Ulf Brefeld
Jesse Davis
Jan Van Haaren
Albrecht Zimmermann (Eds.)

Communications in Computer and Information Science 2460

Machine Learning and Data Mining for Sports Analytics

11th International Workshop, MLSA 2024
Vilnius, Lithuania, September 9, 2024
Revised Selected Papers

Springer

Communications in Computer and Information Science 2460

Rationale
The CCIS series is devoted to the publication of proceedings of computer science conferences. Its aim is to efficiently disseminate original research results in informatics in printed and electronic form. While the focus is on publication of peer-reviewed full papers presenting mature work, inclusion of reviewed short papers reporting on work in progress is welcome, too. Besides globally relevant meetings with internationally representative program committees guaranteeing a strict peer-reviewing and paper selection process, conferences run by societies or of high regional or national relevance are also considered for publication.

Topics
The topical scope of CCIS spans the entire spectrum of informatics ranging from foundational topics in the theory of computing to information and communications science and technology and a broad variety of interdisciplinary application fields.

Information for Volume Editors and Authors
Publication in CCIS is free of charge. No royalties are paid, however, we offer registered conference participants temporary free access to the online version of the conference proceedings on SpringerLink (http://link.springer.com) by means of an http referrer from the conference website and/or a number of complimentary printed copies, as specified in the official acceptance email of the event.

CCIS proceedings can be published in time for distribution at conferences or as post-proceedings, and delivered in the form of printed books and/or electronically as USBs and/or e-content licenses for accessing proceedings at SpringerLink. Furthermore, CCIS proceedings are included in the CCIS electronic book series hosted in the SpringerLink digital library at http://link.springer.com/bookseries/7899. Conferences publishing in CCIS are allowed to use Online Conference Service (OCS) for managing the whole proceedings lifecycle (from submission and reviewing to preparing for publication) free of charge.

Publication process
The language of publication is exclusively English. Authors publishing in CCIS have to sign the Springer CCIS copyright transfer form, however, they are free to use their material published in CCIS for substantially changed, more elaborate subsequent publications elsewhere. For the preparation of the camera-ready papers/files, authors have to strictly adhere to the Springer CCIS Authors' Instructions and are strongly encouraged to use the CCIS LaTeX style files or templates.

Abstracting/Indexing
CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago, Scopus. CCIS volumes are also submitted for the inclusion in ISI Proceedings.

How to start
To start the evaluation of your proposal for inclusion in the CCIS series, please send an e-mail to ccis@springer.com.

Ulf Brefeld · Jesse Davis · Jan Van Haaren ·
Albrecht Zimmermann
Editors

Machine Learning and Data Mining for Sports Analytics

11th International Workshop, MLSA 2024
Vilnius, Lithuania, September 9, 2024
Revised Selected Papers

Springer

Editors
Ulf Brefeld
Leuphana University
Lüneburg, Germany

Jesse Davis
Katholieke Universiteit Leuven
Leuven, Belgium

Jan Van Haaren
Katholieke Universiteit Leuven
Leuven, Belgium

Albrecht Zimmermann
Université de Caen Normandie
Caen Cedex 5, France

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-86691-3 ISBN 978-3-031-86692-0 (eBook)
https://doi.org/10.1007/978-3-031-86692-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The Machine Learning and Data Mining for Sports Analytics workshop aims to bring together a diverse set of researchers working on Sports Analytics in a broad sense. In particular, it aims to attract interest from researchers working on sports from outside of machine learning and data mining. The 11th edition of the workshop was co-located with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery 2024.

Sports Analytics has been a steadily growing and rapidly evolving area over the last decade, both in US professional sports leagues and in European football leagues. The recent implementation of strict financial fair-play regulations in European football will definitely increase the importance of Sports Analytics in the coming years. In addition, there is the popularity of sports betting. The developed techniques are being used for decision support in all aspects of professional sports, including but not limited to:

- Match strategy, tactics, and analysis
- Player acquisition, player valuation, and team spending
- Training regimens and focus
- Injury prediction and prevention
- Performance management and prediction
- Match outcome and league table prediction
- Tournament design and scheduling
- Betting odds calculation

The interest in the topic has grown so much that there is now an annual conference on Sports Analytics at the MIT Sloan School of Management, which has been attended by representatives from over 70 professional sports teams in eminent leagues such as Major League Baseball, National Basketball Association, National Football League, National Hockey League, Major League Soccer, English Premier League, and the German Bundesliga. Furthermore, sports data providers such as Statsbomb, Stats Perform, and Wyscout have started making performance data publicly available to stimulate researchers who have the skills and vision to make a difference in the sports analytics community. Moreover, the National Football League has been sponsoring a Big Data Bowl where they release data and a concrete question to try to engage the analytics community.

There has been growing interest in the Machine Learning and Data Mining community about this topic, and the 2024 edition of MLSA built on the success of prior editions at ECML PKDD 2013, ECML PKDD 2015 – ECML/PKDD 2023.

Both the community's on-going interest in submitting to and participating in the MLSA workshop series and the fact that you are reading the sixth volume of MLSA proceedings show that our workshop has become a vital venue for publishing and presenting sports analytics results. While other researchers occasionally organize workshops co-located with other conferences, and sports analytics papers appear more often in main

conferences, MLSA remains a primary venue for publishing sports analytics research from a machine learning and data mining perspective.

In fact, when discussion turned to whether there was interest in continuing the series, participants were overwhelmingly in favor, with several volunteering to take on organizational tasks.

In 2024, the workshop received 21 submissions, in line with historical trends but off the record amount of 2023, of which 13 were selected after a single-blind reviewing process involving at least three program committee members per paper. Four of the papers were extended abstracts based on journal publications. In terms of the sports represented, this year's workshop contained a healthy mix, with team sports continuing to be heavily represented. Topics included tactical analysis, outcome predictions, data acquisition, performance optimization, and player evaluation.

Deviating from previous years, we did not have an invited presentation this year but instead used a panel discussion to hear from practitioners in the field – Kuba Michalczyk, Sportec Solutions AG, Hugo Rios-Neto, Gemini Sports Analytics, Henrik Biermann, Borussia Dortmund – who exchanged ideas with academic researchers.

Further information about the workshop can be found on the workshop's website at https://dtai.cs.kuleuven.be/events/MLSA24/.

December 2024

Ulf Brefeld
Jesse Davis
Jan Van Haaren
Albrecht Zimmermann

Organization

Workshop Co-chairs

Ulf Brefeld	Leuphana University Lüneburg, Germany
Jesse Davis	KU Leuven, Belgium
Jan Van Haaren	Club Brugge, Belgium & KU Leuven, Belgium
Albrecht Zimmermann	Université de Caen Normandie, France

Program Committee

Gennady Andrienko	Fraunhofer IAIS, Germany
Adrià Arbués Sangüesa	Zelus Analytics, Spain
Joris Bekkers	UnravelSports, Netherlands
Harish S. Bhat	University of California Merced, USA
Lotte Bransen	SciSports, Netherlands
Arie-Willem de Leeuw	Universiteit Leiden, Belgium
Martin Eastwood	UK
Clement Gautrais	Brightclue, France
Kilian Hendrickx	KU Leuven/Siemens Digital Industries Software, Belgium
Mehdi Kaytoue	Infologic, France
Leonid Kholkine	University of Antwerp, Belgium
Hyunsung Kim	Fitogether Inc., South Korea
Arno Knobbe	Leiden University, Netherlands
John Komar	Nanyang Technological University, Singapore
Patrick Lambrix	Linköping University, Sweden
Steven Latre	University of Antwerp, Belgium
Etienne Lehembre	Université de Caen Normandie, France
Alexis Mortelier	Université de Caen Normandie, France
Konstantinos Pelechrinis	University of Pittsburgh, USA
Pegah Rahimian	Budapest University of Technology and Economics, Hungary
François Rioult	GREYC CNRS UMR6072 - Université de Caen Normandie, France
Pieter Robberechts	KU Leuven, Belgium
Nathan Sandholtz	Brigham Young University, USA
Robert Seidl	Stats Perform, Germany

Maaike Van Roy	KU Leuven, Belgium
Nicolas Vergne	Université de Rouen Normandie, France
Steven Verstockt	University of Ghent, Belgium
Josh Weissbock	Carleton University, Canada

Contents

Individual Sports

Characterizing Serves in Table Tennis

Aymeric Eradès[1,2], Thomas Papon[1,2], and Romain Vuillemot[1,2(✉)]

[1] École Centrale de Lyon, Lyon, France
[2] LIRIS CNRS UMR 5205, Lyon, France
romain.vuillemot@ec-lyon.fr

Abstract. In table tennis, serves play a crucial role as they are the first and only shot during which players have full control of the game. In this paper, we explore serve techniques and tactics using a dataset of 9 games and a total of 510 serves collected semi-automatically for 5 players. We first provide a descriptive analysis and group the data into clusters to identify what we refer to as a serve repertoire, which are serve categories specific to players. We then identify which serve is used during the game and the tactic players employ during competitions. In particular, we provide a better understanding of the notion of variation, a concept often used in table tennis, and we show the differences in serves that are used in key moments of a game (e. g., during score domination, decisive points).

Keywords: Sports Analytics · Table Tennis · Serves

1 Introduction

Serves are the very first action in racket sports, in particular in table tennis. According to Larry Hodges [7] they are *the most strategic and tactical part of your game*. In table tennis, serves consist of a ball bouncing on the two sides of the table, following a set of specific rules. There are several types of serves, each with unique techniques and tactical purposes. Examples of techniques are numerous. The *topspin serve*, which generates forward spin to make the ball bounce higher and faster. The *backspin serve*, which produces reverse spin, causing the ball to slow down and stay low. The *side-spin serve*, which gives a horizontal spin to make the ball curve sideways. And the *no-spin serve* which aims to surprise the opponent by having minimal to no spin. Serves can also be grouped by ball placements with *short* and *long serves*, and *pivot serves* between the forehand and backhand area. Further nuances can be added by adjusting the speed, angle, ball toss of the serve, offering players a wide palette of tactical options to initiate play effectively. Some serves also have colorful names such as the Ghost Serve (heavy backspin), Tomahawk (the racket is swung like a tomahawk), Pendulum (the server swings the paddle in a pendulum-like motion generating a lot of side-spin), or Lollipop (slow shot with little spin). In general, there is no official or formal classification in particular related to data that can be collected

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 3–13, 2025.
https://doi.org/10.1007/978-3-031-86692-0_1

for this sport. Regarding the tactical aspect of serves, the goal is to find efficient serves according to the opponent and the score of the game, with variations to make the opponent's prediction and anticipation difficult.

In this paper, we aim to characterize serves in a deeper way. Indeed, while both technique and ball placement can characterize them, they are limited for advanced tactical analysis. The first limitation is that the current classification of techniques and ball placement does not capture the intent of the player to anticipate the next shot. The second is that this classification is too coarse, so similar serves may be grouped in different categories, and different serves can be grouped in the same category. Besides making it impossible to accurately classify serves, the key tactical notion of variations cannot be fully explored. If accurate categorization is achieved, then many longstanding open questions in table tennis can be answered: What is the total number of serves?[1] What are the signature serves and the ones used for critical points of the game? What are the main profiles of servers?

To achieve this, we first collected a detailed dataset of serves that includes both the players' positions and techniques, ball placement in an accurate way, and the opponent's returning stroke. This provided a way to group serves by second ball bounce placements to reveal clusters on the opponent's side. To refine such clusters, we split them by the first ball bounce and the technique of the server as a style prior [15]. This led us to a specific number of clusters we referred to as the *serve repertoire* from which the server can pick. We then conducted an analysis of the use of this repertoire to identify when they are used first, their variations, and their use in key moments. We released our code and data as an open-source project to foster more research in this area.

2 Background and Related Work

Serves follow specific rules beyond the two bounces. Each player serves two consecutive points, alternating throughout the game, with adjustments during the final game at deuce. The server must place the ball on their open hand, throw it vertically at least 16 cm, and strike it so it bounces on their side before crossing the net to the opponent's side. Players have to hit the ball behind the edge of their side of the table. The serve must be visible to the opponent, and the free hand must not obstruct the view. Let serves, which touch the net but land correctly, are replayed, while fault serves result in a point for the opponent.

In table tennis, most analyses are based on *rallies*, which can be seen as multivariate sequences and are the focus of most analytical methods. In TacticFlow [16], a mining method is proposed to mine frequent patterns and detect how these patterns change over time using MDL (Minimum Description Length). In [5], sequence mining is achieved using the SPADE algorithm to determine the most frequent tactics and the WRAcc measure to keep the most

[1] Some players claim they have around 100 serve variations https://www.canalplus.com/sport/extrait-interieur-sport-la-nouvelle-dynastie-les-freres-lebrun/h/24369553_50001 (in French).

relevant ones. In [17], they reveal the hit correlations and hit sequences. In [14], a Markov chain model is used to characterize and simulate the sequence process in table tennis. In these works, the serve is only considered as a particular technique attribute. They eventually serve as a descriptor of the type of hit in [8], with explicit use of the pendulum, reverse, and tomahawk serves. However, as far as we know, there does not exist a detailed analysis of serves, particularly due to the categorization of the ball placement on the 3×3 grid [2,5,17] or in a 3×3 grid in 3D [4], which does not capture placement subtleties. Additionally, none of them focus on the exact ball position and first rebound analysis.

Other similar racket sports (e. g., tennis, badminton) share similarities as two players alternate hitting the ball, with one player serving and one player winning the point. Still there is little work focusing on serves analysis. [19] introduces a serve prediction method in table tennis. They extensively review the current performance of prediction models. However, they only rely on shot-by-shot data and do not leverage the potential of tracking data [11], particularly what Hawk-Eye-like systems could provide [10]. [4] analyze rallies in badminton solely based on trajectories reconstructed from the shuttle's kinematic features. Both works are difficult to translate to table tennis, where there is a first bounce and heavy ball spin that is difficult to detect.

3 Data Collection and Exploratory Data Analysis

Our first step was to collect a detailed and representative dataset of serves. We used TV broadcast videos available on the ITTF Channel on YouTube to collect data from professional games. We employed a combination of automated (e. g., OpenPose [3]) and semi-automated computer vision techniques to extract players' positions, ball hit locations, and event characterizations. We picked this approach as there is currently no fully automated and accurate table tennis game reconstruction despite recent advances in this area [13]. We collected full rallies, but in this paper, we focus only on the analysis of the first stroke (serve) and the return technique used (not its placement). We also collected context data such as the outcome of the point (winning/losing) and contextual elements such as scores and players' names. We collected a total of 9 games and up to 510 serves for 5 players. The dataset has been released publicly and is available online[2]. We removed data from left-handed players as they provided spatial inconsistencies, so we only investigated games against right-handed players. We augmented the collected dataset using several post-processing steps to calculate the score and various metrics [2] for the section related to tactical analysis. We also reconstructed 3D ball trajectories [1] to gather details on the server's technique. We stored both the collected data and the augmented data in a database that can be quickly queried and joined with other metadata.

We then conducted an exploratory data analysis [12] of serves to grasp the distribution of serves and identify the attributes and parameters that could separate different types of serves. This step also helped to diagnose any data inaccuracies

[2] https://github.com/centralelyon/table-tennis-services/.

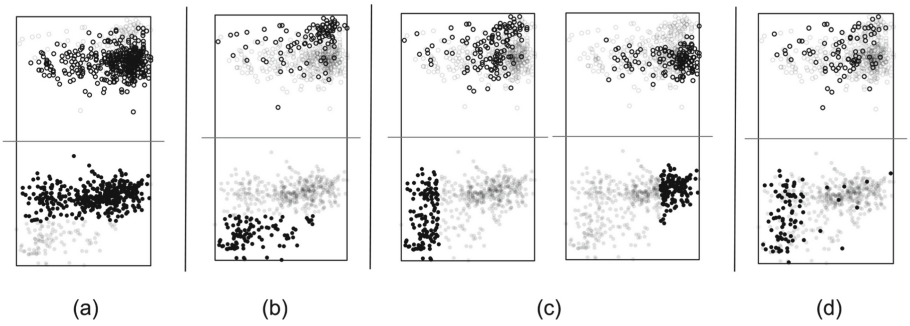


Fig. 1. Spatio-temporal representation of serves we collected, normalized, and filtered. The server is at the top; the dots on the upper part are the first bounce, the ones below are the second bounce, and the receiver is at the bottom. Gray dots represent all the serves, and the black ones: (a) short serves, (b) long serves, (c) side serves, and (d) pivot serves.

and inconsistencies. Figure 1 shows that standard ways to explore ball placement (e. g., by position on the table) are not sufficient to distinguish between types of serves. However, it revealed some interesting patterns, such as regions that are physically impossible to reach due to ball rebound physics (e. g., close to both sides of the net). Also, many serves are not attempted because their trajectory would lead to a fault if both bounces occur on the same side. Some possible bouncing regions are intentionally avoided for tactical purposes, such as those close to the receiver's forehand, which is the most efficient stroke. Another tactical pattern that was revealed was outliers, which are uncommon and may surprise the opponent. These outliers can be easily observed as their ball placement and trajectories are distinctive.

4 Serves Categorization

We detail in this section our method to group serves in a way to reveal players repertoire by grouping similar serves as a combination of second-bounce ball placement and technique.

Our goal is to group balls that land on the other side of the table into meaningful clusters. Similarly to [15] relied upon K-means clustering to automatically find a way to partition 2D space based as the clusters we are looking for are roughly circular, centered around an aiming placement. Figure 2 (b) shows a clear separation with centroid representing the center of the target point and the spread around it (likely due to inaccuracies). Other methods could be used such as HDBSCAN [9] used in Badminton [18] or in Tennis [19]. But those methods are trajectory based and we consider that the current method we used was not sufficiently accurate to reconstruct them. We used K-means combined with the *elbow method* (an heuristic aiming at maximizing similarity within clusters while minimizing similarity between clusters) to determine the number of

clusters. Despite the simplicity of this method, we visually assessed the resulting clusters and found them relevant in both the number, shape and centroid of the groups, to globally separate serves. We also found those clusters relevant to group serves from different games with the same players.

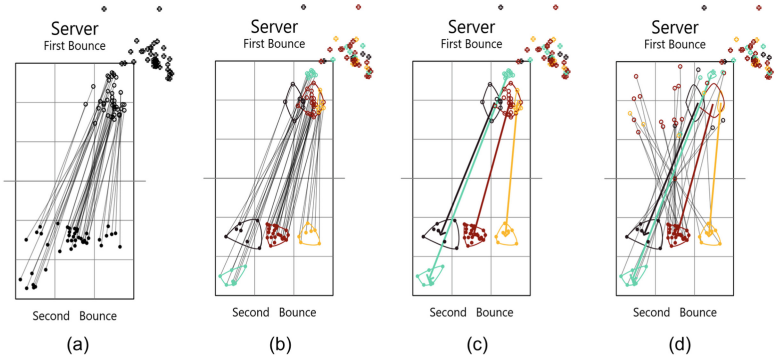


Fig. 2. Steps to create serves clusters (a) a single game is picked for a player, (b) K-means clusters are created based on the second-bounce, (c) the clusters are connected to their corresponding first-bounces, (d) we show the corresponding returns by the other players with lines.

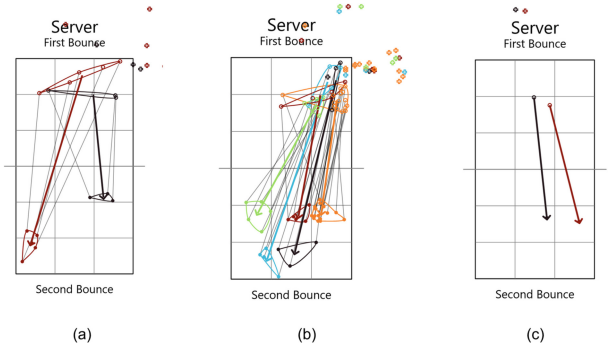


Fig. 3. Each clusters were then separated using serving technique (a) forehand serve with right side-spin, (b) forehand serve with left side-spin (c) backhand serve with left side-spin.

We refined the grouping based on the technique used when serves as two different techniques could lead to the same second-bounce cluster but with a different incoming trajectory (Fig. 3). Also, as the clusters were sometimes too large, and differences were caused by the effect in the serve. Figure 2(c) shows such deeper separation by including the serve technique. Finally we found a total of N clusters which will be identified later as $C_1, C_2, \ldots, C_N$, ranked by number

of serves in each. Each cluster C_i is represented by an area, a density and a distance to the first-bonce cluster (to indicate if this is a short or a long serves). We displayed on Fig. 4 the results from the application of the methodology we previously used by picking a single game for our collection of players.

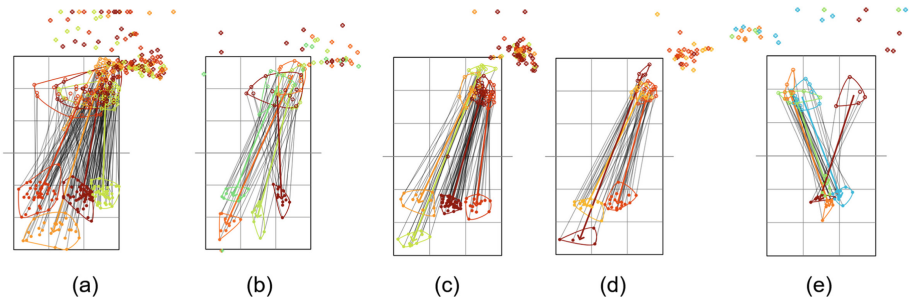


Fig. 4. Server's repertoire for top players: a small first bouncing area ((d) **Player4**, (c) **Player3**), which is explained by a static position; spreading the first bounce across the width ((e) **Player5**); the same serve but from different initial positions ((a) **Player2**, (b) **Player1**).

5 Servers Tactics

In this section we define as tactics the use of efficient serves, and our goal is to reveal which serves are concerned and when they are used.

5.1 Serves Repertoire

Table 1 provides a shorter, more descriptive summary of all the serves clusters we identified for top ranked players. Those are only from male, right-handed players to obtain some comparable results. This enabled to characterize the general signature of players showing either that they limit to same serves in general or have a high level of diversity.

Table 1. Details of the 4 most frequent serves from each of the top-player repertoire we picked. The area % is calculated compared to one half of the table. The distance (*dist*) is in cm (length of a Table is 274 cm and diagonal 313 cm).

Player	$C_1(area, density, dist)$	C_2	C_3	C_4
Player1	(13%, 0.0196, 154)	(8%, 0.0139, 229)	(6%, 0.0160, 174)	(4%, 0.0192, 195)
Player2	(15%, 0.0127, 144)	(8%, 0.0085, 237)	(7%, 0.0079, 156)	(1%, 0.0437, 146)
Player3	(18%, 0.0056, 259)	(18%, 0.0017, 160)	(9%, 0.0034, 166)	(14%, 0.0018, 155)
Player4	(36%, 0.0023, 175)	(34%, 0.0036, 159)	(36%, 0.0007, 246)	
Player5	(3%, 0.0289, 145)	(2%, 0.0383, 160)	(1%, 0.0319, 131)	(1%, 0.0241, 218)

5.2 Serves Similarity

We define a way to group serves based on their similarities. An example of low similarity would be **Player3** who does a long serve C_1 and then a short serve C_3. This aims to capture the level of difference between clusters (we assume inter-cluster serve variations are due to how serves were executed, with a certain level of imperfection). To formalize this intuition, we extend the *shot diversity* metric introduced in [2]. The metric used below follows the following order of priority: serves with different lateralities will be more dissimilar than serves with different side-spins, which in turn will be more dissimilar than serves that bounce a second time in different clusters. We will also consider the distance between different clusters, starting from the centroids of these clusters. Finally, to highlight the variations induced by long serves, we will introduce a higher coefficient for distances along the length of the table than for the width. Here is the exact definition of the distance between two serves used:

$$D(S_1, S_2) = d_{lat}(S_1, S_2) + d_{sidespin}(S_1, S_2) + d_{cluster}(S_1, S_2) \tag{1}$$

with

- $d_{lat}(S_1, S_2) = 0.3$ if both serves have different lateralities (say, one is a forehand while the other one is a backhand), and $d_{lat}(S_1, S_2) = 0$ otherwise.
- $d_{sidespin}(S_1, S_2) = 0.22$ if both serves have different sidespins (say, one is a Left side serve while the other one is a Right side serve), and $d_{sidespin}(S_1, S_2) = 0$ otherwise.
- $d_{cluster}(S_1, S_2) = \|C_2 - C_1\|_1$ if both serves have their second bounce inside a different cluster, and $d_{cluster}(S_1, S_2) = 0$ otherwise. We define C_1 and C_2 the centroids of the first serve's cluster and the second serve's cluster respectively.

Then again, we must define the L1-norm used here: let (x_1, y_1) and (x_2, y_2) be the centroids of two different clusters, C_1 and C_2. Therefore,

$$\|C_2 - C_1\|_1 = \frac{|x_2 - x_1|}{76} \times 0.1 + \frac{|y_2 - y_1|}{137} \times 0.38$$

If 76 and 137 are the dimensions of half a table tennis table in centimeters, meant to homogenize the x and y distances, the coefficients 0.1 and 0.38 are chosen arbitrarily to allow the distance D between two serves to fall within [0,1], and to highlight the variation introduced by a long serve within a series of serves. Such a definition then results in typical distances of 0.12 between a cluster of serves on the opponent's forehand side and another on the backhand side, and 0.16 between a short serve and a long serve, both in the middle of the table.

Finally, the term *diversity* is introduced to help quantify the diversity of serves used by a player over the course of a set. It is defined as:

$$diversity = \sum_{i=1}^{n-1} \left(\frac{D(S_i, S_{i+1})^2}{n-1} \right) \times 100$$

with S_i being the i^{th} serve of a predetermined player in a set, and n being the number of times he served this set. Below is the *diversity* results found for **Player2** and **Player3**, depending on the set :

Table 2. Diversity across sets for each player

Player/Set	Set 1	Set 2	Set 3	Set 4	Set 5
Player 3	0.77	1.09	2.2	0.95	0.45
Player 2	10.20	14.47	3.03	4.96	4.39

5.3 Serving Tactics

We are now interested in understanding the tactics that motivate the change of serves. We will explore such tactics based on diversity and variations of serves, e. g., to understand which players prefer to create surprise or the ones that stick to efficient serves.

Figure 5 illustrates the distances and variation metrics of serves during a single game between **Player2** and **Player3**. It clearly shows that **Player2** tends to use a more diverse range of serves compared to **Player3**. Due to his extensive repertoire, **Player2** explores various serves while occasionally sticking to effective ones before introducing new tactics. Table 2 supports this observation, showing that **Player2** won the second set by employing a wide variety of serves, surprising his opponent with new tactics rather than relying on a single effective serve. In terms of the score's influence, a notable trend is that dominating players often adopt a conservative approach, preferring to stick with reliable serves, which aligns with the tactic outlined by [7]: *"Keep using what works"*. On the other hand, **Player3** utilizes fewer serves, resulting in less variation overall, particularly in crucial points. Further investigation is necessary to refine the identified clusters and potentially uncover subtle variations.

6 Discussion and Perspectives

This paper is a preliminary attempt to characterize table tennis serves using detailed data on bouncing positions from both sides of the table. However, it has several limitations primarily related to the level of details and volume of the data used. Regarding the volume, we aim to collect more serves to build a representative repertoire, as the games we analyzed likely constitute only a subset of all serves used, particularly the less frequently observed ones. The reconstruction of detailed ball trajectories remains an ongoing challenge in computer vision [13]. To address this, we plan to advance research in serve detection by contributing our annotated dataset to computer vision benchmarks [6]. Nonetheless, numerous influential factors, such as players' racket surfaces, deceptive body movements, and external conditions, are challenges to accurately predict ball trajectories.

We also conducted quantitative evaluations of both the resulting clusters and the tactics employed during games. Our future plans include interviewing players from our dataset to gather feedback on the clusters and influential tactical factors (e. g., preparation, fatigue, risk-taking). Looking ahead, we intend to expand our dataset to include doubles matches, which offer more diverse combinations of servers and returners. Additionally, we aim to explore serves used by left-handed players, as well as how players adapt based on their handedness (left or right) or grip style (shake-hand or pen-hold). Further investigation into larger datasets could address questions such as: Do players have invariant routines or recurring serve sequences based on game context? Are there differences in serve tactics between top-ranked players and lower-ranked players? Lastly, do players possess secret serves reserved for critical moments in a match?

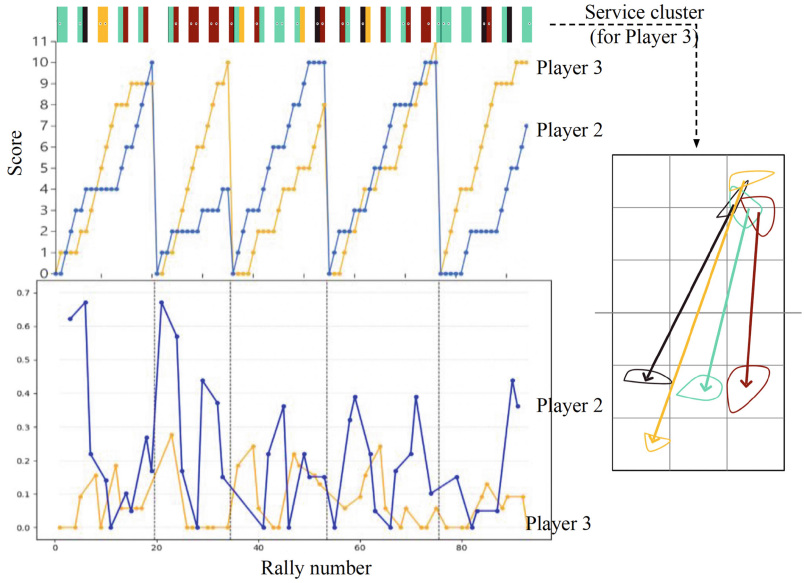


Fig. 5. From top to bottom: serves used by **Player3** against **Player2** (colors represent the point's winner); score evolution; and distances between 2 consecutive serves (colors represent the server).

References

1. Calandre, J., Peteri, R., Mascarilla, L., Tremblais, B.: Extraction and analysis of 3D kinematic parameters of table tennis ball from a single camera. In: 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, pp. 9468–9475. IEEE (2021). https://doi.org/10.1109/ICPR48806.2021.9412391, https://ieeexplore.ieee.org/document/9412391/

2. Calmet, G., Erades, A., Vuillemot, R.: Exploring table tennis analytics: domination, expected score and shot diversity. In: Link, S. (ed.) Machine Learning and Data Mining for Sports Analytics. Communications in Computer and Information Science, Turin, Italy (2023). https://hal.science/hal-04240982
3. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. **43**(1), 172–186 (2021). https://doi.org/10.1109/TPAMI.2019.2929257
4. Chu, X., et al.: TIVEE: visual exploration and explanation of badminton tactics in immersive visualizations. IEEE Trans. Vis. Comput. Graph. **28**(1), 118–128 (2021)
5. Duluard, P., Li, X., Plantevit, M., Robardet, C., Vuillemot, R.: Discovering and visualizing tactics in a table tennis game based on subgroup discovery. In: Machine Learning and Data Mining for Sports Analytics, vol. 1783, pp. 101–112. Springer Nature Switzerland, Cham (2023)
6. Erades, A., et al.: SportsVideo: a multimedia dataset for event and position detection in table tennis and swimming. In: MediaEval Workshop 2023 (2023)
7. Hodges, L.: Table Tennis Tactics for Thinkers. CreateSpace Independent Publishing Platform, Leipzig, February 2013
8. Lan, J., Wang, J., Shu, X., Zhou, Z., Zhang, H., Wu, Y.: RallyComparator: visual comparison of the multivariate and spatial stroke sequence in table tennis rally. J. Vis. **25**(1), 143–158 (2022). https://doi.org/10.1007/s12650-021-00772-0
9. McInnes, L., Healy, J.: Accelerated hierarchical density based clustering. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 33–42. IEEE (2017)
10. Mecheri, S., Rioult, F., Mantel, B., Kauffmann, F., Benguigui, N.: The serve impact in tennis: first large-scale study of big Hawk-Eye data. Stat. Anal. Data Min. ASA Data Sci. J. **9**(5), 310–325 (2016). https://doi.org/10.1002/sam.11316, https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.11316
11. Perin, C., Vuillemot, R., Stolper, C.D., Stasko, J.T., Wood, J., Carpendale, S.: State of the art of sports data visualization. Comput. Graph. Forum **37**(3), 663–686 (2018) (EuroVis'18). https://doi.org/10.1111/cgf.13447, https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13447
12. Tukey, J.W.: Exploratory Data Analysis, vol. 2. Reading, Mass (1977)
13. Voeikov, R., Falaleev, N., Baikulov, R.: TTNet: real-time temporal and spatial video analysis of table tennis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 884–885 (2020)
14. Wang, J., Wu, J., Cao, A., Zhou, Z., Zhang, H., Wu, Y.: Tac-Miner: visual tactic mining for multiple table tennis matches. IEEE Trans. Vis. Comput. Graph. **27**(6), 2770–2782 (2021). conference Name: IEEE Transactions on Visualization and Computer Graphics. https://doi.org/10.1109/TVCG.2021.3074576, https://ieeexplore.ieee.org/document/9411869
15. Wei, X., Lucey, P., Morgan, S., Carr, P., Reid, M., Sridharan, S.: Predicting serves in tennis using style priors. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2207–2215 (2015)
16. Wu, J., Liu, D., Guo, Z., Xu, Q., Wu, Y.: TacticFlow: visual analytics of ever-changing tactics in racket sports. IEEE Trans. Vis. Comput. Graph. **28**, 835–845 (2022). https://doi.org/10.1109/TVCG.2021.3114832, https://ieeexplore.ieee.org/document/9552436

17. Wu, Y., et al.: iTTVis: interactive visualization of table tennis data. IEEE Trans. Vis. Comput. Graph. **24**(1), 709–718 (2018). conference Name: IEEE Transactions on Visualization and Computer Graphics. https://doi.org/10.1109/TVCG.2017.2744218, https://ieeexplore.ieee.org/document/8017600
18. Ye, S., et al.: ShuttleSpace: exploring and analyzing movement trajectory in immersive visualization. IEEE Trans. Visual Comput. Graph. **27**(2), 860–869 (2021). https://doi.org/10.1109/TVCG.2020.3030392, https://ieeexplore.ieee.org/document/9222313/
19. Zhu, Y., Naikar, R.: Predicting tennis serve directions with machine learning. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds.) Machine Learning and Data Mining for Sports Analytics, pp. 89–100. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-27527-2_7

Large Language Models on Race Commentary: Towards Granular Data in Cycling Analytics

Bram Janssens[1,2,3(✉)], Matthias Bogaert[1,2], and Steven Verstockt[4]

[1] Department of Marketing, Innovation, and Organization, Ghent University, Tweekerkenstr. 2, 9000 Ghent, Belgium
bram.janssens@ugent.be
[2] FlandersMake@UGent–corelab CVAMO, Ghent, Belgium
[3] Research Foundation Flanders, Brussel, Belgium
[4] Ghent University - Imec, IDLab; Technologiepark-Zwijnaarde 122, 9052 Gent, Belgium

Abstract. Current cycling analytics studies are limited to data about the eventual race results. This study searches how online commentary can be used to capture information about in-race dynamics by harnessing the power of large language models. The results show that the direct application of these models is already promising but not accurate enough to base end-to-end machine learning applications on the generated data. Our results show the tendency of these models to use information from previous queries in its generation step, which indicates data leakage and might hamper the scientific validation of approaches comparing various techniques. To capture overall rider behavior we suggest using graph representation learning. Our results indicate that this method is capable of identifying similar rider behavior, which to date was not yet feasible.

Keywords: Cycling Analytics · Text Mining · Data Collection · Large Language Models · Graph Representation Learning

1 Introduction

Cycling analytics has seen an increase in popularity in the last couple of years, with applications such as race outcome estimation (Kholkine et al., 2021), talent identification (Van Bulck et al., 2023; Janssens, Bogaert & Maton, 2023), performance evaluation (de Leeuw et al., 2023; Janssens & Bogaert, 2023), and team roster optimization (Sagi et al., 2024).

Despite this surge in interest, one large issue is the fact that currently no data is available about what happens during the races. This is unfortunate, as this information should ideally be included in all the mentioned applications to get an objective race and rider evaluation. For instance, in talent identification a rider who is attacking-oriented but met a lot of malfortune due to crashes and did not finish the race will not be identified by the current models. However, this rider can still be an interesting prospect for scouting purposes.

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 14–25, 2025.
https://doi.org/10.1007/978-3-031-86692-0_2

Other sports such as soccer saw a similar revolution occurring recently. Historically, soccer analytics was heavily focused on match outcome (i.e., goals), which resulted in the majority of research being directed towards the valuation of goal attempts (Decroos et al., 2019). This limited the reach of the studies, after which researchers started collecting and sharing more fine-grained data sets about what happened during the match (Pappalardo et al., 2019). This new information allowed for totally new applications such as action evaluation (Decroos et al., 2019), player vectorizations (Decroos & Davis, 2019), pressing evaluation (Robberechts, 2019) and player chemistry evaluation (Bransen & Van Haaren, 2020).

Unfortunately, the situation is a bit more complicated in road cycling. First, computer vision-based approaches (Okuma et al., 2004) popular in team sports on a fixed playing field (e.g., hockey field, soccer field) with fixed cameras are much more difficult to succeed in road cycling races, where up to 200 participants ride very close to each other, in a landscape and camera angle which changes the entire time. This explains that attempts to develop a tracking technique (Verstockt et al., 2020) which was developed for both road cycling and cyclocross only matured into a methodology to create fine-grained data for cyclocross segments (De Bock & Verstockt, 2021). A second option, the usage of tracking spatio-temporal data using wearable devices, currently popular in soccer (Pappalardo et al., 2019), also did not result into a useful publicly available data source to date. Knowing the exact physical abilities of your opponents can result into large competitive gains in endurance sports such as cycling, which hinders many competitors from being willing to share the data from these devices.

Rather, current in-race data gathering approaches focus on the timing of when riders' devices pass by some hardware which is installed next to the race course (Kolaja & Ehlerova, 2019; Decorte et al., 2024). To install this hardware across all races, a serious investment would be required, while it also does not bear the potential to retrieve information about past races, which limits the retroactive abilities of this technique. Accordingly, this study sets out to explore an alternative in-race data collection methodology. A lot of online commentary on professional road cycling races is provided on the online platform X, formerly known as Twitter. This textual information bears a lot of potential to capture in-race dynamics and events.

We deploy popular large language models (i.e., GPT-3.5-Turbo and GPT-4o) and demonstrate that these models can be used to retrieve structured information about who is positioned where during the race after the usage of few-shot learning. The results are promising, but only after some manual cleaning. GPT-4.o is demonstrated to outperform its predecessor on our task, but there is an additional cost associated to this when deployed on a large sample of races. Finally, we show that representing the uncleaned large language model output as a graph combined with node representation learning is demonstrated as an interesting avenue to automate in-race rider representation which might be useful for downstream tasks.

2 Literature Review

The application of data mining and machine learning in road cycling has remarkably increased in recent years. While the range of applications has grown, they have been focused on a relatively narrow list of data sources, as depicted in Table 1. Historically, two types of data have been used: race results databases and sensor data. Sensor data is typically utilized when an application is developed for a single team, in collaboration with a team. Recent examples are the automation of rider selection per race at Israel – Premier Tech (Sagi et al., 2024), and the pre-race estimation of a single rider's race form at Team Jumbo-Visma (de Leeuw et al., 2023). These applications work very well when a single team is involved, but are difficult to deploy across multiple teams as many teams are not willing to share their sensor data to other teams, which renders applications such as talent identification or race outcome estimation infeasible with sensor data, as one would need the sensor data for practically every participating rider. Rather, these studies use race results databases. These databases, such as ProCyclingStats[1] or CQ ranking[2], typically contain detailed information about the race results, as well as some limited structured information about the course (e.g., total distance, total elevation gain), with the recent addition (Janssens & Bogaert, 2023; De Bock & Verstockt, 2023) of detailed course information into the applications. However, just looking at the end result in a sport with races taking up to 6–7 h, while race tactics heavily influence race outcome, entails an unavoidable loss of information. Current applications give no value to valuable team mate contributions, dangerous attacks, or rider malfortune (e.g., crashes or flat tires). Up until today, only one study used in-race information. When building a methodology to score how dangerous different road segments are to improve rider safety, De Bock and Verstockt (2023) use a database about in-race crashes, which is created through the manual input by stakeholders in the sport.

Such a manual database is a necessity, as currently no public data set is available on in-race information. Current approaches (Kolaja & Ehlerova, 2019; Decorte et al., 2024) to capture in-race information need the deployment of hardware installation near the race course, and then register when each rider passes next to these installations to have some in-race information. Such an approach cannot gather information about past historic races, while optical tracking (Verstockt et al., 2020) did not yield sufficient accuracy. Therefore, this study explores another avenue: the processing of unstructured textual race commentary.

[1] https://www.procyclingstats.com/.
[2] https://cqranking.com/men/asp/gen/start.asp.

Table 1. Previous Studies in Cycling Analytics. Lack of in-race data has resulted in ignorance of this aspect in previous research. RRD = Race Results Databases; SD = Sensor Data; DCI = Detailed Course Information; IRI = In-Race Information.

Study	Application	RRD	SD	DCI	IRI
Hilmkil et al. (2018)	Heart Rate Prediction		x		
Kataoka & Gray (2018)	Power Output Prediction		x		
De Spiegeleer (2019)	Race Outcome Estimation	x			
Karetnikov (2019)	Power Output Prediction	x	x		
Kholkine et al. (2021)	Race Outcome Estimation	x			
Van Bulck et al. (2023)	Talent Identification	x			
Janssens, Bogaert & Maton (2023)	Talent Identification	x			
Baron, Janssens & Bogaert (2023)	Rider Representation	x			
de Leeuw et al. (2023)	Form Estimation		x		
Kholkine et al. (2023)	Peak Age Estimation	x			
De Bock & Verstockt (2023)	Crash Prevention			x	x
Janssens & Bogaert (2023)	Performance Evaluation	x		x	
Sagi et al. (2024)	Team Roster Optimization	x	x		
de Leeuw & Kholkine (2024)	Peak Age Estimation	x			

3 Methodology

3.1 Data

The goal of this study is to investigate whether it is feasible to structure Tweets (i.e., messages on X) into structured information which can aid current cycling analytics solutions. As a case study, we have selected the Dutch WorldTour Amstel Gold Race 2024. This race was selected as its course lends itself towards various attacks, which results into a dynamic race progression, while it is also as one of the more important one-day races on the calendar. This ensures sufficient coverage by the accounts discussing professional road cycling races on X.

A list of X accounts was created who frequently discuss road cycling races on the platform. From these accounts, all Tweets posted between the race starting (10:45:00 CET) and ending (16:43:17 CET) times were collected, resulting in a sample of 336 Tweets. While this sample would still be humanly annotatable, this would no longer be feasible if the process was re-iterated for the thousands of races reported in other studies (Janssens & Bogaert, 2023). Figure 1 gives an example of such a Tweet which

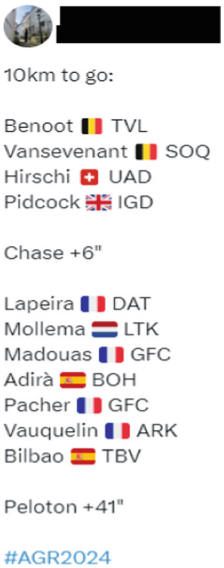


Fig. 1. Example Tweet

clearly contains valuable information about the positioning during the race. Unfortunately, the information is in an unstructured format (i.e., text) which often differs across tweets. Moreover, this is even further complicated by the fact that that riders can be addressed through family names, given names, full names, and nicknames. Accordingly, a method needs to be formulated to re-structure this information into machine-readable formats and link the correct riders.

3.2 Prompt Engineering

Large Language Models (LLMs) are quickly becoming the new standard in Natural Language Processing. These models are trained on billions of data points and parameters and are capable of capturing complex contextual information hidden in natural language and text. This pre-training on vast amounts of data makes them effective in various tasks without a lot of training data if the model is provided with some contextual intermediate reasoning steps (Wei et al., 2022), also called prompt engineering. This can be especially useful in our case, given the situation we encounter. For example, in Fig. 1 the tweet mentions 11 unique riders from the starting list of 175 riders. This translates to a 175-class problem with only 336 observations, with the class options (i.e., participating riders) also changing from race to race. Moreover, the task is even further complicated by the fact that besides rider, we also need to retrieve the mentioned time gaps per rider, making it a multi-label task. Using more traditional fine-tuning machine learning approaches would require vast amounts of training data. Accordingly, we use a few-shot learning approach where we ask the model to act as a Tweet annotator (i.e. system role) for a cycling observation task: "You are a helpful assistant specialized in interpreting Tweets, which has to go through a large chunk of text on updates about a bicycling race.". The final prompt (i.e. user role) is:

"You are a helpful assistant specialized in interpreting Tweets. You are a huge cycling fan and want to get information about what happens during a cycling race, as currently you only have information about the results at the end of the race. To do so, you have to analyze a concatenated heap of text, and restructure this into information about who is positioned where. If no explicit information is given, you must return 'No Information'. Otherwise you should return a list with the mentioned rider(s) and their reported time gaps. Note that the time gaps can be both negative (chasing groups) as well as positive (leading group) and a distinction should be made. The Tweets are written in multiple languages, but your response should always follow the same English structure, regardless of the Tweet. If informative, ONLY return the list, do not explain your reasoning.

Some examples:

Input: 'The lead has an advantage of 1 minute 30.', rider list: ***example_rider_list***.
Response: '['Unknown', '1:30'];
Input: 'Schmid fighting for his 15 second lead.', rider list: ***example_rider_list***. Response: '['SCHMID Mauro', '0:15'];
Input: 'Oomen and Gregoire 30 seconds behind, while the peloton is also closing in.', rider list: ***example_rider_list***. Response: '[['GRÉGOIRE Romain', '-0:30'], ['OOMEN Sam', '-0:30']];
Input: 'The gap to the peloton has been reduced to just over 2 minutes now', rider list: ***example_rider_list*** . Response: '['Peloton', '-2:00']' ;
Input: 'Attack Oomen', rider list: ***example_rider_list***. Response: 'No Information' ;

Now do the same for following input: ***actual_input*** and rider list: ***rider_list***."

With ***actual_input*** being the actual Tweet text (i.e., 336 unique values), ***rider_list*** being a textual list representing the starting list of riders (and 'Peloton') at the Amstel Gold Race (known upfront, retrieved from ProCyclingStats.com), and ***example_rider_list*** being a random list of 10 riders (and 'Peloton'), which includes the riders included in the responses (e.g., Sam Oomen). The few-shot approach allows the model to identify different structures of giving information about the race update, how to perform named entity recognition from a list of starting riders (which differs from the new task), what to do when no riders can be identified, and what to do when no race situation information is provided. This approach proved to be superior over other approaches during preliminary testing which aggregated tweets per time bin, used zero shot learning, and other prompts. Two popular OpenAI models (i.e., GPT-3.5-turbo and GPT-4o) are evaluated against each other. As no ground truth is available, we will compare the stability and known hallucinations of both approaches.

4 Results

An important first observation to make is the large difference in sample size generated by both methods. GPT-3.5-turbo generates 723 rider-timing tuples, while GPT-4o only creates 279 of these tuples. This is even more surprising as the number of Tweets which did not return 'No Information' or 'Unknown' is much more similar (i.e., 85 and 82, respectively). This means that GPT-3.5-turbo is either capable of retrieving more information per Tweet, or that it hallucinates more. Moreover, only 50 of these Tweets overlap for both methods (i.e., only about 60% of each set identified as informative). This already indicates large differences between both method outcomes. To identify differences, we inspect the first five overlapping Tweets (Tweets 1–5), the first five only identified by GPT-3.5-turbo (Tweets 6–10), and the first five only identified GPT-4o (Tweets 11–15) in Table 2. The table already shows clear issues with the results. A first issue is to also include women's race results (which happens around the same time) despite the fact that these riders are not included in the ***rider_list*** variable. Moreover, both models create false information. GPT-3.5-turbo has the tendency to 'hallucinate' large rider lists when no exact information is provided, explaining the larger list of tuples. Interestingly, this list often contains information about riders who are actually involved in race dynamics but not mentioned in the original tweet. For example, the twelve rider list from Tweet 4 (i.e., ['PIDCOCK Thomas', 'HIRSCHI Marc', 'BENOOT Tiesj', 'VANSEVENANT Mauri', 'LAPEIRA Paul', 'MADOUAS Valentin', 'MOLLEMA Bauke', 'PACHER Quentin', 'BILBAO Pello', 'MATTHEWS Michael', 'VAN DEN BERG Marijn', 'ADRIÀ Roger']) actually did contain the riders who were riding at the front, despite the fact that the Tweet itself does not contain this information. Rather, the model probably infers this from earlier calls by the model. For example, Tweet 7 generates a random list of 155 rider (+ peloton), and Tweet 8 generates the actual male top-10 (which was not known at time of posting while Tweets are fed chronologically), while the Tweet actually included an embedded picture of the women's race top-10. This has two main implications. First, too much noise is induced in this application to use the generated output on the single observation level. Second, it highlights an understudied problem in current studies on zero- and few-shot learning studies, that the model learns from earlier calls, which may lead to data leakage when different prompts are compared. The output for GPT-4o is of higher quality, but several issues remain, such as the inclusion of women's race results, instability due the indication of time differences both relative to the front group (e.g., ['PIDCOCK Thomas', '0:00'] in Tweet 4 and ['Peloton', '−3:00'] in Tweet 2) as well as relative to peloton (e.g., ['BENOOT Tiesj', '0:30']) in Tweet 1, and the lack of adhering to the required structure at times (e.g., Tweets 5 and 10). While these might be processed after extensive data cleaning, we are less confident that this process can be done in a high quality way when deployed on hundreds or thousands of races, as other issues might be surfacing. This is especially worrisome given the large costs associated with using GPT-4o. At time of writing, GPT-4o is billed at a price which is ten times higher than the price of GPT-3.5-turbo, which in our study resulted in a cost of around $ 4 for a call deployed to all 336 individual Tweets for GPT-4o. This could amount to a very high cost when deployed on a large set of races while still being of questionable quality.

Table 2. Differences between GPT-3.5-turbo and GPT-4o. * Denotes a large list of names with the number of riders reported between brackets with no time differences reported unless stated otherwise.

Id	Tweet	GPT-3.5-turbo	GPT-4o
1	NL #AGR2024 Tiesj is part of the 12-rider front group after Keutenberg. 💪 ⏱ 30" 🏁 27 km	['BENOOT Tiesj', '0:30']	['BENOOT Tiesj', '0:30']
2	NL #AGR2024 | #TVLSlive 104KM Het peloton lijkt toch gas te geven. De laatste 20min hebben ze 0:48 van de voorsprong afgereden. De 4 hebben nog 3:00 over	['Peloton', '-0:48']	['Peloton', '-3:00']
3	NL #AGR2024 Tiesj is part of the 12-rider front group after Keutenberg. 💪 ⏱ 30" 🏁 27 km	['BENOOT Tiesj', '0:30']	[['BENOOT Tiesj', '0:30']]
4	Sigue en directo la Amstel Gold Race 2024: Tom Pidcock, Pello Bilbao y Roger Adriá, entre los 12 corredores al frente de la carrera a 26 km del final #Ciclismo	* (12)	[['PIDCOCK Thomas', '0:00'], ['BILBAO Pello', '0:00'], ['ADRIÀ Roger', '0:00']]
5	Amstel Gold Race donne: Vos beffa Wiebes	['VOS Marianne', 'WIEBES Lorena']	['VOS Marianne', 'WIEBES Lorena']
6	Auch die Gruppe mit Bilbao macht die Lücke zur Spitze wieder zu! Stark. #AGR2024	['BILBAO Pello', '0:00']	[['BILBAO Pello', 'No Information']]
7	#AGR2024 - 27 km, ora in 12 in testa... 😊 🚴♂ #AGR #AGR24 #Ciclismo #Cyclisme #Cycling	* (156)	'No Information
8	¡Así quedó el top 10 de una carrera que tuvo todos los condimentos! 😷 #AmstelGoldRace #StarPlusLA	* (10)	'No Information'
9	RT @Eurosport_FR: La cruauté de la photo-finish : Cosnefroy annoncé vainqueur en 2022... avant l'énorme désillusion et le triomphe de Kwiat…	['COSNEFROY Benoît', '0:00']	'No Information'
10	40 Sekunden – die Gruppe hat nun richtig gute Karten. #AGR2024	['Peloton', '0:40']	['Unknown', '0:40']
11	VIDEO: Dramatic finale to Amstel Gold Race Ladies 2024 as Marianne Vos punishes Lorena Wiebes' premature celebration #Cycling	['No Information']	['VOS Marianne', 'WIEBES Lorena']
12	Tiesj Benoot valt aan op de Keutenberg en Mathieu van der Poel moet in de achtervolging! Nog 25 kilometer, volg de finale van de Amstel Gold Race in ons liveblog!	['BENOOT Tiesj', 'Unknown']	[['BENOOT Tiesj', '0:00'], ['VAN DER POEL Mathieu', '-0:00']]
13	¡Total white para #MVDP en la #AmstelGoldRace! ♡ ✨ Sintoniza toda la acción por la pantalla de #StarPlusLA	['No Information']	['VAN DER POEL Mathieu', '0:00']
14	RT @UCI_WWT: @marianne_vos wins the 10th édition of @Amstelgoldrace ! 🥇🏘 #UCIWWT #AmstelGoldRace	['MARIANNE VOS', 'No Information']	['MARIANNE Vos', '0:00']
15	#cyclisme 🚴♂ #AmstelGoldRace NL 12 à l'avant : Hirschi, Honoré, Pidcock, Benoot, Vansevenant, Adria, Mollema, Bilbao, Lapeira, Vauquelin, Madouas et Pacher #velo	* (11, including 'No Information')	* (12, indicated at 0:00 time gap)

Rather, we suggest an approach to aggregate these faulty observations in an automated manner, which can be used as an overall representation of rider behavior during the race through graph representation learning. We create a bipartite graph based upon the structured data generated by GPT-3.5-turbo, with the two node types being 'reported

group timing' and 'rider'. Consider Table 3 and Fig. 2 to make this clear. Table 3 visualizes our data after re-structuring it to a traditional table format, with for each rider the time gaps, as well as the original posting of the Tweet. The data is faulty, as the top-4 riders were actually all riding together with an 8 s gap on the group with Lapeira at the time. However, our graph in Fig. 2 shows how this can be 'smoothed' using graph structures. The 16:41:13 observation in Table 3 does not link Pidcock to his three fellow attackers and it will be very hard to identify on a large scale when an LLM makes these types of mistakes. However, as long as there is any other time of posting/gap reported together (i.e., 16:42:02), the overall graph structure will link riders together who are often reported in the same group, which should theoretically give smoothed information about who exhibits similar riding behavior during the race.

Table 3. Example of Structured Information.

Time of Posting	Rider Name	Gap
2024-04-14 16:41:13	PIDCOCK Thomas	0:02
2024-04-14 16:41:13	HIRSCHI Marc	0:08
2024-04-14 16:41:13	BENOOT Tiesj	0:08
2024-04-14 16:41:13	VANSEVENANT Mauri	0:08
2024-04-14 16:41:13	LAPEIRA Paul	0:16
2024-04-14 16:42:02	PIDCOCK Thomas	0:00
2024-04-14 16:42:02	HIRSCHI Marc	0:00
2024-04-14 16:42:02	BENOOT Tiesj	0:00
2024-04-14 16:42:02	VANSEVENANT Mauri	0:00
2024-04-14 16:42:02	LAPEIRA Paul	0:08

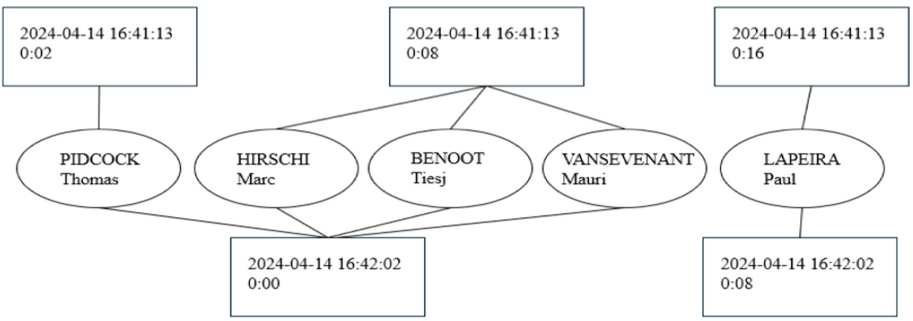


Fig. 2. Graph derived from Table 3. If any other timing is reported together, then Pidcock will still be placed together with the other attacking riders.

To capture this information in a machine-readable format which can easily be used in other tasks, we deploy Node2Vec (Grover & Leskovec, 2016). This algorithm creates random sequences of nodes by 'walking' along the edges after which the skip-gram Word2Vec algorithm is deployed to learn the contextual embedding of each node in the network. We create 20 sequences (i.e., 'random walks') per rider in the network, each of length 10, window size is set to 10, and the dimensionality of the resulting vectors (i.e., the hidden layer) is set to 32. This results into a 32-dimensional representation of each participating rider. To see whether each embedding is a good representation of in-race behavior, we check for each rider node what nodes have the most similar embeddings based on cosine similarity. Table 4 compares (1) the eventual winner Tom Pidcock, which placed a successful attack on the Keutenberg, (2) the pre-race favorite Mathieu van der Poel, (3) Paul Lapeira, who anticipated with an attack before the Keutenberg, and (4) Tosh Van der Sande, who was in the early breakaway. Each rider gets matched to the riders which displayed the most similar racing behavior. Tom Pidcock is matched to his three fellow attackers who rode away on the Keutenberg climb, but also to the random value >30. This is a consequence of the data quality issues stated above. Luckily, this is much easier to filter out based on the rider starting list. Similarly, Paul Lapeira is matched to four of the riders who also anticipated before the Keutenberg and ended up in the chasing group after the Keutenberg, while Tosh Van der Sande is matched to the three other riders who were also in the early break away, but also to Tomáš Kopecky, who was not actively involved in the breakaway, but interestingly joined the breakaway in many race leading up to this race, and finally race favorite Mathieu van der Poel is matched to four apparently random riders. This nicely reflects the fact that van der Poel had a rather anonymous race despite him ending in the second finishing group. These are all dynamics which would be difficult to capture based purely on eventual race results as currently is typically done.

Table 4. Most Similar Riders after Node2Vec

Base Rider	Most Similar Riders	
PIDCOCK Thomas	>30	HIRSCHI Marc
	VANSEVENANT Mauri	BENOOT Tiesj
VAN DER POEL Mathieu	SHAW James	OOMEN Sam
	DE POOTER Dries	CHZHAN Igor
LAPEIRA Paul	PACHER Quentin	MOLLEMA Bauke
	MADOUAS Valentin	MATTHEWS Michael
VAN DER SANDE Tosh	LEIJNSE Enzo	KYFFIN Zeb
	HAJEK Alexander	KOPECKÝ Tomáš

5 Conclusion

This study assesses whether it is feasible to use large language models to structure textual commentary on X into a useful and machine-readable structure. Results are disappointing with regard to the direct application of these models. GPT-4o outperforms GPT-3.5-turbo but the performance is too low to validate the currently large cost associated with using the model. Rather, we suggest a graph representation learning approach on the generated data which seems to capture complex in-race dynamics which are currently not captured in the used data sources, despite the errors made by the LLMs. This seems promising for future applications which want to acknowledge rider performances which are not observable in the end results. Future work should focus on how this information should be encapsulated in these applications and how information across races should ideally be captured.

References

Baron, E., Janssens, B., Bogaert, M.: Bike2Vec: Vector Embedding Representations of Road Cycling Riders and Races (2023). arXiv preprint arXiv:2305.10471

Bransen, L., Van Haaren, J.: Player chemistry: Striving for a perfectly balanced soccer team (2020). arXiv preprint arXiv:2003.01712

De Bock, J., Verstockt, S.: Video-based analysis and reporting of riding behavior in cyclocross segments. Sensors **21**(22), 7619 (2021)

De Bock, J., Verstockt, S.: Road cycling safety scoring based on geospatial analysis, computer vision and machine learning. Multimedia Tools Appl. **82**(6), 8359–8380 (2023)

Decorte, R., De Bock, J., Taelman, J., Slembrouck, M., Verstockt, S.: Fully automatic camera for personalized highlight generation in sporting events. Sensors **24**(3), 736 (2024)

Decroos, T., Bransen, L., Van Haaren, J., Davis, J.: Actions speak louder than goals: valuing player actions in soccer. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1851–1861 (2019)

Decroos, T., Davis, J.: Player vectors: Characterizing soccer players' playing style from match event streams. In: Joint European conference on machine learning and knowledge discovery in databases, pp. 569–584. Springer, Cham (2019)

de Leeuw, A.W., Heijboer, M., Verdonck, T., Knobbe, A., Latré, S.: Exploiting sensor data in professional road cycling: personalized data-driven approach for frequent fitness monitoring. Data Min. Knowl. Disc. **37**(3), 1125–1153 (2023)

de Leeuw, A.W., Kholkine, L.: The relationships between age and race performance in women's road cycling. Int. J. Perform. Anal. Sport 1–13 (2024)

De Spiegeleer, E.: Predicting cycling results using machine learning. MA thesis. Ghent University, Ghent (2019)

Grover, A., Leskovec, J.: node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 855–864 (2016)

Hilmkil, A., Ivarsson, O., Johansson, M., Kuylenstierna, D., van Erp, T.: Towards machine learning on data from professional cyclists. CoRR abs/1808.00198 (2018)

Janssens, B., Bogaert, M.: Imputation of non-participated race results. In: Machine Learning and Data Mining for Sports Analytics: 8th International Workshop, MLSA 2021, Virtual Event, September 13, 2021, Revised Selected Papers, pp. 155–166. Cham: Springer International Publishing (2022)

Janssens, B., Bogaert, M.: Performance measurement 2.0: towards a data-driven cyclist specialization evaluation. In: International Workshop on Machine Learning and Data Mining for Sports Analytics, pp. 179–190. Cham: Springer Nature Switzerland (2023)

Janssens, B., Bogaert, M., Maton, M.: Predicting the next Pogačar: a data analytical approach to detect young professional cycling talents. Ann. Oper. Res. **325**(1), 557–588 (2023)

Kataoka, Y., Gray, P.: Real-time power performance prediction in Tour de France. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds) Machine Learning and Data Mining for Sports Analytics. MLSA 2018. LNCS, vol 11330. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17274-9_10

Karetnikov, A.: Application of data-driven analytics on sport data from a professional bicycle racing team. Eindhoven University of Technology, The Netherlands (2019)

Kolaja, J., Ehlerova, J.K.: Effectivity of sports timing RFID system, field study. In: 2019 IEEE International Conference on RFID Technology and Applications (RFID-TA), pp. 220–223. IEEE (2019)

Kholkine, L., Latré, S., Verdonck, T., de Leeuw, A.W.: Age of peak performance in professional road cycling. J. Sports Sci. **41**(3), 298–306 (2023)

Kholkine, L., Schepper, T.D., Verdonck, T., Latré, S.: A machine learning approach for road cycling race performance prediction. In: International Workshop on Machine Learning and Data Mining for Sports Analytics, pp. 103–112. Springer, Cham (2020)

Kholkine, L., et al.: A learn-to-rank approach for predicting road cycling race outcomes. Front. Sports Act. Living **3** (2021)

Okuma, K., Taleghani, A., De Freitas, N., Little, J.J., Lowe, D.G.: A boosted particle filter: multitarget detection and tracking. In: Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11–14, 2004. Proceedings, Part I 8, pp. 28–39. Springer Berlin Heidelberg (2004)

Pappalardo, L., et al.: A public data set of spatio-temporal match events in soccer competitions. Sci. Data **6**(1), 236 (2019)

Robberechts, P.: Valuing the art of pressing. In: Proceedings of the StatsBomb Innovation in Football Conference, pp. 1–11. StatsBomb (2019)

Sagi, M., Saldanha, P., Shani, G., Moskovitch, R.: Pro-cycling team cyclist assignment for an upcoming race. PLoS ONE **19**(3), e0297270 (2024)

Steyaert, M., De Bock, J., Verstockt, S.: Sensor-based performance monitoring in track cycling. In: International Workshop on Machine Learning and Data Mining for Sports Analytics, pp. 167–177. Springer, Cham (2022)

Van Bulck, D., Vande Weghe, A., Goossens, D.: Result-based talent identification in road cycling: discovering the next Eddy Merckx. Ann. Oper. Res. 1–18 (2023)

Verstockt, S., Van Vooren, B., De Smul, S., De Bock, J.: Data-driven summarization of broadcasted cycling races by automatic team and rider recognition. In: icSPORTS 2020, the 8th International Conference on Sport Sciences Research and Technology Support, pp. 13–21. SCITEPRESS (2020)

Wei, J., et al.: Chain-of-thought prompting elicits reasoning in large language models. Adv. Neural Inf. Process. Syst. **35**, 24824–24837 (2022)

Basketball

GraphEIV: A Framework for Estimating the Expected Immediate Value in Basketball Using Graph Neural Networks

Bruno M. Sá-Freire[1,2], Gabriel R. G. Barbosa[1,2], João Lucas L. Gonçalves[1,2], Jake Schuster[2], and Hugo Rios-Neto[1,2(✉)]

[1] Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
{bruno290103,grgb,jllgoncalves23,hugoriosneto}@ufmg.br
[2] Gemini Sports Analytics, Raleigh, NC 27603, USA
jake@geminisports.ai

Abstract. Basketball is a fast-paced and strategic sport where each possession involves a series of complex decisions and actions. Analyzing and understanding these intricacies is essential for effective performance evaluation. This study introduces a novel framework, Expected Immediate Value, for evaluating basketball possessions from tracking data using Graph Neural Networks. We take inspiration from the Expected Possession Value framework for predicting points scored as an immediate consequence of the current game state. We develop four specific models to enhance the interpretability and composition of metrics: xFG (probability of a shot being successful), xNAT (next action - pass or shot), xR (probability of a player receiving the ball), and xTO (likelihood of a turnover). Our approach provides a comprehensive evaluation of player movements and decisions. This framework offers deeper insights into possession dynamics and supports strategy optimization in basketball.

Keywords: Basketball analytics · Deep learning · Graph neural networks

1 Introduction

Basketball is a dynamic sport where small actions during each possession can significantly impact larger outcomes. While traditional metrics such as box score statistics and play-by-play data fail to capture such contributions, optical tracking data creates opportunities to evaluate player contributions to game-action outcomes more comprehensively. This paper presents a framework by which tracking data can be leveraged to optimize team and player strategies in basketball [1,2,5,6,8].

B.M. Sá-Freire, G.R.G. Barbosa and J.L.L. Gonçalves—These authors contributed equally to this work.

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 29–40, 2025.
https://doi.org/10.1007/978-3-031-86692-0_3

Challenges arise when estimating the value of actions in basketball. The Expected Possession Value (EPV) in basketball represents the expected number of points that will be scored as an immediate consequence of the current game state. Our work takes inspiration from the foundational EPV framework introduced by [2], which predicts the outcomes of basketball possessions through a stochastic process model.

The original EPV framework [2] models the probabilities associated with each possible state transition by considering data of each individual player, their tendencies, and style of play. While this design decision allows for detailed and personalized models, it also introduces limitations. For instance, the need for individual player data and cooperative stats, such as "probability passing maps" for each position, can be restrictive due to the fluidity of positions and team tactics in modern basketball. Moreover, the original EPV framework did not utilize deep learning techniques, which have advanced significantly since 2016. This presents an opportunity to build upon modern techniques to develop new frameworks for evaluating actions in basketball.

In this study, we propose a modular approach, using Graph Neural Networks (GNNs) to create four separate models to estimate the necessary probabilities that compose the Expected Immediate Value (EIV) framework. EIV calculates the immediate expected points a team can score at any given moment, assuming that the player currently in possession of the ball will either shoot or pass to a teammate who will then shoot. The decision to create the EIV framework, rather than improving the EPV, was in large part due to the data limitations we faced. These are thoroughly discussed in Sect. 3 and Sect. 7.

The main contributions of this paper are:

- Introduction of GNNs applied to the estimation of EIV in basketball, providing a novel approach to capture the intricate spatial and temporal relationships of the game.
- Development of new basketball metrics by aggregating predictions from our framework composed of four distinct models, enabling new advanced statistics for performance analysis.
- Given better data annotation, we provide a clear path for others to build upon our work to create an EPV framework that leverages the base architecture and model components we utilized.

2 Related Work

In [2], the first implementation of EPV in basketball, the authors define the metric as the "expected number of points the offense will score on a particular possession conditional on that possession's evolution up to time t". Their approach was stochastic, with multi-resolution modeling, to predict possession states in macro level using Markov chains (to maintain stochastic consistency) and a full resolution model for sensitivity to fine-grained details of data.

The inclusion of various player-specific features in the estimator was justified by improved accuracy when predicting decisions and movement. However, when it comes to general models for evaluating player decision-making and efficiency, there may be more utility in developing a player-agnostic model capable of pinpointing outliers among a given population of players.

Deep Hoops [8] is an architecture that uses deep learning to assign value to actions in basketball. A Long Short-Term Memory (LSTM) network was fed with player embeddings to predict the probability of the final action of each possession, where the possibilities are field goal attempts, shooting fouls, non-shooting fouls, and turnovers. The value of each final action was fixed as the average points which it led to. Finally, game states were evaluated by weighing the average value of each final action by the probability of each outcome.

A limitation presented in [8] is the lack of assigning values to shots that are dependent on the current game state, rather than a pre-defined average value. By fixing shot values, the expected value from a shot depends only on the probability of a shot being attempted, rather than upon its value or difficulty.

Similar to [2], a framework for evaluating the expected value of possessions in soccer was presented [4] in 2021. The authors consider the probability of a pass to all possible receivers and the probability of possible ball-drives and shots. These probabilities are then used as weight-averages to sum outcome values and determine a possession value. The framework uses low-level spatial data as input, from which 13 layers of sparse field-wide matrix representations are created and fed into a Convolutional Neural Network (CNN) that also estimates a field-wide matrix representation.

This approach may lead to problems when converting low-dimensional tracking data into high-dimensional sparse data [9]. To address this issue, GNNs have been used to represent game states of soccer in predictive models [7,9,10]. By representing players as nodes and the relationships between them as edges, it becomes possible to feed neural networks with representative information.

3 Data

For this study, we utilized a tracking data dataset consisting of 576 games from the 2015–16 NBA season, provided by SportVU[1]. This dataset includes the positions of all ten players on the court, along with the ball's coordinates, at a resolution of 25 Hz. The high resolution of the tracking data enables detailed analysis of player movements and interactions during each possession.

However, the data lacks annotations of the actions that happen throughout the game, which makes it challenging to accurately interpret the context and meaning behind the data points. This lack of annotations means that the precise actions being taken by players (such as passes, shots, screens, or defensive maneuvers) are not directly recorded in the dataset. Consequently, identifying and classifying these actions becomes a complex task, often requiring

[1] https://github.com/linouk23/NBA-Player-Movements.

additional processing or manual annotation to generate labels that can be used for supervised learning tasks. This adds a significant layer of complexity to the data analysis process.

Without clear labels, it is challenging to train models that can recognize specific actions or events with high accuracy. Heuristic methods, clustering techniques, or semi-supervised learning approaches to infer the actions from the raw positional data could help mitigate the challenge. Still, these methods introduce noise and uncertainty into the analysis, as these inferred labels might not always correspond perfectly to the actual events on the court.

We found an external play-by-play dataset[2] with annotations of some types of actions. Nevertheless, they were not synchronized to the timestamp of when the player started performing the action, but rather to the identifier of the play. This required us to develop rule-based algorithms to accurately identify the timing of certain events. The algorithms were designed to detect shots, passes, and pass turnovers, achieving a 71% average detection rate for all events.

In Sect. 7, we more thoroughly discuss how the lack of annotations limited our modeling approach for this work, and how one with better data could extend our work to build an EPV framework.

4 EIV Framework

This section defines the proposed GraphEIV framework for basketball. The primary task of this framework is to estimate the immediate expected amount of points a team can score at a given moment, considering the player with the ball will shoot or pass to someone who will catch and shoot. To do so, we estimate the chances of each of the actions happening and their chances of success.

Let G_t be the game state at time t. Let S be the number of points scored as a direct consequence of an action A taken at time t. The EIV of a basketball possession at time t is defined as $EIV_t = \mathbb{E}[S|G_t]$. We can expand it to:

$$\mathbb{E}[S|G_t] = \sum_{a \in A} \mathbb{E}[S|A = a, G_t] \cdot \mathbb{P}(A = a|G_t) \tag{1}$$

In this work, due to limited data annotation, we define the set of possible actions to be taken by the player with the ball to only passes and shots, $A = \{\rho, \zeta\}$, where ρ represents a pass and ζ a shot. Also, let $P(l)$ be a function that returns the points value of a field goal from location l. We can further expand Eq. 1:

$$\begin{aligned}\mathbb{E}[S|G_t] = {} & P(l_c) \cdot \mathbb{P}(O = o_+|A = \zeta, G_t) \cdot \mathbb{P}(A = \zeta|G_t) \\ & + \sum_{r_i = r_1}^{r_4} P(l_{r_i}) \cdot \mathbb{P}(O = o_+|A = \rho_{r_i}, G_t) \cdot \mathbb{P}(A = \rho_{r_i}|G_t) \cdot \mathbb{P}(A = \rho|G_t)\end{aligned} \tag{2}$$

[2] https://github.com/airalcorn2/baller2vecplusplus/blob/master/events.zip.

where o_+ represents a positive outcome O, r_i the receiver i and ρ_{r_i} a pass to receiver r_i. The EIVs of the four potential receivers are summed.

4.1 Base Architecture

To derive the total EIV value of any given game state, the probabilities from Eq. 2 are calculated with four classifier models, which will be described in detail in the next section. The four models share the same base GNN architecture and use fully connected graphs as input, differing only in which players are included in the input graphs. Each model was trained separately. The base architecture can be visualized in Fig. 1.

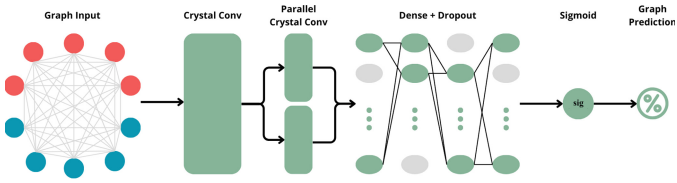


Fig. 1. Architecture of the base GNN model: one Crystal Convolutional Layer (CCL) followed by two Parallel CCLs. This is succeeded by four Dense Layers, with a 0.5 dropout applied after each layer, and a Sigmoid output.

4.2 Individual Models

The four individual models developed are: expected next action taken (xNAT), expected receiver (xR), expected turnover (xTO), and expected field goal (xFG). xR and xTO are combined to create the expected pass outcome (xPO) component, while xFG is also applied to obtain the expected field goal receiver (xFGR). The predictions generated by each model can be visualized through plots like the ones in Fig. 2.

The xNAT model is designed to predict whether the player will pass or shoot as their next action. In Eq. 2, this is represented by the term $\mathbb{P}(A|G_t)$, where $A = \{\rho, \zeta\}$. This is achieved through supervised binary graph classification, where each input graph represents the game state immediately before a shot or pass is made. The input of this model is a fully connected graph, where all 10 players are connected to each other. The target label is 1 if it is a shot and 0 if a pass.

The expected pass outcome (xPO) value is a composite of two separate models: xR (expected receiver), which predicts the probability of passing to each of the four potential receivers; and xTO (expected turnover), which predicts the probability of the pass resulting in a turnover. The outputs of these two models are subsequently merged and normalized to create the xPO prediction. In Eq. 2, the probability of a pass to receiver r_i is given by $\mathbb{P}(A = \rho_{r_i}|G_t)$.

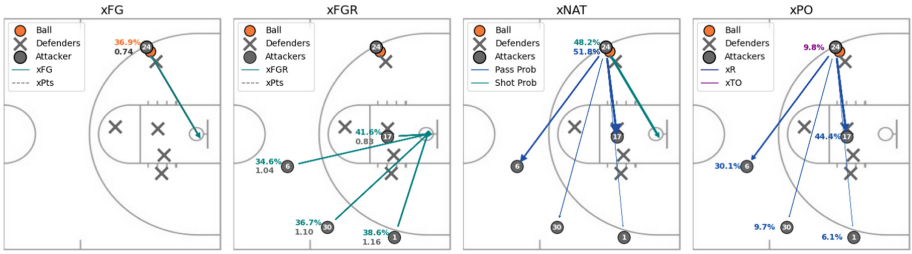


Fig. 2. Visualization of the four expected values that compose the EIV framework. Blue indicates passing probabilities; teal indicates shooting probabilities; purple represents turnover probability; and black text represents the expected points from a shot (xFG). (Color figure online)

The xR model represents the game state on a per-receiver basis, each frame yielding four different inputs. The input graph for xR consists of a fully connected graph with all five defenders, the player with the ball, and one potential receiver. The label is 1 if the potential receiver is the actual receiver and 0 otherwise. Also, The input graph for the xTO model contains all players on the court, with the label being 1 if the outcome is a turnover and 0 otherwise.

The xFG model predicts the probability of a shot being successful, regardless of whether it is a 2-point or 3-point shot. The objective of the model is to analyze the interaction between the shooter and defenders to determine the likelihood of the shot being successful. In Eq. 2, it is the term $\mathbb{P}(O = o_+|A = \zeta, G_t)$. This model represents the game state as a graph, with the shooter and five defenders as nodes. The graph label is 1 if the shot is made and 0 if it is missed.

Initially, we built a separate model, xFGR, to predict the probability that a receiver would score a catch-and-shoot shot, given they received a pass. However, training a model exclusively with plays which ended in a catch-and-shoot heavily biased the model towards high scoring probabilities. To obtain more realistic results, the xFG model was used by using each potential receiver, instead of the player in possession, as a node in the input graph. In Eq. 2, such probability of a receiver r_i scoring after a pass is given by $\mathbb{P}(O = o_+|A = \rho_{r_i}, G_t)$.

This modular approach allows for better calibration of task-specific models. It also allows for combining different models' predictions to create standalone metrics such as the ones explored in Sect. 6.2. Table 1 details the features of each model. Only frames where the team in possession is in the attacking half were included in the training and test sets.

Table 1. Model features. Each model uses a subset of the node features: the players' x and y coordinates, their velocities in x and y directions, distance to the basket, a boolean to indicate if the player is from the attacking team, and a boolean to indicate whether the player is in possession of the ball. The edge feature can be either the distance between players or a boolean that indicates whether the connected nodes are teammates.

Model	Node Features					Edge Features	
	Position (x,y)	Velocity (x,y)	Distance to Basket	Is Attacking (bool)	Is Player with Ball (bool)	Are Teammates	Distance Between Players
xNAT	✓	✓	✓	✓	✓	✓	
xR	✓	✓	✓	✓	✓	✓	
xTO	✓	✓		✓	✓	✓	
xFG	✓	✓	✓	✓			✓

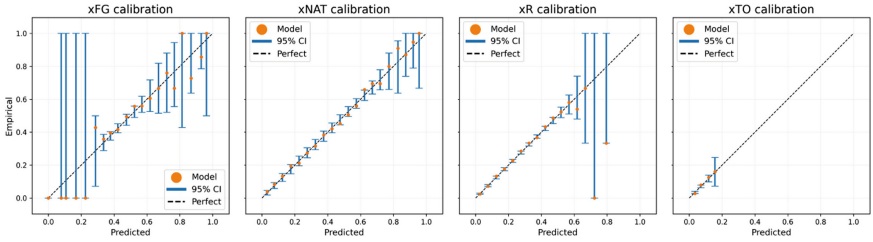


Fig. 3. Calibration plots for the four predictive models (xFG, xNAT, xR, xTO) used in the EIV framework. Each plot shows the relationship between predicted probabilities (x-axis) and empirical probabilities (y-axis) with 95% confidence intervals. The dashed line represents perfect calibration. Orange points indicate model predictions. (Color figure online)

5 Model Validation

Table 2. Model performance metrics

	xFG	xNAT	xR	xTO
BCE	0.58	0.51	0.60	0.67
ROC-AUC	0.75	0.78	0.73	0.65
Accuracy	0.68	0.78	0.67	0.58
F1-score	0.68	0.78	0.67	0.56
NBS	0.97	0.80	0.83	0.95

Since all models are binary classifiers whose training loss function was the Binary Cross Entropy (BCE), the same performance metrics can be used to compare

each model, as shown in Table 2. Although all models performed better than a naive baseline that predicts the class distribution [3] (Normalized Bier Score (NBS) < 1), xNAT and xR were the most performant out of the four models.

In Fig. 3, the calibration curve of all models is shown. The results indicate well-calibrated models, with predictions within the 95% confidence interval. Note that the confidence intervals are narrower for prediction ranges with more samples and are, therefore, more relevant. Also, note that prediction ranges without calibration points have no samples.

6 Applications

This section presents two possible applications of GraphEIV, in different granularities. First, by analyzing EIV on a single-play basis, then by aggregating metrics derived from the framework.[3]

6.1 Single-Play

EIV quantifies the shot and catch-and-shoot values of each moment in a possession, allowing us to analyze the probability and value of potential actions to determine if players are making valuable decisions on a per-play basis.

In Fig. 4, we observe a sequence of frames where player #24 possesses the ball and dribbles toward the basket. In the first two frames (A and B), #24 is far from the three-point line, resulting in a low probability and value for a shot. The most probable action here is passing to the closest teammate, #17. Players on the defensive court have zero probability of receiving a pass, as it would be a violation. Consequently, the EIV starts to drop due to the low-value potential actions. As #24 dribbles closer to the three-point line in frames C to E, the probability of taking a shot increases, with the value of a potential shot increasing more rapidly. By Frame E, the shot value peaks at 1.10 as #24 is closest to the three-point line, making a three-point shot highly valuable. During this movement, the increase in shot value for both #24 and his potential receivers leads to a rise in EIV. As #24 moves into the two-point zone in Frame F, the EIV experiences a slight drop. This occurs because the value of #24's potential shot decreases, while his teammates' shot values increase as they approach the three-point line.

This approach enables a detailed examination of how well players optimize their actions to maximize scoring opportunities. By analyzing each decision in the context of its probability and value, the EIV framework helps identify whether players are making high-value choices that enhance the team's overall performance. This real-time evaluation of decision-making allows identifying strengths and areas for improvement in both individual and team strategies.

[3] https://graph-epv-bc5247578fd6.herokuapp.com/.

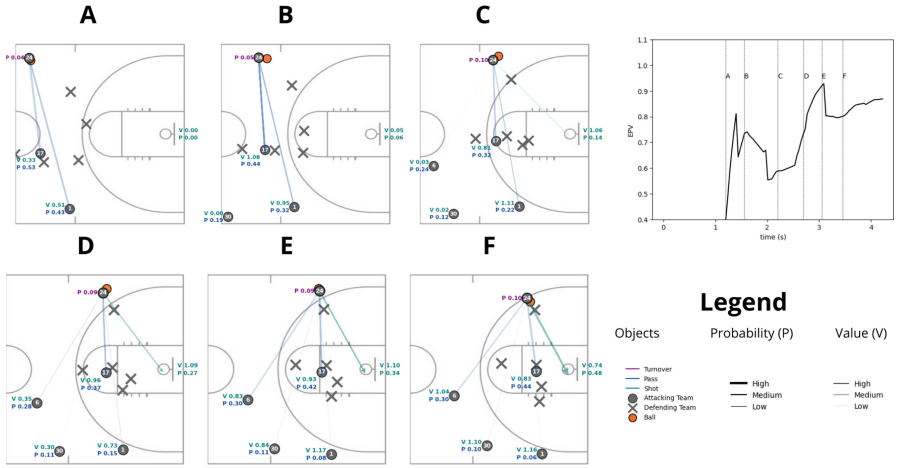


Fig. 4. Application of EIV in evaluating a single play. The sequence (A-F) shows some moments of a basketball possession with player positions, probabilities (P), and values (V). The EIV graph tracks the expected possession value over time, reflecting the impact of each action on the likelihood of scoring. The legend explains the symbols and color coding used for players and actions.

6.2 Aggregated Metrics

This section delves into the synthesis of individual model outputs to form comprehensive performance metrics, thereby providing a holistic view of a player's contribution over multiple possessions. These aggregated metrics are vital in bridging the gap between granular, moment-to-moment analysis and overarching strategic insights, enabling teams to make data-driven decisions that enhance overall performance and drive scouting strategies.

The first two metrics presented derive from the fact that the total EIV value is given by the sum of the expected value of a shot, $\mathbb{E}[S|A = \zeta, G_t]$, and the expected value of a pass, $\mathbb{E}[S|A = \rho, G_t]$. Therefore, it is possible to define Shot Value Above Expected as the difference between the points scored from a shot and the expected value of a shot, $P(l) \cdot \mathbf{1}_{\{O=o_+\}} - EIV_\zeta$. Another metric derivable from the EIV definition is Pass Value Wasted, given by the $\mathbb{E}[S|A = \rho, G_t]$ when the player in possession opts for a shot.

Comparing both metrics in Fig. 5, it is possible to evaluate which players add more value with their shots than they waste by choosing to shoot. Stephen Curry, for example, is positioned well above the identity line, indicating that his average shot adds more value than expected, greatly compensating for the potential value of passes he forgoes.

Lastly, we present Dribbling Touch Value (DTV), given by $EIV_{t+\Delta t} - EIV_t$, the difference in EIV from the moment a player makes a shot/pass decision, at time $t + \Delta t$, and the EIV from the moment he first touched the ball, at time t. By aggregating DTV throughout the season, it is possible to determine which

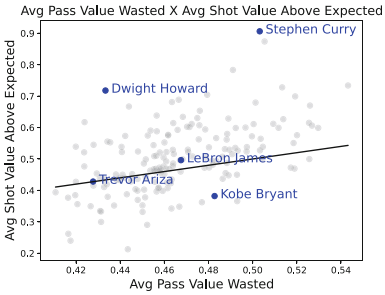


Fig. 5. Scatter plot of players' season average stats, depicting pass value wasted versus shot value above expected.

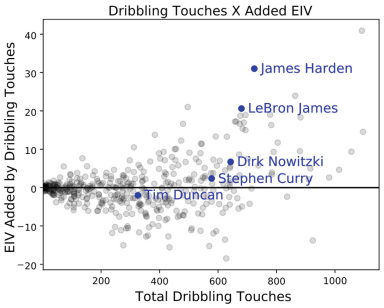


Fig. 6. Scatter plot of players' season aggregate stats, depicting total dribbling touches versus EIV added.

players bring the most value when dribbling and driving. Figure 6 shows players' contribution through dribbling touches. James Harden and LeBron James, known for their ball-driving skills, are highlighted on the upper right.

7 Challenges and Discussion

As aforementioned, the dataset presented several limitations. Primarily, the annotated event data did not align with the game and shot clocks of the tracking data, which required the development of rule-based algorithms. If the tracking data was synchronized, the need for rule-based algorithms would be eliminated, leading to more accurate models due to training with ground-truth target variables.

Additionally, the data lacked crucial information on dribbling and driving. Therefore, the decision was made to take only the immediate expected value of the current game state into account. As a consequence, the xNAT model considers only shot and pass probability. However, dribbles were identified via rule-based algorithms and evaluated using variation in EIV (Fig. 6). Reliable data on fouls was also missing, which prevented other types of turnovers from being considered as possible outcomes of a game state, such as offensive fouls or traveling turnovers, or to consider other possible possession outcomes, such as shooting fouls.

A point of discussion might be not including the shot technique as a feature of the xFG model. From a basketball standpoint, it seems obvious that the type of shot a player performs greatly influences their probability of success. For instance, a floater and a fadeaway from the same location inside the paint should have different average conversion rates. However, from a modeling perspective, it makes sense to represent the game state as the last moment before the decision of performing the action was made, not before the action was taken. This means that a game state whose outcome is a shot does not come right before the ball

leaves the shooter's hand but the moment they start bringing the ball upwards from dribbling or receiving a pass. At that time, it is not possible to know precisely which technique a player will use, or even if they will shoot at all. So, the game state as a whole is evaluated before the action itself is taken.

The estimation of $\mathbb{E}[S|A = \rho, G_t]$, that is, the expected value of passing the ball in the current game state, is another interesting point of discussion and modeling decision we had to make. While estimating the long-term possession value of passing to each of the players could make more sense from a theoretical standpoint, in practice, we would have various scenarios where a pass does not have any impact on the possession outcome, meaning the current game state passed on as input would not be very representative of what contributed to scoring, its associated target. This, in turn, would cause training examples to be very noisy, that is, being uninformative regarding the expected value it would generate. Given all the noise already created by the required rule-based algorithms and not having data on dribbling and driving, we decided on an approach where the current game state would be sufficiently representative of the possible outcome.

We acknowledge that having a model that estimates the probability of a catch-and-shoot, that is, that a player receiving a pass would shoot without dribbling the ball, makes sense within the EIV framework. However, considering the probability of a shot attempt would also require considering the expected value of other possible actions, such as a pass. To do so, the EIV framework would need to be applied for every possible sequential outcome. This approach is both computationally unfeasible and would require too many assumptions regarding the game state in future actions. Therefore, it was determined to attribute value only to the immediate shot following the pass.

If the data limitations above are not an issue, the EIV framework can be easily extended to an EPV framework by considering more elaborate sets of possible actions and outcomes. In this case, Eq. 1 would include the additional possible actions, and Eq. 2 be extended with the additional possession outcomes.

Finally, even though EPV generates more metrics, EIV yields useful metrics by itself, such as points above expected, value of dribbling touches, and decision evaluation. The representation of frames as graphs in a GNN architecture is also an asset that brings new possibilities to action evaluation in sports. This approach is used here in the EIV framework but could also serve as the foundation for new EPV frameworks in the future.

8 Conclusion and Future Work

In this study, we introduced GraphEIV, a novel framework employing GNNs to estimate the EIV in basketball. Our modular approach not only enhances the interpretability of our metrics but also allows for a more granular analysis of player decisions and game dynamics, which can also be aggregated to evaluate teams and players over longer periods. The validation results indicate that our models are statistically relevant and well-calibrated, demonstrating their robustness and reliability.

In future works, with better-annotated data, the inclusion of fouls and rebounds would allow for the framework to be decomposed in a way that better represents possible outcomes of basketball possessions, thus, leading to EPV estimation. Also, utilizing a sequence of past frames up to the current one to represent the current game state, rather than only the current frame, could improve model performance given the appropriate GNN architecture.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alarcón Román, A.: Development of value metrics for specific basketball contexts: evaluating player contribution by means of regression. B.S. thesis, Universitat Politècnica de Catalunya (2022)
2. Cervone, D., D'Amour, A., Bornn, L., Goldsberry, K.: A multiresolution stochastic process model for predicting basketball possession outcomes. J. Am. Stat. Assoc. **111**(514), 585–599 (2016)
3. Decroos, T., Davis, J.: Interpretable prediction of goals in soccer. In: Proceedings of the AAAI-20 Workshop on Artificial Intelligence in Team Sports (2019)
4. Fernández, J., Bornn, L., Cervone, D.: A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. Mach. Learn. **110**(6), 1389–1427 (2021). https://doi.org/10.1007/s10994-021-05989-6
5. Lucey, P., Bialkowski, A., Carr, P., Yue, Y., Matthews, I.: How to get an open shot: analyzing team movement in basketball using tracking data. In: Proceedings of the 8th Annual MIT SLOAN Sports Analytics Conference (2014)
6. Papalexakis, E., Pelechrinis, K.: thoops: A multi-aspect analytical framework for spatio-temporal basketball data. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 2223–2232 (2018)
7. Sahasrabudhe, A., Bekkers, J.: A graph neural network deep-dive into successful counterattacks. In: Proceedings of the 17th Annual MIT SLOAN Sports Analytics Conference (2023)
8. Sicilia, A., Pelechrinis, K., Goldsberry, K.: DeepHoops: evaluating micro-actions in basketball using deep feature representations of spatio-temporal data. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2096–2104 (2019)
9. Stöckl, M., Seidl, T., Marley, D., Power, P.: Making offensive play predictable-using a graph convolutional network to understand defensive performance in soccer. In: Proceedings of the 15th MIT Sloan Sports Analytics Conference (2021)
10. Wang, Z., et al.: TacticAI: an AI assistant for football tactics. Nat. Commun. **15**(1), 1906 (2024)

Mathematical Models for Off-Ball Scoring Prediction in Basketball

Rikako Kono[1] and Keisuke Fujii[2,3,4](✉)

[1] Research School of Physics, Australian National University, Canberra, Australia
rikako.kono@anu.edu.au
[2] Graduate School of Informatics, Nagoya University, Nagoya, Japan
fujii@i.nagoya-u.ac.jp
[3] Center for Advanced Intelligence Project, RIKEN, Osaka, Japan
[4] PRESTO, Japan Science and Technology Agency, Saitama, Japan

Abstract. In professional basketball, the accurate prediction of scoring opportunities based on strategic decision-making is crucial for spatial and player evaluations. However, traditional models often face challenges in accounting for the complexities of off-ball movements, which are essential for comprehensive performance evaluations. In this study, we propose two mathematical models to predict off-ball scoring opportunities in basketball, considering pass-to-score and dribble-to-score sequences: the Ball Movement for Off-ball Scoring (BMOS) and the Ball Intercept and Movement for Off-ball Scoring (BIMOS) models. The BMOS model adapts principles from the Off-Ball Scoring Opportunities (OBSO) model, originally designed for soccer, to basketball, whereas the BIMOS model also incorporates the likelihood of interception during ball movements. We evaluated these models using player tracking data from 630 NBA games in the 2015–2016 regular season, demonstrating that the BIMOS model outperforms the BMOS model in terms of team scoring prediction accuracy, while also highlighting its potential for further development. Overall, the BIMOS model provides valuable insights for tactical analysis and player evaluation in basketball.

Keywords: player evaluation · tracking data · invasion sports

1 Introduction

In the highly competitive world of professional basketball, the precise and comprehensive evaluation of player performance is essential for identifying player potential and capability, as well as enabling teams to make well-informed decisions, such as scouting new talent or determining salary allocations. Traditionally, box score statistics, such as points, assists, and rebounds, have been powerful tools for these evaluations [6,10,16,18]. However, the emergence of tracking technology around 2010 has significantly shifted the way player performance is analyzed. By combining this technology with box score statistics, teams and analysts can glean insights that were previously inaccessible.

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 41–54, 2025.
https://doi.org/10.1007/978-3-031-86692-0_4

Most studies using spatiotemporal tracking data have focused on investigating players' movements around the ball, known as "on-ball movements". For instance in [24], the authors predicted near-future events such as shooting and passing in basketball based on current game state. Furthermore, basketball shot selections were analyzed using Non-negative matrix factorization, a dimensionality reduction technique [15], which was extended to consider additional dimensions [17]. Another study introduced a framework to identify on-ball screens [14], extended to a deep learning model [2], its defense classification [13], and player prediction for obtaining rebounds [9]. Recently, shot trajectories were analyzed to understand the impact of tight defensive contesting on shot accuracy [5].

While on-ball movements certainly deserve attention, movements not directly linked to ball possession, known as "off-ball movements", are also pivotal in understanding the complete dynamics of the basketball game as is the case for many other team sports [8,11,12,21]. Despite the importance of off-ball movements, analyzing the spatiotemporal movements of off-ball players is highly complex due to the multiple potential actions that may interplay with the overall team strategy. Several research attempts have previously been undertaken to handle this complexity. For instance in [23], the authors introduced an extensive model for evaluating off-ball movements in shooting and passing situations, which also accounts for player profiles. Another researches analyzed team and player performances using an expected possession value (EPV) model [3,4].

Although these models offer deep insights, they are generally data-driven approaches that require large quantities of training data and sometimes provide uninterpretable results. This poses a challenge when evaluating continuous changes in player performance from season to season or assessing new players, underscoring the need for models grounded in theory rather than empirical data. Here we propose theory-based basketball models, inspired by the probabilistic mathematical soccer model called OBSO (off-ball scoring opportunities) [19]. They are constructed using physical parameters of players and the ball, providing intuitive and interpretable results. One of our models demonstrates significant improvements in team scoring prediction accuracy, potentially providing valuable insights for tactical analysis and player evaluation in professional basketball. However, this initial model development work also highlights some limitations in our model application. In the following sections, we describe our theoretical framework in Sect. 2 and experimental methods in Sect. 3. Next, we present the results and application in Section 4 and conclude this paper in Sect. 5.

2 Theoretical Framework

The OBSO model [19] estimates the probability that a pass to a target position on the field will lead to a score at a subsequent moment of the possession in soccer, with its applicability limited to the "pass-to-score" sequence. However, in compact court invasion sports like basketball, the "dribble-to-score" sequence also cannot be overlooked. To address this, we extended the OBSO model to encompass dribbling situations and adapted it to basketball, naming it the Ball

Movement for Off-ball Scoring (BMOS) model.Additionally, we further developed the Ball Intercept and Movement for Off-ball Scoring (BIMOS) model, which considers the possibility of the ball interception during transitions, as factor not expressly covered by the OBSO and BMOS models.

Prior to explaining our models in following sections, Fig. 1 provides an overview of the OBSO, BMOS and BIMOS model structures. These models are basically constructed as the multiplication of several step-by-step sub-models. In the OBSO and BMOS models, there are three sub-models: the current spatial occupancy by the attacking team, represented as PPCF (Potential Pitch Control Field); the probability of delivering the ball to a particular location, represented as Transition Model; and the scoring probability, represented as the Score Model. There are two slightly different sub-models in the BIMOS model: the current ball occupancy by the attacking team, represented as PBCF (Potential Ball Control Field); and the Score Model, which is common to all three models. Briefly, PBCF can be interpreted as a combination of PPCF and the Transition Model, with the added consideration of the ball interception. Moreover, in the BMOS and BIMOS models, we considered dribbling situations for PPCF and PBCF, as separate components alongside the pass model.We detail and construct the BMOS and BIMOS models, and demonstrate through comparison that the BIMOS, which explicitly incorporates the interception possibility, represents a more advanced approach.

2.1 BMOS Model

We segment ball movement sequences from its current position to an arbitrary target position $\mathbf{r} \in R \times R$ into three phases according to the OBSO theory [19]: Transition ($\mathrm{T_r}$), Control ($\mathrm{C_r}$), and Score ($\mathrm{S_r}$) such that:

Transition: The ball is delivered to point $\mathbf{r}$,
Control: The attacking team keeps possession at point $\mathbf{r}$,
Score: The attacking team scores from point $\mathbf{r}$.

Considering the instantaneous game state D which encompasses elements such as players' positions and velocities, the BMOS model calculates the joint probability of these phases and breaks it into a series of conditional probabilities as follows:

$$\mathrm{P_{BMOS}} = \sum_{\mathbf{r}\in \mathrm{R}\times \mathrm{R}} \mathrm{P}(\mathrm{S}_{\mathbf{r}} \cap \mathrm{C}_{\mathbf{r}} \cap \mathrm{T}_{\mathbf{r}} | \mathrm{D}) = \sum_{\mathbf{r}} \mathrm{P}(\mathrm{S}_{\mathbf{r}} | \mathrm{C}_{\mathbf{r}}, \mathrm{T}_{\mathbf{r}}, \mathrm{D}) \mathrm{P}(\mathrm{C}_{\mathbf{r}} | \mathrm{T}_{\mathbf{r}}, \mathrm{D}) \mathrm{P}(\mathrm{T}_{\mathbf{r}} | \mathrm{D}). \quad (1)$$

Here, the conditional probabilities, $\mathrm{P}(\mathrm{S_r}|\mathrm{C_r}, \mathrm{T_r}, \mathrm{D})$, $\mathrm{P}(\mathrm{C_r}|\mathrm{T_r}, \mathrm{D})$, and $\mathrm{P}(\mathrm{T_r}|\mathrm{D})$ are defined as the Score Model, PPCF, and Transition Model, respectively.

PPCF. The original PPCF [19] estimates the probability of the attacking team maintaining ball possession at point $\mathbf{r}$ following a successful ball transition, under the condition D. In other words, it can be considered as the space occupancy by

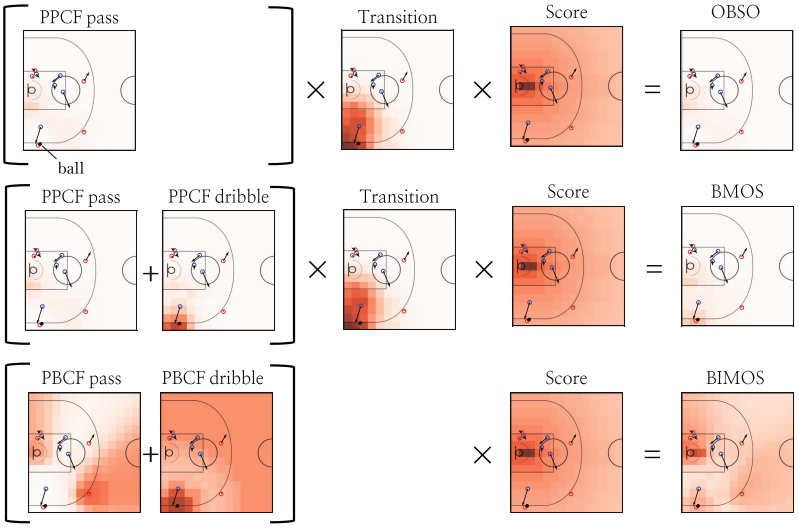


Fig. 1. Overview of OBSO, BMOS, and BIMOS structures. Attacking and defending players are represented in red and blue, respectively. An attacker at the bottom left currently possesses the ball. The OBSO and BMOS models consist of three components: spatial occupancy (PPCF), ball delivery probability (Transition Model), and scoring probability (Score Model). The BIMOS model consists of two components: ball occupancy (PBCF) and scoring probability, common to all three models. In the BMOS and BIMOS models, dribbling situations for PPCF and PBCF are incorporated as separate components alongside the pass model.

each team. The PPCF value for player i is numerically calculated as:

$$\int_{t=\mathrm{T}_{calc}}^{\infty} \mathrm{dPPCF}_i(\mathbf{r}, t|\mathrm{D}) = \int_{\mathrm{T}_{calc}}^{\infty} \left(1 - \sum_k \mathrm{PPCF}_k(\mathbf{r}, t - dt|\mathrm{D})\right) f_i(\mathbf{r}, t|\mathrm{D})\lambda_i dt. \quad (2)$$

Here, t is the time elapsed since the ball starts its movement and T_{calc} is the expected time for the ball to reach point $\mathbf{r}$. The PPCF value is presumed to be zero when $t < \mathrm{T}_{calc}$. The expression $(1-\sum_k \mathrm{PPCF}_k)$ denotes the probability that no player has gained possession before t. The function $f_i(\mathbf{r}, t|\mathrm{D})$ estimates the probability that player i will reach point $\mathbf{r}$ before t, while the parameter λ_i reflects the player's ball-handling ability.

In the BMOS model, we separate the PPCF into pass and dribble components by adjusting ball speeds and the attackers involved in the calculations. Detailed explanations and our modifications from the original theory to T_{calc}, $f_i(\mathbf{r}|\mathrm{D})$, and λ_i are presented in the following paragraphs: Time of Flight, Time to Intercept, and Control Rate, respectively.

Time of Flight. The original PPCF determines T_{calc} by considering both parabolic trajectories and aerodynamic drag, selecting the most suitable trajectory that best matches the anticipated arrival time of the nearest attacking

player [19]. Our adaptation to basketball simplifies this approach by calculating T_{calc} under the assumption that the ball moves at a constant speed v_{ball} as a function of travel distance, based on the type of ball movement (pass or dribble).
Time to Intercept. To establish $f_i(\mathbf{r}, t|D)$, the expected time τ_{exp} taken for player i to reach point $\mathbf{r}$ is calculated. In the original PPCF, this calculation simplifies a game scenario by assuming that players move directly to point $\mathbf{r}$ at uniform acceleration $\mathbf{a}$ with initial velocity $\mathbf{v}_{ini}$, aligned in the same direction with a realistic upper velocity limit v_{max} [19]. We modified this calculation by incorporating player reaction time, which is the duration taken for a player's cognitive processing to change direction, using the approach implemented in [1] (a different PPCF implementation). We set the ball possessor's reaction time to zero and separately accounted for the reaction times of attackers and defenders. Although τ_{exp} calculations do not include more intricate factors that may influence the true time taken τ_{true}, such as tracking data inaccuracies and tactical decision making, the distribution of residuals $\tau_{true} - \tau_{exp}$ can offer insights into these factors because its normalized cumulative distribution as a function of $t - \tau_{exp}$ yields the probability that a player expected to arrive in τ_{exp} will actually arrive before t. Therefore, its fitted function can be expressed as $f_i(\mathbf{r}, t|D)$. In this study, we set t to the expected ball arrival time T_{calc} to adapt to more realistic basketball situations such that players often catch the ball before it hits the floor. As described in Sect. 3.3, we used a cumulative skewed Cauchy distribution, whereas the original PPCF uses a cumulative logistic function.

Control Rate. The parameter $\lambda_i = \lambda$ is defined as the inverse of the time required for player i to gain control of the ball upon reaching point $\mathbf{r}$ [19]. Therefore, higher λ_i values indicate a player's quicker ball control abilities. Attackers seek complete control to either lead a subsequent shot or maintain possession, whereas defenders are satisfied with diverting the ball's direction. To differentiate between these objectives, a scaling parameter κ (≥ 1) is employed for defenders as $\lambda_i = \kappa\lambda$ [19].

Score Model. This model estimates the scoring probability from point $\mathbf{r}$ following a successful transition and control at point $\mathbf{r}$ under the condition D. Since the successful transition $T_{\mathbf{r}}$ and control $C_{\mathbf{r}}$ conditions are already satisfied during the shooting phase, the scoring probability S_d can be defined as a function of the distance to the goal $\mathbf{r}_{goal}$. For simplicity, the condition D is not directly factored into the model [19]. The original Score model employs the adjustment parameter β to reflect the players' decision-making influence (i.e., $P(S_{\mathbf{r}}, \beta) = [S_d(|\mathbf{r} - \mathbf{r}_{goal}|)]^{\beta}$) as a tendency to avoid shots under heavy defensive pressure. Although this is a plausible assumption, our scoring probability is obtained by fitting the model to a game dataset (see Sect. 3.2), and further adjustments based on the BMOS value may lead to overfitting and reduce interpretability. Therefore, we have excluded this factor.

Transition Model. This model estimates the probability of the ball transitioning to point $\mathbf{r}$ under the condition D [19]. Our model is constructed using the displacement (Δx, Δy) distribution of passes and dribbles, specifically those leading directly to either shots or turnovers. Although the OBSO attempts to incorporate the decision-making factor (i.e., $\left[\sum_{k \in att} \mathrm{PPCF}_k\right]^{\alpha}$), as the ball possessor is reluctant to deliver the ball to a lower-controlled position, we also omitted this factor for the same reason as in the Score Model; to avoid overfitting and maintain the Transition Model's independence for enhanced interpretability. Since the resulting distribution does not follow the two-dimensional normal distribution adopted in the OBSO, we used a histogram-based distribution with Gaussian filtering (see Sect. 3.2). Note that the Transition phase in the OBSO and BMOS disregards the possibility of ball interception, and the transition probability should be 100%. Thus, this model can be interpreted as the location where the ball possessor is inclined to deliver the ball.

2.2 BIMOS Model

One important limitation in the OBSO and BMOS models is that they do not explicitly account for the possibility that the ball is intercepted en route to target position $\mathbf{r}$. Although this aspect was partially addressed in a preceding study [20], it is assumed that all relevant players head to point $\mathbf{r}$ rather than selecting alternate paths that may offer a higher chance for ball interception. However, this assumption does not always hold in basketball and potentially oversimplifies game scenarios. To address it, we have rewritten Eq. 1 as:

$$\mathrm{P}_{\mathrm{BIMOS}} = \sum_{\mathbf{r}} \mathrm{P}(\mathrm{S}_{\mathbf{r}} | \mathrm{C}_{\mathbf{r}}, \mathrm{T}_{\mathbf{r}}, \mathrm{D}) \mathrm{P}(\mathrm{C}_{\mathbf{r}} \cap \mathrm{T}_{\mathbf{r}} | D). \tag{3}$$

Here, $\mathrm{P}(\mathrm{S}_{\mathbf{r}} | \mathrm{C}_{\mathbf{r}}, \mathrm{T}_{\mathbf{r}}, \mathrm{D})$ represents the same Score Model mentioned in Sect. 2.1, and $\mathrm{P}(\mathrm{C}_{\mathbf{r}} \cap \mathrm{T}_{\mathbf{r}} | D)$ is defined as the PBCF, which estimates the probability of the attacking team delivering the ball and maintaining possession at point $\mathbf{r}$, under the condition D. In other words, it can be considered as the ball occupancy during the delivering from its current position to target position. Unlike the PPCF, the PBCF value is not zero where $t < \mathrm{T}_{calc}$ due to the consideration of ball interception. To account for the possibility that neither the attacker nor the defender can gain possession of the ball, we set an upper integration limit to T_{calc} rather than integrating to infinity. The function f_i^b is then redefined as the probability that player i reaches point $\mathbf{r}_{mid}(t) = \mathbf{r}_{ball} + \mathbf{v}_{ball} \times t$ before t, where $\mathbf{r}_{ball}$ is the starting location of the ball, and $\mathbf{r} - \mathbf{r}_{ball}$ and $\mathbf{v}_{ball}$ are aligned in the same direction. Accordingly, Eq. 2 is then modified as

$$\int_{t=0}^{\mathrm{T}_{calc}} \mathrm{dPBCF}_i(t | \mathbf{r}, \mathrm{D}) = \int_{0}^{\mathrm{T}_{calc}} \left(1 - \sum_k \mathrm{PBCF}_k(t - dt | \mathbf{r}, \mathrm{D})\right) f_i^b(t | \mathbf{r}, \mathrm{D}) \lambda_i dt. \tag{4}$$

Note that we solely modified the target position to be calculated over time; the methods for separating pass and dribble components for Time of Flight, Time to Intercept, and Control Rate described in Sect. 2.1 remain unchanged.

3 Methods

The codes used for this work are available on GitHub at https://github.com/Run-summer/bimos_bmos_basketball.

3.1 Dataset

We used a tracking dataset from STATS SportVU, which covers 630 NBA games from the 2015–2016 regular season. To streamline the analysis, all game sequences were standardized to appear as if they occurred on the left half of the court. Scenes with incomplete data, such as those that were obviously truncated, were excluded from the analysis. We allocated 580 games for training our models and set aside the remaining 50 games for testing. However, due to computational constraints, only 4,000 shot scenes and 600 turnover scenes were used for the likelihood estimation process (see Sect. 3.3).

3.2 Transition and Score Model Construction

The Transition Model for the BMOS model was constructed based on the distribution of displacements in pass and dribble transitions leading directly to a shot or turnover. Whereas the OBSO model assumes a two-dimensional normal distribution, our resulting distribution does not follow this assumption, exhibiting a skew towards the negative x-axis, as we only use transitions directly related to shots or turnovers. Therefore, we used a histogram-based distribution with Gaussian filtering ($\sigma = 0.4$) to reduce noise and smooth the data (Fig. 2 left).

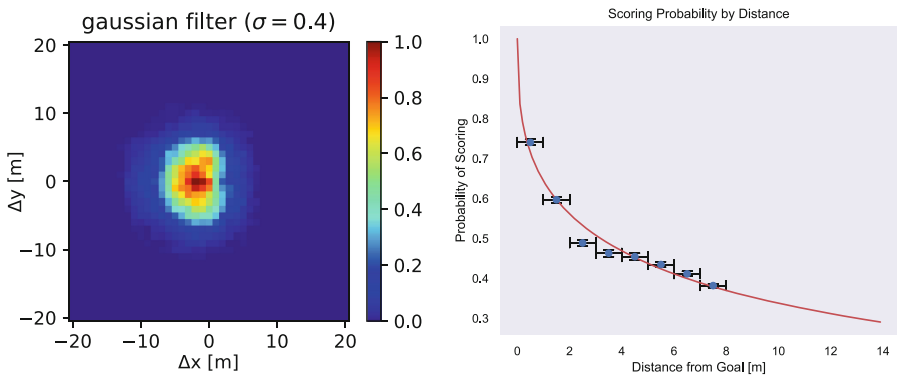


Fig. 2. Transition Model and Score Model. Transition Model distribution after applying a Gaussian filter for noise reduction (left). This model captures the tendency of the ball possessor to choose shorter-range ball movement. Scoring probabilities at various distances (right). The exponential function, depicted in red solid line, models the decreasing trend in scoring probability with increasing distance.

To construct the Score Model, which is common to the BMOS and BIMOS models, scoring probability was calculated based on distance from the goal, considering only distances with more than 2,000 attempts. As expected, the scoring probability decreases with increasing distance, as illustrated in Fig, 2 right. To model this trend, an exponential function was employed to fit the data; in contrast, the original study [19] used Gaussian kernel smoothing for a continuous distribution. The fitted function was then extended to cover the entire left half of the basketball court, providing a comprehensive scoring probability map (see Fig. 1 for the resulting Score Model).

3.3 PPCF/PBCF Model Parameter Estimation

Seven parameters need to be determined for the PPCF/PBCF, including the ball velocity v_{ball} for Time of Flight, player maximum velocity v_{max}, constant acceleration $|\mathbf{a}|$, reaction times of the attacker and defender for Time to Intercept, and Control Rate parameters λ and κ. Whereas v_{ball} and v_{max} can be directly estimated from the data, the other parameters were fitted using the maximum likelihood estimation technique, as they were challenging to approximate from the data alone. The initial values for these parameters were set according to prior studies [7,19,20] or our assumptions.

The player maximum velocity v_{max} was set to 5.0[m/s]. The ball velocity v_{ball} varies between cases of pass and dribble, as well as with distance traveled. Reasonably, ball speed increases with travel distances, and the speed of passes changes more drastically, as shown in Fig. 3 left. To combine the pass and dribble PPCF/PBCF, the rate at which the ball possessor decides to pass or dribble was calculated as a function of travel distance, as illustrated in Fig. 3 right. As expected, dribbling was more likely for shorter distances, while passing was the preferred choice for longer distances.

Initial values of $|\mathbf{a}|$ and κ were set to 7.0 [m/s^2] and 1.72, respectively, in accordance with the original values [19]. We assumed the time required for a player to gain control of the ball to be much shorter in basketball than soccer, and set κ to 30. The attacker and defender reaction times were both set to 0.32 [s] following the outcome of previous study [7]. The resulting $\tau_{true} - \tau_{exp}$ distribution using these parameters suggests a skewed Cauchy distribution with a heavy tail, where $\tau_{true} - \tau_{exp} > 0$ fits better than a logistic function, as shown in Fig. 4. This can be attributed to players not actually moving at constant acceleration, taking longer than expected. Consequently, we decided to utilize the accumulated skewed Cauchy distribution as $f_i(\mathbf{r}, t|D)$ and $f_i^b(\mathbf{r}, t|D)$. Using the models constructed above and Eq. 1 or 3, the total BMOS/BIMOS likelihoods are formulated as:

$$L(\theta) = \prod_{i=1}^{N} \left[P(\theta|D)^{y_i} \times (1 - P(\theta|D))^{(1-y_i)} \right].$$

Here, N represents the number of scenes in the training dataset, y_i indicates whether the i^{th} scene results in a score $y_i = 1$ or not $y_i = 0$, and $P(\theta|x_i)$

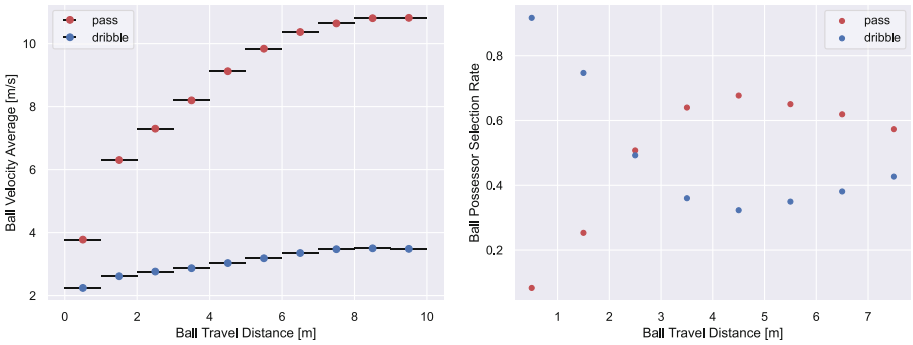


Fig. 3. Ball Velocity and Ball Handler Selection Rate. The left figure illustrates how the ball speed varies based on the type of ball movement (pass or dribble) and the distance traveled. For distance exceeding 10 m, ball speed is assumed constant. The right figure shows the rate at which the ball possessor chooses to pass or dribble based on the travel distance.

denotes the BMOS or BIMOS probability at the position of a shot attempt, given condition D and parameters θ. We used values calculated when players passed the ball in pass-to-score sequences, as well as those when players began dribbling in dribble-to-score sequences. Consequently, the maximum likelihood estimation for both the BMOS and BIMOS was performed using the Nelder-Mead method[1], with boundary conditions set as: $1 \leq |\mathbf{a}| \leq 8$, $1 \leq \lambda$, $1 \leq \kappa$, and $0 \leq$ reaction time ≤ 1. For the BIMOS, the fitted values were $|\mathbf{a}| = 7.76$, $\lambda = 36.6$, $\kappa = 1.02$, attacker reaction time$= 0.157$, and defender reaction time$= 0.495$. The divergence in reaction time between attackers and defenders can be attributed to differences in tactical understanding, as attackers share tactics in advance, allowing for quicker responses. In the case of the BMOS, the parameter values did not converge under the same maximum likelihood conditions applied to the BIMOS. Therefore, we adopted the estimated parameters of the BIMOS. This is justified because the fitted values obtained through the BIMOS model are both reasonable and realistic, and can therefore be reliably applied to the BMOS model.

4 Results

Figure 5 shows the BMOS and BIMOS distribution examples, where the ball possessor at the bottom left passed to an attacker at the bottom middle, leading to a successful 3-point shot. The BIMOS captures a reasonable trend, with areas around the current ball possessor, the shooter, and below the goal exhibiting higher scoring probabilities. In contrast, the BMOS failed to describe a higher probability around the shooter.

[1] The Nelder-Mead algorithm was selected because our model equations are nondifferentiable.

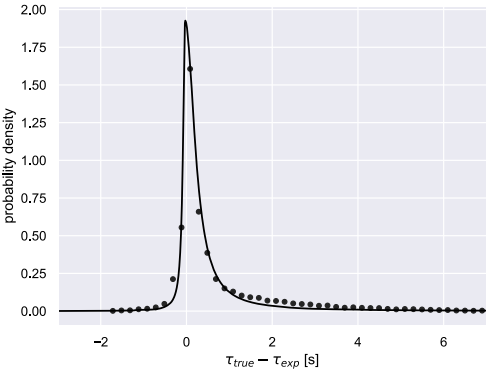


Fig. 4. $\tau_{true} - \tau_{exp}$ **Distribution**. The figure shows the $\tau_{true} - \tau_{exp}$ distribution, highlighting the better fit of an asymmetric Lorentzian function compared to a logistic function. The heavy tail on the positive side suggests that players might not always move at a constant acceleration, resulting in longer than expected times to reach the target position.

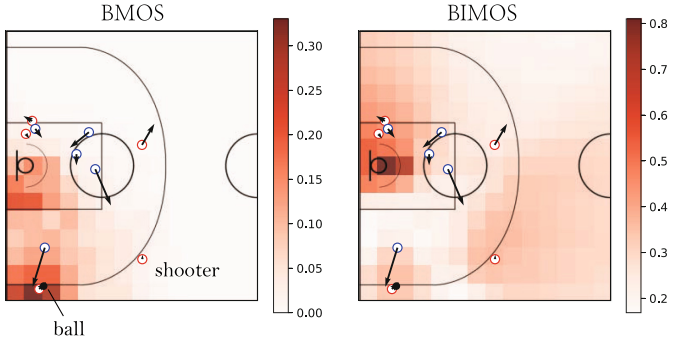


Fig. 5. BMOS and BIMOS distribution. The ball possessor at the bottom left passed to a shooter at the bottom middle, resulting in a successful 3-point shot. The BIMOS effectively captures the higher scoring probability in proximity of the shooter.

Next, we quantitatively examined correlations between each team's actual and expected score per game predicted by the BMOS and BIMOS. Figure 6 shows the results obtained using 50 test games. Additionally, we divided them into pass-to-score (middle) and dribble-to-score (right) sequences. It should be noted that, since some scenes were not recorded in the original dataset and we also filtered out incomplete data, the actual score does not represent the total score per game, but only the accumulated score from the scenes we used.

The coefficients of determination R^2 for these points with respect to the Expected Score = Actual Score line, depicted in red, are -4.22, -1.75, and -2.05 for the BMOS All, Pass, and Dribble sequences, respectively. In the case of the BIMOS, the respective values are 0.548, 0.576, and 0.547. These results indicate

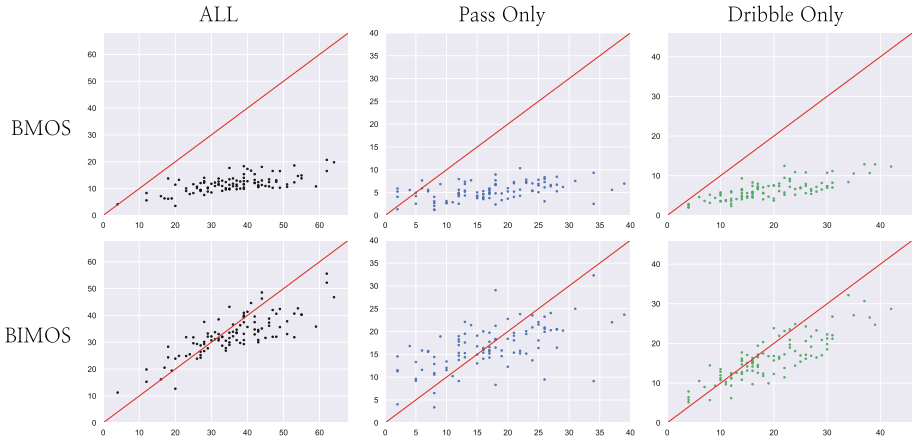


Fig. 6. Actual vs. Expected score. Actual total score per game for each team (x-axis) and expected score (y-axis) predicted by the BMOS (top) and BIMOS (bottom) are shown. The middle and right graphs consider only pass-to-score and dribble-to-score sequences, respectively. The red line in each graph represents the identity line.

that the BIMOS provides more accurate scoring predictions, especially in pass-to-score sequences. The lower effectiveness in dribble-to-score sequences suggests the need to develop a model specifically designed for dribbling, as both the pass and dribble models were constructed using nearly identical structures for convenience.

Applications. Potential applications of the BIMOS model range from tactical reviews to player evaluations. For instance, we show two examples on player evaluations: assessing whether a player maintains position in an area with a high scoring possibility, and determining whether a player creates scoring opportunities by drawing defenders away and moving into other areas. Figure 7 illustrates the first use case, showing the correlation between expected and actual scores per scene for players who played more than 600 scenes across 630 games. For each scene, we selected the individual's maximum expected score and averaged these values.

Contrary to our expectation, the overall trend between actual and expected scores is roughly inversely proportional. In terms of expected scores, the top five players played in Center (C) or Power Forward (PF) positions, which are generally responsible for areas around the basket. Conversely, in terms of actual scores, the top five players played in positions other than Center. This result suggests a limitation of our model: positions around the basket are likely overestimated because our Score Model does not take into account defensive pressure. This limitation should be addressed in future work. Additionally, the second use case—creating scoring opportunities—can be analyzed by computing counterfactuals as described in [22]. Such counterfactual analyses are also expected to

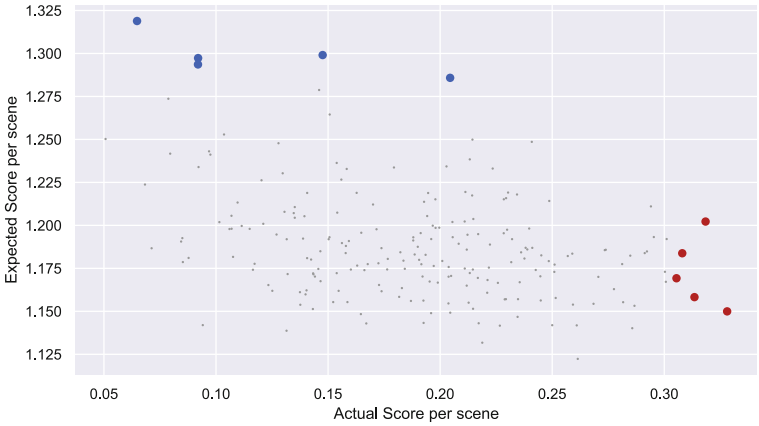


Fig. 7. Actual vs. Expected score for individuals. The x-axis represents the actual score per scene for individuals who participated in more than 600 scenes across 630 games, while the y-axis shows the expected scores. The top five players in terms of expected and actual scores are denoted in blue and red, respectively.

enhance tactical reviews by identifying more effective potential tactics. These analyses will also be part of our future work.

5 Conclusions

In this study, we extended the OBSO soccer model [19] to the basketball context, designing two mathematical basketball spatial models, the BMOS and BIMOS models, to predict off-ball scoring opportunities considering both pass and dribble situations. The BIMOS model, which also accounts for the possibility of ball interception, significantly improves scoring prediction accuracy compared to the BMOS model, suggesting effective functionality with a limited dataset.

Future work will address current limitations of the BIMOS model, including its lower scoring prediction accuracy in dribble-to-score situations discussed in Sect. 4 and its tendency to overestimate scoring opportunities around the basket described in the results of applications. Additionally, the applications of our model can be further examined in combination with box score statistics and/or computing counterfactuals, as discussed in the results of applications.

Acknowledgments. This study was financially supported by JST PRESTO JPMJPR20CA.

References

1. Friends-of-tracking-data-fotd/laurieontracking. https://github.com/Friends-of-Tracking-Data-FoTD/LaurieOnTracking. Accessed 28 May 2024
2. Ai, S., Na, J., De Silva, V., Caine, M.: A novel methodology for automating spatio-temporal data classification in basketball using active learning. In: 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML), pp. 39–45. IEEE (2021)
3. Cervone, D., D'Amour, A., Bornn, L., Goldsberry, K.: Pointwise: predicting points and valuing decisions in real time with NBA optical tracking data. In: Proceedings of the MIT Sloan Sports Analytics Conference (2014)
4. Cervone, D., D'Amour, A., Bornn, L., Goldsberry, K.: A multiresolution stochastic process model for predicting basketball possession outcomes. J. Am. Stat. Assoc. **111**(514), 585–599 (2016). https://doi.org/10.1080/01621459.2016.1141685
5. Daly-Grafstein, D., Bornn, L.: Using in-game shot trajectories to better understand defensive impact in the NBA. J. Sports Anal. **6**, 1–8 (2020). https://doi.org/10.3233/JSA-200400
6. Deshpande, S.K., Wyner, A.: Estimating an NBA player's impact on his team's chances of winning. J. Quant. Anal. Sports **12**, 51–72 (2016)
7. Fujii, K., Shinya, M., Yamashita, D., Oda, S., Kouzaki, M.: Superior reaction to changing directions for skilled basketball defenders but not linked with specialised anticipation. Eur. J. Sport Sci. **14**(3), 209–216 (2014). https://doi.org/10.1080/17461391.2013.780098
8. Hojo, M., Fujii, K., Inaba, Y., Motoyasu, Y., Kawahara, Y.: Automatically recognizing strategic cooperative behaviors in various situations of a team sport. PLoS ONE **13**(12), e0209247 (2018)
9. Hojo, M., Fujii, K., Kawahara, Y.: Analysis of factors predicting who obtains a ball in basketball rebounding situations. Int. J. Perform. Anal. Sport **19**(2), 1–14 (2019)
10. Jaime Sampaio, E.J.D., Leite, N.M.: Effects of season period, team quality, and playing time on basketball players' game-related statistics. Eur. J. Sport Sci. **10**(2), 141–149 (2010). https://doi.org/10.1080/17461390903311935
11. Lamas, L., Santana, F., Heiner, M., Ugrinowitsch, C., Fellingham, G.: Modeling the offensive-defensive interaction and resulting outcomes in basketball. PLoS ONE **10** (2015)
12. Lucey, P., Bialkowski, A., Carr, P., Yue, Y., Matthews, I.: How to get an open shot: analyzing team movement in basketball using tracking data. In: Proceedings of the MIT Sloan Sports Analytics Conference (2014)
13. McIntyre, A., Brooks, J., Guttag, J., Wiens, J.: Recognizing and analyzing ball screen defense in the NBA. In: Proceedings of the MIT Sloan Sports Analytics Conference, pp. 11–12 (2016)
14. McQueen, A., Wiens, J., Guttag, J.: Automatically recognizing on-ball screens. In: Proceedings of the MIT Sloan Sports Analytics Conference (2014)
15. Miller, A., Bornn, L., Adams, R., Goldsberry, K.: Factorized point process intensities: a spatial analysis of professional basketball. In: International Conference on Machine Learning, pp. 235–243 (2014)
16. Page, G.L., Fellingham, G.W., Reese, C.S.: Using box-scores to determine a position's contribution to winning basketball games. J. Quant. Anal. Sports **3** (2007)
17. Papalexakis, E., Pelechrinis, K.: tHoops: A multi-aspect analytical framework for spatio-temporal basketball data. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 2223–2232 (2018)

18. Sampaio, J., Janeira, M.: Statistical analyses of basketball team performance: understanding teams' wins and losses according to a different index of ball possessions. Int. J. Perform. Anal. Sport **3**(1), 40–49 (2003). https://doi.org/10.1080/24748668.2003.11868273
19. Spearman, W.: Beyond expected goals. In: Proceedings of the 12th MIT Sloan Sports Analytics Conference, pp. 1–17 (2018)
20. Spearman, W., Basye, A., Dick, G., Hotovy, R., Pop, P.: Physics-based modeling of pass probabilities in soccer. In: Proceeding of the 11th MIT Sloan Sports Analytics Conference (2017)
21. Teranishi, M., Tsutsui, K., Takeda, K., Fujii, K.: Evaluation of creating scoring opportunities for teammates in soccer via trajectory prediction. In: International Workshop on Machine Learning and Data Mining for Sports Analytics. Springer (2022)
22. Umemoto, R., Fujii, K.: Evaluation of team defense positioning by computing counterfactuals using StatsBomb 360 data. In: StatsBomb Conference (2023)
23. Wu, Y., et al.: OBTracker: visual analytics of off-ball movements in basketball. IEEE Trans. Vis. Comput. Graph. **29**(1), 929–939 (2023). https://doi.org/10.1109/TVCG.2022.3209373
24. Yue, Y., Lucey, P., Carr, P., Bialkowski, A., Matthews, I.: Learning fine-grained spatial models for dynamic sports play prediction. In: 2014 IEEE International Conference on Data Mining, pp. 670–679 (2014). https://doi.org/10.1109/ICDM.2014.106

Soccer

An Analysis of the Influence of Game Context on Team Playing Style

Deniz Can Oruç, Lorenzo Cascioli, Luca Stradiotti, Maaike Van Roy, Pieter Robberechts, and Jesse Davis(✉)

Department of Computer Science; Leuven.AI, KU Leuven, Leuven, Belgium
{DenizCan.Oruc,Lorenzo.Cascioli,Luca.Stradiotti,
maaike.vanroy,Pieter.Robberechts,Jesse.Davis}@kuleuven.be

Abstract. Soccer teams frequently adapt their playing style based on the game context. For example, once a team takes the lead it may sit a little deeper and take fewer offensive actions. This paper focuses on automatically inferring how a team adjusts its in-possession playing style to the game context. Our approach hinges on analyzing passing dynamics as a proxy for a team's in-possession playing style and involves two components. First, we enhance the SoccerMap architecture by incorporating game state variables and team identities, resulting in a more context-aware model that can predict passing tendencies for individual passes. Second, we introduce an encoder-decoder architecture to isolate the impact of game state on passing decisions, providing insights into how teams adapt their strategies in various contexts. By studying passing patterns and decisions made by teams in different game contexts, we aim to provide clubs with concrete insights that can inform tactical adjustments and optimize gameplay in various scenarios.

1 Introduction

Preparing for a soccer match requires an extensive analysis of the opposing team's playing style, tendencies, and tactics. Such analysis enables coaches and analysts to formulate well-informed game plans and tactical adjustments. Previous studies have therefore explored techniques to measure various offensive and defensive factors that delineate a team's playing style such as ball possession [12], passing tendencies [24], pressing behavior [3], team formations [1], attacking patterns [6], etc.

However, contextual variables such as match venue, opposition strength, and game state were previously found to be associated with changes in teams' dynamics within and between matches [2,10,11,13,22]. Typically, methods for tactical analysis do not take these contextual variables into account and aggregate the analysis over an entire game or season [4,16]. Alternatively, they perform separate analyses for small epochs of a game [24], which reduces sample sizes. Our research aims to provide a broader perspective and more comprehensive understanding of the influence of game context on a team's tactical behaviors,

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 57–68, 2025.
https://doi.org/10.1007/978-3-031-86692-0_5

decision-making, and overall effectiveness by integrating contextual variables in the analysis.

Specifically, in this study, we will analyze the effect of the game state (i.e., time remaining, scoreline, and venue) on the passing decisions that players make in soccer. Here, passing decisions serve as a proxy towards team style and, more pertinently, how that style changes depending upon the game state. Indeed, previous research has shown that losing or drawing teams prefer a direct playing style (characterized by instances of play where teams attempt to move the ball quickly toward the opposition's goal through the use of riskier direct or long passes), whereas winning teams prefer shorter passing sequences [7,14,17].

Nonetheless, these analyses offer only a limited breakdown of pass types and the specific phase of play in which they are executed. We hypothesize that the situational context manifests itself differently across different phases of play (e.g., build-up vs. final third) and in terms of the types of pass options available to a player (e.g., opportunities for through balls vs. crosses). Therefore, a more detailed framework is needed for analyzing context-dependent styles of play.

Our approach is two-fold. First, we extend the SoccerMap architecture [8] by introducing limited game state information into the model training process. This game state information is reshaped into surfaces, which are used as additional channels for the prediction architecture. The spatial dissimilarity of the predicted passing distribution across different game states provides an interpretable visualization of how game state affects passing tendencies in specific situations. Second, we propose an encoder-decoder architecture for isolating the effect of game state factors on a player's pass decision. The resulting latent space represents the learned relationship between game state factors and pass decisions, providing insights into how teams are expected to react to specific game state changes.

2 Modeling the Passing Dynamics of a Soccer Team

In this section, we explain how a team's passing dynamics and the game context can be modeled independently based on tactical event stream data. Then, in the next sections, we discuss two complementary approaches for integrating the analysis of passing dynamics and game context.

2.1 Data Description

This work uses StatsBomb 360 tactical event stream data.[1] In this type of data, each match is described by the regular human-annotated event stream data, augmented with partial spatial context extracted from broadcast video. Namely, each event comes with a snapshot providing the location and relationship to the ball carrier (i.e., teammate or opponent) of all the players appearing in the broadcast video at the time of the event. We convert the original event stream to its SPADL representation [5].

[1] See https://statsbomb.com/what-we-do/soccer-data/360-2/.

We focus exclusively on passes to analyze teams' playing styles. Additionally, we filter out passes not performed by foot, passes from dead-ball situations (i.e., corners, free-kicks, goal-kicks, kick-offs, and throw-ins), and passes for which the origin or destination location falls outside the visible area of the 360 snapshot.

Applying the aforementioned filters, we construct a training dataset comprising 162,991 passes from the 2021/22 Premier League, and 164,423 passes from the 2022/23 Premier League. These two datasets only contain games involving a team that finished in the top-10 of the league in the considered season. Additionally, a dataset of 19,544 passes extracted from EURO2020 data has been set aside for the purpose of evaluating the models and developing the use cases.

2.2 Modeling Passing Dynamics

Setting aside individual player decision-making tendencies and team-specific tactics, we posit that passing dynamics in soccer are primarily shaped by two key factors: (1) the phase of play, which encompasses distinctions such as build-up play versus chance creation; and (2) the available passing options, which involves assessing opportunities for various types of passes, such as through balls or crosses. The former is closely tied to the location of the ball and the latter is intricately linked to the positioning of other players on the field. Consequently, pass selection models [8,20,23] effectively capture the passing dynamics of a generic team in a generic match context.

We use the SoccerMap model [8], which is a modern deep convolutional neural network architecture specifically designed to analyze spatiotemporal data in soccer. As input, we use a 9-channel spatiotemporal representation constructed from the 360 snapshot of the relevant pass and its two preceding actions [21], consisting of: (1–2) a sparse matrix with locations of the attacking team and the defending team, (3–4) a dense matrix with distance to the ball and the goal for every location, (5–6) a dense matrix with the sine and cosine of the angle between every location and the ball, (7) a dense matrix with the angle in radians between every location and the goal, (8–9) a sparse matrix with the x and y-components of the ball's velocity vector, derived from timestamps and ball location in the event data during the two preceding actions.

These channels are processed by convolutional layers that create a feature hierarchy with scales of 1x, 1/2x, and 1/4x. These different scaled spatial-aware features are upsampled nonlinearly and merged using fusion layers. Finally, a sigmoid activation layer model estimates the probability of passing to each position on the pitch. The model is trained using the log-loss. The cell corresponding to the destination of the pass is selected by multiplying the probability surface with a binary masked matrix.

2.3 Game State Factors

We analyze three factors that affect team dynamics during a soccer match. In the following, we refer to these three factors as game state factors.

1. **Venue:** Teams often adapt their playing style based on whether they are playing at home or away. When playing in the comfort of their home stadium, teams frequently exhibit a more assertive and attacking style of play. Conversely, when facing an away fixture, teams often adopt a more cautious and strategic approach [18]. We encode the venue as a binary variable, indicating whether the team in possession plays at home or away.
2. **Scoreline:** The scoreline of the match plays a role in shaping teams' risk-aversive behavior. When a team is losing a game, they may take more risks to create scoring opportunities. On the other hand, when a team is winning, they often become more focused on defending their lead. Teams tend to prioritize maintaining their advantage, often adopting a "parking the bus" strategy to ensure defensive solidity [9]. We categorize the scoreline into six distinct groups: ≤ -3, -2, -1, draw, $+1$, $+2$, ≥ 3.
3. **Time remaining:** The remaining time on the clock also affects teams' dynamics. When time is running out, teams may adopt a more aggressive and direct style of play. They might employ tactics such as quick counters, long balls, and all-out attacks in an attempt to score goals rapidly. Conversely, if a team holds a comfortable lead late in the match, they may prioritize ball retention and time-wasting strategies to run down the clock and protect their advantage [15]. We represent the remaining time as a discrete variable by dividing the game into quarters and subdividing the last quarter into 5-minute intervals.

We use a one-hot encoding to represent these three game state factors as a 16-dimensional vector (1 bit indicating whether the in-possession team plays at home or away, 7 bits for the scoreline, and 8 bits for the time remaining).

3 Predicting How the Game State Will Impact a Specific Pass Decision

A team strongly adapts its passing tendencies depending upon the game state and the team's corresponding objectives. The original SoccerMap architecture lacks the capacity to capture the contextual playing styles of a team. It consistently generates the same probability surface for a given spatiotemporal snapshot of the game, irrespective of the team in ball possession and the game state. Therefore, in this section, our goal is to incorporate the team identity and aforementioned game state factors into the SoccerMap input to obtain probability surfaces that are aware of the actual game situation.

Adapting SoccerMap to include game state and team identity requires deciding (i) how to input the game state into the model and (ii) how to adapt the structure of the network to integrate the spatiotemporal information and the game state.

In our approach, we transform the original SoccerMap input by concatenating the game state and team identity to the input channels (Fig. 1). We use the aforementioned one-hot encoding of the game state and concatenate a 23-dimensional one-hot encoding of the team that executes the pass. The game state

and team information are subsequently compressed to a 21-dimensional vector using a fully connected layer. However, this dimensionality does not align with the sizes of the input channels representing the pitch. Therefore, we treat each variable in this vector as a 68×104 channel and replicate its value across each cell. Subsequently, we concatenate this matrix with the original SoccerMap input matrix, creating a $30 \times 68 \times 104$ matrix.

With this approach, we can embed categorical information into the spatiotemporal input independently of the spatial location. The intuition behind this idea is that the information provided by the game state remains consistent across all pitch coordinates. We then apply 1×1 convolutional filters to blend the original SoccerMap information with the game state and reduce the dimensions of the input channels to match the original $9 \times 68 \times 104$ SoccerMap input size.

Our method of integrating additional information into the SoccerMap differs from the approach employed by Pleuler [19]. That work used a parallel network to encode the individual decision-making tendencies of players into latent space embeddings. These embeddings are subsequently reshaped into surfaces and used as additional channels within the prediction architecture. Reshaping those embeddings, concatenating them with spatial input, and using convolutional layers to process them has two problems. Firstly, reshaping the embeddings and concatenating them with a matrix with spatial information creates a spatial behavior for this embedding which does not exist naturally. Secondly, processing it with convolutional networks with a kernel size larger than one infuses a locality for player information which is again not correct. In contrast, our approach integrates information into all locations of the pitch, which seems more appropriate for the game state because it intuitively has a global influence. We concatenate the same embedding with each location and use 1×1 convolutions to integrate it with the spatial data without considering neighboring locations.

4 Describing How the Game State Alters a Team's Passing Strategy

While our modified SoccerMap architecture allows us to investigate individual situations, it would also be informative to take a more holistic view to summarize how teams typically respond to changes in the game state. To address this problem, we employ a neural network-based encoder-decoder architecture. Given the SoccerMap probability surface for a specific situation, the actual passing choice (the so-called mask), and the team that performed the pass, the model will predict the corresponding game state. The encoder will learn a small, latent representation that captures how teams react to different game states.

A key design decision in an encoder-decoder approach is choosing the appropriate prediction task. We consider two possibilities. First, we train a model that jointly predicts all three components of the game state. This has the advantage of capturing synergies among these components. However, it makes the prediction task more difficult and training the network more complicated, which we will

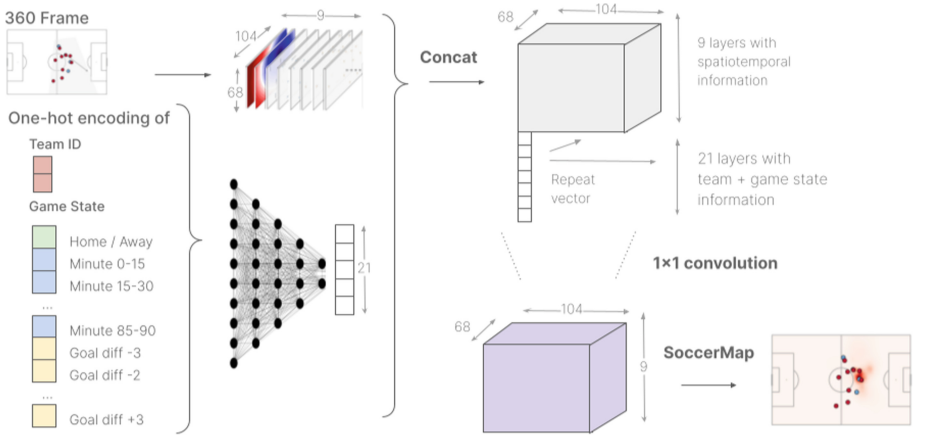


Fig. 1. Illustration of our approach for adding the game state and team identity to the SoccerMap input.

discuss below. Second, we train separate models to predict each component of the game state. This simplifies learning at the expense of missing any interactions that arise among the time remaining, goal difference and home/away. Regardless of the approach, we use the same encoder model. However, each task has a slightly different decoder.

The encoder compresses the probability surfaces, mask, and team information into a 16-dimensional latent space. It has two parts: a Convolutional Encoder and a Combined Encoder. The first applies 3 convolutional layers to the 2D inputs (the probability surface for the pass-selection and its corresponding mask) to obtain a flattened representation. This representation is then concatenated with the team information and fed into the second encoder with the purpose of integrating the pass and team information. This Combined Encoder uses three fully connected layers to compress the data into the 16-dimensional latent space.

First, the decoder expands the latent vector to a 256-dimensional vector by using three fully connected layers. Second, we employ a predictor network to predict the game state. The predictor network depends on the considered setting. When jointly predicting all elements of the game state, the model consists of three parallel networks, one for each component of the game state, each containing 3 fully connected layers. In the separate setting, the predictor network simply has 3 fully connected layers.

To train the network, we use the following loss functions: binary cross entropy for home/away, categorical cross entropy for the time period, and categorical cross entropy for the goal difference. There are two key challenges here: Certain game states, namely those involving a goal difference of zero, occur way more often. We compare two ways to approach this. First, for each game state, we sample a maximum of 2000 passes (if there are <2000 passes then all are selected). Second, we inversely weigh each game state by its frequency. This has the effect

of making it seem that each game state occurs the same number of times in the data. The loss functions are on different scales, which is only relevant when training the joint model. If not accounted for, the model may simply focus on improving the predictions for the component of the loss that has the largest value. We address this by normalizing the losses and assigning an equal weight to each component.

5 Results

5.1 How Game State Impacts a Specific Pass Decision

We used data from both the 2021/22 and 2022/2023 Premier League seasons to train the model. To facilitate model selection, we designated a random 20% subset of the passes as a validation set. We train both the baseline SoccerMap model and the modified SoccerMap model using a learning rate of 1e-6 and a batch size of 32. We use early-stopping with patience set to 10 epochs and a delta of 1e-5. The maximum number of epochs is set to 500.

Table 1 compares the performance and model properties of the vanilla SoccerMap architecture and our modified version. The modified version performs better both in log-loss and Brier score, which indicates that the addition of team identity and game state context allows the model to make more accurate predictions. Also, the number of parameters increases only marginally due to our 1×1 convolutional approach which does not need as many parameters as fully connected layers.

Table 1. Results for the benchmark models and the modified team- and game state aware SoccerMap model.

Model	Log-loss	Brier Score	# params
SoccerMap	5.989	0.988	271,000
SoccerMap + Game state + Team ID	5.961	0.987	275,000

Our modified SoccerMap allows us to visualize how a team will respond to different game states. To illustrate this, we explore the influence of the venue on Arsenal's passing dynamics between minutes 15–30 in a tied game. Figure 2 shows the difference in pass probability surfaces between home and away. When playing away, Arsenal has a stronger preference for a patient build-up through the center. However, at their home ground, they are more likely to directly play the ball to the flanks.

5.2 How Game State Alters a Team's Passing Strategy

To train the encoder-decoder model, we first train SoccerMap to obtain probability surfaces that will be used as one of the inputs for the encoder. For training

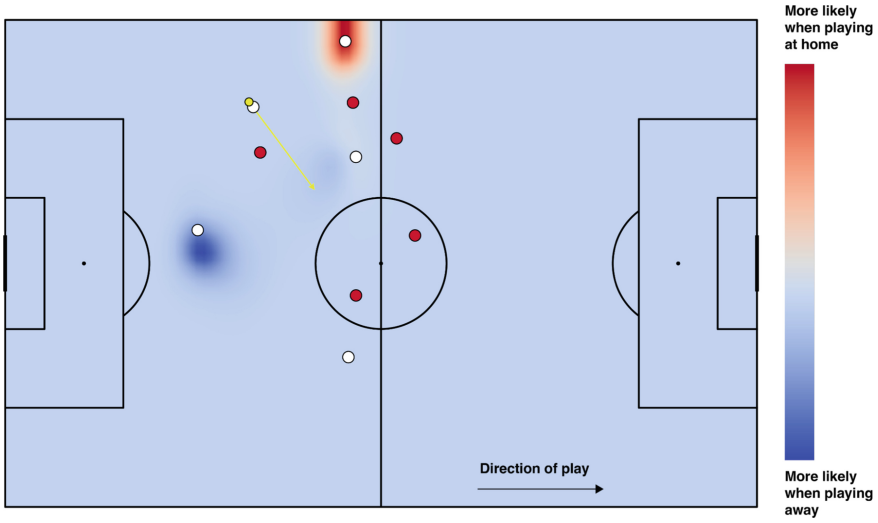


Fig. 2. When playing away, Arsenal has a greater preference for a patient build-up through the center. At their home ground, they are more likely to directly seek the flanks. The other game state factors are fixed to the score being tied in the second quarter of the game.

both the vanilla SoccerMap and the encoder-decoder model, we used the PyTorch Lightning framework with the adaptive moment estimation (ADAM) algorithm for optimizing the weights. For the vanilla SoccerMap [8], we use a learning rate of 5e-6 and a batch size of 32. For the encoder-decoder model, we use a learning rate of 1e-6 and a batch size of 32.

For training the SoccerMap model and the encoder-decoder model itself, we used 2021/22 Premier League data. We use data from the 2022/23 Premier League for evaluating the encoder-decoder model and computing the quantitative results. Again, we split 20

Table 2 shows the accuracy for predicting each of the three game-state components for both the joint model and the separate models. We also report the accuracy of random guessing as a baseline. Overall, we can see that training a separate encoder-decoder to predict each component of the game state has better predictive performance, except for on home vs. away. The joint model is even worse than random guessing for two of the tasks. This happens because the joint prediction task is too complicated and therefore the model is not able to infer the actual game state from the teams' passes.

Moreover, we evaluate our design choices to cope with the imbalance in the dataset. To verify the effectiveness of our approach, we compare it with two other models: the first samples the passes without weighting them according to their frequency, while the latter weighs each sample but the training is done using the whole dataset. Overall, the model with only sampling has the highest average accuracy. This probably happens because rare game states are so infrequent that

Table 2. Accuracy for the three game state components for the model with separate prediction tasks and the joint model.

Model	Home/away	Goal diff.	Remaining time
Separate model	0.5056	0.1291	0.1824
. . . + sampling	0.5108	0.3066	0.1421
. . . + weighting	0.5085	0.0916	0.1421
. . . + sampling + weighting	0.5056	0.1291	0.1824
Joint model	0.5122	0.1027	0.1081
Random guess	0.5000	0.1430	0.1250

they have a high weight, meaning that a small improvement in predicting their value can have a large effect on improving the loss function. Thus the model focuses on these at the expense of the more frequent ones.

The latent space of the encoder-decoder model effectively captures the relationship between a team's passing decision (model input) and the game state (model output) for each pass in the dataset. Hence, we can gain insight into how teams adapt their passing strategy to specific game state changes by analyzing how a team's vectors in this latent space change according to the various game state factors. As an example, we analyze the extent to which team strategies are affected by the match venue (Fig. 3). Here, we compute the mean latent vectors for each team when playing at home and when playing away. Aston Villa is the least dependent on home/away conditions. Remarkably, Liverpool and Bournemouth are again among the teams for which the passing tendencies are most affected.

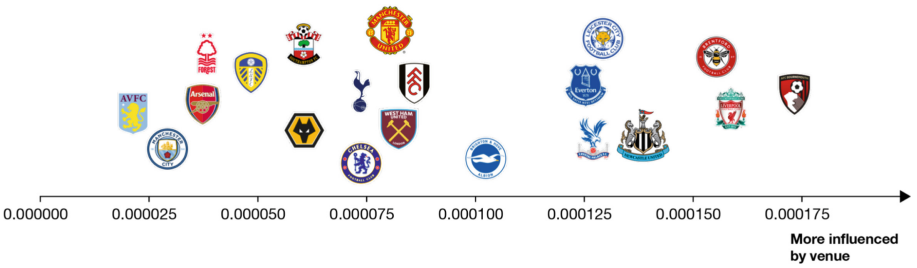


Fig. 3. The degree to which each team varied their playing style when playing at home vs away in the 2022/23 Premier League season.

6 Conclusions

Both of the proposed methods clearly showed that a team's playing style can be strongly affected by the game state. The modified SoccerMap model makes it possible to observe this effect in a specific context. We showed that it is possible to estimate how a certain team will adapt their decision for a specific pass as a function of the game state. This can be especially useful for coaches and players to adapt their actions in different states of a game. The Encoder-Decoder model proposes a method to describe how the game state will alter the team passing strategy, which is a useful tool for both the technical team and the media.

Acknowledgements. This research received funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme. We thank StatsBomb for providing the data used in this research.

References

1. Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., Matthews, I.: Identifying team style in soccer using formations learned from spatiotemporal tracking data. In: 2014 IEEE International Conference on Data Mining Workshop, pp. 9–14 (2014). Journal Abbreviation: 2014 IEEE International Conference on Data Mining Workshop. https://doi.org/10.1109/ICDMW.2014.167
2. Almeida, C.H., Ferreira, A.P., Volossovitch, A.: Effects of match location, match status and quality of opposition on regaining possession in UEFA champions league. J. Hum. Kinet. **41**, 203–214 (2014). https://doi.org/10.2478/hukin-2014-0048
3. Bojinov, I., Bornn, L.: The pressing game: optimal defensive disruption in soccer. In: Proceedings of the 10th MIT Sloan Sports Analytics Conference, SSAC '16. Boston, USA, pp. 1–8 (2016)
4. Clijmans, J., Van Roy, M., Davis, J.: Looking Beyond the Past: analyzing the intrinsic playing style of soccer teams. In: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part VI, pp. 370–385. Springer-Verlag, Berlin, Heidelberg (2023). https://doi.org/10.1007/978-3-031-26422-1_23, https://doi.org/10.1007/978-3-031-26422-1_23
5. Decroos, T., Bransen, L., Van Haaren, J., Davis, J.: Actions speak louder than goals: valuing player actions in soccer. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '19, Anchorage, AK, USA, pp. 1851–1861. ACM Press (2019). https://doi.org/10.1145/3292500.3330758, http://dl.acm.org/citation.cfm?doid=3292500.3330758
6. Decroos, T., Van Haaren, J., Davis, J.: Automatic discovery of tactics in spatiotemporal soccer match data. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '18', New York, NY, USA, pp. 223–232. Association for Computing Machinery (2018). https://doi.org/10.1145/3219819.3219832, https://doi.org/10.1145/3219819.3219832
7. Fernandez-Navarro, J., Fradua, L., Zubillaga, A., McRobert, A.P.: Influence of contextual variables on styles of play in soccer. Int. J. Perform. Anal. Sport **18**(3), 423–436 (2018). https://doi.org/10.1080/24748668.2018.1479925

8. Fernández, J., Bornn, L.: SoccerMap: a deep learning architecture for visually-interpretable analysis in soccer. arXiv:2010.10202 [cs], vol. 12461, pp. 491–506 (2021). https://doi.org/10.1007/978-3-030-67670-4_30, arXiv: 2010.10202
9. García-Rubio, J., Gómez, M.Á., Lago-Peñas, C., Ibáñez, J.S.: Effect of match venue, scoring first and quality of opposition on match outcome in the UEFA Champions League. Int. J. Perform. Anal. Sport **15**(2), 527–539 (2015). https://doi.org/10.1080/24748668.2015.11868811
10. Konefał, M., Chmura, P., Zacharko, M., Chmura, J., Rokita, A., Andrzejewski, M.: Match outcome vs match status and frequency of selected technical activities of soccer players during UEFA Euro 2016. Int. J. Perform. Anal. Sport **18**(4), 568–581 (2018). https://doi.org/10.1080/24748668.2018.1501991
11. Lago, C.: The influence of match location, quality of opposition, and match status on possession strategies in professional association football. J. Sports Sci. **27**(13), 1463–1469 (2009). https://doi.org/10.1080/02640410903131681
12. Link, D., Hoernig, M.: Individual ball possession in soccer. PLoS ONE **12**(7), e0179953 (2017). https://doi.org/10.1371/journal.pone.0179953
13. Marcelino, R., Mesquita, I., Sampaio, J.: Effects of quality of opposition and match status on technical and tactical performances in elite volleyball. J. Sports Sci. **29**(7), 733–741 (2011). https://doi.org/10.1080/02640414.2011.552516
14. Monte, J.: Using xPass to Measure the Impact Of Gamestate on Team Style (2023). https://statsbomb.com/articles/soccer/using-xpass-to-measure-the-impact-of-gamestate-on-team-style/
15. Morgulev, E., Galily, Y.: Analysis of time-wasting in english premier league football matches: evidence for unethical behavior in final minutes of close contests. J. Behav. Exp. Econ. **81**, 1–8 (2019). https://doi.org/10.1016/j.socec.2019.05.003, https://www.sciencedirect.com/science/article/pii/S2214804319301430
16. Muller, J.: Who never fouls? Who presses? Who hoofs it? What our data experiment says about your team. https://theathletic.com/3050603/2022/01/05/who-never-fouls-who-presses-who-hoofs-it-what-our-data-experiment-you-about-your-team/
17. Paixão, P., Sampaio, J., Almeida, C.H., Duarte, R.: How does match status affects the passing sequences of top-level European soccer teams? Int. J. Perform. Anal. Sport **15**(1), 229–240 (2015). https://doi.org/10.1080/24748668.2015.11868789
18. Pedro Santos, C.L.P., García-García, O.: The influence of situational variables on defensive positioning in professional soccer. Int. J. Perform. Anal. Sport **17**(3), 212–219 (2017). https://doi.org/10.1080/24748668.2017.1331571
19. Pleuler, D.: Player Masks: Encoding Soccer Decision Making Tendencies (2021). https://www.nessis.org/nessis21
20. Rahimian, P., Hyunsung, K., Schmid, M., Toka, L.: Pass receiver and outcome prediction in soccer using temporal graph networks. In: International Workshop on Machine Learning and Data Mining for Sports Analytics (2023)
21. Robberechts, P., Van Roy, M., Davis, J.: un-xPass: measuring soccer player's creativity. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD '23, ACM, New York, NY, USA (2023). https://doi.org/10.1145/3580305.3599924
22. Taylor, J.B., Mellalieu, S.D., James, N., Shearer, D.A.: The influence of match location, quality of opposition, and match status on technical performance in professional association football. J. Sports Sci. **26**(9), 885–895 (2008). https://doi.org/10.1080/02640410701836887
23. Vercruyssen, V., De Raedt, L., Davis, J.: Qualitative spatial reasoning for soccer pass prediction. In: CEUR Workshop Proceedings, vol. 1842. CEUR-WS.org (2016)

24. Wang, Q., Zhu, H., Hu, W., Shen, Z., Yao, Y.: Discerning tactical patterns for professional soccer teams: an enhanced topic model with applications. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, New York, NY, USA, pp. 2197–2206. Association for Computing Machinery (2015). https://doi.org/10.1145/2783258.2788577

Augmented Intelligence for FIFA Predictions

Kunal Jethuri(✉), Sai Charan Emmadi, Satya Samudrala, Jayakrishna Natarajan, Piyush Ghotkar, and Maitreya Natu(✉)

Digitate, Tata Consultancy Services, Pune, India
{kunal.jethuri,sai.e1,satya.samudrala,jayakrishna.n,
piyush.ghotkar,maitreya.natu}@digitate.com

Abstract. In today's data-centric world, businesses strive to monitor, manage, and optimize their operations efficiently. While data-driven solutions are increasingly popular, they often fall short due to challenges such as limited data availability, poor data quality, or high computational costs. Moreover, relying solely on data overlooks the valuable intuition and expertise of domain experts. To address this gap, we propose augmenting the power of data-driven analysis with human intuition. We demonstrate this approach through our FIFA match prediction engine, which seamlessly integrates data-driven analysis with fan intuition, resulting in a significant boost in accuracy from 65% with traditional data-driven approaches to 81%, effectively adapting to diverse match scenarios.

1 Introduction

Organizations today have increasing focus on developing data-driven solutions to monitor, manage, and optimize their business and operations. However, data-driven solutions are often not sufficient in many real-world setups. This happens due to a range of factors such as data availability issues, data quality issues, or costly compute power. Many solutions also inherently rely on the tacit knowledge and the intuition of the experts which they develop over years of experiences. Data-driven solutions often fail to capture this intuition. In such scenarios, the power of data-driven solutions can be best leveraged when it is augmented with natural intelligence.

Analytics for many sports events face such challenges. A football match prediction is an excellent example of a challenge that demands an augmented intelligence solution. In this paper, we present our attempt to create an augmented intelligence solution for predicting the FIFA world cup. Below are some reasons to select this problem:

1. The inherent nature of the game of football holds a certain level of surprise element and non-determinism. While data provides a solid base to form opinions but there are often scenarios for intuition and instincts to play a role.
2. Match results are highly impacted by a wide variety of variables such as team composition, team ranks, location, team dynamics, team morale, player form, fitness, and many such subtle aspects that are difficult to capture just by data. They need to be augmented with fan intuition.

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 69–79, 2025.
https://doi.org/10.1007/978-3-031-86692-0_6

3. There is historical data available of past 90 years of FIFA world-cup matches. Different data sources are available that capture different types of information such as player profile, demography, league match statistics, team statistics, etc.
4. Different data sources contain information of different quality, completeness, and fidelity. Older FIFA matches capture minimal statistics, but recent FIFA matches capture detailed statistics such as ball possession, pass accuracy, on target shots, etc.
5. There is a massive fan base of FIFA world-cup, and fans do extensive research to form their opinions.

Several aspects must be carefully considered when developing an augmented intelligence solution, i) how to translate data-driven observations into actionable insights and recommendations ii) how to motivate humans to contribute, iii) how to assist in forming viewpoints to make informed decisions iv) when to trust humans and when to rely on data, along with other related factors. By navigating these aspects, augmented intelligence systems can leverage the combined strengths of data-driven analytics and human intuition to deliver accurate insights.

In this paper, we address these aspects while creating a fan engagement experience app by leveraging the power of artificial intelligence and natural intelligence with the following features: (1) analyze historical data to mine trivia and generate data-driven prediction, (2) engage fans through polls and leader-boards and learn from fans' responses to generate intuition-driven predictions, (3) use both data and intuition to generate final winning odds of each FIFA football match.

Our innovation is an improvement over the existing solutions for football predictions and is a contribution in the space of augmented intelligence. This approach uses the data-driven approach of machine learning algorithms as well as the intuition-driven approach of football fans to boost the quality of the derived insights.

The proposed solution was deployed on our organization's website as a marketing campaign and hosted for one of our customers in Qatar as an HR initiative for employee engagement during FIFA 2022. Overall, the solution was used by more than 1000 fans. We successfully demonstrated the effectiveness of the augmented intelligence approach, achieving superior prediction accuracy compared to many state-of-the-art-predictors.

There are many real-world scenarios that can benefit by harnessing the power of augmented intelligence such incident resolution, procure-to-pay processes, retail promotions planning, etc. Augmented intelligence solutions can truly make a difference in making analytics-driven solutions a lot more real and usable. They not just make the insights accurate, but they also increase the trust and adoption of the insights.

2 Related Work

Several approaches exist for predicting football matches. In [7], various machine learning algorithms and features were explored, and found that a combination of features such as "home team factor," "away team," "first twenty-two players," and "goal difference," combined with a K-NN classifier, produced optimal results. [8] employed logistic regression to highlight significant variables such as "home defense" and "away defense",

while [9] utilized a Naive Bayes. Although, each of this approaches works well in different scenarios, there is no one-size-fit-all solution, and they fail to incorporate contextual understanding that fan intuition offers, developed over years of observing the sport.

The solutions in [2, 3] developed machine learning models relying on large volume of data and heavy compute power for performing football match predictions. In contrast, [9] introduces a multi-level weighted ensemble model utilizing LIWC and ElMa word embeddings from Twitter data to predict team support and match outcomes. While this approach captured fan sentiment, it may introduce biases and overlook the analytical depth provided by comprehensive historical data analysis, potentially leading to less precise predictions. [10] uses textual pre-game reports and numerical data to create classifiers for predicting success in the NHL. However, their approach overlooks bias and trust, and doesn't account for recent predictor accuracy while making predictions.

3 Solution Approach

Our prediction engine comprises of three essential components 1) Data-driven prediction model 2) Fan-intuition prediction model 3) Ensemble algorithm. The data-driven and fan-intuition prediction models each generate multiple prediction odds for win, loss, and draw outcomes. The ensemble component integrates these data-driven and human-driven odds to provide consolidated prediction odds. The following subsections detail the methodology and functionality of each model.

3.1 Data-Driven Predictions

As part of data driven predictions, we have adopted two approaches: a black box and a white box to predict the outcome of a football match. We have chosen this dual strategy because the available data sources are very diverse. Some datasets cover a wide range of matches but offer limited information, while others provide more detailed statistics but cover only a short period. Also, a match prediction involves simple known factors such as on-shot targets, passing accuracy, and not-so-known factors such as how the ball possession of one team will depend on tackles made by the other team.

Black-Box Approach
For datasets with a large number of matches but limited details, we have implemented a black-box approach. We have used openly available data of 43,000 matches that include all international and World Cup games. Although these datasets lack in-depth information, they provide a comprehensive overview of match outcomes over a period of 120 years. We utilized this data and trained a neural network based classification engine. We have performed feature engineering to extract relevant information from the dataset, including temporal and historical features about the participating teams. Specifically, we have extracted features such as winning streak, losing streak, last match outcome, and last won match, along with representing team strength such as rankings and ratings etc., for both teams. We fed these features along with team details and goals scored as input to the model and provided the match outcome as output. This engine is designed to identify hidden patterns and relationships within the data, even when the available information

is sparse. These patterns include various aspects such as (i) how a team's performance changes after a series of losses/wins, (ii) how a team's performance in current stage is effected by previous stages or previous international matches, (iii) identification of teams with long-term dominance, etc.

White-Box Approach
Every team has different set of attributes that play a crucial role in their match outcome. To identify those attributes for each team, we used classification-based rule mining technique [1]. We considered various attributes such as ball possession, shots on target, first few goals timings, etc., as features and the match outcome as the target variable, and extracted rules that characterize the winning behavior of teams. An example rule mined from historical patterns could be as follows:

When team = England & shots on target > 2 & Passing accuracy is > = 78% then Outcome = Won with a confidence of 80%

We apply such rules in real-time during the match to predict and fine-tune the winning odds of a team.

In addition to utilizing the black box and white box techniques, we have also included prediction odds from various open-source predictors [2], acknowledging the potential diversity in their data sources and methodologies. The individual odds generated by the black box and white box models, along with those from open-source predictors, are fed to ensemble algorithm (Sect. 3.3).

3.2 Fan-Intuition Predictions

While data-driven approaches excel in revealing hidden patterns, they struggle with intangible factors such as team morale, player psychology, tactical nuances, etc. To address this, we integrate human intuition into our solution by user engagement. To enhance user engagement, we implemented an interactive fan leaderboard featuring prediction scores, streak bonuses, bonus incentives, etc. Additionally, we provided users with diverse range of insights through trivia to aid in making informed predictions. However, human intuitions are inherently subjective and prone to biases, noise, and inconsistencies. To mitigate these, we have developed an approach to compute trustworthiness score to assess bias and knowledge of each fan.

Bias: Fans can be biased for a team, a player, a geography, etc. Consider a fan who is biased towards Argentina. This fan might predict Argentina will win every match irrespective of any conditions. This lack of variation is often a good indication of bias. For each of the potential source of bias, we analyze the variation in the probabilities of a fan choosing a team in the matches where the team wins and the team loses. Lesser the difference in the probability in the fan's response higher is the bias. We measure this bias by computing mutual information between fan's predictions and actual outcomes.

$$I(mout, fpred) = \Sigma_{mout=win,lose, fpred=win,lose} P(mout, fpred)\left(\log \frac{P(mout, fpred)}{P(mout)P(fpred)}\right) \tag{1}$$

$$B_{score} = normalized(I(mout, fpred)) \tag{2}$$

where *mout, fpred* corresponds to actual match outcome, fan prediction.

A low mutual information indicates that the fan is highly biased. If a fan's bias score (B_{score}) is below a certain threshold, then we consider their predictions to be biased and do not consider for predictions.

Knowledge: Not all fans hold the same knowledge about the game. Some are causal fans while some are serious researchers of the game! We measure the knowledge score of each fan to assess appropriate weight to their predictions. We have developed a structured questionnaire to categorize the knowledge of fans into three distinct levels. *Beginners* possess a basic understanding of the game, including fundamental rules and gameplay mechanics, allowing them to answer straightforward questions about player numbers and match duration. *Intermediates* are more engaged and regularly follow the game, offering a deeper and more informed perspective on ongoing events, such as player performance statistics and recent match outcomes. *Experts,* demonstrate a profound comprehension of the game. They have the ability to critically analyze and anticipate future events, providing valuable insights into likely game developments such as potential scorelines and the timing of critical events like the first goal. To quantify the knowledge of fans, we calculate a Knowledge Score (K_{score}) using a weighted formula that considers the accuracy of responses across the three levels. The formula is structured as follows:

$$K_{score} = \Sigma_{i=level}(w_i \, x \, Correctness_i) \tag{3}$$

$$Correctness_i = \frac{number\ of\ \ correct\ answers\ in\ level\ i}{total\ number\ of\ \ questions\ in\ level\ i} \tag{4}$$

where the weights $w_1 = 0.2$, $w_2 = 0.3$, and $w_3 = 0.5$ for levels 1, 2, and 3 respectively.

Fans' predictions are incorporated into our system only if their K_{score} exceeds a predetermined threshold, thereby ensuring the integration of only valuable and informed contributions into our predictive models.

Trustworthiness Score: Next, we compute the trustworthiness score for each fan, utilizing the previously calculated bias and knowledge scores. The trustworthiness score is calculated as follows:

$$T_{score} = average(B_{score}, K_{score}) \tag{5}$$

To derive the match predictions, we multiply the trustworthiness score by the prediction odds provided by the fan. This adjustment ensures that the final predictions reflect both the fan's credibility and the likelihood of their predictions.

$$Fan\ prediction\ odds = T_{score} \, x \, Fan\ prediction \tag{6}$$

3.3 Ensemble Prediction Engine

The ensemble prediction engine encompasses a diverse library of prediction algorithms comprising (1) in-house data-driven prediction algorithms (2) open-source prediction algorithms (3) fan predictions. While each predictor type offered unique insights, they

also had inherent limitations. Data driven predictions, although rooted in statistical analysis, may not encompass all variables. Likewise, fan-intuitive predictions, while driven by sentiments, might lack sufficient analysis required for accurate predictions. To address these challenges, we developed a method to dynamically adjust the influence of each predictor in the form of weights which are based on their historical performance. The final prediction odds are generated for each outcome – win, loss and draw, by combining the weighted predictions odds. The final prediction is represented as:

$$Prediction = \frac{w_1 pred_1 + w_2 pred_2 + w_3 pred_3 + \ldots}{w_1 + w_2 + w_3 + \ldots} \tag{7}$$

To dynamically adapt the weights, we explored various approaches:

1. Additive Increase Multiplicative Decrease (AIMD): Weights are increased additively for correct predictions and decreased multiplicatively for incorrect ones.

$$W_{t+1} = \begin{cases} W_t + af * pred(t), & if\ prediction\ is\ correct \\ W_t * mf * pred(t), & if\ prediction\ is\ wrong \end{cases} \tag{8}$$

2. Additive Increase, Subtractive Decrease (AISD): Weights are increased additively for correct predictions and decreased subtractive for incorrect one.

$$W_{t+1} = \begin{cases} W_t + af * pred(t), & if\ prediction\ is\ correct \\ W_t - sf * (1 - pred(t)), & if\ prediction\ is\ wrong \end{cases} \tag{9}$$

3. Inverse Increase Multiplicative Decrease (IIMD): Weights are increased inversely proportional to the magnitude of incorrect predictions and decreased multiplicatively for incorrect ones.

$$W_{t+1} = \begin{cases} W_t \div if * (1 - pred(t)), & if\ prediction\ is\ correct \\ W_t * mf * pred(t), & if\ prediction\ is\ wrong \end{cases} \tag{10}$$

4. Inverse Increase Subtractive Decrease (IISD): Weights are increased inversely proportional to the magnitude of incorrect predictions and decreased subtractive for incorrect one.

$$W_{t+1} = \begin{cases} W_t \div if * (1 - pred(t)), & if\ prediction\ is\ correct \\ W_t - sf * (1 - pred(t)), & if\ prediction\ is\ wrong \end{cases} \tag{11}$$

where af, mf, sf, if are additive, multiplicative, subtractive and inverse factor respectively, while W_t and W_{t+1} are the weights before and after adjusting for match results, while $pred(t)$ is the prediction percentage assigned to the actual outcome. We present an evaluation of these approaches later in the Experimental Evaluation section where we concluded that Additive Increase Multiplicative Decrease demonstrated the best performance.

The final step involves pretraining the weights using historical data. Through iterative refinement and optimization, we fine-tuned the weights and identified optimal parameter configuration for adjustment factors (af and mf) to get optimal results. With pretrained weights, the prediction engine is enabled to dynamic adjust weights to ensure accurate predictions.

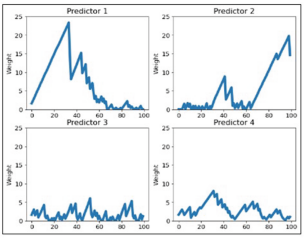
4 Experimental Evaluation

In this section, we demonstrate that our proposed ensemble approach produces promising results and can be successfully applied to several predictors with varying characteristics. We first present an evaluation of the effectiveness of the proposed ensemble approach and demonstrate that the ensemble approach outperforms individual predictors. We then present the evaluation of our approach's predictions of FIFA 2022.

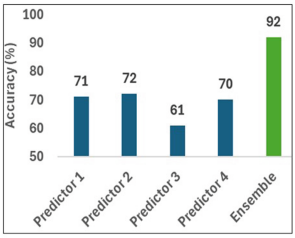
4.1 Sensitivity Analysis

Method	Correct Predictions (100 matches)	Correct Predictions (last 10 matches)
AIMD	92	9
AISD	81	4
IIMD	74	4
IISD	75	4
Mean	72	4

(a)



(b)



(c)

Fig. 1. Experiment on sensitivity analysis

One of the most important aspects of our ensemble prediction approach is to assign the right weights to the right predictor. We discussed 4 ways to tune predictor weights in the previous section. In this section, we present an evaluation of these methods in their effectiveness in improving prediction accuracy.

To perform these experiments, we created synthetic data of match predictions by 4 different predictors for 100 matches with each predictor exhibiting distinct behavior. Predictor 1 shows initial strong performance followed by steady decline. Predictor 2 starts poorly but improves consistently as matches progressed. Predictor 3 exhibits inconsistent performance with no discernible trend. Predictor 4 begins with good accuracy but declines over time. Synthetic data enabled precise control over these behaviors, helping us evaluate adaptability and robustness to different predictor patterns, which may not be evident in historical match data.

Experiment 1: Comparison of Various Weight Tuning Techniques
For each match, we assign weights to each predictor using the four methods explained in the previous section: AIMD, AISD, IIMD, IISD, and compute the overall prediction accuracy. We also compute the accuracy by taking a simple mean of all predictors. Figure 1(a) shows the prediction accuracy using each of these methods. For each method, we computed the total number of correct predictions and the number of correct predictions in the last 10 matches to evaluate both overall performance and consistency towards the end of predictions.

All four weight tuning approaches outperformed the simple mean approach. AIMD demonstrated the best results with 92 out of 100 correct predictions and 9 out of 10 correct

predictions in the last 10 matches. Detailed analysis revealed that AISD did not sufficiently penalize incorrect predictions, resulting in high weights across predictors. IIMD and IISD overly rewarded correct predictions, leading to skewed weight distributions. AIMD provided the most effective performance, with weights gradually increasing for correct predictions and sharply decreasing for incorrect ones. This evaluation highlights AIMD as the most effective weight tuning technique for enhancing prediction accuracy and confidence in our ensemble model.

Experiment 2: Effect of Additive Increase and Multiplicative Decrease
We next demonstrate how AIMD method changes the weights of these predictors and leads to a much higher accuracy of the ensemble algorithm.

Figure 1(b) shows how the weights of the 4 predictors change over the course of 100 matches using the AIMD method. Predictor 1 initially receives higher weights, which gradually decline over time. Predictor 2 starts with minimal weights, progressively increasing throughout the matches. Predictor 3 exhibits fluctuating low weights with no discernible trend. Predictor 4 shows an initial increase in weights followed by a gradual decrease.

Figure 1(c) presents a bar plot showing the accuracy of each predictor and the ensemble weight tuning approach. Individual predictors achieve accuracies in range 61% and 72%. In contrast, our ensemble approach achieved an accuracy of 92%.

4.2 FIFA 2022 Case-Study

The proposed solution was implemented as a fan engagement application, hosted on our organization's website and at one of our customers in Qatar as part of their HR initiative for employee engagement during FIFA 2022. In this section we present our findings from this live experiment.

For the data-driven component of our approach, we collected data spanning 120 years, including detailed records of more than 17000 players. These records encompassed performance metrics such as overall rating, work rate, crossing/passing/dribbling/goalkeeping/defending ability, rating at different positions. Historical data of international and FIFA matches were also included, with records of over 43000 matches detailing goals scored, venue, attendance, stage of the tournament, and more. Additionally, data of 2018 FIFA world cup matches with fine details of offsides, possession, pass accuracies, shots on target and saves were collected.

Through our fan engagement application, we collected various fan opinions. These included their demographic details, favorite team, player, and predictions on the World Cup winner. To further engage fans and gather predictions for each match, as well as to analyze bias, trust and knowledge among fans, we posed specific question such as:

1. Which team is going to win?
2. When will the first goal be scored?
3. What will be the final scoreline?

The application registered over 1000 users and collected over 10,000 votes across matches through fan polls.

Experiment 3: Comparison of Ensemble Algorithm with Data-Driven and Fan-Based Predictions

Table 1 presents a comparison of accuracy of data-driven, intuition-driven, and ensemble approaches in predicting the 64 matches of FIFA world cup 2022. It can be seen that the proposed approach prediction outperformed individual predictors, achieving an accuracy of 69% with 44 out of 64 correct predictions. In 6 matches, fan predictions augmented data-driven predictions, resulting in correct final predictions. Notably, fans accurately predicted 3 draws. In 2 matches, data-driven predictions were chosen over fan predictions, yielding correct outcomes. This experiment demonstrates the effectiveness of combining fan intuition with data-driven insights, enhancing the overall accuracy of match predictions.

Table 1. Accuracy comparison of Predictions

Total Matches	Data-driven Predictions	Fan-based Predictions	Ensemble Predictions
64	38 (59%)	37 (58%)	**44 (69%)**

Experiment 4: Comparison with the State-of-the-art

In this experiment, we evaluated the performance of our prediction engine against state-of-the-art predictors namely Kashef [5] and CBS [4]. Table 2 presents the comparison of the 3 approaches.

In the round of 16, the proposed solution correctly predicted the outcomes of 7 out of 8 matches, matching Kashef's performance and outperforming CBS, which correctly predicted 6 out of 8 matches. In quarter finals, the proposed solution predicted 2 out of 4 matches correctly, which was on par with Kashef and better than CBS, which only predicted 1 out of 4 matches correctly. In semi-finals and finals, the proposed solution achieved a perfect record, accurately predicting all 4 matches, whereas Kashef predicted 3 out of 4 matches correctly, and CBS predicted 2 out of 4 matches correctly.

Table 2. Comparative analysis of different prediction engines

	Proposed approach	Kashef	CBS
Round of 16	**7 of 8**	7 of 8	6 of 8
Quarterfinal	**2 of 4**	2 of 4	1 of 4
Semi-final & final	**4 of 4**	3 of 4	2 of 4

Interesting Observations

FIFA 2022 predictions were an interesting exercise giving many insights into the effectiveness of different approaches in different scenarios. Below we present some stories.

Observation 1: In the final match, data-driven insights strongly matched with what actually happened. Argentina didn't let in any goals in the first half, and conceded 3 goals near the end, similar to what data-driven insights predicted in the below rules.

- *Argentina is yet to concede a goal in the first half! whereas France scored its 1st goal between 30 to 60 min 4 out of 7 times in the Football World Cup 2018!*
- *Argentina has conceded 60% of its goals between 75th to 90th minute of the match! and France has scored 63% of its goals between 60th to 90th minute of the match!*

Observation 2: In the Serbia Vs. Switzerland match, the data-driven predictions failed to account for injuries within Serbian squad, incorrectly predicting Serbia to win with 34% odds. In contrast, the intuition-driven predictor effectively identified the right set of fans who considered the impact of injuries. Consequently, the final prediction was favoring Switzerland with 41% odds, highlighting the limitations of relying solely on data-driven methodologies and the complementary role of intuition-driven approaches in enhancing prediction accuracy.

5 Conclusion

In this paper, we have proposed an approach that bridges the gap between data-driven solutions and human intuition, aiming to enhance the effectiveness of predictive analytics. By seamlessly integrating these two components in our FIFA prediction engine, we improved the accuracy of predictions that dynamically adapted to diverse match scenarios. This approach addressed the shortcomings of relying solely on data-driven solutions, which often encountered challenges such as limited data availability, poor data quality, or high computational costs. Through our approach, we had showcased the potential of harnessing both data-driven analysis and human intuition to optimize operational efficiency in today's data-centric business landscape.

References

1. Győrödi, C., Gyorodi, R., Prof, D., Ing, S., Stefan, H.: A Comparative Study of Association Rules Mining Algorithms (2004). https://doi.org/10.13140/2.1.1450.3365
2. Majumdar, A., Kaur, R., Kulkarni, T., Jiruwala, M., Shah, S., Pise, N.: Football match prediction using exploratory data analysis & multi-output regression. In: 2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, pp. 1–6 (2022). https://doi.org/10.1109/IATMSI56455.2022.10119340
3. Luo, X., Liu, K., Yuan, X.: Research on football team ability based on random forest and ranking method. In: 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Shanghai, China, pp. 212–216 (2021). https://doi.org/10.1109/AINIT54228.2021.00050
4. 2022 FIFA World Cup predictions, picking every game: Lionel Messi leads Argentina to glory; USMNT move on - CBSSports.com. https://www.cbssports.com/soccer/news/2022-fifa-world-cup-predictions-picking-every-game-lionel-messi-leads-argentina-to-glory-usmnt-move-on/

5. Jazeera, A.: World Cup predictions: How many games did our AI get right? Al Jazeera. https://www.aljazeera.com/news/2022/12/19/world-cup-predictions-did-our-ai
6. Maitama, J., Mohammed, M., Raj, R.: Predicting the Outcomes of Football Matches Using Machine Learning Approach (2022). https://doi.org/10.1007/978-3-030-95630-1_7
7. Prasetio, D., Harlili, D.: Predicting football match results with logistic regression. In: 2016 International Conference on Advanced Informatics: Concepts, Theory and Application (ICAICTA), Penang, Malaysia, pp. 1–5 (2016). https://doi.org/10.1109/ICAICTA.2016.7803111. keywords: {Logistics;Predictive models;High definition video;Training data;Testing;Training;Games;regression coefficients;variables;logistic regression;prediction accuracy}
8. Venkatachalam, A., Devarajan, R., Sree Nandha S.S.: Football prediction system using gaussian naïve bayes algorithm. In: 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, pp. 1640–1643 (2023). https://doi.org/10.1109/ICEARS56392.2023.10085510. keywords: {Renewable energy sources;Games;Machine learning;Prediction algorithms;Classification algorithms;Bayes methods;Decision trees;football;Gaussian naïve bayes;machine learning;Big data;goals;outcomes},
9. Rabbi, M.F., Mukta, M.S.H., Jenia, T.N., Islam, A.K.M.N.: Predicting fans' FIFA world cup team preference from tweets. In: Bhuiyan, T., Rahman, M.M., Ali, M.A. (eds.) Cyber Security and Computer Science. ICONCS 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 325. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-52856-0_22
10. Weissbock, J., Inkpen, D.: Combining textual pre-game reports and statistical data for predicting success in the national hockey league. In: Sokolova, M., van Beek, P. (eds.) Advances in Artificial Intelligence. Canadian AI 2014. Lecture Notes in Computer Science(), vol 8436. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06483-3_22

Transformer-Based Framework for Versatile Analysis of Events Data in Soccer

Aviv Rovshitz(✉) and Rami Puzis

Software and Information Systems Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel
{avivrov,puzis}@post.bgu.ac.il

Abstract. In sports analytics, machine learning models are used for players' appointment analysis, strategic planning, etc. Multiple rating systems were developed to rank the player performance based on statistical measurements, machine learning, and deep learning analysis of actions and sequences. However, capturing the complex relationships between actions within open play sequences, affected by tactical game planning and team deployment, is a major challenge. Inspired by language models, we developed SoccerTransformer, a transformer-based deep learning framework tailored to soccer analytics to address this challenge. The proposed framework includes (1) self-supervised pre-training to capture generic in-game dynamics, (2) a downstream sequence classification task to predict the attack phase outcomes, and (3) a player rating system. Empirical results demonstrate that SoccerTransformer accurately captures player roles and predicts goals with $F1$ of 0.814 to 0.862 on previously unseen games. SoccerTranformer framework stands out by providing a greater correlation with market values than other state-of-the-art rating systems.

Keywords: sports analytics · soccer · transformers · language model · market value

1 Introduction

In 2022, Premier League clubs spent over €3 billion on player acquisitions during the transfer windows, a 70% increase compared to the previous year. Between 2013 and 2022, Premier League clubs collectively spent €18.26 billion on player transfers [3]. These substantial financial investments underscore the necessity for competitive teams to maintain high spending levels to stay relevant. However, to address the issues of financial wastage and competitive imbalance, the Fédération Internationale de Football Association (FIFA) implemented new financial regulations that limit spending on wages, transfers, and agent fees to 70% of club revenue [21]. This regulatory framework necessitates that teams accurately assess how new players will meet their strategic needs when making acquisitions.

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 80–92, 2025.
https://doi.org/10.1007/978-3-031-86692-0_7

Market value plays a critical role in the financial sustainability and strategic planning of soccer clubs. The player's market value (PMVs) is often perceived as an indicator of their potential impact on the team, influencing decisions related to player recruitment and transfers, and salary negotiations. Therefore, creating a player rating system that correlates well with PMV is essential. Such a system can provide evidence of players with a high rating score and a correspondingly high PMV, indicating their worth. However, PMV can sometimes misrepresent a player's true intrinsic value. A well grounded generic scoring system can help identifying undervalued players or new talents increasing the potential return on investments. Furthermore decisive perspective for clubs is identifying the optimal time to sell overpriced players or buy underpriced players to maximize profits on available assets. Hence the seasonal market value difference (SMVDs) are a critical aspect of recent regulatory policies for profit-driven organizations.

For many years, pundits in journalism and analysts within soccer clubs have relied on general key performance indicator (KPIs) to define the traditional roles of soccer players. These KPIs include metrics such as total distance covered, average running speed, number of accelerations, number of duels, and percentage of duels won [14]. In recent years, advanced metrics have been developed based on ball-events data and tracking data to allow for a comprehensive evaluation of players and enhance the effectiveness of scouting and recruitment processes through a nuanced evaluation of specific types of players [5,11,19,20]. While these metrics correspond to PMV there is a large space for improvement.

We believe that most existing metrics achieve non-optimal performance because they struggle with capturing the complex relationships between players' actions affected by tactical game planning and team deployment. Transformer models such as BERT [10] and GPT [17] capture the intricate dependencies between words in natural language relying on an attention mechanism. Today, transformers are used in a variety of domains to analyze sequence data with complex relationships [2,4, etc.].

In this article, we introduce SoccerTransformer – deep learning framework for assessing soccer players. Empirical results demonstrate that a model trained on event data of the top-five European leagues covering the 2015/2016 season [22], performs well not only on a hold-out test set but also on seasons recorded 4–8 years after the training data. Our contributions are threefold: (1) We provide the first transformer model for ball-events data (2) The pre-trained transformer is generic and capable of: (a) detecting players' types and sub-types and (b) the fine-tuning model predicting goal scoring outcomes of a partial sequences of actions ($F1 > 0.81$). (3) We show that the goal scoring and goal prevention probabilities are useful for developing a player rating system with state-of-the-art correlation to PMV (Pearson $r. > 0.6$). Employing advanced transformer-based deep learning architectures for analyzing action sequences has a great potential to assist clubs in making more informed decisions, characterize players' attributes based on their actions, and estimate similarities between players using contextual features rather than relying solely on statistics.

2 Related Work

Recent advancements in ball game analytics are driven by extensive data collection by companies like Statsbomb, Wyscout, Stats Perform, and SkillCorner. These firms use both human expertise and machine learning to capture two main types of data: tracking data and *on-the-ball-events data* [16]. Tracking data records the positions of all players and the ball, capturing up to 25 data points per second for off-ball activities. Such data is expensive and not publicly available, often utilized only by private firms and well-resourced organizations. On-the-ball-events data (or events for short) records every action performed with the ball during a game, including the location, action type, involved player, etc. Such data is more accessible and much cheaper to produce than tracking data. In this article, we focus on the analysis of events.

Representation learning: Vector representations of players, actions, trajectories, etc., are useful for a variety of downstream tasks. For example, player representations can be created using matrix factorization techniques [6]. Gyarmati and Hefeeda [12] characterized players based on their movement patterns. Rudolph and Brefeld [18] employed transformers to learn representations of multi-player trajectories from tracking data using a masked language model (MLM). To the best of our knowledge, transformers have not been applied so far for representation learning from events data.

Tactics and Predictive Analytics: Segmented match event streams were analyzed in the past using dynamic time warping and the CM-SPADE algorithm to identify common team tactics [8]. Decroos et al. [9] examined defensive strategies by applying soft clustering on a non-fixed zone grid, utilizing a mixture model and a Von Mises distribution to analyze playing style, team formations, and movements. POGBA algorithm [7] used machine learning to estimate the conditional probability of scoring a goal based on events within a fixed horizon. In a similar vein, we utilize transformers to predict goal scoring.

Players Rating: Pappalardo et al. [15] leveraged machine learning to rate players and discussed multiple applications for an accurate players' rating system, such as scouting, estimating players' versatility, or database retrieval. Other nuanced ratings were developed to assess the performance of players taking different roles. expected goals (xG) for forwards [20], expected assists (xA) for midfielders [1], expected goals against (xGA) for goalkeepers, and expected threat (xT) for offensive players [19]. While xG, xA, and xT offer insights into offensive concepts, they overlook other actions that are performed during the game. The Expected Possession Value (EPV) [11] expands the evaluation to include defensive maneuvers. It estimates the potential outcome of a possession, indicating whether the attacking team is likely to score a goal or a turnover will take place. valuing actions by estimating probabilities (VAEP) [5] assesses actions by their impact on either scoring a goal or reducing the likelihood of conceding one within a fixed period. VAEP excels at evaluating risk-reward balances, while xT is more stable [24], and both are considered state-of-the-art. In this article, we use xT and VAEP as baselines to evaluate SoccerTranformer on the players' rating task.

3 Soccer Transformer

The proposed SoccerTranformer pipeline includes five phases: (1) data pre-preprocessing, (2) self-supervised pre-training to capture player roles, (3) a downstream sequence classification task to predict the attack phase outcomes, (4) action value estimation, and (5) a player rating system to derive seasonal PMVs. Figure 1 summarizes the relationships between the pipeline components and the organization of the main sections of this article.

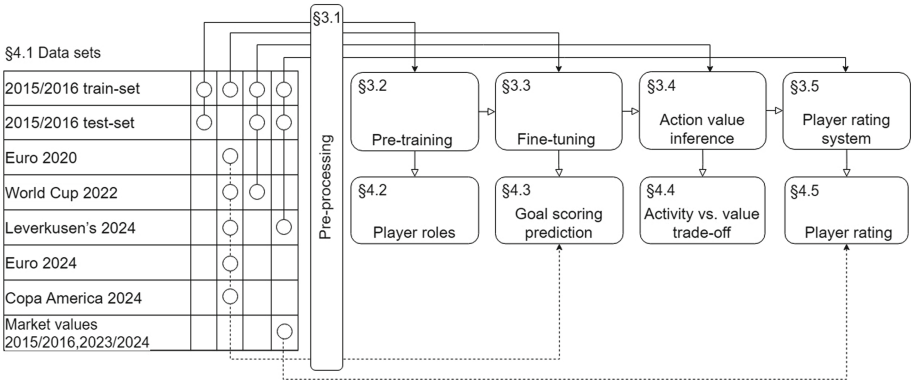


Fig. 1. The SoccerTransformer pipeline and datasets. Solid arrows represent data used to create the models. Dashed arrows represent validation data.

3.1 Data Pre-processing

We employ the SoccerAction framework [5] to convert StatsBomb events data into a uniform format of on-the-ball player actions (actions for short). We extract eight attributes to characterize any action: the action type, the player who performs it, the start and end x,y pitch coordinates (four attributes), the body part, and the action result. For simplicity, coordinates are discretized into zones within a 12×16 grid.

Next, we build a custom tokenizer, which is an integral part of a transformer pipeline. Eight attributes comprising an action correspond to eight types of tokens, each characterized by a unique prefix. All unique values are indexed and appended to the respective prefixes. In total, we defined a vocabulary of 3477 tokens including special tokens (start, end, mask, padding, etc.) and all values of action attributes found in the 2015/2016 season data.

We split the stream of actions into sequences, attempting to capture the natural flow of the game. We adopt an approach commonly referred to as open-play sequences while breaking a sequence if one of the following events occurs: (1) Actions ending with ball out of play, always initiate a new sequence. (2) Successful interceptions and clearances represent a definitive turnover in possession.

(3) Period time change naturally reset many elements of the game state. (4) We consider more than three consecutive touches by an opponent team as an indicator of control change, signifying a shift in possession. (5) A goal ends a successful attacking phase.

We label all sequences as either goal (in the case of event 5) or non-goal (in all other cases).

3.2 Pre-Training for Player Representations

In this phase, we employ a self-supervised learning approach similar to large language model (LLMs) pre-training, specifically through MLM, which is instrumental in developing a language model without the need for labeled data. MLM is a prevalent approach where the model is trained to reconstruct masked sequence elements from their contexts. The attention mechanism of transformers allows the models to learn sparse contexts and capture complex relationships between sequence elements.

In textual data, a common practice is to mask random words. Following preliminary experiments, we decided to mask only random player tokens without masking action type, body part, etc. This approach encourages the model to focus on subtle distinctions and similarities between players. It facilitates inference of the masked players' identities from their actions, location zones, and the entire progression of the open-play sequences. By the way, with such an approach, the model implicitly learns player roles and sub-role as will presented in subsequent.

In our experiments, we have used a BERT-like architecture comprising six attention heads, six hidden layers, a maximum sequence length of 512 tokens, and an embedding size of 96. We have pre-trained the model for 20 epochs on the 2015/2016 season train and test sets.

3.3 Fine-Tuning for Goal Prediction

We develop a goal-scoring predictor that classifies sequences of soccer actions based on their potential to lead to a goal. We have attached a classification head comprising one randomly initialized fully connected layer with two softmaxed outputs to the pre-trained transformer. We fine-tuned the entire model for 15 epochs using the 2015/2016 train set. The predictor should assess the likelihood of any partial sequence of actions resulting in a goal. To achieve this, we discard random suffixes of the input sequences during training. For example, if a full open-play sequence consists of four events (32 tokens), we keep it intact with 25% probability, replace the last action with eight padding tokens with a 25% probability, etc., while retaining at least one action in every sequence[1]. By understanding which sequences are most likely to lead to goals, a transformer model can value individual tokens or actions (8-tuples) according to their contribution to the goal or non-goal outcomes.

[1] We omit technicalities such as the correct handling of EOS and SEP special tokens.

3.4 Action Value

A trained goal-scoring predictor allows assessing action values based on changes in goal-scoring likelihood. Specifically, consider the i'th action in a sequence. We apply the trained predictor to derive the goal-scoring likelihood provided the first $i-1$ actions of the sequence and provided the first i actions. A positive difference between the latter and the former indicates that i'th action contributed to a goal outcome. A negative difference indicates that it contributed to the non-goal outcome[2].

This approach highlights the impact of individual actions within the broader context of team play. We cover both offensive and defensive contributions of individual actions. Offensive contribution measures the extent of actions contributing to increasing the scoring likelihood if a player is in the attacking team. Defensive contribution measures the extent to which an action reduces the goal-scoring likelihood if a player is on the defending team.

3.5 Player Rating System

We conclude the SoccerTranformer pipeline by determining the individual contributions of each player over a long period, such as a game or a season. Here, we simply aggregate the attack and defense contributions of all individual actions performed by the player. The sum of these components provides a total rating for each player, offering a view of their effectiveness in both attacking and defending tasks within the game.

4 Experiments and Evaluation

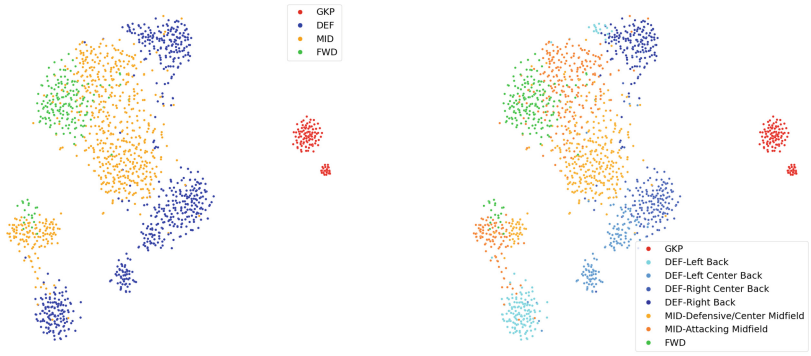
4.1 Data Sets

The main dataset utilized in this study comprises on-the-ball action events meticulously recorded by StatsBomb [22]. This dataset covers the 2015/2016 season from the top five European soccer leagues: Serie A (Italy), Premier League (England), La Liga (Spain), Ligue 1 (France), and Bundesliga (Germany). It includes detailed records from 1,823 matches, featuring 98 teams and documenting the activities of 2,640 players. This dataset is widely recognized among researchers for its utility in training and evaluating predictive models related to soccer performance analysis. During pre-processing we constructed 427,434 open play sequences. All the data from the 2015/2016 season was used for pre-training and to demonstrate player roles and sub-roles.

We used 90% of the open play sequences from the 2015/2016 season for training a goal predictor and 10% for testing. Additional hold-out test sets used include Euro 2020, World Cup 2022, Leverkusen's 2024 data, Euro 2024, and Copa America 2024, also released by StatsBomb. These datasets were pre-processed in the same manner and were not used for pre-training. We used the

[2] This is a naive approach which can be enhanced using explainable AI techniques such as Shapley values [13].

2015/2016 season with World Cup 2022 for action value inference. The 2015/2016 season and Leverkusen's 2024 data were chosen for player rating evaluation due to the availability of the respective market value data sourced by Transfermarkt [23]. We fetched start and end-of-season market values to derive the ground truth PMV at the end of the season and SMVD. The datasets and their mapping to results are summarized in Fig. 1.



(a) Player roles (b) Player sub-roles

Fig. 2. Two-dimensional t-SNE visualization of player embeddings based on the 2015/2016 season data for 1638 players who have played more than 900 min.

4.2 Player-Representations Correspond to Their Roles

We extracted vector representations of player tokens from the model pre-trained with open play sequences as described in §3.2. In order to understand whether player representations contain meaningful information, we used t-SNE to visualize them in Fig. 2. This figure shows 1638 players who accumulated over 900 min of playtime color-coded by their roles and sub-roles. We can observe a compelling clustering of the players with clearly distinguished goalkeepers, multiple clusters of defensive players, and two clusters with a mixture of midfielders and forward players. At this stage, the reasonable relative position of the clusters is sufficient to conclude that player embedding may be useful for downstream tasks. Since the main objective of this paper is to present the player rating system, we do not investigate here the subtle nuances of the role distribution across clusters marking it as a topic for future research.

4.3 Goal Scoring Prediction

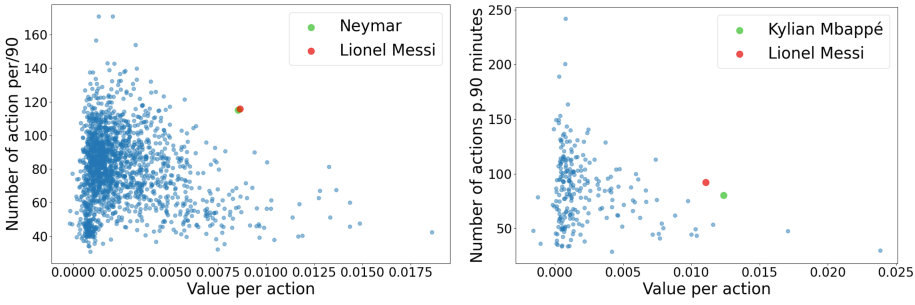
Table 1. Goal prediction results.

Test set	Seqs.	F1	ROC	PR
Euro 2020	10450	0.814	0.992	0.828
World Cup 2022	13103	0.834	0.992	0.846
Leverkusen's 2024	6335	0.853	0.992	0.876
Euro 2024	9442	0.825	0.992	0.819
Copa America 2024	6374	0.862	0.995	0.893

The SoccerTranfromer model with a classification head was fine-tuned for 15 epochs using the 2015/2016 train set. We evaluated the fine-tuned model on five different test sets as depicted in Fig. 1. Table 1 details the test sets and the respective prediction results. The test data underwent the same preprocessing as described in §3.1. Although the challenge is posed by a heavily imbalanced dataset, where only 1% of the sequences are labeled as positive. The area under the precision-recall curve (PR) and, $F1$ scores of 0.814 to 0.862 for goal detection on previously unseen games showcase the model's potential in dissecting sequences of actions into those that end with a goal and those that do not. These results further confirm that SoccerTransformer can understand sequences of soccer actions. Next, we focus on individual actions within the sequences and their potential values.

4.4 Activity Vs. Value Trade-Off

Similar to previous work, we present the average attack and defense contribution of players' actions vs the activity of the players in Fig. 3. These results are consistent with previous work showing a clear trade-off between the number of actions per 90-minute game time and the value of these actions. The best players are found on the Pareto frontier in this plot. For example, Messi consistently excels during the critical phases of games during both the 2015/2016 season and the World Cup 2022 tournament. While this result is not novel, it is important to ensure that SoccerTransformer can assess the value of actions before we proceed to rate players.



(a) 2015/2016 season (>900 minutes) (b) World Cup 2022 (>270 minutes)

Fig. 3. Activity vs. value per action trade-off. Red dot highlights Lionel Messi. (Color figure online)

4.5 Player Ratings

The SoccerTransformer player rating system can be applied for arbitrary time frames across games and seasons. Here, we applied the proposed rating system to the 2015/2016 season and Leverkusen's 2024 dataset and compared it with state-of-the-art represented by xT and VAEP ratings. We implemented VAEP models using CatBoost since it achieved the best performance in terms of area under the receiver operating characteristic curve (AUC) in preliminary experiments, consistent with the original publication [5]. We used the original 8X12 grid weights published for xT [19] and used them to calculate aggregated action values. Finally, to make the ratings comparable across playtime, we normalize all rating scores by 90 min. All three ratings were compared to PMV, and SMVD extracted from TransferMarkt, using Pearson and Spearman's correlation coefficients, as well as the coefficient of determination (R^2). The results, summarized in Table 2 and Fig. 4, display favorable performance of SoccerTransformer by a large margin according to all three coefficients.

Table 2. Correlation analysis between player ratings, who played more that 900 min in 2015–2016, related to PMV and SMVD

(a)PMV

Rating	Pearson	Spearman	R^2
xT	0.508	0.225	0.258
VAEP	0.456	0.378	0.208
Soccer- Transformer	**0.625**	**0.432**	**0.391**

(b) SMVD

Rating	Pearson	Spearman	R^2
xT	0.135	0.083	0.018
VAEP	0.047	0.136	0.002
Soccer- Transformer	**0.284**	**0.227**	**0.081**

To further evaluate the effectiveness of the ratings, we evaluate recall@k and precision@k while attempting to retrieve the top 100 players with the highest PMV. Recall@k and precision@k are commonly used metrics for evaluating the effectiveness of recommendation systems. Recall@k measures the model's ability to retrieve all relevant instances within the top-k results, reflecting its completeness. Precision@k, on the other hand, assesses the accuracy of these predictions by indicating the proportion of relevant results. According to both metrics VAEP and SoccerTransformer are consistently better than xT (see fig. 5). SoccerTransformer shows better retrieval performance up to $k = 60$. Consequently, SoccerTransformer identifies more relevant players among both the top 100 most valuable players overall and within smaller subsets of k candidates, a signal of the potential efficacy of financial decisions.

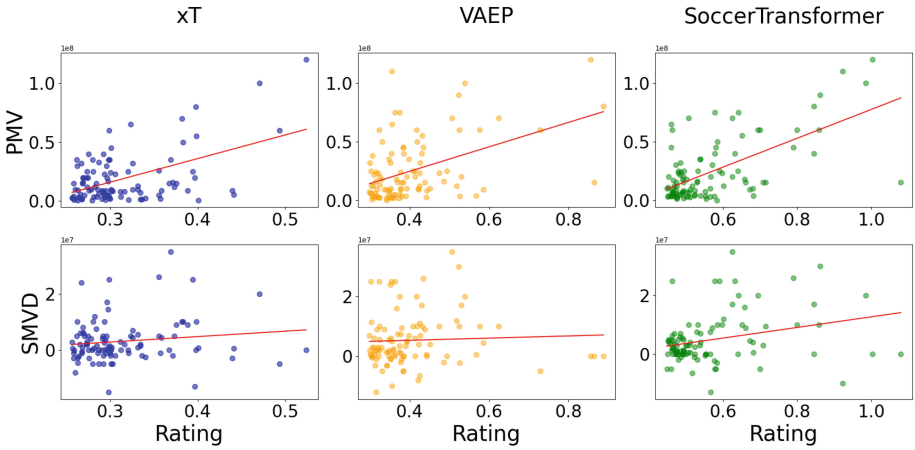
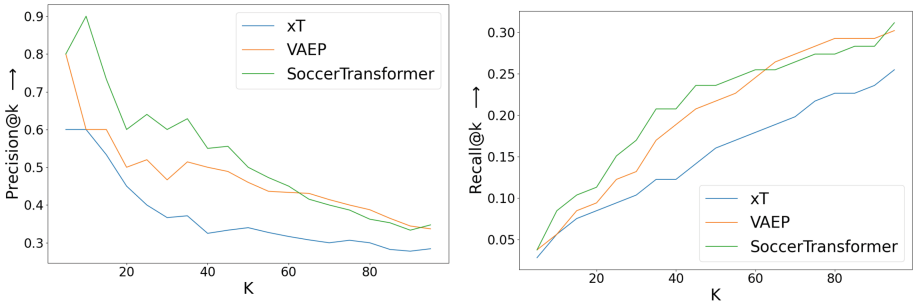


Fig. 4. A scatter plot depicts the relationships between player ratings, who played more that 900 min in 2015–2016, and PMV in the top row, and SMVD in the bottom row. Red line represents R^2.



(a) Precision@k (b) Recall@k for top-100

Fig. 5. Performance of rating systems for retrieval of 2015/2016 season players with top 100 regarding PMV.

Finally, similar to [5,24] we showcase the top players according to each rating system in Table 3. Each rating preferred different types of players. In the 2015/2016 season (table 3a) xT and VAEP favored attacking midfielders, while SoccerTransformers highlighted the standout season of Luis Suárez, who won European Golden Shoe, and Cristiano Ronaldo, who ranked second in contributing to goals and assists, following Suárez and preceding Messi and Neymar. In the Leverkusen's 2024 data (table 3b), all ratings recognized the Bundesliga's Player of the Season, Florian Wirtz. Notably, xT identified more creative players, VAEP provided a balanced representation of both attacking and defensive players, and SoccerTransformer highlighted contributors from across the pitch.

Table 3. Top-5 players with respective market values according to xT, VAEP, and SoccerTransformer rating systems.

(a) The 2015-2016 season

Method	Rank1	Rank2	Rank3	Rank4	Rank5
xT	Messi 120M	Di María 60M	Neymar 100M	Montero 5.5M	Deulofeu 9M
VAEP	Bale 90M	Zlatan 15M	Messi 120M	Di María 60M	Rodríguez 70M
Soccer-Transformer	Zlatan 15M	Messi 120M	Neymar 100M	Ronaldo 110M	Suárez 90M

(b) The Bayer Leverkusen's 2024 dataset (effective date 05/29/2024)

Method	Rank1	Rank2	Rank3	Rank4	Rank5
xT	Wirtz 130M	Frimpong 50M	Grimaldo 45M	Adli 30M	Tella 23M
VAEP	Wirtz 130M	Tella 23M	Stanisic 28M	Andrich 17M	Xhaka 20M
Soccer-Transformer	Wirtz 130M	Boniface 40M	Schick 22M	Grimaldo 45M	Stanisic 28M

5 Conclusion

In this paper, we introduced the SoccerTransformer framework specifically tailored to analyze complex game dynamics. This transformer-based framework shows value through its four analytic phases: (1) Self-supervised pre-training phase learns in-game dynamics and adequately assesses the roles of players. (2) Downstream prediction of attack phase outcomes can be useful for in-game analytics and strategic planning. (3) The action value assessment is useful for analysts to identify key contributors in match scenarios. (4) Finally, the proposed player rating system shows a high correlation with market value and with seasonal changes in market value facilitating talent hunting and recruitment. Our results are consistent with related work while outperforming state-of-the-art rating systems (xT and VAEP) in terms of correlation to market value. This article demonstrates the benefits of transformers for analysis of on-the-ball-events data in soccer while highlighting the large space for future improvement.

This work can be extended to examine the impact of MLM pre-training on downstream tasks. In addition, SoccerTransformer can be useful for exploring the similarities in playing styles between players and teams based on learned embeddings and enhancing pre- and post-match tactical analyses via learned attacking goal-scoring predictors. Future investigations may incorporate player performance metrics with additional features such as age, national team, league, and contract length to improve PMV predictions.

References

1. Analyst, O.: What Are Expected Assists (xA)? (2021). https://theanalyst.com/eu/2021/03/what-are-expected-assists-xa/
2. Chen, S., Liao, H.: BERT-Log: anomaly detection for system logs based on pre-trained language model. Appl. Artif. Intell. **36**(1), 2145642 (2022)
3. CIES: Financial analysis of big-5 league clubs' transfers (2022). https://football-observatory.com/IMG/sites/mr/mr77/en/
4. Dang, W., Zhou, B., Wei, L., Zhang, W., Yang, Z., Hu, S.: TS-BERT: time series anomaly detection via pre-training model BERT. In: Paszynski, M., Kranzlmüller, D., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds.) ICCS 2021. LNCS, vol. 12743, pp. 209–223. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77964-1_17
5. Decroos, T., Bransen, L., Van Haaren, J., Davis, J.: Actions speak louder than goals: valuing player actions in soccer. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1851–1861 (2019)
6. Decroos, T., Davis, J.: Player vectors: characterizing Soccer players' playing style from match event streams. In: Brefeld, U., Fromont, E., Hotho, A., Knobbe, A., Maathuis, M., Robardet, C. (eds.) ECML PKDD 2019. LNCS (LNAI), vol. 11908, pp. 569–584. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46133-1_34
7. Decroos, T., Dzyuba, V., Van Haaren, J., Davis, J.: Predicting Soccer highlights from spatio-temporal match event streams. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (2017)
8. Decroos, T., Van Haaren, J., Davis, J.: Automatic discovery of tactics in spatio-temporal soccer match data. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 223–232 (2018)
9. Decroos, T., Van Roy, M., Davis, J.: SoccerMix: representing soccer actions with mixture models. In: Dong, Y., Ifrim, G., Mladenić, D., Saunders, C., Van Hoecke, S. (eds.) ECML PKDD 2020. LNCS (LNAI), vol. 12461, pp. 459–474. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67670-4_28
10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
11. Fernández, J., Bornn, L., Cervone, D.: Decomposing the immeasurable sport: a deep learning expected possession value framework for soccer. In: 13th MIT Sloan Sports Analytics Conference (2019)
12. Gyarmati, L., Hefeeda, M.: Analyzing in-game movements of soccer players at scale. arXiv preprint arXiv:1603.05583 (2016)
13. Kokalj, E., Škrlj, B., Lavrač, N., Pollak, S., Robnik-Šikonja, M.: BERT meets Shapley: extending SHAP explanations to transformer-based classifiers. In: Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation, pp. 16–21 (2021)
14. Konefa, M., Chmura, P., Zaj, T., Chmura, J., Kowalczuk, E., Andrzejewski, M.: A new approach to the analysis of pitch-positions in professional soccer. J. Hum. Kinet. **66**(1), 143–153 (2019)
15. Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., Giannotti, F.: PlayeRank: data-driven performance evaluation and player ranking in Soccer via a machine learning approach. ACM Trans. Intell. Syst. Technol. (TIST) **10**(5), 1–27 (2019)

16. Pappalardo, L., et al.: A public data set of spatio-temporal match events in Soccer competitions. Sci. Data **6**(1), 1–15 (2019)
17. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al.: Improving language understanding by generative pre-training (2018)
18. Rudolph, Y., Brefeld, U.: Masked autoencoder pretraining for event classification in elite soccer. In: International Workshop on Machine Learning and Data Mining for Sports Analytics, pp. 24–35. Springer (2023). https://doi.org/10.1007/978-3-031-53833-9_3
19. Singh, K.: Introducing expected threat (xt). https://karun.in/blog/expected-threat.html
20. Spearman, W.: Beyond expected goals. In: Proceedings of the 12th MIT Sloan Sports Analytics Conference, pp. 1–17 (2018)
21. Sports, S.: UEFA's new financial sustainability regulations to replace FFP: all you need to know (2022). https://www.skysports.com/football/news/11095/12584543/uefas-new-financial-sustainability-regulations-to-replace-ffp-all-you-need-to-know
22. statsBomb: Statsbomb open data (2015-2024). https://statsbomb.com/what-we-do/hub/free-data/
23. Transfermarkt: Transfermarkt. https://www.transfermarkt.com/
24. Van Roy, M., Robberechts, P., Decroos, T., Davis, J.: Valuing on-the-ball actions in soccer: a critical comparison of XT and VAEP. In: Proceedings of the 2020 AAAI Workshop on AI in Team Sports, pp. 1–8 (2020)

Other Team Sports/e-Sports

Automated Detection of Shot Events in Game Phases Using GNSS Data from a Single Team

Dermot Sheridan[1], Valerio Antonini[1(✉)], Michael Scriney[1,2], and Mark Roantree[1,2]

[1] School of Computing, Dublin City University, Dublin 9, Dublin D09 V209, Ireland
{dermot.sheridan36,valerio.antonini3}@mail.dcu.ie,
{michael.scriney,mark.roantree}@dcu.ie
[2] Insight SFI Research Centre for Data Analytics, School of Computing, Dublin City University, Dublin 9, Dublin, Ireland

Abstract. Recent advancements in wearable Global Navigation Satellite Systems (GNSS) devices and optical tracking technologies have significantly expanded the availability of spatiotemporal data in sports analytics. Combined with the common practice of manually recording game event logs, these sources provide unprecedented levels of data that describe in detail what happens during a game. However, the manual collection and tagging of events within a game can be costly and time-consuming, requiring domain experts to repeatedly view and annotate footage to identify discrete events. This study presents a novel method for automatically detecting shooting events using spatiotemporal metrics using GNSS tracking data from a single team to achieve event identification.

Keywords: Spatiotemporal Data · GNSS Tracking · Shooting Events Classification · Sports Analytics

1 Introduction

In recent years, the availability of tracking data has significantly increased due to the use of wearable GNSS devices and advancements in optical object tracking technology [1]. The unprecedented increase in data availability has significantly enhanced data analytics in sports, enabling machine learning (ML) algorithms to evaluate player and team performance more effectively, providing deeper insights and improving overall performance [2]. This increase availability of data has also led to a rise in spatial sports research within invasion team sports. In this domain, researchers commonly utilise two types of spatiotemporal data: player tracking data and event log data [3]. A recent framework was outlined to help categorise these diverse research methods and models used in team sports analysis. One category highlighted is data mining, which involves using representations and

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 95–105, 2025.
https://doi.org/10.1007/978-3-031-86692-0_8

structures for more advanced analysis. The data collected can serve as inputs for sophisticated algorithms and probabilistic methods to predict and automatically label events more quickly [4].

Gaelic Football, the most popular Irish national sport, has been described as a hybrid of soccer and Australian Rules and is played with a round ball on a rectangular pitch [5]. Elite inter-county competition typically begins in January with the National League and concludes in September after the All-Ireland Championship [6]. The sport maintains an amateur status; however, the training regimes of elite players mirror those of professional athletes. A typical week for an elite player includes one to two resistance training sessions, two field sessions, and a match or training at the weekend [7]. Performance analysis has shown that winning teams in elite Gaelic football often demonstrate superior offensive capabilities, including higher total scores, shot efficiency, and counterattacking ability ([6,8]). Strong defensive performance, particularly in generating turnovers, is crucial for success [6].

The intense competition in sports has led to the recruitment of dedicated sports scientists within team staff to optimise player and team performance throughout the season [9]. This professional approach has resulted in a significant increase in the use of player-tracking technologies at club and elite inter-county levels during both games and training. These technologies have revolutionised training methodologies, providing valuable insights that enhance overall performance [10].

The extensive data generated from GNSS tracking devices offers a rich source of information that details player activity profiles during games. However, detecting specific events still necessitates extensive manual work by human annotators. The potential exists for this data to aid in the automatic tagging of events, streamlining the process and reducing the reliance on manual annotation [8].

Contribution. This paper introduces a novel approach for automatically detecting shot events during phase of game using spatiotemporal data from a single team's GNSS tracking data. Our method effectively identifies shot events even with data from just a few games.

Paper Structure. The remainder of this paper is structured as follows: in §2, we discuss related research in the evolution of spatiotemporal data mining in the area of sports analytics; in §4, we present a 3-step methodology to process event logs and GNSS data to extract game phases and features based on distance and similarity of speed between players; in §5, we present the results of our experiments, while in §6, we discuss our results and implications; and finally in §6, we present our conclusions along with future directions.

2 Related Research

Recent advancements in sports analytics have leveraged spatiotemporal data for analysing player and team performances. Technologies like optical tracking and GNSS are crucial in capturing detailed movement data in team sports such as soccer and Gaelic Football ([11,12]). Recent studies have focused on automating game event tagging, providing valuable feedback to players and coaches [13].

The shift from manual event tagging to ML has utilised rule-based systems and advanced neural networks. One study demonstrated the use of algorithms like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forests to detect football events from positional data [14]. Another study used a data-driven approach to identify offensive tactics, highlighting the importance of algorithm choice for strategic insights [15]. A rule-based algorithm was proposed for identifying basic soccer events, achieving high accuracy but noting limitations in capturing complex game dynamics [16]. Further advancement was made with a deterministic decision tree-based algorithm, achieving over 90% detection accuracy but relying on detailed data, which may not always be available [17].

GNSS data and heatmaps were employed in the 6MapNet model to identify players based on movement styles, demonstrating GNSS data's potential for detailed analysis [18]. However, a gap remains in using GNSS data from a single team for event labeling. Our research addresses this by proposing a novel method for automatically labeling shot events using players' spatiotemporal data. This approach is particularly relevant for Gaelic football, where data collection challenges exist due to the sport's amateur status. We aim to analyse player movement during the phase of play and identify a shot event.

This novel method not only fills a significant gap in the literature but also provides a framework for applying advanced ML techniques to GNSS data, offering practical solutions for amateur sports teams to gain deeper insights into performance and strategies without extensive resources.

3 Data

This study utilises two distinct datasets: GNSS data and event log data. The GNSS data were collected on the same team during 11 competitive Gaelic Football inter-county games across the 2019–2021 seasons. Athletes of the analyzed team wore micro 10 Hz GNSS sensor devices (STATSports Apex 10 Hz, Northern Ireland, UK), which recorded ten observations per second for each of the following variables: latitude, longitude, speed, 3D acceleration, and 3D gyroscope (which provide information about the angular velocity). Data were filtered and pre-processed using STATSports software (version 4.5.19). Previous research [19] has demonstrated that the error margin of the 10 Hz Apex unit, which is between 1−2% of the measured distances, is negligible.

Event logs are generated from video analysis, and teams use these computerised notational analysis statistics to enhance their performance and seek to analyse their own and their opponents' gameplay systematically [6]. Event logs are not continuous; they are only recorded when an event occurs. game footage from internal team video recordings and external media broadcasters was imported and coded using a custom-built tagging panel. All games were coded using Nacsport Scout Plus software (Scout Plus V 6.5.0, New Assistant for Coach Sport, S.L). During an analysis of a game, the coder had the ability to pause and rewind the video to watch events more than once; it was also possible to use a slow-motion feature to allow for a clearer view of fast dynamic

events. Following data validation, the coding events were then exported into Microsoft Excel (Microsoft, USA) to facilitate analysis.

4 Methodology

This paper focuses on classifying game phases using features extracted from GNSS data of one of the two playing teams. Initially, we utilise event log data to isolate game phases, which we define as intervals starting when a team gains possession of the ball and ending when the same team loses possession or attempts a shot. We conduct two distinct experiments. In the first experiment, we identify game phases that conclude with a shot by the analysed team (Team A) using GNSS data from that team. In the second experiment, we identify game phases that end with a shot by the opposing team (Team B) using GNSS data from Team A. The objective of this study is to explore the feasibility of classifying different game phases based on GNSS data from one team, without relying on event log data or data from the opposing team.

The methodology for this study follows a sequence of three steps:

1. Phases Extraction from Event Logs Data. The first step involves extracting game phases from the event log data. The event log data describes a series of sequential events occurring during the game, such as 'Throw in', 'Attack Team B', 'TB Shot', 'TA Kickout', and 'TA Possession from KO', along with their corresponding timestamps. We isolate phases of ball possession for the two teams, which start when one of the two teams gains the ball (it happens after a Throw in, Kickout, Turnover, etc.) and end when the team loses control of the ball or makes a shot. We record the moment when the ball is gained (start time) and when the ball is lost or a shot is made (end time). This step takes as input a data frame of sequential events and transforms it into a data frame of game phases describing which team is in possession of the ball, along with the corresponding timestamp. Each phase is labeled based on the target of the analysis. When classifying shots during the attacking phase, phases ending with a "TA Shot" are labeled as 1, while all other phases are labeled as 0. On the contrary, when classifying shots performed by the opponent team, phases ending with a "TB Shot" are labeled as 1. The resulting data frame for the game's phases will display the corresponding events' start and end times. This process is repeated for each of the analysed games. A sample of the resulting data frame of the phases of game is shown in Table 1.

2. Features Extraction from GNSS Data. After identifying the game phases and their respective starting and ending timestamp, we created a data frame containing features extracted from GNSS data. The 10 Hz GNSS device collects data at 10 observations per second, including latitude, longitude, speed, 3D acceleration, and 3D gyroscope. Given a game phase with a defined starting and ending timestamp, we computed metrics such as the average and standard deviation of distances, and the average and standard deviation of the absolute differences in speed, 3D acceleration, and 3D gyroscope between each pair of players in that

Table 1. Sample data frame illustrating the phases of the game. Columns include *gameID* (ID of the corresponding game), *Team* (team in possession of the ball), *Start Event* (event initiating the phase), *End Event* (event concluding the phase), *Start Time* (game time in seconds at the beginning of the phase), and *End Time* (game time in seconds at the end of the phase).

GameID	Team	Start Event	End Event	Start Time	End Time
1	A	Team A Pos frm TO	TA Shot(P)	5	62
1	B	TB Kickout	TB Kickout	63	71
1	A	TA Pos frm KO	TA Shot(P)	72	98
1	B	Team B Pos frm TO	Attack TB	99	149

interval of time. This process is repeated for each phase of the game. Metrics were collected for every player on the team to ensure consistency, with non-participating players' values recorded as 0. These metrics indicate team movement and coordination during the phase. Each metric corresponds to a column in the final data frame, resulting in columns for the average and standard deviation of distance, speed, 3D acceleration, and 3D gyroscope between each pair of players. This step is summarised in Fig. 1. The input GNSS data is transformed into a data frame of similarity scores (average and standard deviation) for each variable and for each couple of players Table 2. The resulting data frame con-

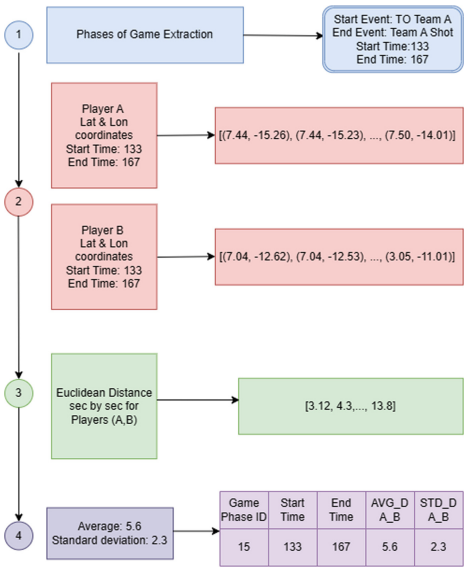


Fig. 1. Computation of average and standard deviation of distance between a pair of players for a selected game phase. This process is repeated for each pair of players and for each extracted game phase.

tains 441 rows (one for each game phase) and 13,921 columns. Due to varying player participation, some features may be sparse, necessitating a feature selection step to eliminate noise and identify the most relevant features for each ML model tested.

Table 2. Resulting data frame after feature extraction from GNSS data. Each row represents a game phase, with columns for game ID, start/end time, distances, speed, and acceleration (x) metrics between player pairs (only shown between player 1 and player 2). Other features that are not shown include additional acceleration (y and z) and gyroscope (x, y, and z) differences for each player pair.

EventID	Players Pair	Avg Dist	Std Dist	AvG D Sp	Std D Sp	Avg D ACx	Std D ACx	Avg D ACy	Std D ACy	Avg D ACz	Std D ACz
1	(1,2)	19.3	2.3	2.9	2.3	1.1	0.9	0.8	0.4	1.3	0.3
1	(1,3)	32.1	11.3	3.1	1.9	0.5	1.3	0.6	0.3	1.1	0.2
1	(1,4)	5.7	5.1	1.3	0.5	0.4	0.1	1.4	0.6	1.0	0.1
1	(1,5)	8.9	2.3	4.4	1.7	0.6	0.4	1.7	0.8	0.9	0.3
1	(1,6)	22.1	6.3	3.3	1.6	1.1	0.3	2.1	0.9	0.8	0.3
1	(1,7)	31.1	8.1	2.9	1.4	0.9	0.5	2.0	0.8	0.8	0.3

3. Feature Selection and Modelling. Our approach utilises an automated feature selection method to identify the most significant features for predictive modeling from a data frame containing thousands of features. We use the Recursive Feature Elimination with Cross-Validation (RFECV) method, which iteratively removes the least important features and evaluates model performance using k-fold (k = 3) cross-validation [20]. K-fold cross-validation is a method used to assess the performance of a ML model by dividing the data into k equal subsets. The model is trained on k-1 subsets and validated on the remaining subset, and this process is repeated k times, with each subset used exactly once as the validation set. The RFECV method reduces the risk of overfitting and provides a reliable estimate of the model's generalisability. RFECV involves training a model on the full set of features, ranking them based on importance scores, removing the least important feature, and evaluating the model. This cycle repeats until only one feature remains, and the set of features achieving the best cross-validation performance is selected. The outcome of the recursive feature elimination process is a refined and optimised set of features tailored by each ML model during the training phase. In the testing phase, each model is evaluated using this optimised feature set to assess its performance. The motivation for using RFECV in our study is supported by several key factors and references. Firstly, RFECV is known for its ability to effectively handle high-dimensional datasets by identifying the most relevant features, thereby improving model performance and interpretability [21]. RFECV method helps in mitigating the risk of overfitting by incorporating cross-validation, which ensures that the selected features generalize well to unseen data [22]. Lastly, RFECV has been widely used in various domains, including malicious intrusion detection in computer networks [23], diagnosis of Alzheimer's disease [24], and fraud detection [25]; demonstrating its versatility and robustness in feature selection. By

using RFECV, we aim to leverage its strengths to enhance the predictive power of our models, ensuring that the features selected contribute significantly to the performance and reliability of the predictions.

5 Results

In this section, we present the performance evaluation of ML models on two distinct classification tasks: identifying game phases culminating in a shooting event for Team A (Task A) and for the opposing Team B (Task B). Due to the relative rarity of shooting events compared to other events, both datasets are inherently imbalanced. To address this issue, we employed undersampling of the majority class in both tasks, training the models on datasets balanced with an equal number of shooting and non-shooting events.

This paper aims to demonstrate the effectiveness of ML models in recognising partial game phases, potentially leading to the future automation of event labeling based solely on spatiotemporal data from a single team. Both tasks utilise GNSS data collected on the same team during the first halves of 11 GF games across the years 2019, 2020, and 2021. The analysis includes data from all players participating in these games. The ML models' performances are evaluated using the following metrics [26]:

- **Accuracy** measures the proportion of correct predictions (both true positives TP and true negatives TN) out of the total number of predictions, which includes true positives, true negatives, false positives (FP), and false negatives (FN).
- **Precision** is the proportion of true positive results out of all positive predictions made by the model.
- **Recall** (or True Positive Rate) is the proportion of true positive results out of all actual positive cases.
- **F1 Score** is the harmonic mean of precision and recall. It balances the two metrics and provides a single measure of a model's performance, especially useful when the class distribution is imbalanced.
- **AUC** (Area Under the ROC Curve) measures the ability of the classifier to distinguish between positive and negative classes. It is a single scalar value that summarizes the performance of a classifier across all possible classification thresholds.

5.1 Classifying Shooting Events for Team A

In the first experiment, we evaluated the performance of four ML models in classifying shooting events performed by Team A: Random Forest, Decision Tree, XGBoost, and Logistic Regression. The results for each model are presented in Table 3. The Random Forest model demonstrated the highest performance, with an F1 score of 0.69. Decision Tree and Logistic Regression models showed lower

performance. Achieving a high F1 score indicates that the model not only accurately predicts shooting events (high precision) but also effectively identifies most of the actual shooting events (high recall). Random Forest, which selected 11,136 features, and Decision Tree, which selected 13,786 features, highlight the benefits of feature-rich models. However, the XGBoost model achieved similar performance with a significantly smaller set of features. This indicates that while a high number of features can enhance performance, the efficiency and power of the algorithm are also crucial. These findings highlight the importance of balancing feature selection with model complexity to achieve optimal performance.

Table 3. Performance metrics of ML models in classifying Team A shooting events. Precision and recall are weighted by the number of true instances per class and averaged.

Model	Accuracy	F1 Score	AUC	Precision	Recall	Selected Features
Logistic Regression	0.52	0.52	0.53	0.52	0.52	4
Decision Tree	0.52	0.54	0.52	0.52	0.53	13786
Random Forest	**0.63**	**0.69**	0.61	**0.65**	**0.63**	11136
XGBoost	0.61	0.67	**0.67**	0.62	0.61	3

5.2 Classifying Shooting Events for Opposing Team (Team B)

Table 4 displays the performances of the ML models in classifying shooting events for the opposing team (Team B). Recall, as this is the opposing team, these data are based on Team A' spatiotemporal data. In this evaluation, XGBoost outperformed the other models across all metrics (sharing the highest accuracy with Random Forest) while using a smaller set of features. An AUC of 0.78 indicates that there is a 78% chance that the model will correctly distinguish between a randomly chosen positive instance (shooting event) and a randomly chosen negative instance (non-shooting event). In practical terms, this level of performance is useful for sports analysts because it provides a reliable way to identify shooting events from the available GNSS data.

6 Discussion

Four ML models were developed to predict shooting events using pair-wise similarity metrics of players' movements. The Random Forest model performed best for predicting Team A shooting events (Task A), while XGBoost perfomed best for the opponent team's events (Task B). In Task A, the Random Forest model achieved a higher F1 score and recall, effectively identifying true positives (actual Team A shooting events) but also producing many false positives, indicating a tendency to overestimate shots (Fig. 2 Left). In contrast, the XGBoost model in

Table 4. Performance metrics of ML models in predicting opponent (Team B) shooting events.

Model	Accuracy	F1 Score	AUC	Precision	Recall	Selected Features
Logistic Regression	0.67	0.68	0.70	0.67	0.67	1695
Decision Tree	0.47	0.45	0.47	0.47	0.47	8294
Random Forest	**0.69**	**0.68**	0.73	0.68	0.68	11325
XGBoost	**0.69**	0.67	**0.78**	**0.69**	**0.69**	15

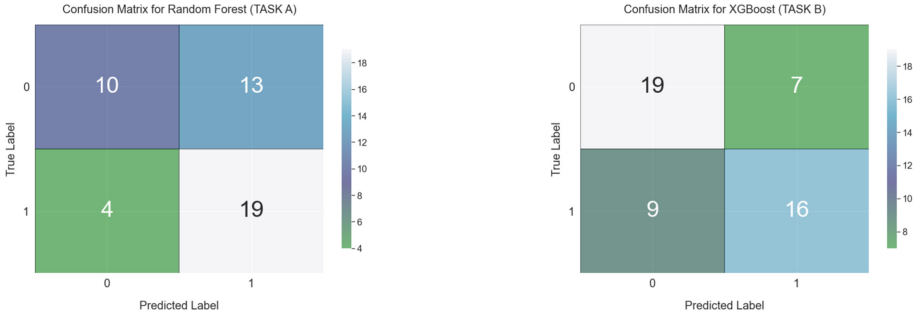


Fig. 2. Confusion matrix for Random Forest predicting Team A (Left) and Team B (Right) shooting events.

Task B showed higher accuracy and AUC score, demonstrating superior overall performance and a better ability to distinguish between classes (Fig. 2 Right). XGBoost achieved better precision with fewer false positives, though it had more false negatives, reflecting a more balanced performance but missing more actual Team B shots (Fig. 2 Right).

Classyfing game phases events from GNSS data can offer valuable insights into how players' spatiotemporal data influences game phases and leads to specific events. Currently, these event labels are manually assigned by watching game videos, a time-consuming and labour-intensive process. By developing ML algorithms that automatically assign labels to events, we can significantly speed up this process and improve efficiency, enabling more timely and accurate analysis of player movements and game dynamics [4].

7 Conclusion

This study represents the first attempt to identify shooting events in Gaelic Football using player tracking data of a single team. We leverage spatiotemporal data (speed, distance, 3D acceleration, and 3D gyroscope) to compute similarity metrics between players during play phases. Our experimental findings demonstrate that integrating similarity scores, automatic feature selection, and ML models enables the identification of phases of play that result in shot events using GNSS

tracking data. Future research should explore the use of additional features to identify the movements that positively impact the outcome of a phase of game.

Acknowledgments. This work was funded by Science Foundation Ireland through the Centre for Research Training in Artificial Intelligence (SFI/18/CRT/6223), Machine Learning (SFI/18/CRT/6183) and Insight Centre for Data Analytics (SFI/12/RC/2289_P2) .

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Lutz, J., Memmert, D., Raabe, D., Dornberger, R., Donath, L.: Wearables for integrative performance and tactic analyses: opportunities, challenges, and future directions. Int. J. Environ. Res. Public Health **17**(1), 59 (2020)
2. Goes, F.R., et al.: Unlocking the potential of big data to support tactical performance analysis in professional soccer: a systematic review. Eur. J. Sport Sci. **21**(4), 481–496 (2021)
3. Memmert, D., Rein, R.: Game Analysis, Big Data and Tactics: current trends in elite soccer. German J. Sports Med. Deut. Z. fur Sportmed. **69**(3) (2018)
4. Gudmundsson, J., Horton, M.: Spatio-temporal analysis of team sports. ACM Comput. Surv. (CSUR) **50**(2), 1–34 (2017)
5. Cullen, B.D., et al.: Fitness profiling of elite level adolescent Gaelic football players. J. Strength Conditioning Res. **27**(8), 2096–2103 (2013)
6. Gamble, D., Bradley, J., McCarren, A., Moyna, N.M.: Team performance indicators which differentiate between winning and losing in elite Gaelic football. Int. J. Perform. Anal. Sport **19**(4), 478–490 (2019)
7. Beasley, K.J.: Nutrition and Gaelic football: review, recommendations, and future considerations. Int. J. Sport Nutr. Exerc. Metab. **25**(1), 1–13 (2015)
8. Carroll, R.: Team performance indicators in Gaelic football and opposition effects. Int. J. Perform. Anal. Sport **13**(3), 703–715 (2013)
9. Malone, S., Collins, K., McRobert, A., Doran, D.: Quantifying the training and match-play external and internal load of elite Gaelic football players. Appl. Sci. **11**(4), 1756 (2021)
10. Mangan, s, et al.: The relationship between technical performance indicators and running performance in elite Gaelic football. Int. J. Perform. Anal. Sport **17**(5), 706–720 (2017)
11. Beernaerts, J., De Baets, B., Lenoir, M., Van de Weghe, N.: Spatial movement pattern recognition in soccer based on relative player movements. PLoS ONE **15**(1), e0227746 (2020)
12. Sheridan, D., Brady, A.J., Nie, D., Roantree, M.: Predictive analysis of ratings of perceived exertion in elite Gaelic football. Biol. Sport **41**(4), 61–68 (2024)
13. Salim, F.A., Postma, D.B., Haider, F., Luz, S., Beijnum, B.J.F.V., Reidsma, D.: Enhancing volleyball training: empowering athletes and coaches through advanced sensing and analysis. Front. Sports Act. Living **6**, 1326807 (2024)
14. Richly, K., Bothe, M., Rohloff, T., Schwarz, C.: Recognizing compound events in spatio-temporal football data. In: International Conference on Internet of Things and Big Data, vol. 2, pp. 27–35 (2016)

15. Decroos, T., Van Haaren, J., Davis, J.: Automatic discovery of tactics in spatio-temporal soccer match data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 223-232 (2018)
16. Khaustov, V., Mozgovoy, M.: Recognizing events in spatiotemporal soccer data. Appl. Sci. **10**(22), 8046 (2020)
17. Vidal-Codina, F., Evans, N., El Fakir, B., Billingham, J.: Automatic event detection in football using tracking data. Sports Eng. **25**(1), 18 (2022)
18. Kim, H., Kim, J., Chung, D., Lee, J., Yoon, J., Ko, S.K.: 6MapNet: representing soccer players from tracking data by a triplet network. In: International Workshop on Machine Learning and Data Mining for Sports Analytics, pp. 3–14 (2021)
19. Beato, M., Coratella, G., Stiff, A., Iacono, A.D.: The validity and between-unit variability of GNSS units (STATSports Apex 10 and 18 Hz) for measuring distance and peak speed in team sports. Front. Physiol. **9**, 1288 (2018)
20. Misra, P., Yadav, A.S.: Improving the classification accuracy using recursive feature elimination with cross-validation. Int. J. Emerg. Technol **11**(3), 659–665 (2020)
21. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Mach. Learn. **46**(1), 389–422 (2002)
22. Priyatno, A.M., Widiyaningtyas, T.: A systematic literature review: recursive feature elimination algorithms. JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer) **9**(2), 196–207 (2024)
23. Ustebay, S., Turgut, Z., Aydin, M.A.: Intrusion detection system with recursive feature elimination by using random forest and deep learning classifier. In: 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), pp. 71–76 (2018). IEEE
24. Richhariya, B., Tanveer, M., Rashid, A.H., Initiative, A.D.N.: Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). Biomed. Signal Process. Control **59**, 101903 (2020)
25. Mustaqim, A.Z., Adi, S., Pristyanto, Y., Astuti, Y.: The effect of recursive feature elimination with cross-validation (RFECV) feature selection algorithm toward classifier performance on credit card fraud detection. In: 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), pp. 270–275 (2021). IEEE
26. Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. Int. J .Data Min. Knowl. Manage. Process **5**(2), 1 (2015)

Team Dynamics in DotA2 Through Attention Mechanism

Alexis Mortelier(✉), Sébastien Bougleux, and François Rioult

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
alexis.mortelier@unicaen.fr

Abstract. We analyse team dynamics in the popular e-sports game DotA2 using an approach that combines convolutional and LSTM networks with a *feature-temporal* attention mechanism. Our goal is to identify strategic behaviours that lead to successful goals, such as scoring kills, during World Championship matches. Each team's formation is represented by a polygon, which feeds an RNN that learns *kill* events from this polygon under its area, diameter, and moments around the centroid. By exploiting the attention mechanism, our network highlights the most relevant features at each time step, providing insights into strategic team movements and formations. Our results demonstrate the effectiveness of our approach in capturing critical dynamics that influence the outcome of engagements in DotA2.

Keywords: DotA2 · e-sport · deep learning · CNN-LSTM · attention · multivariate time series

1 Introduction

The e-sport phenomenon has emerged as a new era of sport, highlighting the opportunities it offers for sports scientists [13]. The way in which traditional sports and e-sport interact, particularly with regard to audiences, sponsors, revenues, and other economic and socio-cultural aspects, implies that e-sport fits into the wider landscape of professional sport and can benefit from the experience and infrastructure of traditional sports [5].

We are here mostly interested in analysing collective behavior by the way of a team of players. Data analysis can indeed be used to study the performance of the team as a whole, as well as that of its opponents. Analyzing historical match data can provide valuable insights into game trends, player and team performance, strategies used, and so on.

In DotA2, the *events* that punctuate the match occur when the heroes embodied by the players are temporarily neutralized, referred to as death then rebirth or *lifeline*. These events determine the score, which can be analyzed using machine learning techniques to determine the reasons behind the occurrence of an event, in terms, for example, of the geometric configuration of the set of

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, pp. 106–118, 2025.
https://doi.org/10.1007/978-3-031-86692-0_9

teammates. This involves analyzing the geometric configurations that a network detects as emblematic of an event occurrence.

Attention mechanisms in recurrent networks are recognized for their ability to enrich the contextualization of a decoder, by learning a complementary path of temporal dependence. Our aim is to show that the analysis of attention weights offers a strategic perspective for the study of DotA2 matches.

The interpretation of the attention mechanism to obtain explanations for the decisions made by the network is a matter of debate. The interested reader may refer to [3], which livers a broad synthesis on the subject. Thus, attention would not provide an explanation and the use of Total Variation Distance (TVD), Jensen-Shannon Divergence (JSD) or Kullback Leibler divergence should be preferred [15] as attention is not correlated with more traditional gradient-based measures of attribute importance [33]. Experiments show that mixing attention weights does not affect the final decision. This statement is qualified by [35], which indicates that attention provides *one* explanation, but not *the only one*. The use of attention as a flux could also lead to the emergence of explanations [1].

Here, we are content to draw out *interpretations* of the weights of the attention layer of an RNN (Reccurent Neural Network), trained to predict the occurrence of a death event from the geometric features of the players' polygon. After a quick presentation of DotA2 in Sect. 2, Sect. 3 relates the state of the art, Sect. 4 describes our process. The article concludes with the presentation of experiments in Sect. 5.

2 What Is DotA2?

DotA2, or Defense of the Ancients 2, is a MOBA (Multiplayer Online Battle Arena) video game developed by Valve Corporation. Two teams of five players compete to destroy the opposing team's Ancient, located in their base. Figure 1 shows a representation of the map of DotA2. Each player chooses a hero from a vast selection. Each hero has unique skills and different roles (such as support, carry, tank, etc.). The map is symmetrical, with three main lines (top, mid, bottom) bordered by jungles and a river linking the elders of each team. The heroes compete in three different lines, based on the principle of invasion sports like rugby and American football.

Team coordination and strategy are crucial : players must plan their attacks, defend their towers, and collectively take on objectives. DotA2 is very popular in esports competitions, with tournaments like The International[1] offering millions of dollars in prizes. DotA2 combines real-time action, strategy and role-playing, requiring both individual skills and strong team collaboration to succeed.

3 Related Work

The literature survey shows numerous contributions to sports analysis using RNNs with attention mechanisms, possibly on geometric features.

[1] see https://www.dota2.com/esports/ for this year's edition.

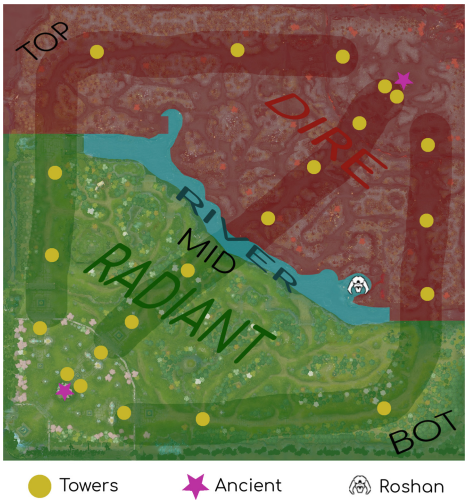


Fig. 1. DotA2 map.

The popularity of video games in general and e-sports in particular has led to a number of articles studying the mechanics of these games, using a data science approach. MOBA games are very relevant for comparative studies because they offer a complex but relatively formatted game, both in terms of the playground and the way the games are played. Themes range from pick recommandation [28, 32], studies of individual behaviour [4,26] to game summaries [2,23], outcome prediction [6], the use of geometric features and RNN [20].

The chronological nature of a video game match and the popularity of neural network methods are leading to a significant use of recurrent neural network techniques, for example to predict the outcome of the match [16,22,27]. This type of network can help to understand the strategic preparation phases of the game (pick/ban of heroes) [37]. As the game progresses, an important task is the recommendation of equipment items, supported by an RNN in [36] that models player interest from historical match sequences. A RNN can also predict a player's performance based on previous performances [34].

The theme of geometric features on trajectories is not new. The analysis of a collective practice on a given field naturally lends itself to studies of the trajectories and geometry of player organizations. Thus, for football, a work similar to the one presented in this article is proposed in [10], with the aim of predicting the outcome of the match. For DotA2, the geometric features allow to correctly predict the outcome of a match and this relevance is confirmed on 6 billion matches of League Of Legends (another instance of MOBA) via a RandomForest system with an F-measure of 92%. Regularities in the trajectories, discriminating in winning of the match, could also be extracted [6]. Behavior on a MOBA court is indeed crucial and differences between levels of play are revealed by positional

features: change of zone, distance between players, which allows a grouping of matches according to the similarity of movements [9].

Using RNNs carefully to learn models from trajectories is a natural task for time series data from sporting activities. In particular, in the sports domain, there are many approaches proposing models to predict or complete trajectories [11,19,31,39], model sequences to predict the outcome of an event (e.g. a basketball shot [25]) or the outcome of a match [38]. RNNs are also used to analyze complex spatio-temporal dependencies in order to recognize individual events or detect group activities [12,18,24].

Attention mechanism in RNN is useful for capturing temporal dependencies. So it's not surprising to find many uses for attention in sports data analysis, for example for movement prediction [29] or detecting similarity between movements [17].

Attention contributes to the design of the network to be learned, a classic task in machine learning. In this article, however, we interpret attention to analyze sports situations. This approach, also found in [40] to identify relevant portions of trajectories or understand differences between classes and determine correlations between variables and labels, is unusual in the literature.

4 Method

We start by explaining how the spatial configuration of the players is considered by the neural network. Its architecture is then described, with a focus on the attention layer and its analysis for tactical purposes.

4.1 Polygon Features

DotA2 provides a record of game activity in the form of a *replay* file, which can be retrieved after each game. Hero movement data is then available, and death events for each hero are recorded. For the 258 matches of one of the game's major tournaments, The International 2018, we therefore have the temporal evolution of players' positions and their lifelines.

One objective would be to train a network to predict the lifelines of all the players based on their movements, on the assumption that this network will carry knowledge about the relationships between field occupation and lifelines. A single network would be common to all matches and would require comparable data. If fed with all ten trajectories, it would treat the first player's trajectory identically across matches. However, the first player varies in playing style, hero, and team composition from match to match. The order of players also depends on their position on the pitch - in DotA2, the *Radiant* team is on the lower half, including the first five players, while the *Dire* team is on the upper half, without exchanging positions at half-time.

In order to feed the network with comparable feature from match to match, we therefore consider, for each team and each time step in the sequence, the features of the polygon defined by the convex hull of all the players in a team,

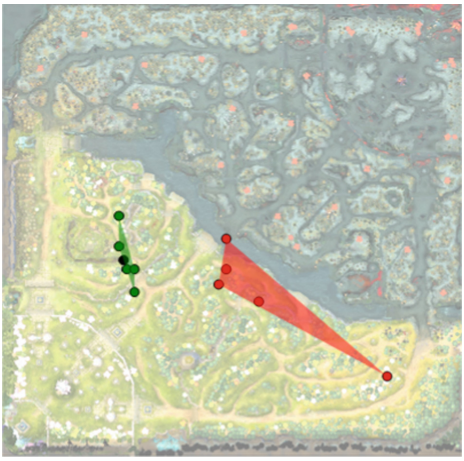


Fig. 2. Polygon visualization (one for each team).

as illustrated in Fig. 2, rather than individual trajectories. For each polygon, we calculated :

- *area:* the polygon surface;
- *diameter:* the maximum distance between two players;
- *grouping:* average of distances to the barycenter;
- *inertia:* second-order moment of the distances to the barycenter.

These simple measurements are a fine reflection of the spatial configuration of the players and their ability to carry out actions in a grouping configuration. For example, a high *diameter* indicates high dispersion (at least one player is far from the others), while *grouping* capacity is a clue of the average distance from the barycenter. *Inertia* is relative to the amplitude of variation in *grouping* capacity. We have already shown that these geometric features are sufficient to estimate the complexity of this game [21]: individual trajectories do not need to be observed.

4.2 Network Architecture

Figure 3 shows a representation of the neural network architecture. Given a match, the input consists of a sequence of t values per polygonal feature for one team, represented by a matrix of size $t \times f$, with f the number of input features (4 in our case). The objective is to predict whether all the players on the opposing team are alive (*no kill*) or not (*kill*). This is a binary classification problem. In order to decide, the network first extract spatio-temporal information from the input features by a 1D depthwise separable convolution layer [7] followed by two LSTM layers [14], reducing the input sequence of $f = 4$ features to an hidden sequence of 1 feature per time step. Then, $f = 4$ independent

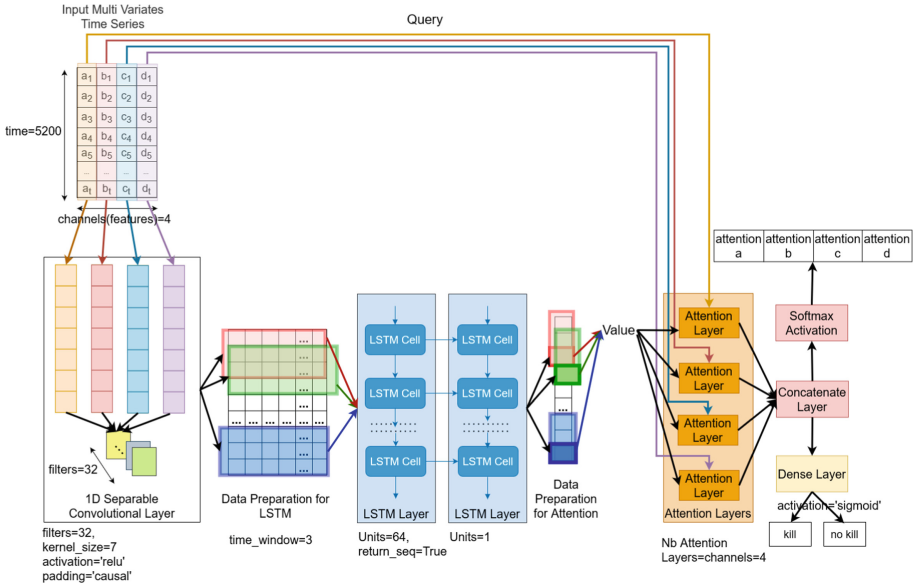


Fig. 3. Network architecture.

dot-attention mechanisms [30] are applied on this sequence (*key* and *value*) to focus on information important for the final prediction, according to each input feature (a *query*) at each time. The resulting 4 feature sequences, one per input feature, are concatenated, and fed to a dense layer with bias and a sigmoid activation function, which produces the binary prediction for each time step of the input sequence. We briefly describe each main component of the network.

The 1D depthwise separable convolution layer applies a separate 1D temporal convolution per input feature sequence with a kernel of size 7, followed by 32 independent pointwise convolutions (1×1 convolutions) to learn correlations between features, and the ReLu activation function. The output of this layer is a $t \times 32$ feature matrix.

Then, both short and long term feature dependencies are captured by two LSTM layers. To this end, the sequence of 32 features is decomposed into overlying sub-sequences of 60 time steps, giving $t/60$ sub-sequences of 32 features. Each sub-sequence is fed to a first LSTM layer composed of 60 cells (one per time steps), a cell being composed of 64 units to produce a 60×64 intermediate feature matrix. The $t/60$ intermediate feature matrices are then fed to a second LSTM layer, producing $t/60$ feature matrices of size 60×1. Then, the last time step of the $t/60$ sub-sequences are concatenated into a feature sequence of length t ($t \times 1$ matrix), denoted H, which is used by the attention mechanism.

Attention allows to relate the input sequence to output sequence H of the LSTM, capturing the specific influences of each feature. To this end, the dot-attention mechanism calculates similarity scores between the input sequence

(*query*) and H (*keys*), with a dot-product followed by a softmax, to weight H (*value*), for the four attention block independently. This enables the network to focus on the most relevant parts of the sequence for each feature. The concatenation of all attention layers is normalized with a softmax and leads to a *spatio-temporal attention*, offering a detailed view of the importance of each feature at each time step, enabling a fine-grained and nuanced analysis of the impact of each feature on the overall prediction.

5 Experiments

5.1 Protocol

The dataset was constructed from the *replays* of the 258 DotA2 matches played during The International 2018[1]. It contains a temporal sequence of the 4 polygon features described in Sect. 4.1 for each match and each team, i.e., 516 examples. As all sequences do not have the same length, we standardized them to 5 200 time steps (maximum duration of a match in the dataset) by 0-padding, so that they can be processed in a batch. A time step corresponds to 1 second.

The 516 examples were divided randomly (with random_state = 69) into training (60%), validation (20%) and test (20%) sets. For each example, the lifelines of the opposing players provide, at each time step, the ground truth class *kill* (at least 1 player from the opposing team is dead) or *no kill* (all players from the opposing team are alive). As the *no kill* class has a large majority (*kill*: 33% and *no kill*: 67% in the training set), we considered a *cost-sensitive learning* technique which weights each class in the loss function. Class weights were calculated from the inverse of the class distributions present in the training set (*kill*: 1.52, *no kill*: 0.74), and applied during model training. The network was trained according to the cost-sensitive binary cross-entropy loss, with the Adam optimizer, 0.001 learning rate and a batch size of 32.

The results are presented in the next section for the test set (used as a batch).

5.2 Results

Table 1. Network performances for the test set.

	with attention	without attention
accuracy	0.87	0.74
precision	0.86	0.74
recall	0.87	0.73
f1 score	0.86	0.73

First, we are interested in controlling whether or not the attention mechanism improves the learning process. Table 1 shows the results of the two CNN-LSTM

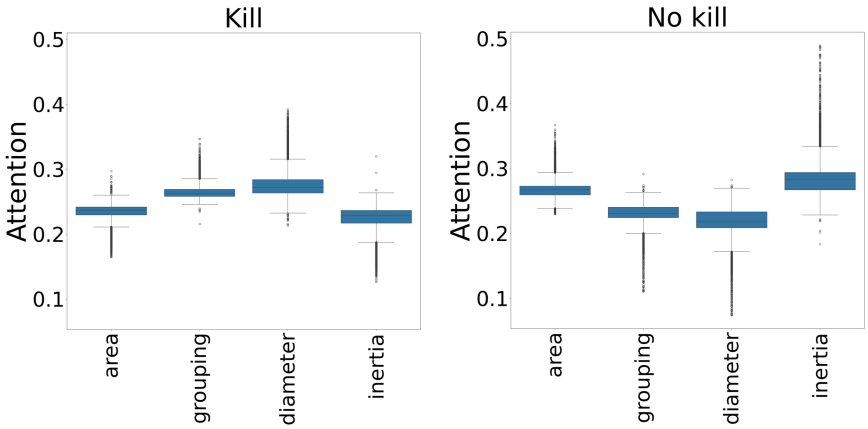


Fig. 4. Feature importance with Attention.

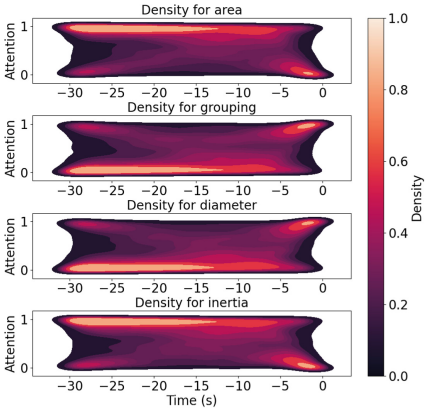


Fig. 5. Attention distribution (KDE plots) – the horizontal axis indicates the time before the kill occurs.

networks with and without the attention mechanism for the test set. The network with attention outperforms the network without attention, with scores around 0.86 while the network without attention peaks at 0.74. On the one hand, these results demonstrate the robustness of the CNN-LSTM network for processing the dataset, but more importantly they confirm the contribution of the attention mechanism. The attention layer is therefore useful for learning, and its interpretation is capable of providing information specifying the relationship between the input features and the prediction of an event.

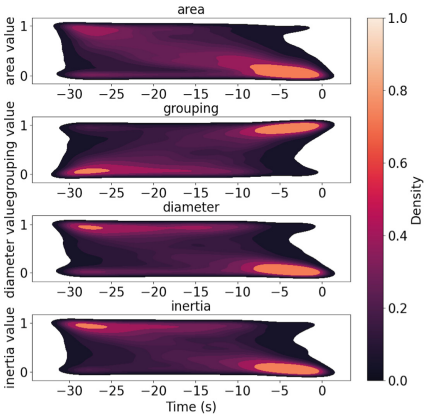


Fig. 6. Feature value distribution (KDE plots) – the horizontal axis indicates the time before the kill occurs.

Figure 4 shows the distribution of attention between polygon features, depending on whether the event to be predicted is *kill* or *no kill*. During a *kill* event, *diameter* and *grouping* show higher attention values with a wide distribution. *area* and *inertia* register a lower value. For *no kill* events, the trend is reversed. To predict a *kill* event, attention is drawn to the features *grouping* and *diameter*, which are clues of the shape of the polygon, depending on whether it is more or less stretched. To predict a *no kill* event, the network is more attentive to the polygon's inertia, relative to the variation in its shape.

Then, we analyzed the distribution of the attention weights in a 30-second interval before the number of live players on the opposing team decreased. Figure 5 shows that a change in attention occurs around 15 s before a death event (confirming [8] in establishing 15 s as enabling the best detection rate of important events). In addition, there is a clear variation pattern in the attention, with *grouping* and *diameter* attention increasing as one gets closer to the death event. Figure 6 shows that feature values change over time. In addition, we can clearly see that there is a tendency for the team to group together to make a *kill*.

We can observe these trends from a more general point of view with Fig. 7, which shows high attention as *area* increases with high *grouping*. This is indicative of the need for the team to group together to make a kill, surrounding the target within reach of teammates to help each other.

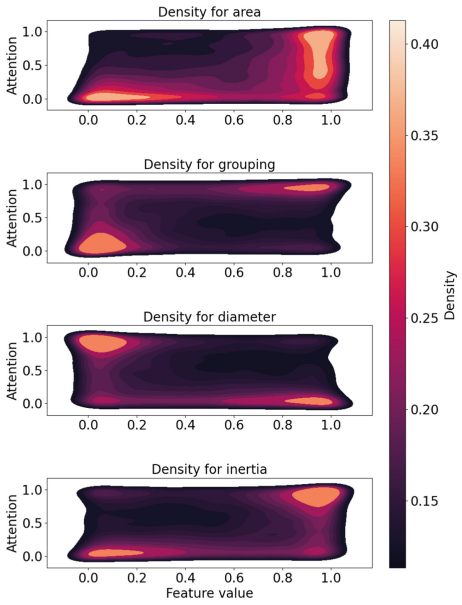


Fig. 7. Correlation distribution between Attention and Feature value (KDE plots).

6 Conclusion

Using a CNN-LSTM network with attention not only improved the accuracy of predictions by focusing on the relevant information, it also provided a way for interpretation. By examining the attention weights, it was indeed possible to identify which geometric features influenced most the network. This information highlighted the moments and actions that impacted the number of active players on the opposing team. Integrating an attention mechanism into a CNN-LSTM network therefore offered the opportunity to interpret the dynamics of collective play.

Acknowledgements. The research described here is supported by the French National Research Agency (ANR) and Region Normandie under grant HAISCoDe.

References

1. Abnar, S., Zuidema, W.: Quantifying attention flow in transformers. arXiv preprint arXiv:2005.00928 (2020)
2. Ahmad, S., Bryant, A., Kleinman, E., Teng, Z., Nguyen, T.H.D., Seif El-Nasr, M.: Modeling individual and team behavior through spatio-temporal analysis. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, pp. 601–612 (2019)

3. Bibal, A., et al.: Is attention explanation? An introduction to the debate. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 3889–3900 (2022)
4. Block, F., et al.: Narrative Bytes: data-driven content production in Esports. In: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video, pp. 29–41 (2018)
5. Castronova, E.: The surprising impact of professional sports on Esports. J. Sports Econ. **21**(7), 705–717 (2020)
6. Cavadenti, O., Codocedo, V., Boulicaut, J.F., Kaytoue, M.: What did i do wrong in my MOBA game?: mining patterns discriminating deviant behaviours. In: International Conference on Data Science and Advanced Analytics, Montréal, Canada (2016). https://hal.archives-ouvertes.fr/hal-01388321
7. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
8. Chu, W.T., Chou, Y.C.: Event detection and highlight detection of broadcasted game videos. In: Computational Models of Social Interactions, pp. 1–8 (2015)
9. Drachen, A., et al.: Skill-based differences in spatio-temporal team behaviour in defence of the ancients 2 (DotA 2). In: 2014 IEEE Games Media Entertainment, pp. 1–8. IEEE (2014)
10. Frencken, W., Lemmink, K., Delleman, N., Visscher, C.: Oscillations of centroid position and surface area of soccer teams in small-sided games. Eur. J. Sport Sci. **11**(4), 215–223 (2011)
11. Hauri, S., Djuric, N., Radosavljevic, V., Vucetic, S.: Multi-modal trajectory prediction of NBA players. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1640–1649 (2021)
12. Hauri, S., Vucetic, S.: Group activity recognition in basketball tracking data–neural embeddings in team sports (NETS). arXiv preprint arXiv:2209.00451 (2022)
13. Himmelstein, M., Shapiro, J.: Esports: a new era of sport with opportunities for sports scientists. Int. J. Sports Physiol. Perform. **15**(4), 485–487 (2020)
14. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
15. Jain, S., Wallace, B.C.: Attention is not explanation. arXiv preprint arXiv:1902.10186 (2019)
16. Lan, X., Duan, L., Chen, W., Qin, R., Nummenmaa, T., Nummenmaa, J.: A player behavior model for predicting win-loss outcome in MOBA games. In: International Conference on Advanced Data Mining and Applications. pp. 474–488. Springer (2018). https://doi.org/10.1007/978-3-030-05090-0_41
17. Mao, W., Liu, M., Salzmann, M.: History repeats itself: human motion prediction via motion attention. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16, pp. 474–489. Springer (2020)
18. Mehrasa, N., Zhong, Y., Tung, F., Bornn, L., Mori, G.: Learning person trajectory representations for team activity analysis. arXiv preprint arXiv:1706.00893 (2017)
19. Monti, A., Bertugli, A., Calderara, S., Cucchiara, R.: Dag-Net: double attentive graph neural network for trajectory forecasting. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 2551–2558. IEEE (2021)
20. Mora-Cantallops, M., Sicilia, M.Á.: MOBA games: a literature review. Entertainment comput. **26**, 128–138 (2018)

21. Mortelier, A., Rioult, F.: Addressing the complexity of MOBA games through simple geometric clues. In: International Conference on Computational Collective Intelligence, pp. 643–649. Springer (2022). https://doi.org/10.1007/978-3-031-16210-7_52
22. Qi, Z., Shu, X., Tang, J.: DotaNet: two-stream match-recurrent neural networks for predicting social game result. In: 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), pp. 1–5 (2018). https://doi.org/10.1109/BigMM.2018.8499076
23. Ringer, C., Nicolaou, M.A.: Deep unsupervised multi-view detection of video game stream highlights. In: Proceedings of the 13th International Conference on the Foundations of Digital Games, pp. 1–6 (2018)
24. Sanford, R., Gorji, S., Hafemann, L.G., Pourbabaee, B., Javan, M.: Group activity detection from trajectory and video data in soccer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 898–899 (2020)
25. Shah, R., Romijnders, R.: Applying deep learning to basketball trajectories. arXiv preprint arXiv:1608.03793 (2016)
26. Sifa, R., Drachen, A., Bauckhage, C.: Profiling in games: understanding behavior from telemetry. In: Social Interactions in Virtual Worlds: An Interdisciplinary Perspective (2018)
27. Silva, A.L.C., Pappa, G.L., Chaimowicz, L.: Continuous outcome prediction of league of legends competitive matches using recurrent neural networks. In: SBC-Proceedings of SBC Games, pp. 2179–2259 (2018)
28. Summerville, A., Cook, M., Steenhuisen, B.: Draft-analysis of the ancients: predicting draft picks in DotA 2 using machine learning. In: Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference (2016)
29. Tang, Y., Ma, L., Liu, W., Zheng, W.: Long-term human motion prediction by modeling motion context and enhancing motion dynamic. arXiv preprint arXiv:1805.02513 (2018)
30. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc. (2017)
31. Victor, B., Nibali, A., He, Z., Carey, D.L.: Enhancing trajectory prediction using sparse outputs: application to team sports. Neural Comput. Appl. **33**, 11951–11962 (2021)
32. Wang, W.: Predicting Multiplayer Online Battle Arena (MOBA) Game Outcome Based on Hero Draft Data. Master's thesis, Dublin, National College of Ireland (2016). http://trap.ncirl.ie/2523/
33. Wang, Y., Zhang, T., Guo, X., Shen, Z.: Gradient based feature attribution in explainable AI: a technical review (2024)
34. White, A., Romano, D.: Scalable psychological momentum estimation in Esports. In: Workshop SUM '20: State-based User Modelling, The Thirteenth ACM International Conference on Web Search and Data Mining (WSDM '20). ACM (2020)
35. Wiegreffe, S., Pinter, Y.: Attention is not not explanation. arXiv preprint arXiv:1908.04626 (2019)
36. Yao, Q., Liao, X., Jin, H.: Hierarchical attention based recurrent neural network framework for mobile MOBA game recommender systems. In: 2018 IEEE International Conference on Parallel and Distributed Processing with Applications, Ubiquitous Computing and Communications, Big Data and Cloud Computing, Social Computing and Networking, Sustainable Computing and Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), pp. 338–345. IEEE (2018)

37. Yu, C., Zhu, W.N., Sun, Y.M.: E-sports ban/pick prediction based on bi-LSTM meta learning network. In: International Conference on Artificial Intelligence and Security, pp. 97–105. Springer (2019). https://doi.org/10.1007/978-3-030-24274-9_9
38. Zhang, Q., Zhang, X., Hu, H., Li, C., Lin, Y., Ma, R.: Sports match prediction model for training and exercise using attention-based LSTM network. Digit. Commun. Netw. **8**(4), 508–515 (2022)
39. Zhao, Y., Yang, R., Chevalier, G., Shah, R.C., Romijnders, R.: Applying deep bidirectional LSTM and mixture density network for basketball trajectory prediction. Optik **158**, 266–272 (2018)
40. Ziyi, Z., Bunker, R., Takeda, K., Fujii, K.: Multi-agent deep-learning based comparative analysis of team sport trajectories. IEEE Access **11** (2023)

Author Index

U. Brefeld et al. (Eds.): MLSA 2024, CCIS 2460, p. 119, 2025.
https://doi.org/10.1007/978-3-031-86692-0