lames, M., & McGarry, T. (2007). On the search for reliable performance indicators in game sports. *International Journal of Performance Analysis in Sport*, *7*(1), 62–79.

Le, H. M., Carr, P., Yue, Y., & Lucey, P. (2017). Data-Driven Ghosting using Deep Imitation Learning. *MIT Sloan Sports Analytics Conference*, 1–15.

Le, H. M., Yue, Y., Carr, P., & Lucey, P. (2017). Coordinated multi-agent imitation learning. *International Conference on Machine Learning (ICML)*, *70*, 1995–2003.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lindström, P., Jacobsson, L., Carlsson, N., & Lambrix, P. (2020). Predicting Player Trajectories in Shot Situations in Soccer. In U. Brefeld, J. Davis, J. Van Haaren, & A. Zimmermann (Eds.), *Machine Learning and Data Mining for Sports Analytics* (pp. 62–75, Vol. 1324). Springer International Publishing. https://doi.org/10.1007/978-3-030-64912-8_6

Linke, D., Link, D., & Lames, M. (2018). Validation of electronic performance and tracking systems EPTS under field conditions. *PLoS ONE*, *13*(7), 1–19. https://doi.org/10.1371/journal.pone.0199519

Linke, D., Link, D., & Lames, M. (2020). Football-specific validity of TRACAB's optical video tracking systems (H. A. Kerhervé, Ed.). *PLoS ONE*, *15*(3), e0230179. https://doi.org/10.1371/journal.pone.0230179

Lippi, G., Banfi, G., Favaloro, E. J., Rittweger, J., & Maffulli, N. (2008). Updates on improvement of human athletic performance: Focus on world records in athletics. *British Medical Bulletin*, *87*(1), 7–15. https://doi.org/10.1093/bmb/ldn029

Long, J. A. (2016). Kinematic interpolation of movement data. *International Journal of Geographical Information Science*, *30*(5), 854–868. https://doi.org/10.1080/13658816.2015.1081909

Lord, F., Pyne, D. B., Welvaert, M., & Mara, J. K. (2020). Methods of performance analysis in team invasion sports: A systematic review. *Journal of Sports Sciences*, *38*(20), 2338–2349. https://doi.org/10.1080/02640414.2020.1785185

Low, B., Rein, R., Raabe, D., Schwab, S., & Memmert, D. (2021). The porous high-press? An experimental approach investigating tactical behaviours from two pressing strategies in football. *Journal of Sports Sciences*. https://doi.org/10.1080/02640414.2021.1925424

Lutz, J., Memmert, D., Raabe, D., Dornberger, R., & Donath, L. (2020). Wearables for integrative performance and tactic analyses: Opportunities, challenges, and future directions. *International Journal of Environmental Research and Public Health*, *17*(1), 1–26. https://doi.org/10.3390/ijerph17010059

Mackenzie, R., & Cushion, C. (2013). Performance analysis in football: A critical review and implications for future research. *Journal of Sports Sciences*, *31*(6), 639–676. https://doi.org/10.1080/02640414.2012.746720

Mahaseni, B., Faizal, E. R. M., & Raj, R. G. (2021). Spotting Football Events Using Two-Stream Convolutional Neural Network and Dilated Recurrent Neural Network. *IEEE Access*, *9*, 61929–61942. https://doi.org/10.1109/ACCESS.2021.3074831

Manafifard, M., Ebadi, H., & Abrishami Moghaddam, H. (2017). A survey on player tracking in soccer videos. *Computer Vision and Image Understanding*, *159*, 19–46. https://doi.org/10.1016/j.cviu.2017.02.002

McCall, A., Fanchini, M., & Coutts, A. J. (2017). Prediction: The Modern-Day Sport-Science and Sports-Medicine “Quest for the Holy Grail”. *International Journal of Sports Physiology and Performance*, *12*(5), 704–706. https://doi.org/10.1123/ijspp.2017-0137

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

Memmert, D., & Raabe, D. (2023, November). *Data Analytics in Football: Positional Data Collection, Modelling and Analysis* (2nd ed.). Routledge. https://doi.org/10.4324/9781003411079

Mitchell, T. M. (1980). The need for biases in learning generalizations.

Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, *5*(4), 213–222. https://doi.org/10.1007/s41060-017-0093-7

Müller, O., Caron, M., Döring, M., Heuwinkel, T., & Baumeister, J. (2022). PIVOT: A Parsimonious End-to-End Learning Framework for Valuing Player Actions in Handball Using Tracking Data. In U. Brefeld, J. Davis, J. Van Haaren, & A. Zimmermann (Eds.), *Machine Learning and Data Mining for Sports Analytics* (pp. 116–128, Vol. 1571). Springer International Publishing. https://doi.org/10.1007/978-3-031-02044-5_10

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(44), 22071–22080. https://doi.org/10.1073/pnas.1900654116

Nassis, G., Verhagen, E., Brito, J., Figueiredo, P., & Krustrup, P. (2023). A review of machine learning applications in soccer with an emphasis on injury risk. *Biology of Sport*, *40*(1), 233–239. https://doi.org/10.5114/biolsport.2023.114283

NFL. (2023). NFL Big Data Bowl.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1356. https://doi.org/10.1002/widm.1356

Omidshafiei, S., Hennes, D., Garnelo, M., Wang, Z., Recasens, A., Tarassov, E., Yang, Y., Elie, R., Connor, J. T., Muller, P., Mackraz, N., Cao, K., Moreno, P., Sprechmann, P., Hassabis, D., Graham, I., Spearman, W., Heess, N., & Tuyls, K. (2022). Multiagent off-screen behavior prediction in football. *Scientific Reports*, *12*(1), 8638. https://doi.org/10.1038/s41598-022-12547-0

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), 236. https://doi.org/10.1038/s41597-019-0247-7

Patel, D., Shah, D., & Shah, M. (2020). The Intertwine of Brain and Body: A Quantitative Analysis on How Big Data Influences the System of Sports. *Annals of Data Science*, *7*(1), 1–16. https://doi.org/10.1007/s40745-019-00239-y

Perl, J. (2011). Net-Based Game Analysis by Means of the Software Tool SOCCER. *International Journal of Computer Science in Sport*, *10*(2), 77–84.

Pettersen, S. A., Johansen, D., Johansen, H., Berg-Johansen, V., Gaddam, V. R., Mortensen, A., Langseth, R., Griwodz, C., Stensland, H. K., & Halvorsen, P. (2014). Soccer video and player position dataset. *Proceedings of the 5th ACM Multimedia Systems Conference*, 18–23. https://doi.org/10.1145/2557642.2563677

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*, *42*(19), 1821–1830. https://doi.org/10.1080/02640414.2024.2414363

Raabe, D., Biermann, H., Bassek, M., Wohlan, M., Komitova, R., Rein, R., Groot, T. K., & Memmert, D. (2022). Floodlight - A high-level, data-driven sports analytics framework. *Journal of Open Source Software*, *7*(76), 4588. https://doi.org/10.21105/joss.04588

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

Randers, M. B., Mujika, I., Hewitt, A., Santisteban, J., Bischoff, R., Solano, R., Zubillaga, A., Peltola, E., Krustrup, P., & Mohr, M. (2010). Application of four different football match analysis systems: A comparative study. *Journal of Sports Sciences*, *28*(2), 171–182. https://doi.org/10.1080/02640410903428525

Redwood-Brown, A., Cranton, W., & Sunderland, C. (2012). Validation of a Real-Time Video Analysis System for Soccer. *International Journal of Sports Medicine*, *33*(08), 635–640. https://doi.org/10.1055/s-0032-1306326

Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus*, *5*(1), 1410–1410. https://doi.org/10.1186/s40064-016-3108-2

Rein, R., Perl, J., & Memmert, D. (2017). Maybe a tad early for a Grand Unified Theory: Commentary on "Towards a Grand Unified Theory of sports performance". *Human Movement Science*, *56*, 173–175. https://doi.org/10.1016/j.humov.2017.04.011

Richly, K., Moritz, F., & Schwarz, C. (2017). Utilizing Artificial Neural Networks to Detect Compound Events in Spatio-Temporal Soccer Data. *ACM Reference format*, (August), 7–7.

Rico-González, M., Pino-Ortega, J., Nakamura, F. Y., Arruda Moura, F., Rojas-Valverde, D., & Los Arcos, A. (2020). Past, present, and future of the technological tracking methods to assess tactical variables in team sports: A systematic review. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, *234*(4), 281–290. https://doi.org/10.1177/1754337120932023

Rodrigues, A. C. N., Pereira, A. S., Mendes, R. M. S., Araújo, A. G., Couceiro, M. S., & Figueiredo, A. J. (2020). Using Artificial Intelligence for Pattern Recognition in a Sports Context. *Sensors*, *20*(11), 3040. https://doi.org/10.3390/s20113040

Rudenko, A., Palmieri, L., Herman, M., Kitani, K. M., Gavrila, D. M., & Arras, K. O. (2020). Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, *39*(8), 895–935. https://doi.org/10.1177/0278364920917446

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Sarmento, H., Anguera, M. T., Pereira, A., & Araújo, D. (2018). Talent Identification and Development in Male Football: A Systematic Review. *Sports Medicine*, 1–25. https://doi.org/10.1007/s40279-017-0851-7

Sarmento, H., Marcelino, R., Anguera, M. T., Campaniço, J., Matos, N., & Leitão, J. C. (2014). Match analysis in football: A systematic review. *Journal of Sports Sciences*, *32*(20), 1831–1843. https://doi.org/10.1080/02640414.2014.898852

Schlenger, J., Wunderlich, F., Raabe, D., & Memmert, D. (2023). Systematic Analysis of Position-Data-based Key Performance Indicators. *International Journal of Computer Science in Sport*, *22*(1), 80–101. https://doi.org/10.2478/ijcss-2023-0006

Schmid, M., Blauberger, P., & Lames, M. (2021). Simulating Defensive Trajectories in American Football for Predicting League Average Defensive Movements. *Frontiers in Sports and Active Living*, *3*, 669845. https://doi.org/10.3389/fspor.2021.669845

Schmid, M., & Lames, M. (2023). Correction of systematic errors in electronic performance and tracking systems. *Sports Engineering*, *26*(1), 30. https://doi.org/10.1007/s12283-023-00421-9

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schrapf, N., Alsaied, S., & Tilp, M. (2017). Tactical interaction of offensive and defensive teams in team handball analysed by artificial neural networks. *Mathematical and Computer Modelling of Dynamical Systems*, *23*(4), 363–371. https://doi.org/10.1080/13873954.2017.1336733

Schulz, H., & Behnke, S. (2012). Deep Learning: Layer-Wise Learning of Feature Hierarchies. *KI - Künstliche Intelligenz*, *26*(4), 357–363. https://doi.org/10.1007/s13218-012-0198-z

Scott, M. T., Scott, T. J., & Kelly, V. G. (2016). The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review. *Journal of Strength and Conditioning Research*, *30*(5), 1470–1490. https://doi.org/10.1519/JSC.0000000000001221

Seidl, T., Cherukumudi, A., Hartnett, A., Carr, P., & Lucey, P. (2018). Bhostgusters: Realtime interactive play sketching with synthesized NBA defenses. *Proceeding of the 12th MIT Sloan Sports Analytics Conference, Boston, MA. Boston: MIT*.

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

Sicilia, A., Pelechrinis, K., & Goldsberry, K. (2019). DeepHoops: Evaluating Micro-Actions in Basketball Using Deep Feature Representations of Spatio-Temporal Data. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2096–2104. https://doi.org/10.1145/3292500.3330719

SkillCorner. (2023). SkillCorner Open Data.

StatsBomb. (2023). StatsBomb Open Data.

Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., Kosub, S., Schreck, T., Keim, D., & Grossniklaus, M. (2017). How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects. *Data*, *2*(1), 2. https://doi.org/10.3390/data2010002

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 (N. Q. K. Le, Ed.). *PLOS ONE*, *18*(1), e0265372. https://doi.org/10.1371/journal.pone.0265372

Stone, P., & Veloso, M. (2000). Multiagent Systems: A Survey from a Machine Learning Perspective. *Autonomous Robots*, *8*(3), 345–383. https://doi.org/10.1023/A:1008942012299

Taberner, M., O'Keefe, J., Flower, D., Phillips, J., Close, G., Cohen, D. D., Richter, C., & Carling, C. (2020). Interchangeability of position tracking technologies; can we merge the data? *Science and Medicine in Football*, *4*(1), 76–81. https://doi.org/10.1080/24733938.2019.1634279

Theiner, J., Gritz, W., Müller-Budack, E., Rein, R., Memmert, D., & Ewerth, R. (2022). Extraction of positional player data from broadcast soccer videos. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 823–833.

Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A., & Clubb, J. (2022). Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis. *Sports Medicine - Open*, *8*(1), 15. https://doi.org/10.1186/s40798-022-00408-z

Tuyls, K., Omidshafiei, S., Muller, P., Wang, Z., Connor, J., Hennes, D., Graham, I., Spearman, W., Waskett, T., Steel, D., Luc, P., Recasens, A., Galashov, A., Thornton, G., Elie, R., Sprechmann, P., Moreno, P., Cao, K., Garnelo, M., … 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

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in neural information processing systems*, (Nips), 5998–6008.

Wagenaar, M., Okafor, E., Frencken, W., & A. Wiering, M. (2017). Using Deep Convolutional Neural Networks to Predict Goal-scoring Opportunities in Soccer. *International Conference on Pattern Recognition Applications and Methods*. https://doi.org/10.5220/0006194804480455

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T.-Y., Manrai, A., … Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60. https://doi.org/10.1038/s41586-023-06221-2

Wunderlich, F., Biermann, H., Yang, W., Bassek, M., Raabe, D., Elbert, N., Memmert, D., & Garnica-Caparrós, M. (2024). Machine learning and data imputation approaches to handle the issue of data sparsity in sports forecasting. *Machine Learning*.

Wunderlich, F., & Memmert, D. (2020). Forecasting the outcomes of sports events: A review. *European Journal of Sport Science*, 1–14. https://doi.org/10.1080/17461391.2020.1793002

Yeh, R. A., Schwing, A. G., Huang, J., & Murphy, K. (2019). Diverse generation for multi-agent sports games. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4610–4619.

Yurko, R., Matano, F., Richardson, L. F., Granered, N., Pospisil, T., Pelechrinis, K., & Ventura, S. L. (2020). Going deep: Models for continuous-time within-play valuation of game outcomes in American football with tracking data. *Journal of Quantitative Analysis in Sports*, *16*(2), 163–182. https://doi.org/10.1515/jqas-2019-0056

Zednik, C. (2019). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy and Technology*. https://doi.org/10.1007/s13347-019-00382-7

A. Published Articles

This appendix contains abstracts and references for all published articles and submitted manuscripts of this thesis. All articles are written and published in English.

Study I

Graph representations for the analysis of multi-agent spatiotemporal sports data

Reference:

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

[Impact Factor 7.04] {Q2 Artificial Intelligence}

Abstract

Analyzing tactical patterns in invasion games using multi-agent spatiotemporal data is a challenging task at the intersection of computer and sports science. A fundamental yet understudied problem in this area is finding an optimal data representation for processing athlete trajectories using machine learning algorithms. In the present work, we address this gap by discussing common representations in use and propose *Tactical Graphs*, an alternative graph-based format capable of producing integrative, contextualized models for machine learning applications. We provide an in-depth, domain-specific motivation of the proposed data representation scheme and show how this approach exploits inherent data traits. We propose *Tactical Graph Networks* (TGNets), a light-weight, hybrid machine learning architecture sensitive to player interactions. Our method is evaluated with an extensive ablation study and the first comprehensive state of the art comparison between standard feature, state vector, and image-based methods on the same dataset. Experiments were conducted using real-world football data containing short sequences of defensive play labelled according to the outcome of ball winning attempts. The results indicate that TGNets are on par with state-of-the-art deep learning models while exhibiting only a fraction of their complexity. We further demonstrate that selecting the right data representation is crucial as it has a significant influence on model performance. The theoretical findings and the proposed method provide insights and a strong methodological alternative for all classification, prediction or pattern recognition applications in the areas of collective movement analysis, automated match analysis, and performance analysis.

Study II

Analysis of contextualized physical performance in elite Men's handball using graph-based deep learning

Reference:

Bassek, M.*, **Raabe, D.***, Banning, A., Memmert, D., & Rein, R. (2023). Analysis of contextualized physical performance in elite Men's handball using graph-based deep learning. *Journal of Sports Sciences, 41*(13), 1299-1308.
https://doi.org/10.1080/02640414.2023.2268366

** both authors contributed equally to this work*

[IF 3.87] {Q1 Sports Science}

Abstract

Manual annotation of data in invasion games is a costly task which poses a natural limit on sample sizes and the level of granularity used in match and performance analyses. To overcome this challenge, this work introduces FAUPA-ML, a Framework for Automatic Upscaled Performance Analysis with Machine Learning, which leverages graph neural networks to scale domain-specific expert knowledge to large data sets. Networks were trained using position data of match phases (counter/position attacks), annotated manually by domain experts in 10 matches. The best network was applied to contextualize $N = 539$ matches of elite handball (2019/20–2021/22 German Men's Handball Bundesliga) with 86% balanced accuracy. Distance covered, speed, metabolic power, and metabolic work were calculated for attackers and defenders and differences between counters and position attacks across seasons analyzed with an ANOVA. Results showed that counter attacks are shorter, less frequent and more intense than position attacks and that attacking is more intense than defending. Findings show that FAUPA-ML generates accurate replications of expert knowledge that can be used to gain insights in performance analysis previously deemed infeasible. Future studies can use FAUPA-ML for large-scale, contextualized analyses that investigate influences of team strength, score-line, or team tactics on performance.

Study III

floodlight - A high-level, data-driven sports analytics framework

Reference:

Raabe, D., Biermann, H., Bassek, M., Wohlan, M., Komitova, R., Rein, R., Kuppens Groot, T., & Memmert, D. (2022). floodlight - A high-level, data-driven sports analytics framework. *Journal of Open Source Software, 7*(76), 4588.
https://doi.org/10.21105/joss.04588

[IF 5.2] {Q1 Software Engineering}

Abstract

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 provider- and 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.

Study IV

The dual problem of space: player positioning determines attacking success in elite men's football

Reference:

Raabe, D., Biermann, H., Bassek, M., Memmert, D., & Rein, R. (2024). The dual problem of space: player positioning determines attacking success in elite men's football. *Journal of Sports Sciences, 42*(19), 1821-1830. https://doi.org/10.1080/02640414.2024.2414363

[IF 3.87]{Q1 Sports Science}

Abstract

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.

Study V

Analysis of Motion Characteristics and Metabolic Power in Elite Male Handball Players

Reference:

Bassek, M., **Raabe, D.**, Memmert, D., & Rein, R., (2023). Analysis of Motion Characteristics and Metabolic Power in Elite Male Handball Players. *Journal of Sports Science and Medicine, 22*(2), 310-316. https://doi.org/10.52082/jssm.2023.310

[IF 3.3] {Q1 Orthopedics and Sports Medicine}

Abstract

While handball is characterized by repeated sprints and changes of direction, traditional player load models do not consider accelerations and decelerations. The aim of this study was to analyze the differences between metabolic power and speed zones for player load assessment with regard to the player role. Position data from 330 male individuals during 77 games from the 2019/20 German Men's Handball-Bundesliga (HBL) were analyzed, resulting in 2233 individual observations. Players were categorized into wings, backs and pivots. Distance covered in different speed zones, metabolic power, metabolic work, equivalent distance (metabolic work divided by energy cost of running), time spend running, energy spend running, and time over 10 and 20 W were calculated. A 2-by-3 mixed ANOVA was calculated to investigate differences and interactions between groups and player load models. Results showed that total distance was longest in wings (3568 ± 1459 m in 42 ± 17 min), followed by backs (2462 ± 1145 m in 29 ± 14 min), and pivots (2445 ± 1052 m in 30 ± 13 min). Equivalent distance was greatest in wings (4072.50 ± 1644.83 m), followed by backs (2765.23 ± 1252.44 m), and pivots (2697.98 ± 1153.16 m). Distance covered and equivalent distance showed moderate to large interaction effects between wings and backs ($p < .01$, ES = 0.73) and between wings and pivots ($p < .01$, ES = 0.86) and a small interaction effect between backs and pivots ($p < .01$, ES = 0.22). The results underline the need for individualized management of training loads and the potential of using information about locomotive accelerations and decelerations to obtain more precise descriptions of player load during handball game performance at the highest level of competition. Future studies should investigate the influence of physical performance on smaller match sequences, like ball possession phases.

Study VI

A Unified Taxonomy and Multimodal Dataset for Events in Invasion Games

Reference:

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*, 1-10.
https://doi.org/10.1145/3475722.3482792

{A* - Computer vision and multimedia computation}

Abstract

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 low-level 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

Study VII

Synchronization of passes in event and spatiotemporal soccer data

Reference:

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.
https://doi.org/10.1038/s41598-023-39616-2

[IF 4.44] {Q1 Multidisciplinary}

Abstract

The majority of soccer analysis studies investigates specific 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 significant 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 refinement 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-flight 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 70%, while our algorithm also detects localization errors (noise) in the position data. 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 significantly less data). Hence, our proposed lightweight framework offers a viable solution that enables groups facing limited access to (recent) data sources to effectively synchronize passes in the event and position data.

Study VIII

Contextualization of Soccer Analysis with Tactical Periodization and Machine Learning

Reference:

Biermann, H., Memmert, D., Petersen, N., & **Raabe, D.** (2024). Contextualization of Soccer Analysis with Tactical Periodization and Machine Learning.

Manuscript under review in Data Mining and Knowledge Discovery.

Abstract

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 time-consuming 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 best-performing 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 80% accuracy while being modular extendable for future work. Moreover, we found a strong relation between semantic complexity of *match phases*, 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.

Study IX

Objectivity of match analysis in football: Testing the level of agreement between coaches' interpretations of video data

Reference:

Furley, P., Mehta, S., **Raabe, D.**, & Memmert, D. (2024). Objectivity of match analysis in football: Testing the level of agreement between coaches' interpretations of video data. *International Journal of Sports Science & Coaching*.
https://doi.org/10.1177/17479541241278603

[IF 2.58]{Q1 Social Sciences (miscellaneous)}

Abstract

Using video data is a widespread procedure in the preparation for an upcoming opponent across all levels of football, but the way coaches interpret this data and use it for player feedback is still not fully understood. Three studies were conducted to investigate the level of agreement between football coaches on tactical questions regarding the opponent when interpreting the same video data. In Study 1 (scouting feed; N = 15) and 2 (tactic view feed; N = 24), different video viewing angles of the same match were provided to coaches, followed by simple questions regarding the viewed team (e.g., team formation, most striking player in the opening play of the attacking team). Response analyses using descriptive statistics and Fleiss-Kappa statistics showed great diversity regardless of the angle of the feed. Study 3 was a replication study (scouting feed; N = 16) using the identical approach as before but used a different match to introduce greater variety of video stimuli. Across all studies there was a high degree of diversity in coach responses and little consensus on basic questions like the adopted formation or the most striking player in the opening play (Fleiss-Kappa coefficients between -.036 [poor agreement] and .236 [fair agreement]). The present research shows that it is problematic to treat information from video feeds as being objective when preparing for the next opponent, as different coaches derive different interpretations from the same data source. Implications for use of video data, and related contributions to coaching research are discussed.

Study X

Time Series Data Mining for Sport Data: a Review

Reference:

Komitova, R., **Raabe, D.**, Rein, R., & Memmert, D. (2022). Time Series Data Mining for Sport Data: a Review. *International Journal of Computer Science in Sport, 21*(2), 17-31.
https://doi.org/10.2478/ijcss-2022-0008

[IF 1.42] {Q3 Computer Science (miscellaneous)}

Abstract

Time series data mining deals with extracting useful and meaningful information from time series data. Recently, the increasing use of temporal data, in particular time series data, has received much attention in the literature. Since most of sports data contain time information, it is natural to consider the temporal dimension in form of time series. However, in sports, the effective use of time series data mining techniques is still under development. The main goal of this paper is therefore to serve as an introduction to time series data mining and a glossary for interested researchers from the sports community. The paper gives an overview about current data mining tasks and tries to identify their potential research direction for further investigation. Furthermore, we want to draw more attention with respect to the importance of mining approaches with sport data and their particular challenges beyond usual time series data mining tasks.

Study XI

Examining how data becomes information for an upcoming opponent in football

Reference:

Mehta, S., Furley, P., **Raabe, D.**, & Memmert, D. (2023). 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

[IF 2.58] {Q1 Social Sciences (miscellaneous)}

Abstract

As the sport industry witnesses a surge in the type and volume of data-driven decisions, the general question of the process of information development remains: how is data used to develop meaningful information? And does the presence of novel quantitative data sources lend greater objectivity to match analysis? Study 1 examines how 12 football analysts use the same qualitative (video) and quantitative (event and position) data to develop information constituting a typical opponent report for an upcoming match, while Study 2 investigates the agreement between grade evaluations of these opponent reports by numerous professional coaches. Findings of Study 1 through independent-samples t-tests ($t(18) = 3.922$, $p = 0.001$) indicate a clear dominance of qualitative video data over quantitative event and position data in all opponent reports. Despite the presence of quantitative data sources, analysts tend to prefer annotated video data. Possible relations to previous experience and familiarity with data, coach–analyst preferences and biases are discussed. Results from Study 2 show extremely weak intra-class correlations (ICC) ($r = 0.147$; $p = 0.011$) between different grades awarded to the same video, depicting a clear lack of agreement in what coaches consider a good opponent report. Furthermore, coaches most valued the comprehensibility and relevance of the report. No significant associations were found between use of either data type and better grades. The subjectivity of the coaching process highlighting preferences regarding data validity and negotiations of adopting new key performance indicators (KPIs) is discussed, alongside limitations of the sample as well as the level of coach–analysts involved.

Study XII

Systematic Analysis of Position-Data-based Key Performance Indicators

Reference:

Schlenger, J., Wunderlich, F., **Raabe, D.**, & Memmert, D. (2023). Systematic Analysis of Position-Data-based Key Performance Indicators. *International Journal of Computer Science in Sport, 22*(1), 80-101. https://doi.org/10.2478/ijcss-2023-0006

[IF 1.42] {Q3 Computer Science (miscellaneous)}

Abstract

In the past 20 years, performance analysis in soccer has accumulated a wide variety of key performance indicators (KPI's) aimed at reflecting a team's strength and success. Thanks to rapidly advancing technologies and data analytics more sophisticated metrics, requiring high resolution data acquisition and big data methods, are developed. This includes many position-data-based KPI's, which incorporate precise spatial and temporal information about every player and the ball on the field.

The present study contributes to this research by performing a large-scale comparison of several metrics mainly based on player positions and passing events. Their association with team's success (derived from goals scored) and team's strength (estimated from pre-game betting odds) is analysed.

The systematic analysis revealed relevant results for further KPI research: First, the magnitude of overall correlation coefficients was higher for relative metrics than for absolute metrics. Second, the correlation of metrics with the strength of a team is stronger than the correlation with the game success of a team. Third, correlation analysis with team strength indicated more positive associations, while correlation analysis with success is most likely confounded by the intermediate score line of a game and revealed more negative associations.

Study XIII

Machine learning and data imputation approaches to handle the issue of data sparsity in sports forecasting

Reference:

Wunderlich, F., Biermann, H., Yang, W., Bassek, M., **Raabe, D.**, Elbert, N., Memmert, D., & Garnica Caparrós, M. (2024). Machine learning and data imputation approaches to handle the issue of data sparsity in sports forecasting.

Manuscript accepted for publication in Machine Learning.

Abstract

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-match-ahead 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.