

Game Plan: What AI can do for Football, and What Football can do for AI

Karl Tuyls, Shayegan Omidshafiei, Paul Muller, Zhe Wang, Jerome Connor, Daniel Hennes

KARLTUYLS@ SOMIDSHAFIEI@ PMULLER@ ZHEWANG@ JEROMECONNOR@ HENNES@ ...GOOGLE.COM
DeepMind

Ian Graham, William Spearman, Tim Waskett, Dafydd Steele

IAN.GRAHAM@ WILLIAM.SPEARMAN@ TIM.WASKETT@ DAFYDD.STEELE@ ...LIVERPOOLFC.COM
Liverpool Football Club

Pauline Luc, Adria Recasens, Alexandre Galashov, Gregory Thornton, Romuald Elie, Pablo Sprechmann, Pol Moreno, Kris Cao, Marta Garnelo, Praneet Dutta, Michal Valko, Nicolas Heess, Alex Bridgland, Julien Pérrolat, Bart De Vylder, S. M. Ali Eslami, Mark Rowland, Andrew Jaegle, Remi Munos, Trevor Back, Razia Ahamed, Simon Bouton, Nathalie Beauguerlange, Jackson Broshear, Thore Graepel, Demis Hassabis

PAULINELUC@ ARECASENS@ AGALASHOV@ GREGTHORNTON@ RELIE@ PSPRECHMANN@ POLC@ KRISCAO@ GARNELO@ PRANEETDUTTA@ VALKOM@ HEES@ ABRIDGLAND@ PEROLAT@ BARTDV@ AESLAMI@ MARKROWLAND@ DREWJAEGLE@ MUNOS@ BACK@ RAHAMED@ BOUTON@ NBEAUGUERLANGE@ BROSHEAR@ THORE@ DEMISHASSABIS@ ...GOOGLE.COM

DeepMind

Abstract

The rapid progress in artificial intelligence (AI) and machine learning has opened unprecedented analytics possibilities in various team and individual sports, including baseball, basketball, and tennis. More recently, AI techniques have been applied to football, due to a huge increase in data collection by professional teams, increased computational power, and advances in machine learning, with the goal of better addressing new scientific challenges involved in the analysis of both individual players' and coordinated teams' behaviors. The research challenges associated with predictive and prescriptive football analytics require new developments and progress at the intersection of statistical learning, game theory, and computer vision. In this paper, we provide an overarching perspective highlighting how the combination of these fields, in particular, forms a unique microcosm for AI research, while offering mutual benefits for professional teams, spectators, and broadcasters in the years to come. We illustrate that this duality makes football analytics a game changer of tremendous value, in terms of not only changing the game of football itself, but also in terms of what this domain can mean for the field of AI. We review the state-of-the-art and exemplify the types of analysis enabled by combining the aforementioned fields, including illustrative examples of counterfactual analysis using predictive models, and the combination of game-theoretic analysis of penalty kicks with statistical learning of player attributes. We conclude by highlighting envisioned downstream impacts, including possibilities for extensions to other sports (real and virtual).

1. Introduction

Recent years have seen tremendous growing interest in sports analytics, not only from an economic and commercial perspective, but also from a purely scientific one, viz. the growing number of publications (Baumer & Zimbalist, 2014; Beal, Norman, & Ramchurn, 2019; Shih, 2017; Szymanski, 2020) and scientific events organized on the topic (e.g., *CVSports International Workshop on Computer Vision in Sports at CVPR*, 2020; *Machine Learning and Data Mining for Sports Analytics*, 2020; *MIT Sloan Sports Analytics Conference*, 2020).

As evident in many different downstream domains that have benefited from applications of artificial intelligence (AI) and machine learning (ML), this is due to important technological advances in data collection and processing capabilities, progress in statistical and in particular deep learning, increased compute resources, and ever-growing economic activities associated with sports and culture (e.g., emergent consultancy ventures revolving around sports data collection and statistics, Beal et al., 2019; ChyronHego, 2020; *InStat*, 2020; Kuper & Szymanski, 2018; *Opta*, 2020; *StatsBomb*, 2020).

Predictive analytics has been investigated and applied in the context of several sports in the past decades, including basketball (Skinner, 2010), tennis (Anderson, Rosen, Rust, & Wong, 2020; Gauriot, Page, & Wooders, 2016), and baseball (Albert, 2010; Albert, Bennett, & Mead, 2002; Baumer & Zimbalist, 2014; Costa, Huber, & Saccoman, 2009; Michael, 2004; Puerzer, 2002; Song, Severini, & Allada, 2017), with data for the latter having been systematically collected since the 19th century. Although statistical analysis of data has led to impressive outcomes in various sports (e.g., Moneyball in baseball, Baumer & Zimbalist, 2014; Michael, 2004), football started participating rather late in this data collection and number-crunching game, with the data science transformation that informs stakeholders (e.g., decisions related to player transfers, scouting, pre- and post-match analysis, etc.) still in its infancy (Kuper & Szymanski, 2018). Several factors influenced this late arrival. Football takes place under far less controllable settings than other sports due to its outdoor and highly dynamic nature, a larger pitch, a large number of players involved, a low number of player changes and longer non-interrupted game sequences than sports such as basketball. As a result, football analytics companies have only relatively recently started collecting so-called big data (e.g., high-resolution videos, annotated event-streams, player tracking and pose information). Concurrently, only recently have major breakthroughs been made in deep learning, yielding techniques that can handle such new high-dimensional data sets (Arel, Rose, & Karnowski, 2010; Bengio, 2009; Goodfellow, Bengio, & Courville, 2016; LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015). Finally, for a long time, credibility in decision-making primarily depended on human specialists such as managers, retired players, and scouts, all of them with track records and experience in professional football, in part due to cultural reasons (Decroos & Davis, 2019; Kuper & Szymanski, 2018). As a result of these various factors, the potential influence and gains of predictive analytics on the football game have also been less obvious, with sports analytics as a game-changing phenomena not realized until recent years. In more philosophical terms, Kuper and Szymanski (2018) highlight a cultural hesitation regarding the integration of data science into football and an overdependence on gut instincts, noting that “until very recently, soccer had escaped the Enlightenment”.



Figure 1: AI research frontiers associated with football analytics. We highlight three foundational areas related to AI research—statistical learning, computer vision, and game theory—which have independently been demonstrated to be effective in analyzing the game of football (with example problems and works from literature shown above per field). We argue that the domain spanned by these research fields is most promising for establishing seminal progress in football analytics in the coming years, along 3 frontiers: Frontier 1 (GT&SL), Frontier 2 (SL&CV), and Frontier 3 (GT&CV). We claim that the above frontiers culminate into a unique microcosm mutually benefiting AI and football analytics.

Despite football’s late adoption of sports analytics, there are a number of early-bird approaches from different areas of AI such as statistical learning (SL), computer vision (CV), and game theory (GT) that are making initial contributions to support decision-making of managers, coaches and players. For example, basic statistical learning tools such as principal component analysis (PCA) already enable automated means of identifying player types

(Decroos & Davis, 2019), training of models predicting trajectories of individual teams or imitating league-average behaviors (H. Le, Carr, Yue, & Lucey, 2017), and valuing individual player decisions (such as passes or tackles) in a series of actions leading up to a goal (Decroos, Bransen, Haaren, & Davis, 2019). The study of interactive decision-making as formalized by game theory plays a critical role in AI for systems involving more than one actor (human or artificial). Game-theoretic tools shed light on players' strategic interactions during scenarios such as penalty kicks, analysis of their decisions in comparison to mathematically-principled baselines, and prediction of their goal-scoring probabilities when playing according to a mixed-strategy Nash equilibrium (Chiappori, Levitt, & Groseclose, 2002; Palacios-Huerta, 2003, 2016). Enriched with empirical game theory (Omidshafiei et al., 2019; Tuyls et al., 2020; Wellman, 2006) the effects of various high-level strategies pitted against each other can also be analyzed. Finally, recent developments in computer vision have been employed for player tracking (Bialkowski et al., 2015; Gade & Moeslund, 2018; Liu, Carr, Collins, & Liu, 2013; Lu, Ting, Little, & Murphy, 2013), pose estimation (Bridgeman, Volino, Guillemaut, & Hilton, 2019; Fastovets, Guillemaut, & Hilton, 2013; Sypetkowski, Kurzejamski, & Sarwas, 2019; Sypetkowski, Sarwas, & Trzcinski, 2019; F. Zhang, Zhu, & Ye, 2019), and automated injury prediction (Kampakis, 2016) based on, e.g., gait and fatigue analysis (Bartlett, 2006; Claudino et al., 2019; Kakavas, Malliaropoulos, Pruna, & Maffulli, 2019; Op De Beeck, Meert, Schütte, Vanwanseele, & Davis, 2018; Ramos et al., 2020).

While these separate areas within AI research have independently been demonstrated to be effective for football analytics, we believe that the most pertinent research problems lie in the underexplored intersection of statistical learning, computer vision, and game theory (see Figure 1). Specifically, we pinpoint several frontiers at the intersections of these fields, and identify the ultimate frontier to be the microcosm requiring integrated approaches across all three fields. A large portion of the state-of-the-art research in football analytics, by contrast, typically falls under the umbrella of one of these areas, with some initial activities taking places at Frontier 2 (SL&CV) (Choi, Kwon, & Lee, 2019; Lu et al., 2013; Mehrasa, Zhong, Tung, Bornn, & Mori, 2018; Mora & Knottenbelt, 2017; Quiroga, Carrillo, Maldonado, Ruiz, & Zapata, 2020), and no notable research activities identified at Frontier 1 (GT&SL) and Frontier 3 (GT&CV) for football or sports in general. At Frontier 1 (GT&SL), game-theoretic analysis is blended with learned predictive models, combining interactive decision-making with predictive modeling to provide more granular analysis tools. We present a detailed case study illustrating this frontier, revisiting the work of Palacios-Huerta (2003) on penalty kicks under this new perspective and illustrate how mixing with SL provides deeper insights into penalty-kick taking. Frontier 2 (SL&CV) focuses on research that integrates statistical learning with computer vision, directly learning from video as the primary input and building predictive models (e.g., forecasting player and team behaviors directly). At Frontier 3 (GT&CV), we classify football research integrating computer vision and game theory, a largely uncharted territory focusing on generative models based on visual inputs, which takes strategic interactions into account.

We claim that the above frontiers culminate into a unique microcosm mutually benefiting both AI and football analytics, to the point it becomes possible to develop, for example, an Automated Video Assistant Coach (AVAC). The AVAC system is an example of what we believe to be the future of human-centric AI research for football, with the aim of

integrating all aspects of the frontiers into a cohesive system enabling both understanding and improvement of human football play. Such an AVAC system is envisioned to improve the overall experience of the game for players, coaches, and spectators alike. Finally, we also point out that jointly these frontiers provide the necessary theoretical background for analytics research to build upon in the microcosm, laying out a foundation to build accurate models and predictions, as advocated for in Szymanski (2020).

In the following, we first provide an overview of the literature associated with football analytics, subsequently highlighting gaps and sketching a long-term vision for research in the football microcosm. We lay out the long-term perspective of how AI can benefit the domain of football analytics, and vice versa, by combining the three identified axes in the literature in manners that have not yet been fully explored. Next, we present illustrative examples that examine penalty-kick taking from a game-theoretic perspective, building on the work of Palacios-Huerta (2003, 2016), and bringing several new insights based on data from the main European leagues. We subsequently demonstrate how this game-theoretic work can be enriched via integration with statistical learning at Frontier 1 (GT&SL), providing several new insights about penalty-kick taking followed by a high-level example of counterfactual trajectory predictions (i.e., a what-if analysis), thus further justifying the football microcosm as a useful AI research domain.

2. Literature Overview: AI for Football Analytics

The following sections outline a long-term research vision for AI applied to football analytics. We first consider the state-of-the-art with respect to each of the respective areas (GT, SL, and CV) applied to the football domain, after which we look into the opportunities that each of the frontiers brings forth to unlocking larger scientific challenges in football analytics. The following sections summarize works that lie in these peripheral fields of Figure 1 and highlight opportunities for further work and added value in each.

2.1 Statistical Learning

Football is arguably one of the most challenging to analyze of all the major team sports. It involves a large number of players with varied roles, few salient events, and minimal scoring. Statistical football analysis attempts to provide quantitative answers to questions that pertain to different aspects of the game. Notably, these include the problem of characterizing players' and teams' styles, evaluation of the impact that such teams have on the pitch, and the temporal and counterfactual predictions of players' actions. When one compares styles of players (and teams, by extension), one usually refers to high level and abstract notions that summarize their unique characteristics. The goal of the statistical learning line of research is to learn features capturing such information, normally to be used by other down-stream tasks. For instance, one means of summarizing information about a football player is through aggregation of their play statistics (e.g., offensive, defensive, or dribbling abilities, shots on target, etc.), and that of their teams (Fernandez-Navarro, Fradua, Zubillaga, Ford, & McRobert, 2016; *Stats Perform*, 2020). While such statistics are typically either handcrafted or rely on simple measures of play outcomes, recent works have analyzed them from a statistical learning perspective, using notions such as Player Vectors (Decroos & Davis, 2019) (detailed in Appendix C) and analogous techniques in sports such

as basketball (Franks, Miller, Bornn, Goldsberry, et al., 2015). Given the growing success of unsupervised learning methods, there is potential for more advanced representations of player traits to be learned directly from the data.

In football, it is particularly difficult to assess which players, or groups of players, deserve credit for favorable outcomes. For high-scoring team sports (e.g., basketball) one can provide a reasonable answer to this question by restricting attention to actions that have immediate impact on scoring. By contrast, few goals are scored in football (e.g., 2.72 goals were scored on average per game of the 2019/2020 Premier League season, Clarke, 2020). Consequently, models considering only actions with immediate impact on goals (e.g. shots or saves) capture a crucial yet narrow view of the game. Moreover, game states in football are significantly more complex than that estimated by current models. Features describing them are mostly hand-crafted and only consider on-ball actions. A given pass might be extremely valuable or a poor choice depending on the disposition of the players. While these methods are able to value on-ball actions, they rely on sparse signals provided by goals. Concurrently, off-ball actions significantly impact each game, as exemplified by player actions covering certain pitch areas to prevent attacks, running to open areas to create space for teammates, and so on. Due to the temporally extended nature of football, the task of inferring the value of actions is an instance of the temporal credit assignment problem in reinforcement learning (RL) (Minsky, 1961). The combination of RL techniques with deep learning has great potential to tackle the idiosyncrasies of football and to close the gap between statistical and human analysis. It is exciting to see recent progress in this direction in sports analytics, including ice hockey (Liu & Schulte, 2018) and football (Liu, Luo, Schulte, & Kharrat, 2020; X. Sun, Davis, Schulte, & Liu, 2020) (with additional related works detailed in Appendix A).

Overall, while the above pieces of work showcase the promise of modern statistical methods for temporal predictions in football, this remains an open and challenging problem that will likely require development of novel methods and means of leveraging the diversity of newly-available football data (as later made more precise in terms of what football can do for AI).

2.2 Game Theory

Game theory plays an important role in the study of sports, enabling theoretical grounding of players' behavioral strategies. Numerous works have applied game-theoretic analysis to sports over recent decades (Lennartsson, Lidstrom, & Lindberg, 2015; Sindik & Vidak, 2008), including football (Azar & Bar-Eli, 2011; Buzzacchi & Pedrini, 2014; Chiappori et al., 2002; Coloma, 2012; Levitt, Chiappori, & Groseclose, 2002; Moschini, 2004; Palacios-Huerta, 2003), tennis (Anderson et al., 2020; Gauriot et al., 2016), volleyball (K. Lin, 2014), basketball (Skinner, 2010), and American football (Emara, Owens, Smith, & Wilmer, 2014). High-level game theory can be applied to team selection and choice of formation. Set pieces, such as corner kicks and penalties, are particularly amenable to game-theoretic analysis, wherein identified high-level strategies can be pitted against one another and ranked in terms of empirical performance (see Section 5 for details). Due to the real-world nature of football, the types of game-theoretic analysis conducted in this domain have typically been driven by the availability of data sources (in contrast to, e.g., simulation-based

domains, where data of varying types and granularity can be synthetically generated). In particular, the volume of available high-level statistics (e.g., match outcomes spanning across different leagues, seasons, and competition levels) makes football a particularly attractive topic from a behavioral and empirical game-theoretic perspective (Camerer, 2011; Camerer, Loewenstein, & Prelec, 2005; Camerer, Loewenstein, & Rabin, 2011; Wellman, 2006). In contrast to traditional game theory, which often assumes rational decision-making players seeking to maximize utilities, behavioral game theory (and the closely-associated field of behavioral economics) studies the interactions of decision-making individuals that either have imperfect knowledge of the underlying game, have bounded rationality, or are simply not purely self-interested (Camerer et al., 2005; Nax, 2015). The complexity of real-world football scenarios involving interacting humans, who may be influenced by a multitude of factors other than purely maximizing their in-game utilities, implies that the use of such techniques may prove invaluable for more accurate understanding of the game.

From a theoretical perspective, the majority of the existing works exploit the fact that various football scenarios can be modeled as two-player zero-sum games. For example, in football, the penalty kick situation may be straightforwardly modeled as a two-player asymmetric game, where the kicker's strategies may be neatly categorized as left, center, or right shots. The controlled nature of these scenarios compounds their appeal from a quantitative analysis perspective; in the penalty example, the penalty taker, goalkeeper, and ball's initial positions are generally static across all dataset trials (albeit, with minor variations, e.g., scenarios where goalkeepers stand slightly off-center to entice the kicker to shoot in one direction). In fact, the majority of the literature that analyzes football under a game-theoretic lens focuses on penalty kicks (Azar & Bar-Eli, 2011; Buzzacchi & Pedrini, 2014; Chiappori et al., 2002; Coloma, 2012; Levitt et al., 2002; Palacios-Huerta, 2003), which we contend is due to this amenability for analysis via classical game-theoretic solution concepts (such as Nash equilibria).

However until the paradigm shifts away from the classical analysis of set pieces and toward live-play settings such as those considered by Moschini (2004), the potential benefits of game-theoretic analysis for football are likely to remain largely untapped. Moving beyond this classical paradigm involves resolution of significant challenges: the number of active players in football is quite large (22 including goalkeepers) and the exponential size of the strategy space (with respect to the number of players) makes it more challenging to analyze than games such as tennis or basketball; the mapping of low-level player actions to strategies is no longer straightforward in durative live plays, due to the variability of player trajectories; the duration of plays is also generally longer than sports played in more controlled environments (e.g., tennis), implying that such scenarios may benefit from analysis in the so-called extensive-form, which explicitly consider each player's knowledge, opportunities, and actions, unlike in the simpler (but more feasible to analyze) simultaneous-move, normal-form approaches typically used for set piece analysis (later detailed in Section 5.1).

There exists potential for interesting team-to-team interactions to occur when use of game-theoretic techniques for prescription of player behaviors becomes prominent. Once such techniques are deployed en-mass to aid teams, one can safely assume that opposing teams are also using such systems. This yields an interesting scenario where an AI system used by one team may be used to, in the language of game theory, best-respond to the other team's system. Even further, one can also consider this best-response behavior emerging

from the human players themselves (e.g., to try to act adversarially to cause the opposing team’s analytics platforms to mispredict or otherwise provide non-useful advice to their players). In recent years, a significant amount of effort has been spent by researchers in attempting to thwart or make less sensitive the predictions of AI and ML algorithms to such adversarial attacks (e.g., Madry, Makelov, Schmidt, Tsipras, & Vladu, 2017), with techniques for analyzing these challenges often inspired by game theory. It is similarly perceived that incorporating such robustness mechanisms is likely to be beneficial for deployed football AI techniques (e.g., by refining training, or by explicitly considering sources of uncertainty to quantify and convey risk levels when making recommendations to coaches and decision-makers). One important distinction to bear in mind, however, is that the nature of football implies that, ultimately, it will be the humans that will make decisions and directly interact with one another on the pitch. Despite the advice received from a combination of coaches and AI systems, these direct on-pitch interactions will remain human-to-human, thus potentially making such adversarial countermeasures more difficult. This stands in contrast, for example, to applications of AI to domains such as self-driving cars, where ultimately the AI is in the driver’s seat and directly interacts with humans on the road.

Overall, the availability of new data types increases the viability of conducting more advanced game-theoretic analysis by bootstrapping to advances in statistical learning and computer vision, as later detailed. Surmounting these challenges will benefit football from a game-theoretic perspective, enabling both a descriptive analysis (i.e., understanding the interactions undertaken by players and teams in the presence of others), and a prescriptive one (i.e., suggesting the actions such individuals should have executed).

2.3 Computer Vision

Computer vision has seen major breakthroughs in the past decade, thanks to the application of deep learning approaches. Progress in tasks such as image classification (Deng et al., 2009), video action recognition (Carreira & Zisserman, 2017), and pose estimation (Alp Güler, Neverova, & Kokkinos, 2018) have unlocked the possibility of automatically extracting complex information from videos. Similarly, in football analytics, computer vision methods can enable enrichment of statistical and game-theoretic approaches, which typically rely on hand-labeled, low-dimensional data. Video is a particularly appealing signal for football analytics: it is a rich data source (likely to contain large amounts of information of interest for decision-making) and cheap to acquire (using only widespread camera sensors). While raw video frames are insufficiently structured to be processed by traditional rule-based processing techniques, computer vision methods enable the extraction of high level, potentially spatially-detailed representations, ready for use by downstream applications. Examples of such extraction processes include human pose estimation (Alp Güler et al., 2018), object detection (Ren, He, Girshick, & Sun, 2015), segmentation (Long, Shelhamer, & Darrell, 2015), tracking (Dong & Shen, 2018), depth estimation (Godard, Mac Aodha, & Brostow, 2017), and event or action detection (Carreira & Zisserman, 2017).

In addition to explicit, human-interpretable representations, learned counterparts can be used by downstream deep learning models. In football, several companies already commercialize tracking information, relating to the players and the ball, automatically extracted from videos recorded by dedicated static cameras, placed to have a view covering the entire

terrain (*Opta*, 2020; *StatsBomb*, 2020). Moreover, immersive sports analytics has been gaining popularity with the increased access to specialized hardware and portability (T. Lin, Yang, Beyer, & Pfister, 2020). Computer vision techniques enable reconstruction of real game scenarios, which provide more feedback to players and coaches than 2D screens, and have been extended to other sports media as well (*StriVR Immersive Sports Analytics*, 2020). Vision-based player tracking, pose estimation, and event detection can improve learning of player skill vectors and predictive modeling, subsequently improving player ranking and game-theoretic analysis of strategies.

While video has been traditionally used as the primary high-dimensional signal for the above applications, other modalities such as audio and text streams can provide extremely rich and complementary information in these settings. Audio commentary and news threads readily provide a human interpretation of the events occurring in a scene, which is highly complementary to the spatially fine-grained information available in videos. Learning-based approaches for generation of text commentary have been previously investigated (D. L. Chen, Kim, & Mooney, 2010), with temporally-aligned sound and text streams seeing recent application in learning rich representations of video signals (Alayrac et al., 2020; Arandjelovic & Zisserman, 2017; Miech et al., 2020), having been shown to be useful in a variety of computer vision tasks. Although such information significantly overlaps with the structured annotations already used by downstream applications (event annotations, etc.), its unstructured nature enables the presence of a greater information content (e.g., a commentator's tone can provide some cues as to an action's value).

Unique challenges arise when seeking to improve the performance and to enlarge the applications of computer vision methods for broadcast video of football games. In general, the camera's field of view is centered on the most interesting action in the field, leaving many players out of the frame. This poses an interesting geometric problem, as broadcast footage shots typically do not overlap, and many state-of-the-art systems are unable to geometrically relate these multiple shots. Other challenges arise due to the multiple occlusions of players by one another, further complicating the detection, tracking, and identification tasks. These problems can be addressed by geometrical approaches in settings where one can intervene to correct camera positions or, ideally, access cameras' extrinsic and intrinsic parameters (that is, the camera's position, orientation, as well as geometrical and optical properties like its focal length). However, approaches that do not assume access to such information have the potential to leverage more data. In contrast with other applications, football broadcast videos present an interesting blend between real, large-scale, complex data and a constrained setting (due to the rules of the game, Giancola, Amine, Dghaily, & Ghanem, 2018), hence providing an attractive setting for developing such approaches.

3. Game Plan: Long-Term Research Vision

In this section we outline a long-term research vision for football analytics in the microcosm frontiers at the intersection of statistical learning, game theory, and computer vision (see Figure 1).

3.1 Frontier 1: Interactive Decision-Making (GT&SL)

Learning to make suitable decisions in the presence of other agents is where game theory and statistical learning can converge. This interdisciplinary area of research has received significant attention in the multi-agent RL community over the past decades, with thorough survey articles (Bloembergen, Tuyls, Hennes, & Kaisers, 2015; Busoniu, Babuska, & Schutter, 2008; Hernandez-Leal, Kartal, & Taylor, 2020; Panait & Luke, 2005; Shoham, Powers, & Grenager, 2007; Tuyls & Weiss, 2012) available. When considering the football domain in particular, it is evident that the potentials of game theory have yet to be fully exploited, let alone its combination with machine learning techniques such as RL.

There are several promising routes in this area that not only are challenging from a football analytics perspective, but also from an AI research one. Two of these routes are studied further in this article in Section 5; one concerns the study of set pieces using the combination of statistical learning with game theory, and a second focuses on predictive modelling for counterfactual analysis. In the former, which has been mostly studied in a game-theoretic setting, we show how augmenting the analysis with player-specific statistics can provide deeper insights in how various types of players behave or take decisions about their actions in a penalty kick scenario. In the latter case, we illustrate how machine learning techniques can facilitate counterfactual analysis in football matches. The possibility to predict, for example, trajectories of players can enable investigation of counterfactual scenarios, (e.g., wherein one would like to know how a specific player or team would respond in a specific match scenario). Doing this enables one to not only learn to generate behaviors, but also leverage game-theoretic techniques for counterfactual analysis. We defer a more detailed discussion of these research lines to Section 5.

Building on the counterfactual prediction of players' behaviors, one can also consider the possibility of using this as a coaching tool. Specifically, one can use counterfactual insights to advise tactics to individual players, and even go further by optimizing the overall team strategy depending on the specific opponent in an upcoming match. This would go beyond the state-of-the-art, which focuses on simply predicting player behaviors; here, one would seek to actively optimize suggested tactics based on the particular behaviors and play style of the opposing team, and upcoming match-ups. Such tactical considerations can also be conducted in an iterative manner (e.g., predictions can be made for the opposing team as conditioned on the best-response behavior of the main team), and effective counter strategies can be learned, for instance, using multi-agent RL. Such an approach opens the door to a slew of interesting research challenges. For instance, use of multi-agent RL entails definition of a reward function for the players; while rewarding goals is an obvious candidate, denser reward signals (e.g., associated with successful passes, intercepts, etc.) may be useful for accelerating learning of such policies. Such reward functions are also likely to depend on the role of a player in the team and their respective skill set (i.e., may even be heterogeneous across players in a given team), and could also be learned using techniques such as inverse RL (Ng & Russell, 2000). Moreover, one may seek to first define an effective 'action space' for players (i.e., more granular or structured actions than 'move left' or 'move right'), before solving the RL problem. Finally, the combination of the previous learnt models with pitch control (Spearman, 2016), a technique to determine which player/team has control over a

specific area of the pitch, will provide additional information on open space, passing and scoring opportunities, yielding a powerful tool to enable in-match tailored coaching tools.

3.2 Frontier 2: Predictive Modeling from Videos (SL&CV)

Several challenges naturally lie in the frontier between statistical learning and computer vision. Statistical learning depends on large quantities of labelled data. Many of the quantities suitable for models of football are the product of hand-labelling data; on the other hand, vision-based models could automatically identify events which could be fed into such models. In addition to the quantity of events that vision-based systems could provide, the quality could also be improved (e.g., with events being accurately registered to the corresponding frame, with minimal temporal error compared to human-labeled annotations).

Furthermore, video is a much richer signal compared to what is traditionally used in predictive tasks, such as forecasting future movement of players or predicting the value of individual actions to the game outcome. The combination of advanced deep learning models and a rich video signal enables learning over subtle clues otherwise not captured in event-stream or tracking data. Capturing more of the partially-observable state of a game will ultimately enable more accurate predictions. Richer information may additionally help to shed light on the intention of players and thus better address the credit assignment problem in action-value estimation.

On the other hand, models that better capture the game dynamics may be necessary to resolve some of the limitations of vision-based tracking approaches, which arise as players are occluded or move off camera. The resulting ambiguities can likely be resolved using predictive model of player dynamics, which may be trained independently and used as a source of additional inputs to the tracking system or trained as one component of a larger pipeline. Explicit or implicit access to the game dynamics will very likely also improve vision-based action labeling. Finally, presenting prediction outcomes by means of synthetic video generation remains an ambitious challenge that combines the task of trajectory prediction with video generation. Presenting predictions of counterfactual futures as video will enable intuitive understanding both by coaching personnel and players (e.g., in the same vein as recent work on video generation for tennis matches) (H. Zhang, Sciutto, Agrawala, & Fatahalian, 2020).

3.3 Frontier 3: Generative Game-Theoretic Video Analysis Models (GT&CV)

Video modeling and game theory can mutually benefit one another. In the simplest application, computer vision can provide new features to drive game-theoretic models. In more advanced applications, game theory can, in turn, guide video generation, as illustrated in Figure 3. Throughout a football match, individual players carry out a series of encounters with one another, which can profitably be viewed through game theory. Each player may have some idea of the likely success of strategies based on past performance and current in-game observations. Penalty kicks are an almost idealized example, with a kicker and goalkeeper observing each other for signs of intent, while simultaneously weighing preconceived strategies. As described previously, computer vision models can be used to automatically extract high-level and potentially spatially-detailed representations that can be complementary to the low-dimensional, hand-collected inputs game-theoretic models typically rely on.



Figure 2: Example of a pose estimation model applied to a penalty kick. The results showcased here follow a multi-phase approach (Papandreou et al., 2017) The first stage applies a Faster-RCNN candidate detector (Ren et al., 2015) which extracts bounding boxes around the players. This is followed by a pose estimation model applied to these bounding box proposals, which identify the player keypoints, and provide the pose estimate.

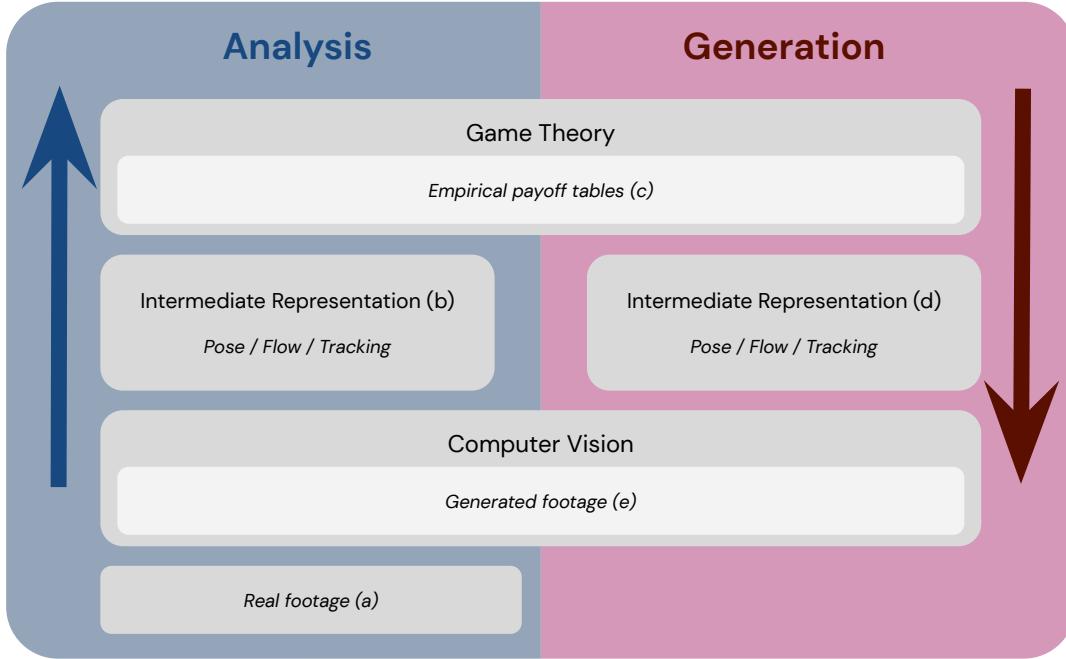


Figure 3: Analysis and generation of video data. From footage video (a), computer vision techniques can extract various representations (b), useful for empirical game-theoretic analysis. In turn, empirical payoff tables (c) can be used in a hierarchical generative process of videos, possibly involving a step in intermediate representation spaces (d). Generated videos (e) could be used for prescriptive analysis, or in the process of capturing the footage videos themselves, closing the cycle between analysis and generation.

We illustrate this representation extraction pipeline in the left portion of Figure 3. In the example of penalty kicks, vision models could identify visual signatures of intent that are difficult to even perceive for humans, let alone annotate; such information includes, for example, pose estimates extracted from broadcast footage (as visualized in Figure 2, with technical details provided in Appendix B), which can enable inference of the intentions of players, providing valuable insights to improve their respective strategies.

In the reverse direction (right portion of Figure 3), game-theoretic outputs can be used to improve the plausibility and usefulness of generated videos. Specifically, generative video models need to precisely capture the data distribution to be useful. The game-theoretic context found within football offers an opportunity to constraint such models. The dynamics of play are complex, with games taking place at the level of opposing individuals to entire teams, though with specific constraints that can inform a generative process. For example, a hierarchical generative process could condition on high-level latent variables drawn from the distribution described by empirical payoff tables, to represent player decisions conditioning a particular sample. One could consider an additional hierarchical level of generation targeting intermediate features, possibly building on approaches for predicting future pose (Chan, Ginosar, Zhou, & Efros, 2019; Villegas et al., 2017), trajectory (Bhattacharyya, Fritz, & Schiele, 2018, 2019; Kitani, Ziebart, Bagnell, & Hebert, 2012; Vu, Choi, Schulter, & Chandraker, 2018), action (Abu Farha, Richard, & Gall, 2018; Vondrick, Pirsiavash, & Torralba, 2016) and shape information (Jin et al., 2017; Luc, Couprie, Lecun, & Verbeek, 2018; Luc, Neverova, Couprie, Verbeek, & LeCun, 2017; J. Sun et al., 2019; Xu, Ni, Li, Cheng, & Yang, 2018).

Game-theoretic analysis could be used to assess the plausibility of game simulations. Specifically, these intermediate outputs could be used directly for prescriptive analysis (e.g., informing players of physical tactics to attempt in a play of interest, based on generated poses) or serve as further conditioning for generation in the raw RGB space. An alternative direction would be the formulation of new game-theory inspired metrics, for example imposing that the payoff tables extracted from generated samples match those obtained from real data. This would be closely related to the Fréchet Inception Distance (Heusel, Ramsauer, Unterthiner, Nessler, & Hochreiter, 2017) for generative modeling of images, and the derived Fréchet Video Distance (Unterthiner et al., 2018) for video, which have successfully driven progress in both fields (Brock, Donahue, & Simonyan, 2019; Clark, Donahue, & Simonyan, 2019; Karras, Aila, Laine, & Lehtinen, 2017; Luc et al., 2020; Miyato, Kataoka, Koyama, & Yoshida, 2018; Weissenborn, Täckström, & Uszkoreit, 2019; H. Zhang, Goodfellow, Metaxas, & Odena, 2019). In turn, such metrics could serve as a basis to design novel losses inspired by game theory. Besides their potential use for prescriptive analysis, generative models of videos could lead to improvements in broadcast data itself, due to automatic anticipation of the most plausible immediate future outcomes. Overall, game-theoretic analysis can inform video generation, in turn enabling influence of the process by which future data will be acquired, hence closing the cycle described in Figure 3.

3.4 Football as an AI Testbed: Microcosm

The development of performative AI algorithms relies on various recurring objectives: learning and acting based on real-world data streams, interpreting the actions of other agents and acting strategically in response, being able to understand and generate natural language for efficient communication with humans, and so on. As discussed in earlier sections, football analytics involves core problems associated with many of the above AI challenges, though in a more well-controlled scope.

The value of the football domain for AI can be observed, at a low level, in the range of useful data available for corresponding analytics (see Figure 4). Real-world data streams

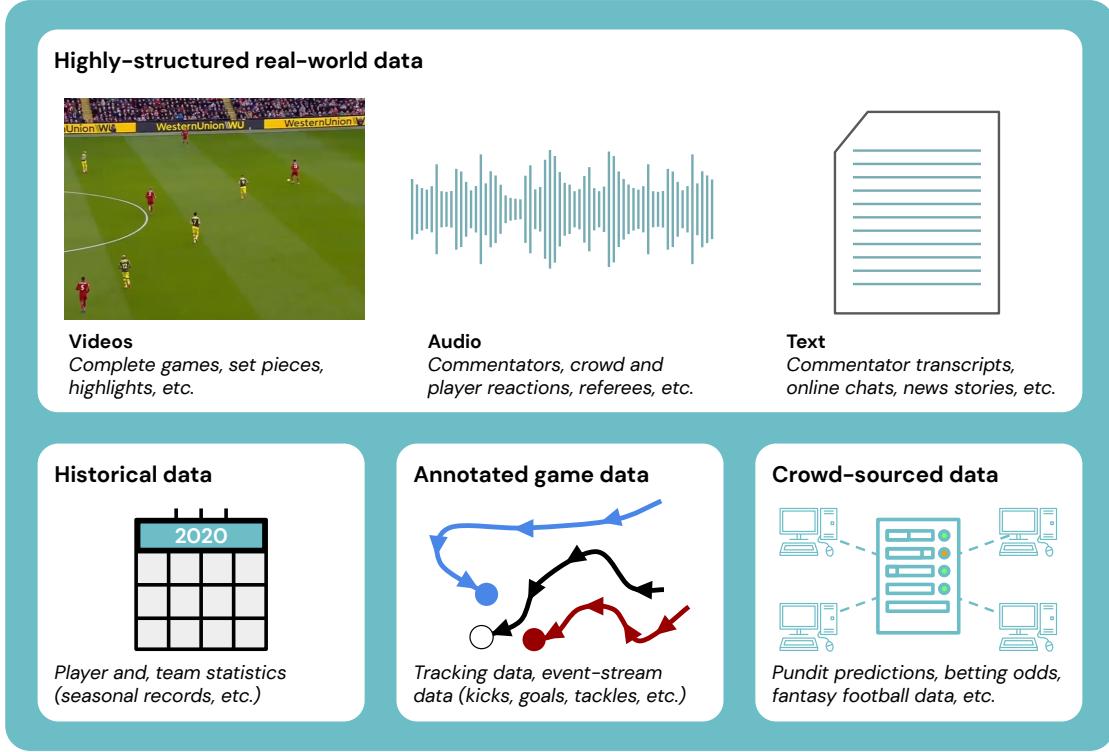


Figure 4: The range of data types relevant to football. At a high-level, highly structured and well-correlated data types are available for each match: video streams, audio streams, and associated text streams. On a more aggregate level, historical play data (e.g., player and team statistics) spanning many decades of play are broadly available for many professional football leagues. Moreover, third-party companies (such as *InStat*, 2020; *Opta*, 2020; *StatsBomb*, 2020; *Wyscout*, 2020) provide annotated player event and tracking data. Finally, there exists a broad range of crowd-sourced data for football, including user preferences and pundit predictions, betting odds, and so on.

such as vision, audio, and text are the mainstay of AI research, and are abundant in football. Crucially, the various data types available in football are well correlated, in the sense that they typically involve a large amount of shared context—a characteristic researchers can take advantage of. For instance: football video feeds always involve two teams and a ball; an enormous amount of text is available, yet it is centered on the current game; the sound of the crowds and commentators can be relied upon to respond to temporally-adjacent events, such as goals and penalties. There is a large amount of crowd-sourced data available such as current betting odds and pundits’ predictions, a novel form of data atypical in other AI application domains. Football offers the opportunity for AI to evaluate multi-modal models on synthesized vision, audio, and text data in a unified (though simpler) domain than the broader real world. Naturally, the abundance of data raises important considerations with respect to the computational efficiency of both training and deploying such models in

a scalable manner, which remains an open challenge. Moreover, despite the availability of large-scale and increasingly-precise data, the issue of unobserved heterogeneity between data and the real-world implies that AI and ML techniques that are robust to data uncertainty (or lack of observations of certain key variables altogether) will play an important role in advising football decision-makers with confidence. Statistical Learning, however, already provides useful techniques and models that can learn with missing and latent variables, and as such provide a foundation for handling such challenges in the football domain (Goodfellow et al., 2016; Murphy, 2013).

Football analytics also currently relies heavily upon hand-labeled data, such as ball and player tracking and identification information. This reliance on hand-labeled data imposes a significant barrier for fast-paced analysis, due to the cost and time needed to generate it. This provides a golden opportunity for AI to accelerate data collection and the subsequent development of novel learning algorithms (by automating the labeling and annotation process) and assisting coaches and decision-makers by instead allowing them to focus their expertise on the tactical analysis of the game itself. As such, a worthy long-term challenge for football analytics is to develop such an assistive agent, which uses minimal hand-labeled data: an Automated Video Assistant Coach (AVAC). A successful AVAC would help players, coaches, and spectators alike. Specifically, it could help the players by analyzing their play for weak points to further develop. A player's performance throughout a game could be analyzed to suggest improvements in position play and assessing the performance overall. Prior to a game, an AVAC could suggest strategies tuned to the opponents of the day. Coaches also seek to get the best out of their players, but have a limited amount of time and many players to observe and provide feedback to. An AVAC would offer coaches many opportunities to help individual players and the team as a whole, suggesting player rosters for a given game, as well as trading or scouting strategies based on counterfactual evaluation of team performance with brand new players. Such an AVAC system would have the ability to automatically sift and label huge quantities of video streams, enabling broadcasters and spectators alike to retrieve key moments. An AVAC could automatically keep a running tally of information the spectator may find interesting based on their reaction and the current state of play. To enhance the spectator experience, the AVAC may automatically generate highlight reels that effectively reflect the flow of the game or summarize the most exciting segments (Mahasseni, Lam, & Todorovic, 2017; Merler et al., 2018; Xiong, Kalantidis, Ghadiyaram, & Grauman, 2019; Yang et al., 2015; K. Zhang, Chao, Sha, & Grauman, 2016); moreover, the AVAC might suggest related games predicted to engage or interest the spectator. For those interested in fantasy football, an AVAC might search for players based on a set of qualities. Overall, the possibilities for such an automated system are open-ended. To make the research objectives and intermediate benefits of developing an AVAC system concrete, we detail three associated research agendas at increasing levels of abstraction in Section 4: representation learning, predictive modeling and decision-making, and human factors.

4. Football for AI Research

We next consider the dual perspective of the potential unique challenges and opportunities associated with the football domain that make it an interesting testbed for AI research. We

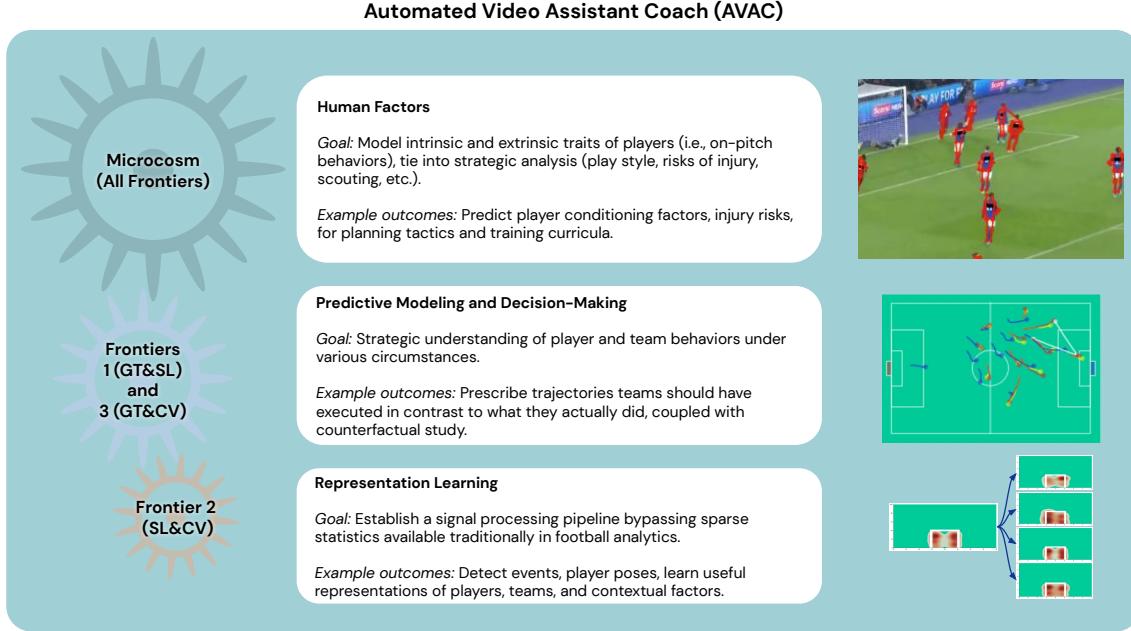


Figure 5: Hierarchy of key research challenges, defined over three layers. The foundational layer targets representation learning over the various input modalities available in football to generate useful representations for more complex learning tasks targeted in the subsequent layers: prescriptive and predictive analysis and modeling of human factors.

introduce here a hierarchy of key challenges associated with football research, illustrated in Figure 5, as defined over three layers: the foundational layer concerns representation learning, operating directly on the various input modalities available in football (e.g., time-synchronized videos, event-streams, and tracking data) to generate useful representations for the more complex learning tasks targeted in the subsequent layers of prescriptive and predictive analysis, and modeling of human factors. We next detail each of these research layers, drawing connections with the three aforementioned frontiers.

4.1 Representation Learning

The variety of hand-labeled football statistics make for natural fodder for machine learning algorithms. These algorithms range from classification and regression tools (e.g., in expected possession value models, Fernández, Bornn, & Cervone, 2019), generative models (e.g., in trajectory generation models, H. Le et al., 2017; H. M. Le, Yue, Carr, & Lucey, 2017; Li, Jiang, Zhu, Che, & Liu, 2020; Yeh, Schwing, Huang, & Murphy, 2019), and variational auto-encoding models (player embeddings). The success of machine learning algorithms generally depends on data representation, as different representations can entangle and obfuscate various explanatory factors of variation behind low-level sensory data. In football analytics, although expert knowledge is widely used to help design existing representations, learning with generic priors bears promise for avoiding such hand-encoded knowledge. Under this view, we identify three unique challenges related to representation learning, detailed next.

The first challenge concerns learning representations with multi-modal football data. Particularly, in football analytics, it remains a fundamental challenge to effectively recognize long-duration playing styles of individual players and teams given the variety of data types available (as detailed earlier). While expert knowledge goes a long way towards analyzing these multi-modal data sources, it remains insufficient to process them efficiently. The increasing multitude of input modalities available to football analysts are likely to challenge existing methods of representation learning, thus driving researchers to develop cohesive models that take these many modalities into account simultaneously.

The second challenge concerns learning contextual representations of individual players. Due to the dynamics and uncertainty of football outcomes, long-term static representations for predictive modeling of in-game events are likely to be beneficial when used in conjunction with representations of individual players. For example, a player passing the ball may take into account the context of the game to estimate the most appropriate receiver that maximizes the probability of scoring. Another concrete example is using contextual representations to identify the dynamic roles of players, which may change given the game context and must be inferred and controlled to tactically counter the opposing team. Finally, player behaviors depend not only on the game context, but also on their and the opposing team's overall strategy (e.g., formations, tactical advice provided by the coaching staff, etc.).

Finally, in addition to learning representations of individual players, identifying an effective means of contextualizing or ranking entire teams is another unresolved challenge. Teams are usually ranked with historical match results and the collective performance of individual players, which can be coarse (i.e., may not reveal the long-term playing styles of teams) and may fail to reflect in-game dynamics. Overall, to tackle these challenges, we aim to achieve two goals: i) learning representations that are able to characterize the long-term playing styles of football teams, ii) learning contextual representations of football teams that are able to depict in-game dynamics.

4.2 Predictive Modeling and Decision-Making

Learning useful representations (i.e., as opposed to hand-coded features) serves as an important means of advancing subsequent predictive and prescriptive analysis of football matches. Specifically, dense embeddings that summarize not only the state of a particular game, but also historical trends evident throughout many games (e.g., across seasons) will serve as enablers of the more accurate, impactful, and longer-horizon predictions of match outcomes. The interaction between predictive-prescriptive models is envisioned to be tightly-linked with game-theoretic analysis, thus coupling this direction of research most closely with Frontier 3 (GT&CV) and Frontier 1 (GT&SL) (see Figure 5).

The combination of these fields with game theory is likely to usher in new opportunities for coaches and decision-makers. For example, predictive models of football players at the trajectory-level (H. Le et al., 2017; H. M. Le et al., 2017; C. Sun, Karlsson, Wu, Tenenbaum, & Murphy, 2019) currently treat the game as a black-box dynamical process (i.e., a system of dynamic entities making decisions solely based on the joint on-pitch state of the teams); such models do not yet account for the game-theoretically driven counterfactual responses of players to one another (e.g., taking into account the current game score, time

remaining, relative strength of the two teams, impact of current game decisions on upcoming matches, etc.). Conducting such an analysis of these models involves identification of high-level strategies typically used by empirical game-theoretic techniques (so-called meta-strategies) (Tuyls et al., 2020; Wellman, 2006). These meta-strategies, for example, could be clusters of on-pitch behaviors correlated with play style, counterattack types, defense schemes (such as zonal vs. man-to-man defense), and so on. While such meta-strategies are typically manually defined, automatically learning them poses an interesting challenge. Appropriate clustering and identification of such meta-strategies involves not only access to a large, representative dataset of plays, but also the aforementioned learned representations that summarize the most relevant context for game theory models.

Synthesis of empirical games over the possible meta-strategies of two opposing teams can be used to forecast the performance of various team tactics when pitted against one another (e.g., investigating the Nash equilibrium of football at a higher level, rather than the typically considered low-level scenarios such as penalty kicks). Moreover, while characterization and ranking of players has received considerable attention in the literature (Bransen & Van Haaren, 2019; Decroos et al., 2019; Decroos & Davis, 2019), automated ranking of tactics has received considerably less attention (Decroos, Van Haaren, & Davis, 2018; Meerhoff, Goes, Knobbe, et al., 2019). Application of game-theoretic analysis techniques here remains unexplored to the best of our knowledge. Analysis of empirical games using meta-strategies conditioned on player identities would be beneficial for counterfactually evaluating player performance in new teams (i.e., for scouting). For training staff, a model that enables accurate evaluation of players' contributions to the team's overall strategy would be valuable, for example, for pinpointing which players to coach or to substitute. For broadcasters, automatic identification of salient, exciting meta-strategies (e.g., those that are typically low in probability yet high in payoff, or games where there is a large difference in terms of the play styles or meta-strategies of the teams) can be used for automatic generation of highlight reels.

Learning the appropriate meta-strategies and associated predictive models are, simultaneously, challenging in football due to the number of players involved on-pitch (and the exponential size of the strategy space with respect to this quantity). Despite this, the development of richer models leveraging more complex input modalities (e.g., video-based representations) is likely to unlock commensurate benefits (in terms of improved predictions and strategic decision-making) for football analysts.

4.3 Human Factors

The human-centric nature of football analytics stems from several factors: coaching and improvement of individual play and coordinated team play through predictive and prescriptive modelling, injury and fatigue prediction, and psychological analysis of players, which can affect the decision-making of players and lead to non-rational behaviors, an aspect modeled by behavioral game theory (Camerer, 2011; Camerer et al., 2005). This focus distinguishes it sharply from, for example, challenges such as RoboCup (*RoboCup project*, 2020). In contrast to the robot-centric focus of RoboCup (which revolves around developing robotic footballing agents, Hanna & Stone, 2017; Hausknecht & Stone, 2016; Kalyanakrishnan & Stone, 2010; Stone, Sutton, & Kuhlmann, 2005; Urieli, MacAlpine, Kalyanakrishnan, Ben-

tor, & Stone, 2011; Visser & Burkhard, 2007), the focus in football analytics is entirely on understanding and improving human gameplay and team coordination based on an integrated approach from the three research areas involved. Another key difference concerns evaluation of the impact said analysis on human play, which is distinct from evaluation of robotic agents in the RoboCup project. Namely, human play is significantly more difficult to realistically simulate and systematically evaluate (in contrast to evaluation on robotics platforms). Moreover, the football analytics frontiers targeted here entail taking into account human factors such as injury and fatigue, but also inter-player relationships and their effects on play efficiency and cooperation, psychological challenges such as pressure or mental state, notably on recent transfers, and their impact on play performance, and overall player discipline and tendency to follow the best plan for the team instead of the best plan for themselves.

Injury prediction is another topic of interest. Injury prediction is the task of predicting the probability that a player will sustain an injury given data on the past and current seasons. Previous studies have investigated the acute-chronic workload ratio (ACWR) as a predictor for sports-related muscular injuries (Gabbett, 2010). Acute workload measures an athlete's workload over 1 week, while chronic workload is the average workload over the past 4 weeks. Rossi et al. (2018) use richer workload measures extracted from Electronic Performance and Tracking Systems (EPTS) to train a decision tree to predict the probability of future injuries. Their approach uses manually designed features that aim to temporally aggregate players' workload histories. Recent work uses Convolutional Neural Networks (CNNs) applied to EPTS time-series data directly, thereby alleviating the need for hand-designed time-aggregated features (Gabbett, 2010). Current injury prediction methods are limited to a binary signal and do not explicitly capture uncertainty of the prediction, the type of injury, the severity, nor projected time to recover. Progress in this direction is likely limited by the availability of data; preventive measures are taken by sports clubs and as a result injuries occur relatively infrequently, although accurate prediction of such injuries (or determination of whether current performance sensors are sufficient for doing so) is a promising avenue for application of AI techniques.

5. Illustrative Examples: Frontier 1 (GT&SL)

In this section, we highlight some of the benefits that the combination of frontiers can yield for football analytics. We focus these examples on Frontier 1 (GT&SL), in particular, given the track record of game theory and statistical learning work done in football analytics in recent years; thus, this section provides a concrete sampling of the types of multi-disciplinary contributions that can be made via the proposed microcosm-centric vision. In the following, we first provide an overview of the necessary game theory background. Subsequently, we conduct an in-depth analysis of real-world football data under the lens of Frontier 1 (GT&SL), providing new insights into penalty kick scenarios by combining statistical learning with game theory.

5.1 Game Theory: Elementary Concepts

Empirical game theory has become an important tool for analysis of large-scale multi-agent settings, wherein either a large amount of data involving agent interactions is readily avail-

$$\begin{pmatrix} a_{11}, b_{11} & a_{12}, b_{12} \\ a_{21}, b_{21} & a_{22}, b_{22} \end{pmatrix}$$

Figure 6: Bimatrix payoff tables (A, B) for a two-player, two-action NFG. Here, the row player chooses one of the two rows, the column player chooses one of the columns, and the outcome of their joint action determines the payoff to each player.

able, or is collected through simulations of the system under study for the construction of the games (Tuyls et al., 2020; Wellman, 2006). Empirical game-theoretic modeling of penalty kick taking and set pieces facilitates strategic understanding of player and team behavior under various circumstances (e.g., play according to a Nash equilibrium), and can assist both in predicting opponent behavior and prescribing how a player (or team) should behave in the presence of other players (teams). These game-theoretic models can be leveraged in pre- and post-match analysis, and can be combined with analysis of dynamic trajectory behavior (e.g., generative trajectory prediction or ‘ghosting’, as later described). Additionally, the models can be enriched by carrying out Empirical Game Theoretic Analysis (EGTA) on meta-game models of set pieces, automatically clustering and identifying useful meta-strategies, and providing insights into higher-level team strategies.

A common representation of a game used for EGTA analysis is a Normal Form Game (NFG), defined in the following.

Definition 1 (Normal Form Games (NFG)) *A game $G = \langle P, (S_i), (u_i) \rangle$ consists of a finite set of players, P , indexed by i ; a nonempty set of strategies S_i for each player; and a utility function $u_i : \times_{j \in P} S_j \rightarrow \mathbb{R}$ for each player.*

In this work, we solely focus on bimatrix games, which are 2-player NFGs, $G = \langle P, (S_1, S_2), (u_1, u_2) \rangle$ with $|P| = 2$. The utility functions (u_1, u_2) can be described in terms of two payoff matrices A and B , wherein one player acts as the row player and the other as the column player. Both players execute their actions simultaneously. The payoffs for both players are represented by bimatrix (A, B) , which gives the payoff for the row player in A , and the column player in B (see Figure 6 for a two strategy example).

A player i may play a pure strategy, $s_i \in S_i$, or a mixed strategy, $\sigma_i \in \Delta(S_i)$, which is a probability distribution over the pure strategies in S_i . In a strategy profile $\sigma = (\sigma_1, \sigma_2)$, each player has a strategy σ_i . We use notation σ_{-i} to denote a strategy profile for all players excluding i . Having defined NFGs, we can model empirical games as an NFG wherein player payoffs are directly computed from data of real-world interactions or simulations. For example, one can construct a win-loss table between two chess players when they both have access to various strategies.

Given an NFG, the traditional solution concept used in game theory is the Nash equilibrium, which selects strategy profiles such that no player can benefit from unilateral deviation:

Definition 2 (Nash Equilibrium) *A strategy profile σ^* is a Nash equilibrium if and only if,*

$$\mathbb{E}[u_i(\sigma_i^*, \sigma_{-i}^*)] \geq \mathbb{E}[u_i(s'_i, \sigma_{-i}^*)] \quad \forall i \quad \forall s'_i \in S_i. \quad (1)$$

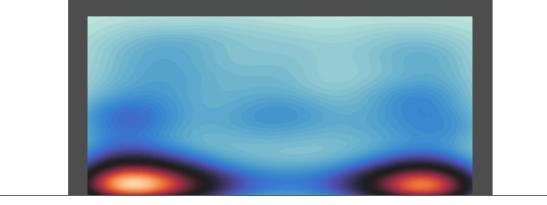


Figure 7: Visualization of shot distribution for the penalty kicks in the considered dataset.

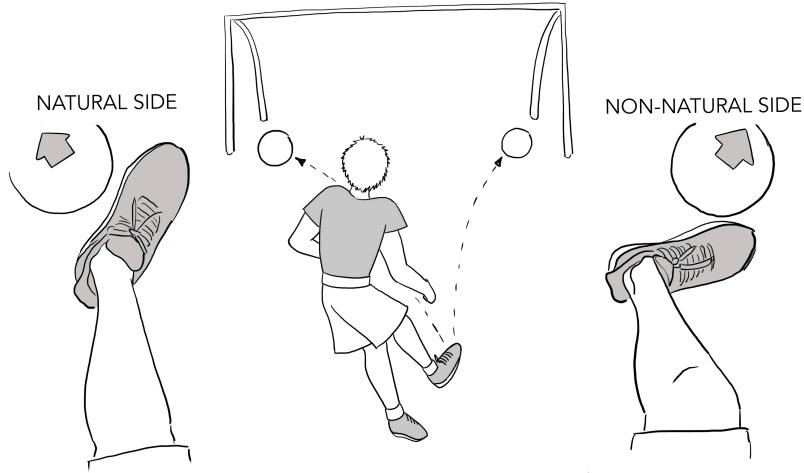


Figure 8: Illustration of natural vs. non-natural sides.

In a so-called ϵ -Nash equilibrium, there exists at least one player who could gain by deviating to another strategy, but that gain is bounded by ϵ . More formally:

Definition 3 (ϵ -Nash Equilibrium) *A strategy profile σ^* is an ϵ -Nash equilibrium if and only if there exists $\epsilon > 0$ such that,*

$$\mathbb{E}[u_i(\sigma_i^*, \sigma_{-i}^*)] \geq \mathbb{E}[u_i(s'_i, \sigma_{-i}^*)] - \epsilon \quad \forall i \quad \forall s'_i \in S_i. \quad (2)$$

5.2 Game Theory for Penalty Kick Analysis

For our analysis we use a data set of 12399 penalty kicks based on Opta data (Opta, 2020). In Figure 7 we show a heatmap of the shot distribution of the penalty kicks in our data set. Table 1 shows the distribution of the penalty kicks over the various leagues we consider.

Palacios-Huerta (2003) examines penalty kick scenarios from a game-theoretic perspective, using empirical payoff tables to determine whether the associated kickers and goalkeepers play a Nash equilibrium. Here we revisit the work of Palacios-Huerta (2003), by first reproducing several of its key results with a substantially larger and more recent data set from the main professional football leagues in Europe, Americas, and Asia (for comparison, the data set used in the work of Palacios-Huerta (2003) consists of 1417 penalty kicks from the 1995-2000 period, whereas ours contains 12399 kicks from the 2011-2017

Table 1: Penalty kick distribution over leagues considered (12399 kicks in total).

League	# Kicks	League	# Kicks
Italian Serie A	607	Chile Primera (Apertura)	151
US Major League Soccer	575	Japanese J-League	149
Engl. Npower Champ.	569	English League 1	139
Spanish Segunda Division	568	English League 2	130
Spanish La Liga	531	Austrian Bundesliga	129
French Ligue 1	497	Danish Superligaen	115
German DFB Pokal	441	European World Cup Qualifiers	108
Brazilian Série A	440	Internationals	93
Engl. Barclays Premier League	436	African Cup of Nations	90
German Bundesliga	409	United Soccer League	80
Dutch Eredivisie	398	Europ. Championship Qualifiers	78
German Bundesliga Zwei	389	Swedish Allsvenskan	74
Portuguese Primiera Liga	352	Coppa Italia	67
Saudi Arabian Profess. League	337	Copa America	51
Russian Premier League	329	FIFA Club World Cup	51
Chinese Super League	324	World Cup	48
Copa Libertadores	322	Europ. Championship Finals	45
Belgian Jupiler League	287	Champions League Qualifying	41
Turkish Super Lig	284	Confederations Cup	39
French Ligue 2	270	UEFA Europa League Qualifying	32
Argentina Primera (Anual)	261	Coupe de France	29
English Capital One Cup	234	Belgian UEFA Europa Play-offs	24
Mexican Primera (Clausura)	234	German 3rd Liga	23
Colombia Primera Apertura	221	Russian Relegation Play-offs	15
Norwegian Tippeligaen	219	Dutch Relegation Play-offs	13
AFC Champions League	193	Copa Sudamericana	9
International Champions Cup	188	Friendly	4
Australian A-League	172	German Bundesliga Playoff	3
Copa Chile	172	German Bundesliga 2 Playoff	3
English FA Cup	153	Swedish Relegation Play-off	1
Copa do Brasil	153		

period). While several results of this earlier work are corroborated, we also find surprising new additional insights under our larger dataset. We then go further to extend this analysis by considering larger empirical games (involving more action choices for both kick-takers and goalkeepers). Finally, we develop a technique for illustrating substantive differences in various kickers' penalty styles, by combining empirical game-theoretic analysis with Player Vectors (Decroos & Davis, 2019) illustrating the added value and novel insights research at Frontier 1 (GT&SL) of the microcosm can bring to football analytics.

As in Palacios-Huerta (2003)'s work, we first synthesize a 2-player 2-action empirical game based on our penalty kick data set. Table 2a illustrates the 2×2 normal form

Table 2: Natural (N) / Non-Natural (NN) payoff tables for Shots (S) and Goalkeepers (G). Here, (c) and (d) compare Nash and empirical (Empir.) probabilities.

(a) Palacios-Huerta (2003) payoff table.

	N-G	NN-G
N-S	0.670	0.929
NN-S	0.950	0.583

(b) Reproduced table.

	N-G	NN-G
N-S	0.704	0.907
NN-S	0.894	0.640

(c) Palacios-Huerta (2003) Nash probabilities.

	NN-S	N-S	NN-G	N-G
Nash	0.393	0.607	0.432	0.568
Empir.	0.423	0.577	0.400	0.600

Jensen–Shannon divergence: 0.049%

(d) Reproduced table Nash probabilities.

	NN-S	N-S	NN-G	N-G
Nash	0.431	0.569	0.408	0.592
Empir.	0.475	0.525	0.385	0.615

Jensen–Shannon divergence: 0.087%

as presented by Palacios-Huerta (2003). The actions for the two players, the kicker and goalkeeper, are respectively visualized in the rows and columns of the corresponding payoff tables, and are detailed below. The respective payoffs in each cell of the payoff table indicate the win-rate or probability of success for the kicker (i.e., a score); for ease of comparison between various payoff tables, cells are color-graded in proportion to their associated values (the higher the scoring probability, the darker shade of green used).

The choice of player actions considered has an important bearing on the conclusions drawn via empirical game-theoretic analysis. The actions used by Palacios-Huerta (2003) in Table 2a correspond to taking a shot to the natural (N) or non-natural (NN) side for the kicker, and analogously diving to the natural side or non-natural side for the goalkeeper. Figure 8 provides a visual definition of natural versus non-natural sides. Specifically, as players tend to kick with the inside of their feet, it is easier, for example, for a left-footed player to kick towards the right (from their perspective); thus, this is referred to as their natural side. Analogously, the natural side for a right-footed kicker is to kick towards their left. The natural side for a goalkeeper depends on the kicker in front of him. Specifically, when facing right-footed kickers, goalkeepers’ natural side is designated to be their right; vice versa, when they face a left-footed kicker, their natural side is to their left. Importantly, shots to the center count as shots to the natural side of the kicker, because, as explained in Palacios-Huerta (2003), kicking to the center is considered equally natural as kicking to the natural side by professional football players (Palacios-Huerta, 2003).

Table 2b shows our reproduction of Table 2a of Palacios-Huerta (2003), computed using 12399 penalty kicks spanning the aforementioned leagues in our Opta-based dataset; importantly, players (goalkeepers and kickers) appear at least 20 times each in this dataset, to ensure consistency with Palacios-Huerta (2003). The trends in these two tables are in agreement: when the goalkeeper and the kicker do not choose the same sides of the goal, shot success rate is high; otherwise, when the keeper goes to the same side as the kicker, success rate is higher for natural shots than for non-natural shots. We also include Nash and empirical probabilities for Palacios-Huerta’s dataset and ours, respectively in Tables 2c

Table 3: Natural / Non-natural game restricted by footedness.

(a) Left-footed players payoff table		(b) Right-footed players payoff table			
	N-G	NN-G			
N-S	0.721	0.939	N-S	0.700	0.898
NN-S	0.903	0.591	NN-S	0.892	0.653

Table 4: Footedness equivalence p-value tables.

(a) Natural / Non-natural game p-values		(b) Left / Center / Right game p-values			
	N-G	NN-G	R-G	C-G	
N-S	0.924566	0.170504	0.000011	0.947369	6.931197e-01
NN-S	0.394900	0.407741	0.592054	0.868407	1.305657e-01

and 2d, enabling us to conclude that payoffs, Nash probabilities, and empirical probabilities are all in agreement between Palacios-Huerta’s results and our reproduction; more quantitatively, the Jensen-Shannon divergence between Palacios-Huerta’s results and ours is 0.84% for the Nash distribution and 1.2% for the empirical frequencies. We also notice that players’ empirical action selection frequencies are quite close to the Nash-recommended frequencies, as measured by their Jensen-Shannon Divergence, and are actually playing an ϵ -Nash equilibrium with a very low ϵ of 0.4%.

Having examined the similarity of payoff tables and distributions, we verify whether the Natural / Non-Natural game is statistically identical for left-footed and right-footed players (Table 3), as assumed in Palacios-Huerta (2003). To do so, we use a t-test to verify whether per-cell scoring rates are identical across footedness types. The t-tests’ p-values are reported in Table 4a, and reveal that the games cannot be proven to be dissimilar across footedness and can, therefore, be assumed to be identical for left-footed and right-footed players. Figure 9 refines this result by representing the relationship between p-values of our t-test and minimal player appearance counts: when we modulate minimal appearance count of players in our test, the Natural Shot / Natural Goalkeeper cell goes from strongly dissimilar across footedness (low p-value) when including all players, to likely non-dissimilar (high p-value) when only including the players appearing the most in our dataset. This could be explained by low-appearance-counts-, which we take here as a proxy for low experience, kickers being less able to control their kicks, resulting in different control effectiveness for different footedness preferences, and in goalkeepers being less proficient in stopping shots going to their less frequently-kicked side (left) than to the other, a preference that we infer has been trained away in professional goalkeepers. To remove potential side-effects of merging data from low- and high-experience players together, Figure 10 shows the relationship between p-values of our t-test and experience category where we allow for some overlap—between 1 and 7 shots, 5 and 12, etc.; the insight drawn from this figure is the

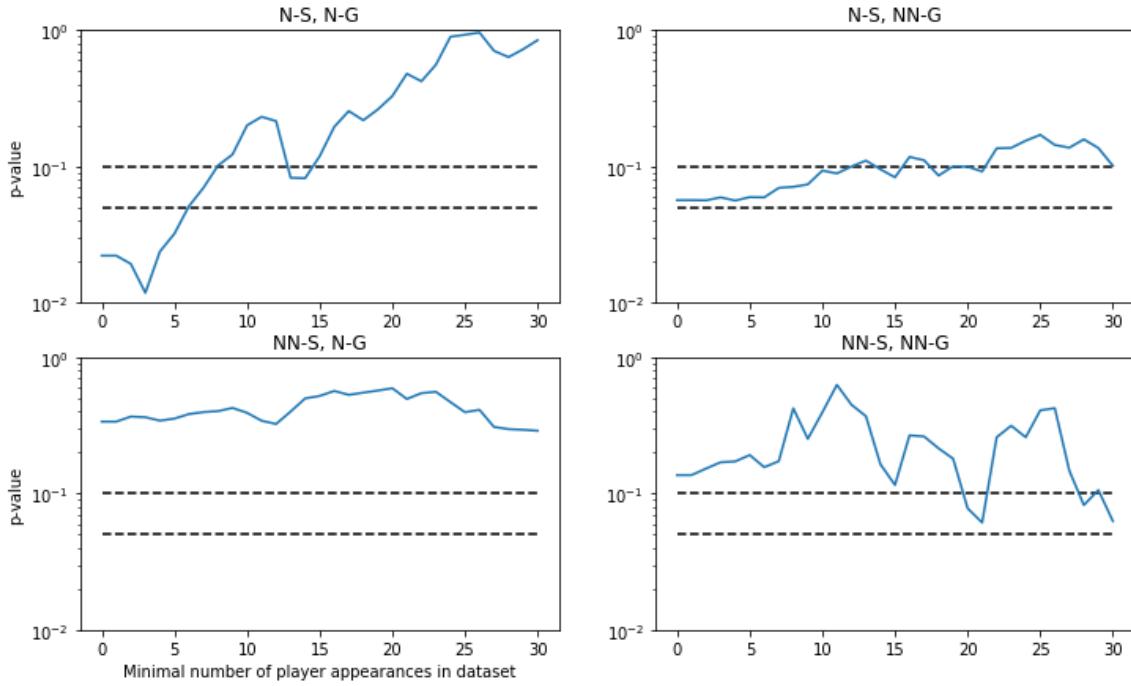


Figure 9: P-value table as a function of minimal experience.

Table 5: Left (L) - Center (C) - Right (R) tables for Shots (S) and Goalkeepers (G), with the three directions of kick/movement defined from the goalkeeper's perspective.

(a) Payoff table.

	R-G	C-G	L-G
R-S	0.684	0.939	0.969
C-S	0.964	0.160	0.953
L-S	0.964	0.960	0.633

(b) Nash probabilities vs. Empirical frequencies corresponding to (a).

	R-S	C-S	L-S	R-G	C-G	L-G
Nash	0.478	0.116	0.406	0.441	0.178	0.381
Empirical	0.454	0.061	0.485	0.475	0.089	0.436

Jensen–Shannon divergence: 0.75%

same as that of Figure 9, supporting the conclusion that experience removes the difference between left- and right-footed penalty kicks.

We also analyzed the game defined by kicking to the left, center, or right, and confirmed Palacios-Huerta's intuition that it is fundamentally different across footedness preferences. Specifically, Table 5a synthesizes the empirical game corresponding to this new choice of

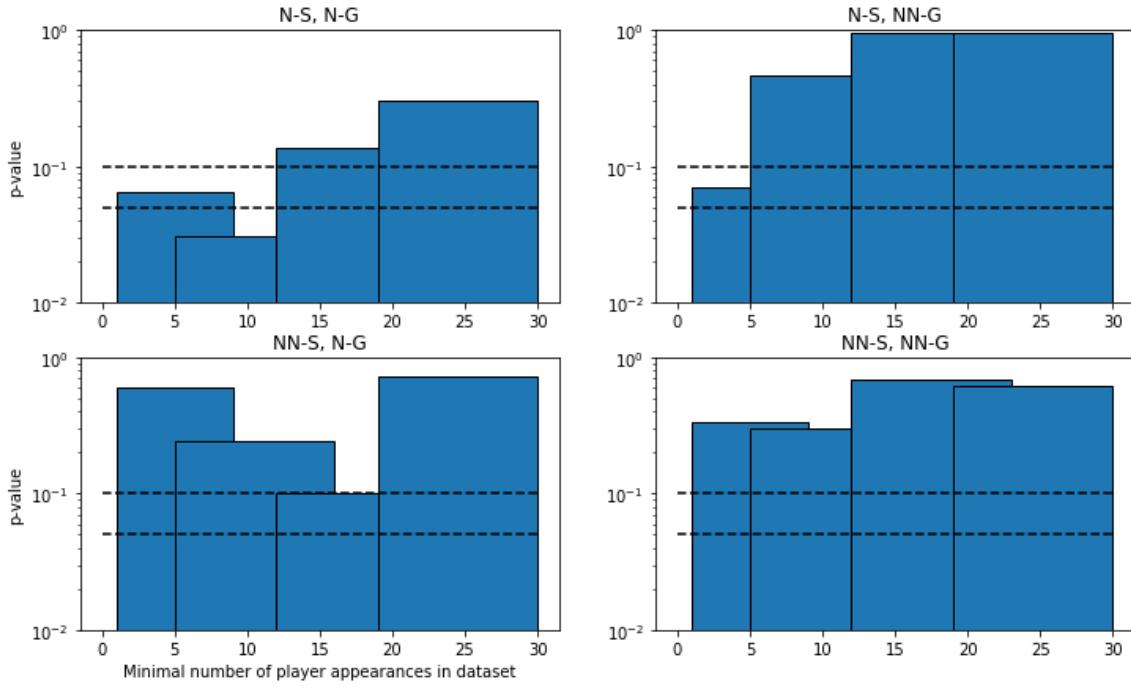


Figure 10: P-value table as a function of player-experience.

actions, with aggregated scoring rates over both feet preferences. Note that in this case, left, center, and right are measured from the goalkeeper’s perspective, such that the natural kick of a right-footed player would be considered a right kick. The per-cell t-tests’ p-values for this game are reported in Table 4b. Interestingly, the game is different when the goalkeeper jumps to the same side as the ball, but is otherwise mostly similar across footedness preference. The empirical play frequencies for kickers, as reported in Table 5b, are also further away from Nash frequencies than observed in the Natural / Non-Natural game (Table 2d), as can be seen from the Jensen-Shannon divergence between empirical frequencies and Nash (0.75%, versus the the 0.087% of the Natural / Non-Natural game) These insights indeed confirm the intuition that such a game is neither correct across footedness, nor the one the players follow.

Overall, these results provide insights into the impacts that the choice of actions have on conclusions drawn from empirical payoff tables, and are illustrative of the practical usefulness of theoretically well-founded principles in application to real-world analytics, which is highlighted as a hallmark of useful theory in Szymanski (2020). However, behavior and shooting styles also vary wildly per-player given footedness. If one is willing to consider several payoff tables (e.g., one per footedness), it seems natural to also take into account kickers’ playing styles, as considered in the next section.

5.3 Augmenting Game-Theoretic Analysis of Penalty Kicks with Embeddings

While the previous section undertook a descriptive view of the penalty kick scenario (i.e., providing a high-level understanding of kicker and goalkeeper play probabilities), here we

Table 6: Cluster statistics.

	# Players	# Goals	# Shots	Success %	Proportion left-foot goals (%)
Cluster 1	197	144	167	86.2	10.4
Cluster 2	216	494	612	80.7	21.9
Cluster 3	52	3	4	75.0	33.3
Cluster 4	82	58	73	79.4	51.7
Cluster 5	87	44	60	73.3	34.1.0
Cluster 6	1	0	0	-	0.0
Total	635	743	916	81.1	25.2

Table 7: Pair-wise comparison for the identified clusters. < indicates that data was missing and minimum true p-value may be lower than the reported minimum p-value in the table. The symbol * indicates we cannot conclude whether clusters are different at the 10% confidence level.

	1 vs. 2	1 vs. 4	1 vs. 5	2 vs. 4	2 vs. 5	4 vs. 5
Min. cell p-value of t-test over table equality	4.49e-2	< 9.56e-2*	< 1.09e-1*	4.49e-2	4.48e-2	< 3.39e-1*
Jensen-Shannon div. between Nash distr. (%)	0.03	0.57	0.09	0.35	0.02	0.21
Jensen-Shannon div. between empirical distr. (%)	0.06	0.01	0.06	0.08	0.24	0.04
Left footedness t-test p-value	3.43e-4	1.37e-7	3.18e-3	4.92e-5	1.07e-1	7.52e-2

investigate whether we can find the best strategy for a player given the knowledge of the kicker’s play style. In game-theoretic terms, we conduct a prescriptive analysis of penalty kicks to enable informed decision-making for players and coaching staff in specific penalty kick situations. Ideally, one would iterate the earlier empirical payoff analysis for every possible combination of goalkeeper and kicker in a given league, thus enabling decision-making at the most granular level; however, the inherent sparsity of penalty kick data makes such an approach infeasible. Instead, we introduce a meaningful compromise here by combining statistical learning with game theory (i.e., Frontier 1 (GT&SL)), first quantifying individual playing styles, then using clustering techniques to aggregate players (i.e., both strikers and goalkeepers) based on said styles, and finally synthesizing empirical games for each identified cluster. We focus our analysis on penalties including all players who participated in Premier League matches from 2016 to 2019.

Table 8: p-values for t-test that empirical action distributions are equal among different clusters. Minimum p-value (across kicker and goalkeeper roles) is indicated in bold for each row.

Kicker clusters compared	Kicker p-value	Goalkeeper p-value
1 vs. 2	0.52	0.05
1 vs. 4	0.85	0.95
1 vs. 5	0.42	0.27
2 vs. 4	0.52	0.14
2 vs. 5	0.51	0.16
4 vs. 5	0.4	0.26

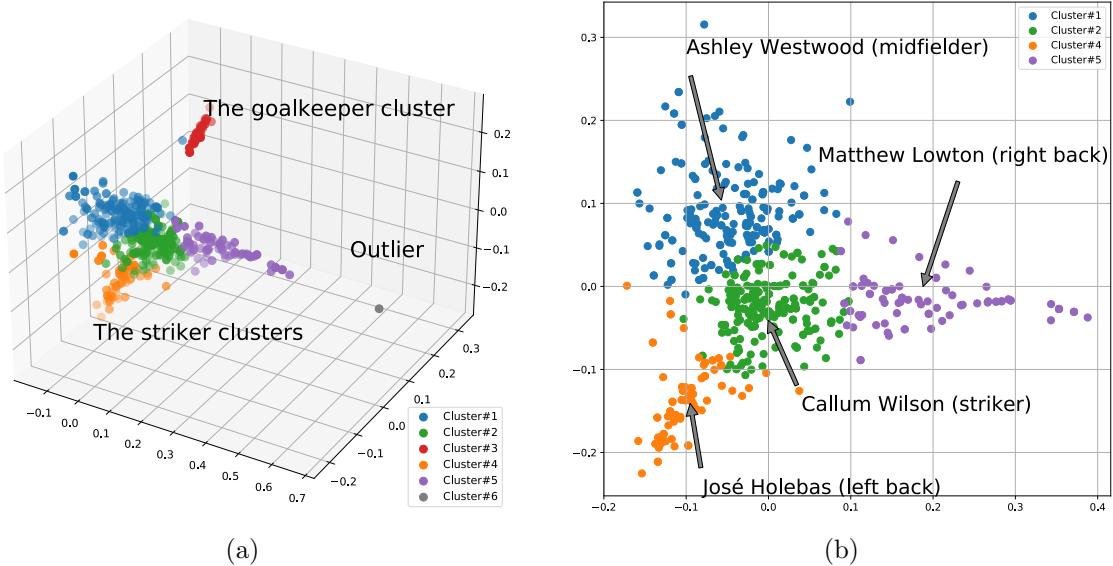


Figure 11: Visualization of the identified player clusters. (a) visualizes the keeper cluster, the kicker clusters and an outlier automatically detected through K-means clustering. To show the separation of the kicker clusters clearly, we visualize them in (b) after removing the keeper and outlier clusters, and we also label each cluster with a Premier League player in it.

On a technical level, our approach consists of the three following steps. First, we characterize the playing style of a player in a manner that can be interpreted both by human experts and machine learning systems. In particular, we use Player Vectors (Decroos & Davis, 2019) to summarize the playing styles of kickers using an 18-dimensional real-valued vector. These Player Vectors are extracted from historical playing trajectories in real matches, with technical details provided in Appendix C. Each dimension of the Player Vector corresponds to individual on-pitch player behaviors (e.g., styles of passes, take-ons, shots, etc.), and the value of each dimension is standardized and quantifies the weight of that particular action style for the considered player. We also filter experienced players with at least 50

Table 9: Nash probabilities and empirical (Empir.) frequencies tables for Shot (S) and Goalkeepers (G) with Natural (N) and Non-Natural (NN) actions. Note that Cluster 3 is omitted due to it consisting of very few shots (taken by goalkeepers).

(a) All players. 916 total shots.

	NN-S	N-S	NN-G	N-G
Nash	0.391	0.609	0.406	0.594
Empir.	0.503	0.497	0.413	0.587

ϵ -Nash equilibrium: $\epsilon = 2.71\%$

(b) Kickers in Cluster 1. 167 total shots.

	NN-S	N-S	NN-G	N-G
Nash	0.423	0.577	0.379	0.621
Empir.	0.485	0.515	0.371	0.629

ϵ -Nash equilibrium: $\epsilon = 0.08\%$

(c) Kickers in Cluster 2. 612 total shots.

	NN-S	N-S	NN-G	N-G
Nash	0.401	0.599	0.430	0.570
Empir.	0.520	0.480	0.418	0.582

ϵ -Nash equilibrium: $\epsilon = 2.89\%$

(d) Kickers in Cluster 4. 73 total shots.

	NN-S	N-S	NN-G	N-G
Nash	0.320	0.680	0.375	0.625
Empir.	0.479	0.521	0.438	0.562

ϵ -Nash equilibrium: $\epsilon = 5.17\%$

(e) Kickers in Cluster 5. 60 total shots.

	NN-S	N-S	NN-G	N-G
Nash	0.383	0.617	0.317	0.683
Empir.	0.450	0.550	0.400	0.600

ϵ -Nash equilibrium: $\epsilon = 4.86\%$

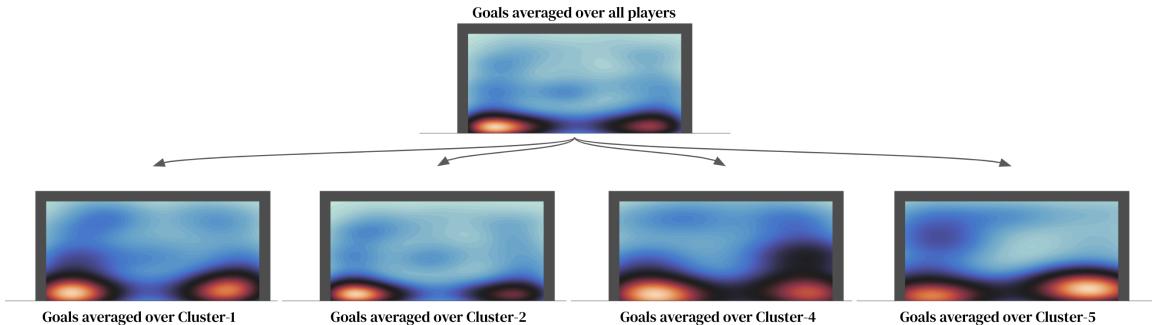


Figure 12: Heatmaps of goals by all kickers and kickers in individual clusters with respect to empirical probabilities. We exclude the goalkeeper cluster (Cluster 3) and the outlier cluster (Cluster 6) because of insufficient samples.

appearances in the Premier League matches from 2016 to 2019. In total, we obtain 635 such vectors for the individual players in our dataset. Second, we cluster players in accordance to their Player Vectors, using K-means with the number of clusters chosen as the value causing the most significant drop in inertia (a standard heuristic). This process yields 6 clusters in total, with statistics summarized in Table 6. In particular, K-means clustering

detects an outlier cluster with only one player (Cluster 6), and we also observe that there are very few shot samples in Cluster 3, as it consists of a cluster of goalkeepers (an interesting artifact illustrating the ability of Player Vectors and K-means clustering to discern player roles). Given the few samples associated with these two clusters, we henceforth exclude them from the game-theoretic analysis. We observe that cluster pairs (1, 2), (1, 4), (2, 4), and (2, 5) are significantly different, with the minimum cell-wise p-values for these cluster pairs smaller than 0.10 in Table 7. We therefore focus our game-theoretic analysis on these cluster pairs. Moreover, we also qualitatively illustrate differences between the clusters in Figures 11a and 11b, which visualize the results of reducing the Player Vectors dimensionality from 18 to, respectively, 3 and 2 via Principal Component Analysis. Here, we observe that the goalkeeper cluster is well-separated from the kicker clusters in Figure 11a, and in order to better visualize the kicker clusters, we project Figure 11a onto its x and y axis after removing the goalkeeper and outlier clusters in Figure 11b. We also identify therein the most representative kicker per-cluster (i.e., the player whose feature vector is closest to the mean of the corresponding cluster)

Finally, we conduct the aforementioned game-theoretic analysis for each cluster. In our earlier Table 6, we observe that the kickers in some clusters have different success rates in penalty kicks. Moreover, a closer behavioral analysis yields deeper insights. We first examine the Nash strategies played by each cluster, and then visualize the actual play behavior with respect to empirical probabilities in Figure 12. Table 9a summarizes the overall Nash distributions for all players considered, with Tables 9b to 9e showing cluster-specific distributions. These tables illustrate that the kickers have the same empirical behavior, an assertion statistically confirmed in Table 8; yet their Nash-derived recommendations are different: although kickers in all clusters are recommended by the Nash to shoot more to their natural sides than to their non-natural sides, the recommended strategy for kickers in Cluster 1 is actually quite balanced between natural and non-natural shots. This greater imbalance is shown by comparing Jensen-Shannon divergence. As we see in Table 7, the Jensen-Shannon divergence of the Nash probabilities between Cluster 1 and 4 (0.57%) is 6-7 times greater than that between Cluster 1 and 5 (0.09%) and 19 times greater than that between Cluster 1 and 2 (0.03%). We also notice that the clusters' players are all playing epsilon Nash equilibria with relatively low epsilon (Table 9). In other words, although their empirical strategies seem to deviate from corresponding Nash strategies action-wise, the expected payoffs of these two strategies are close, and they could still stand to gain in "stability" by switching to corresponding Nash strategy. Nevertheless, most of these Nash recommendations come from very low-sample empirical payoff tables, which entails potentially inaccurate Nash distributions. We nevertheless note that this low-data regime is induced by the restriction of our analysis to players having played in matches of Premier League only from 2016 to 2019. Obtaining Player Vector data for all players in our dataset would allow us to study cluster behavior with greater statistical precision. Nevertheless, the current study leaves no statistical doubt regarding the pertinence of clustering payoff tables using player embeddings—specifically Player Vectors.

Qualitatively, in addition to analyzing the strategies with respect to Nash probabilities, the patterns of positions of the ball of successful goals also vary from clusters to clusters, as visualized in Figure 12. For instance, kickers in Cluster 2 tend to score mostly to the

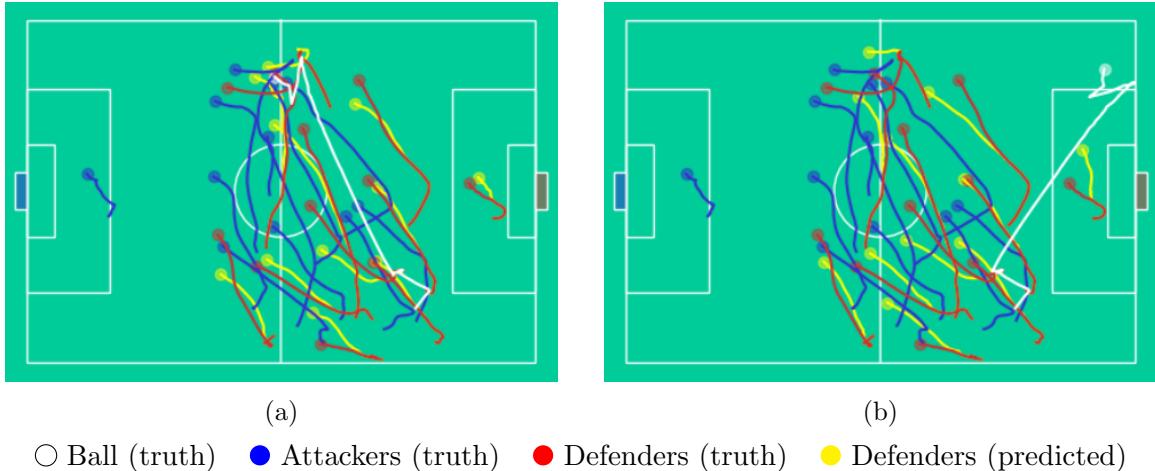


Figure 13: Predictive modeling using football tracking data. (a) visualizes predictions under the original data. Here, ground truth information for all players and the ball is provided to the underlying predictive model, with defender positions truncated and predicted by the model after a cut-off time (as indicated by the yellow traces). (b) illustrates the same scenario, after counterfactual perturbation of the ground truth ball direction to ascertain the predicted reaction of the defending goalkeeper (far right).

bottom left corner of the goalmouth, while the scoring positions in other clusters are more balanced, though these could also be partly due to lower sample sizes for some clusters.

5.4 Generative Trajectory Prediction Models for Counterfactual Analysis

Ghosting refers to the prescription of the trajectories the players in a sports team should have executed, in contrast to what they actually did (Lowe, 2013). Solution of this and the broader problem class of generative trajectory prediction implies benefits spanning from recommendation of trajectories or setups for constrained set pieces, then to short-term plays involving a subset of players, and eventually to long-term strategies/plays for the entire team. Team-level predictions would also strongly benefit from game-theoretic and multi-agent considerations, and is perceived to play a key role in an established AVAC system. We here present an illustrative example to ground the earlier discussion regarding the potential impacts of using learned predictive models to conduct counterfactual analysis of football matches.

For example, one might train a trajectory prediction model on league data (e.g., as done in H. M. Le et al. 2017), provide an input context to such a model (e.g., consisting of the true state of the ball, defenders, and attackers up to some point in time), and subsequently predict future trajectories of players. Figure 13a visualizes league-average predicted behaviors conditioned on such an input context. This illustrative example was trained using a baseline predictive model, similar to that of H. Le et al. (2017). Here we trained a centralized long-short term memory model (of 2 layers, each with 256 units), taking as input the raw trajectories of players and the ball, and predicting as output the step-wise change in trajectory of the defensive players. The model was trained on 240 frames

of 25 fps tracking data, downsampled to 12.5 fps, with half the frames in each play used for providing a prediction context, and the other half occurring at the prediction cut-off. We used the l_2 -loss on the tracking data for training, and randomized the order of attacking and defending players to avoid the role-assignment problem mentioned in H. Le et al. (2017) (similar to one of the baseline approaches of Yeh et al. 2019).

As pointed out in the literature (H. Le et al., 2017; H. M. Le et al., 2017; Li et al., 2020; Yeh et al., 2019), a key advantage of generative predictive models is that they can be used for counterfactual analysis of play outcomes. We illustrate such an example in Figure 13b, where we perturb the trajectory of the ball, inferring the subsequent behaviors of defenders in reaction (noting, e.g., the tendency of the goalkeeper to chase the ball in reaction to it entering the penalty area). While simple, case-by-case counterfactual case studies such as the above have been conducted to some extent in the literature, consideration of responses to more complex perturbations (e.g., changes of one team’s tactics or meta-strategy as a whole, changes in player behavior due to injuries, or changes due to substitutions of individual players) bear potential for significantly more in-depth analysis.

6. Discussion

Football analytics poses a key opportunity for AI research that impacts the real world. The balance of its reasonably well-controlled nature (versus other physical domains beyond sports, e.g., search-and-rescue), considerations associated with human factors (e.g., heterogeneous skill sets, physiological characteristics such as injury risks for players, etc.), and the long-term cause-and-effect feedback loop due to the relative infrequency of scoring even in professional play make it a uniquely challenging domain. Nonetheless, the rapidly-emerging availability of multi-modal sensory data make it an ideal platform for development and evaluation of key AI algorithms, particularly at the intersection of the aforementioned fields of statistical learning, computer vision, and game theory.

In this paper, we highlighted three frontiers at the intersection of the above fields, targeting the simultaneous advancement of AI and football analytics. We highlighted the overlying goal of developing an Automated Video Assistant Coach (AVAC), a system capable of processing raw broadcast video footage and accordingly advising coaching staff in pre-, in-, and post-match scenarios. We subsequently illustrated how the combination of game theory and statistical learning could be used to advance classical results in football analytics, with an in-depth case study using a dataset comprised of over 15000 penalty kicks, and subsequently combined with the Player Vectors analysis of Decroos and Davis (2019) to discern kicking styles.

A notable observation for future work focusing on prescriptive football analytics is that the domain and some of the state-of-the-art research bear key similarities to RL. At a high level, the process of winning football championships can be cast as a sequential decision-making problem, with a concrete reward structure centered on three timescales of increasing abstraction: scoring goals, winning matches, and subsequently winning championships. We illustrate this view in Figure 14. Under this hierarchical view of football, each layer can be considered an RL problem at the designated level of abstraction. For example, at the lowest level, the sequential decisions made by teammates that lead to a goal can be considered a policy mapping states to actions, using the lexicon of RL. Likewise, estimates of the value

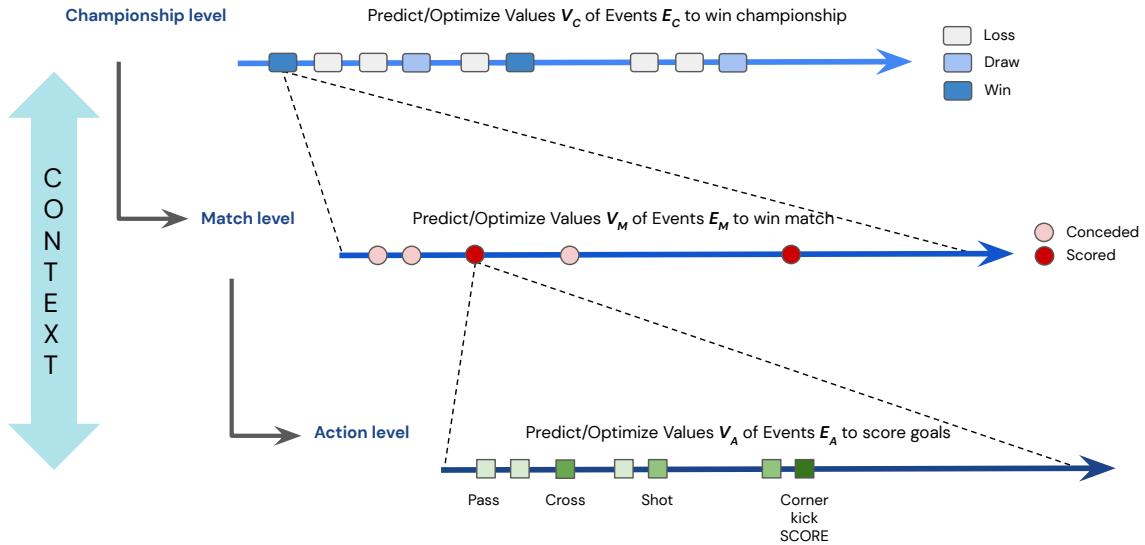


Figure 14: A multi-level view of football analytics cast as a reinforcement learning problem. We discern three levels: the top level aims to learn how to win championships by winning matches; the middle level optimizes for winning a match; finally, the bottom level seeks to optimize goal-scoring. The context between these various level is shared in both a top-down and bottom-up fashion.

of player actions based on the outcomes associated with actions taken in real games (as in VAEP, Decroos et al., 2019) can be considered analogous to those that learn action-values associated with RL policies. Further expanding this analogy, learning to quantify the contribution of individual players to a team’s estimated goal-scoring value can be cast as a so-called credit assignment problem, a key area of research in RL. Finally, given the presence of multiple on-pitch players with both cooperative and competitive incentives, the value function learning problem situates itself in the area of multi-agent RL. Multi-agent RL, critically, seeks to understand and learn optimal policies for agents in such interactive environments, linking also to game theory in providing the appropriate mathematical foundations to model this strategic process. As such, the multi-agent RL approach fits well under Frontier 1 (GT&SL), which considers the game-theoretic interactions of strategic players given specified payoffs, and use of learning techniques for identifying optimal policies. Moreover, this connection also highlights a potential overlap of interest between real-world football and RoboCup, in that the RL paradigm can be used to optimize player and robot policies alike, despite the widely-different player embodiments considered in each of these two fields. Overall, such parallels can be drawn at all levels of abstraction highlighted in the aforementioned hierarchical process modeling football championships, implying the foreseeable importance of the RL paradigm as football analytics shifts from understanding the game to subsequently optimizing player and team decisions at increasingly broader levels.

Moreover, the toolkits developed within the context of football analytics are also likely to have direct benefits for closely-related fields, and could be foreseeably adapted to many other

sports. Specifically, while we focus on football in this paper, analysis of team coordination behavior is also being conducted in a multitude of other sports (Albert, 2010; Albert et al., 2002; Anderson et al., 2020; Baumer & Zimbalist, 2014; Costa et al., 2009; Gauriot et al., 2016; Michael, 2004; Puerzer, 2002; Skinner, 2010; Song et al., 2017; Yee, Lis , & Bowling, 2016). Techniques developed for other sports could, in turn, be reciprocally applied to football. In particular, the element of team play is a key characteristic of football that is shared with sports such as basketball and hockey; in fact, the similarities between these sports are strong enough that for certain problem regimes, the same foundational techniques can be seamlessly applied across datasets for each sport (e.g., the work of Yeh et al. (2019) on player trajectory predictions for basketball and football). A related mapping concerns the application of football analytics techniques to the emerging field of eSports, wherein there is a large amount of data collected (in both raw video form, and structured data formats), e.g., such data streams are available for games such as Dota 2 or StarCraft. In Dota 2, for example, a coaching functionality analogous to that in football is available, wherein an experienced player is connected to the game and advises other players on various strategic tactics. Moreover, several of the most popular eSports games are inherently multi-player, in the sense that their outcomes are not determined by only an individual's skill, but a team's skill, mixing cooperative and competitive behaviors (as in football). Automatic analysis of games could provide insights into weak and strong points of teams, tactics used, and directions for improvement. These related domains could, therefore, provide a low-hanging fruit for football analytics techniques to generalize, in a seamless manner, beyond football.

Overall, the combination of data sources, downstream benefits on related domains, and potentials for impact that AI could have on the football domain are quite evident. Perhaps more importantly, the promising commensurate impacts of football analytics on AI research (through the feedback loop established between the football microcosm to the three foundational fields highlighted in Figure 1) are foreseen to make football a highly appealing domain for AI research in coming years.

Acknowledgments

The authors gratefully thank Thomas Anthony and Murray Shanahan for their helpful feedback during the paper writing process. The authors also thank the editor and three anonymous reviewers for their constructive feedback. Karl Tuyls and Shayegan Omidshafiei are equal contributors. Karl Tuyls is the corresponding author.

Appendix A. Additional Works Related to Statistical Learning in Football

Evaluating the effect of individual actions throughout the game is challenging as they naturally depend on the circumstances in which they were performed and have long-term consequences that depend on how the sequence plays out. Most works have focused on measuring the quality of specific action types in distinct concrete game situations (Barr, Holdsworth, & Kantor, 2008; Bransen & Van Haaren, 2018; Spearman, 2018). More recent work has focused on a unifying view in which actions are valued according to how they increase or decrease the likelihood of the play leading to a goal (Decroos et al., 2019;

Fernández, 2019). The main idea is to estimate the value of a given ‘state’ of the game. Intuitively, the state of a particular game includes everything that happened in the match until this point, including the score, identities of players and associated traits, time left on the clock, all prior actions, position of the players and the ball, etc.; moreover, one may wish to also consider the state of a tournament as a whole (e.g., previous and upcoming matches, the number of yellow cards accrued by players, etc.). A recent method used for assigning values to on-ball actions is known as *Valuing Actions by Estimating Probabilities* (VAEP) (Decroos et al., 2019). Actions are valued by measuring their effect on the game state and in turn the probabilities that a team will score. These scores can then be used to assess contribution of players to a team or measuring the mutual chemistry for a pair of players (Bransen & Van Haaren, 2019).

Finally, a promising application of statistical learning is the development of models that can carry out temporal predictions. This area is closely related to trajectory prediction (Alahi et al., 2016; Deo & Trivedi, 2018; Fernando, Denman, Sridharan, & Fookes, 2018; Gupta, Johnson, Fei-Fei, Savarese, & Alahi, 2018; Wang, Fleet, & Hertzmann, 2007). In the context of sports analytics, such trajectory prediction models can be useful for conducting the form of analysis known as *ghosting*, which, given a particular play, predicts the actions that a *different* team or player would have executed. Beyond just capturing game dynamics, models that can accurately carry out predictions could constitute valuable tools for counterfactual reasoning, which allows us to consider the outcomes of alternative scenarios that never actually took place. So far, such predictive models have been primarily used for predicting the trajectory of the ball (Maksai, Wang, & Fua, 2016) and of players themselves (H. Le et al., 2017; H. M. Le et al., 2017; Li et al., 2020; Su, Hajimirsadeghi, & Mori, 2019; Yeh et al., 2019). Also of importance are models which identify player roles from predicted trajectories (Felsen, Lucey, & Ganguly, 2018).

Appendix B. Pose Estimation

As previously illustrated, multi-person human pose estimation (Cheng, Yang, Wang, Yan, & Tan, 2019; Y. He, Yan, Fragkiadaki, & Yu, 2020; Iskakov, Burkov, Lempitsky, & Malkov, 2019; Lassner et al., 2017; Pavlakos et al., 2019; Pavlakos, Zhou, Derpanis, & Daniilidis, 2017; Pavllo, Feichtenhofer, Grangier, & Auli, 2019) is a central part of vision-based analysis of football video. Methods for this task can be grouped into two types: one the one hand, bottom-up approaches first detect human joints, and group them into pose instances (Fang, Xie, Tai, & Lu, 2017; K. He, Gkioxari, Dollár, & Girshick, 2017; Huang, Gong, & Tao, 2017; Iqbal & Gall, 2016; Papandreou et al., 2017; K. Sun, Xiao, Liu, & Wang, 2019); on the other, top-down approaches first detect body instances and run single-person pose estimation models on each instance (Cao, Simon, Wei, & Sheikh, 2017; Insafutdinov, Pishchulin, Andres, Andriluka, & Schiele, 2016; Kocabas, Karagoz, & Akbas, 2018; Newell, Huang, & Deng, 2017; Papandreou et al., 2018; Pishchulin et al., 2016). The computation cost of top-down methods increases linearly with the number of people in an image, while that of bottom-up methods stays constant. However, in cases where there is significant overlap between instances, top-down approaches are often more accurate (Y. Chen, Tian, & He, 2020).

We experimented with G-RMI (Papandreou et al., 2017), a well-established top-down approach, and give examples of predictions in Figure 2. In the first stage, Faster-RNN (Ren et al., 2015) is used to detect person instances. Inspired by detection methods, the second stage combines classification and regression to process each resulting crop: a fully convolutional network first densely classifies whether each spatial position is in the vicinity of a given keypoint class, and then refines each prediction by predicting an offset. A specialized form of Hough voting (see (Duda & Hart, 1972) for background) is introduced to aggregate these predictions and form highly localized activation maps. A key-point based confidence score and non-maximum suppression procedure further improve results. We plan to build on this approach to develop methods for the previously mentioned challenges.

Appendix C. Player Vectors

In particular, we follow definition of *playing style* in Decroos and Davis (2019), which is defined as a player’s preferred area(s) on the field to occupy and which actions they tend to perform in each of these locations, and generate our player vectors with the method proposed in Decroos and Davis (2019). The procedure of generating player vectors unfolds into four steps. First, we collect the event stream data of all Premier League matches that Liverpool Football Club participated in from 2017 to 2019, and filter the actions of types passes, dribbles, shots and crosses. Secondly, for each pair of player p , who is observed in the event stream dataset, and relevant action type t , we overlay a grid of size 60×40 on the football pitch and count how many times player p performed action t in each grid cell. This procedure yields a matrix which summarizes spatial preference of player p performing action type t . Thirdly, we compress that matrix into a small vector. To do this, we reshape each matrix into a vector and group it together with all other vectors of the same action type, and we then perform non-negative matrix (NMF) factorization to reduce the dimensionality of these matrices. This procedure yields a smaller vector, and the value of each dimension quantifies the preference of player p performing the action type t in the area a . Finally, for each player, we obtain 4 vectors corresponding to the 4 action types, and we generate one final vector of 18 dimensions by concatenating his compressed vectors for relevant action types.

References

- Abu Farha, Y., Richard, A., & Gall, J. (2018). When will you do what?-Anticipating temporal occurrences of activities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Alayrac, J.-B., Recasens, A., Schneider, R., Arandjelović, R., Ramapuram, J., De Fauw, J., ... Zisserman, A. (2020). Self-supervised multimodal versatile networks. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Albert, J. (2010). Sabermetrics: The past, the present, and the future. *Mathematics and*

sports, 3–14.

- Albert, J., Bennett, J., & Mead, C. (2002, 10). Curve ball: Baseball, statistics, and the role of chance in the game. *Physics Today*, 55, 56-57.
- Alp Güler, R., Neverova, N., & Kokkinos, I. (2018). DensePose: Dense human pose estimation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Anderson, A., Rosen, J., Rust, J., & Wong, K.-P. (2020). Disequilibrium play in tennis. *Working paper*.
- Arandjelovic, R., & Zisserman, A. (2017). Look, listen and learn. In *Proceedings of the IEEE International Conference on Computer Vision*.
- Arel, I., Rose, D., & Karnowski, T. (2010, 01). Deep machine learning - a new frontier in artificial intelligence research [research frontier]. *IEEE Comp. Int. Mag.*, 5, 13-18.
- Azar, O. H., & Bar-Eli, M. (2011). Do soccer players play the mixed-strategy Nash equilibrium? *Applied Economics*, 43(25), 3591-3601.
- Barr, G., Holdsworth, C., & Kantor, B. (2008). Evaluating performances at the 2007 cricket world cup. *South African Statistical Journal*, 42(2), 125–142.
- Bartlett, R. (2006, 12). Artificial intelligence in sports biomechanics: New dawn or false hope? *Journal of Sports Science & Medicine*, 5, 474-479.
- Baumer, B., & Zimbalist, A. (2014). *The sabermetric revolution: Assessing the growth of analytics in baseball*. University of Pennsylvania Press.
- Beal, R., Norman, T. J., & Ramchurn, S. D. (2019). Artificial intelligence for team sports: A survey. *The Knowledge Engineering Review*, 34, e28.
- Bengio, Y. (2009, January). Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning*, 2(1), 1–127.
- Bhattacharyya, A., Fritz, M., & Schiele, B. (2018). Long-term on-board prediction of people in traffic scenes under uncertainty. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Bhattacharyya, A., Fritz, M., & Schiele, B. (2019). Bayesian prediction of future street scenes using synthetic likelihoods. In *Proceedings of the International Conference on Learning Representations*.
- Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2015, 01). Identifying team style in soccer using formations learned from spatiotemporal tracking data. *IEEE International Conference on Data Mining Workshops*.
- Bloembergen, D., Tuyls, K., Hennes, D., & Kaisers, M. (2015). Evolutionary dynamics of multi-agent learning: A survey. *J. Artif. Intell. Res.*, 53, 659–697.
- Bransen, L., & Van Haaren, J. (2018). Measuring football players' on-the-ball contributions from passes during games. In *International Workshop on Machine Learning and Data Mining for Sports Analytics*.
- Bransen, L., & Van Haaren, J. (2019). Player chemistry: Striving for a perfectly balanced

- soccer team. In *MIT Sloan Conference*.
- Bridgeman, L., Volino, M., Guillemaut, J.-Y., & Hilton, A. (2019, June). Multi-person 3D pose estimation and tracking in sports. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.
- Brock, A., Donahue, J., & Simonyan, K. (2019). Large scale GAN training for high fidelity natural image synthesis. In *Proceedings of the International Conference on Learning Representations*.
- Busoniu, L., Babuska, R., & Schutter, B. D. (2008). A comprehensive survey of multiagent reinforcement learning. *IEEE Trans. Syst. Man Cybern. Part C*, 38(2), 156–172.
- Buzzacchi, L., & Pedrini, S. (2014). Does player specialization predict player actions? Evidence from penalty kicks at FIFA World Cup and UEFA Euro Cup. *Applied Economics*, 46(10), 1067–1080.
- Camerer, C. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.
- Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, 43(1), 9–64.
- Camerer, C., Loewenstein, G., & Rabin, M. (2011). *Advances in behavioral economics*. Princeton University Press.
- Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Carreira, J., & Zisserman, A. (2017). Quo vadis, action recognition? A new model and the kinetics dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Chan, C., Ginosar, S., Zhou, T., & Efros, A. A. (2019). Everybody dance now. In *Proceedings of the IEEE International Conference on Computer Vision*.
- Chen, D. L., Kim, J., & Mooney, R. J. (2010). Training a multilingual sportscaster: Using perceptual context to learn language. *Journal of Artificial Intelligence Research*, 37, 397–435.
- Chen, Y., Tian, Y., & He, M. (2020). Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding*, 192, 102897.
- Cheng, Y., Yang, B., Wang, B., Yan, W., & Tan, R. T. (2019). Occlusion-aware networks for 3d human pose estimation in video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Chiappori, P.-A., Levitt, S., & Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, 92(4), 1138–1151.
- Choi, J., Kwon, J., & Lee, K. M. (2019). Deep meta learning for real-time target-aware visual tracking. In *Proceedings of the IEEE/CVF International Conference on Com-*

puter Vision.

- ChyronHego. (2020). *ChyronHego*. Retrieved 2020-09-09, from <https://chyronhego.com/>
- Clark, A., Donahue, J., & Simonyan, K. (2019). Adversarial video generation on complex datasets. *arXiv*, arXiv–1907.
- Clarke, A. (2020). *Season trends: Quality on rise as midfield takes focus*. Retrieved 2020-09-17, from <http://www.premierleague.com/news/1745761>
- Claudino, J., Capanema, D., Souza, T., Serrao, J., Pereira, A., & Nassis, G. (2019, 07). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: A systematic review. *Sports Medicine - Open*, 5.
- Coloma, G. (2012). The penalty-kick game under incomplete information. *University of CEMA Economics Serie Documentos de Trabajo*(487).
- Costa, G., Huber, M., & Saccoman, J. (2009). *Practicing sabermetrics: Putting the science of baseball statistics to work*. McFarland, Incorporated, Publishers. Retrieved from https://books.google.fr/books?id=Kkf7gowrH_UC
- CVSports International Workshop on Computer Vision in Sports at CVPR*. (2020). Retrieved 2020-09-07, from <https://vap.aau.dk/cvsports/>
- Decroos, T., Bransen, L., Haaren, J. V., & Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, KDD 2019, anchorage, ak, usa, august 4-8, 2019* (pp. 1851–1861). ACM.
- Decroos, T., & Davis, J. (2019). Player vectors: Characterizing soccer players' playing style from match event streams. In *European Conference on Machine Learning and Knowledge Discovery in Databases*.
- Decroos, T., Van Haaren, J., & Davis, J. (2018). Automatic discovery of tactics in spatio-temporal soccer match data. In *Proceedings of the international conference on knowledge discovery & data mining*.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255).
- Deo, N., & Trivedi, M. M. (2018). Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 1468–1476).
- Dong, X., & Shen, J. (2018). Triplet loss in siamese network for object tracking. In *Proceedings of the european conference on computer vision (ECCV)* (pp. 459–474).
- Duda, R. O., & Hart, P. E. (1972). Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1), 11–15.
- Emara, N., Owens, D. M., Smith, J., & Wilmer, L. (2014). Minimax on the gridiron: Serial correlation and its effects on outcomes in the national football league. *Available at*

- SSRN 2502193.*
- Fang, H.-S., Xie, S., Tai, Y.-W., & Lu, C. (2017). Rmpe: Regional multi-person pose estimation. In *Proceedings of the IEEE international conference on computer vision* (pp. 2334–2343).
- Fastovets, M., Guillemaut, J., & Hilton, A. (2013). Athlete pose estimation from monocular tv sports footage. In *2013 IEEE conference on computer vision and pattern recognition workshops* (p. 1048-1054).
- Felsen, P., Lucey, P., & Ganguly, S. (2018). Where will they go? predicting fine-grained adversarial multi-agent motion using conditional variational autoencoders. In *Computer vision - ECCV 2018 - 15th european conference, munich, germany, september 8-14, 2018, proceedings, part XI* (pp. 761–776).
- Fernández, J. (2019). Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In *Mit sloan conference*.
- Fernández, J., Bornn, L., & Cervone, D. (2019). Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In *13 th annual mit sloan sports analytics conference*.
- Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P. R., & McRobert, A. P. (2016). Attacking and defensive styles of play in soccer: analysis of spanish and english elite teams. *Journal of sports sciences*, 34(24), 2195–2204.
- Fernando, T., Denman, S., Sridharan, S., & Fookes, C. (2018). Gd-gan: Generative adversarial networks for trajectory prediction and group detection in crowds. In *Asian conference on computer vision* (pp. 314–330).
- Franks, A., Miller, A., Bornn, L., Goldsberry, K., et al. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics*, 9(1), 94–121.
- Gabbett, T. J. (2010). The development and application of an injury prediction model for noncontact, soft-tissue injuries in elite collision sport athletes. *The Journal of Strength & Conditioning Research*, 24(10), 2593–2603.
- Gade, R., & Moeslund, T. B. (2018). Constrained multi-target tracking for team sports activities. *IPSJ Transactions on Computer Vision and Applications*, 10(1), 2.
- Gauriot, R., Page, L., & Wooders, J. (2016). Nash at wimbledon: evidence from half a million serves. Available at *SSRN 2850919*.
- Giancola, S., Amine, M., Dghaily, T., & Ghanem, B. (2018). Soccernet: A scalable dataset for action spotting in soccer videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 1711–1721).
- Godard, C., Mac Aodha, O., & Brostow, G. J. (2017). Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 270–279).
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (<http://www.deeplearningbook.org>)

- Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., & Alahi, A. (2018). Social GAN: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.
- Hanna, J., & Stone, P. (2017, February). Grounded action transformation for robot learning in simulation. In *Proceedings of the 31st aaai conference on artificial intelligence (AAAI)*.
- Hausknecht, M., & Stone, P. (2016, May). Deep reinforcement learning in parameterized action space. In *Proceedings of the international conference on learning representations (ICLR)*.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).
- He, Y., Yan, R., Fragkiadaki, K., & Yu, S.-I. (2020). Epipolar transformers. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 7779–7788).
- Hernandez-Leal, P., Kartal, B., & Taylor, M. E. (2020). A very condensed survey and critique of multiagent deep reinforcement learning. In *Proceedings of the 19th international conference on autonomous agents and multiagent systems, AAMAS '20, auckland, new zealand, may 9-13, 2020* (pp. 2146–2148). International Foundation for Autonomous Agents and Multiagent Systems.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in neural information processing systems* (pp. 6626–6637).
- Huang, S., Gong, M., & Tao, D. (2017). A coarse-fine network for keypoint localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 3028–3037).
- Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., & Schiele, B. (2016). Deepcut: A deeper, stronger, and faster multi-person pose estimation model. In *European conference on computer vision* (pp. 34–50).
- InStat*. (2020). Retrieved 2020-09-09, from <https://football.instatscout.com/>
- Iqbal, U., & Gall, J. (2016). Multi-person pose estimation with local joint-to-person associations. In *European conference on computer vision* (pp. 627–642).
- Iskakov, K., Burkov, E., Lempitsky, V., & Malkov, Y. (2019). Learnable triangulation of human pose. In *Proceedings of the IEEE international conference on computer vision* (pp. 7718–7727).
- Jin, X., Xiao, H., Shen, X., Yang, J., Lin, Z., Chen, Y., ... Yan, S. (2017). Predicting scene parsing and motion dynamics in the future. In *Advances in neural information processing systems* (pp. 6915–6924).
- Kakavas, G., Malliaropoulos, N., Pruna, R., & Maffulli, N. (2019, 08). Artificial intelligence a tool for sports trauma prediction. *Injury*.
- Kalyanakrishnan, S., & Stone, P. (2010). Learning complementary multiagent behaviors: A case study. In *RoboCup 2009: Robot soccer world cup XIII* (pp. 153–165). Springer

- Verlag.
- Kampakis, S. (2016). *Predictive modelling of football injuries* (Unpublished doctoral dissertation). UCL (University College London).
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.
- Kitani, K. M., Ziebart, B. D., Bagnell, J. A., & Hebert, M. (2012). Activity forecasting. In *European conference on computer vision* (pp. 201–214).
- Kocabas, M., Karagoz, S., & Akbas, E. (2018). Multiposenet: Fast multi-person pose estimation using pose residual network. In *Proceedings of the european conference on computer vision (ECCV)* (pp. 417–433).
- Kuper, S., & Szymanski, S. (2018). *Soccernomics, why england loses, why germany, spain and france win, and why one day, japan, iraq and the united states will become the kings of the world's most popular sport*. HarperCollins Publishers.
- Lassner, C., Romero, J., Kiefel, M., Bogo, F., Black, M. J., & Gehler, P. V. (2017). Unite the people: Closing the loop between 3d and 2d human representations. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6050–6059).
- Le, H., Carr, P., Yue, Y., & Lucey, P. (2017, 03). Data-driven ghosting using deep imitation learning. In *Mit sloan conference*.
- Le, H. M., Yue, Y., Carr, P., & Lucey, P. (2017). Coordinated multi-agent imitation learning. In *Proceedings of the 34th international conference on machine learning, ICML 2017, sydney, nsw, australia, 6-11 august 2017* (pp. 1995–2003).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Lennartsson, J., Lidstrom, N., & Lindberg, C. (2015, 05). Game intelligence in team sports. *PLoS one*, 10, e0125453.
- Levitt, S., Chiappori, P., & Groseclose, T. (2002, 02). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, 92, 1138-1151.
- Li, M. G., Jiang, B., Zhu, H., Che, Z., & Liu, Y. (2020). Generative attention networks for multi-agent behavioral modeling. In *Aaai* (pp. 7195–7202).
- Lin, K. (2014). Applying game theory to volleyball strategy. *International Journal of Performance Analysis in Sport*, 14(3), 761–774.
- Lin, T., Yang, Y., Beyer, J., & Pfister, H. (2020). Sportsxr - immersive analytics in sports. *ArXiv, abs/2004.08010*.
- Liu, G., Luo, Y., Schulte, O., & Kharrat, T. (2020, 07). Deep soccer analytics: learning an action-value function for evaluating soccer players. *Data Mining and Knowledge Discovery*. doi: 10.1007/s10618-020-00705-9
- Liu, G., & Schulte, O. (2018). Deep reinforcement learning in ice hockey for context-aware player evaluation. In (p. 3442–3448). AAAI Press.

- Liu, J., Carr, P., Collins, R. T., & Liu, Y. (2013). Tracking sports players with context-conditioned motion models. In *2013 IEEE conference on computer vision and pattern recognition* (p. 1830-1837).
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).
- Lowe, Z. (2013, March). *Lights, Cameras, Revolution.* Retrieved 2020-09-07, from <https://grantland.com/features/the-toronto-raptors-sportvu-cameras-nba-analytical-revolution/>
- Lu, W., Ting, J., Little, J. J., & Murphy, K. P. (2013). Learning to track and identify players from broadcast sports videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7), 1704-1716.
- Luc, P., Clark, A., Dieleman, S., Casas, D. d. L., Doron, Y., Cassirer, A., & Simonyan, K. (2020). Transformation-based adversarial video prediction on large-scale data. *arXiv preprint arXiv:2003.04035*.
- Luc, P., Couprie, C., Lecun, Y., & Verbeek, J. (2018). Predicting future instance segmentation by forecasting convolutional features. In *Proceedings of the european conference on computer vision (ECCV)* (pp. 584–599).
- Luc, P., Neverova, N., Couprie, C., Verbeek, J., & LeCun, Y. (2017). Predicting deeper into the future of semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 648–657).
- Machine Learning and Data Mining for Sports Analytics.* (2020). Retrieved 2020-09-08, from <https://dtai.cs.kuleuven.be/events/MLSA20/>
- Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2017). Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*.
- Mahasseni, B., Lam, M., & Todorovic, S. (2017, July). Unsupervised video summarization with adversarial lstm networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.
- Maksai, A., Wang, X., & Fua, P. (2016). What players do with the ball: A physically constrained interaction modeling. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 972–981).
- Meerhoff, L. A., Goes, F. R., Knobbe, A., et al. (2019). Exploring successful team tactics in soccer tracking data. In *Joint european conference on machine learning and knowledge discovery in databases* (pp. 235–246).
- Mehrasa, N., Zhong, Y., Tung, F., Bornn, L., & Mori, G. (2018). Deep learning of player trajectory representations for team activity analysis. In *11th mit sloan sports analytics conference* (Vol. 2, p. 3).
- Merler, M., Joshi, D., Mac, K.-N. C., Nguyen, Q.-B., Hammer, S., Kent, J., ... Feris, R. S. (2018). The excitement of sports: Automatic highlights using audio/visual cues. In *CVPR workshops* (pp. 2520–2523).

- Michael, L. (2004). *Moneyball : the art of winning an unfair game*. New York: W. W. Norton.
- Miech, A., Alayrac, J.-B., Smaira, L., Laptev, I., Sivic, J., & Zisserman, A. (2020, June). End-to-end learning of visual representations from uncurated instructional videos. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (CVPR)*.
- Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1), 8–30.
- MIT Sloan Sports Analytics Conference*. (2020). Retrieved 2020-09-07, from <http://www.sloansportsconference.com/>
- Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*.
- Mora, S. V., & Knottenbelt, W. J. (2017). Deep learning for domain-specific action recognition in tennis. In *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)* (p. 170-178).
- Moschini, G. (2004). Nash equilibrium in strictly competitive games: live play in soccer. *Economics Letters*, 85(3), 365 - 371.
- Murphy, K. P. (2013). *Machine learning: a probabilistic perspective*. MIT press.
- Nax, H. (2015). *Behavioral game theory* (Unpublished doctoral dissertation). ETH Zurich, Zurich.
- Newell, A., Huang, Z., & Deng, J. (2017). Associative embedding: End-to-end learning for joint detection and grouping. In *Advances in neural information processing systems* (pp. 2277–2287).
- Ng, A. Y., & Russell, S. J. (2000). Algorithms for inverse reinforcement learning. In *Proceedings of the seventeenth international conference on machine learning (ICML 2000), stanford university, stanford, ca, usa, june 29 - july 2, 2000* (pp. 663–670). Morgan Kaufmann.
- Omidshafiei, S., Papadimitriou, C., Piliouras, G., Tuyls, K., Rowland, M., Lespiau, J.-B., ... Munos, R. (2019). α -rank: Multi-agent evaluation by evolution. *Scientific reports*, 9(1), 1–29.
- Op De Beéck, T., Meert, W., Schütte, K., Vanwanseele, B., & Davis, J. (2018). Fatigue prediction in outdoor runners via machine learning and sensor fusion. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (p. 606–615). Association for Computing Machinery.
- Opta*. (2020). Retrieved 2020-09-09, from <https://www.optasports.com>
- Palacios-Huerta, I. (2003, 04). Professionals Play Minimax. *The Review of Economic Studies*, 70(2), 395-415.
- Palacios-Huerta, I. (2016). *Beautiful game theory: How soccer can help economics*. Princeton University Press.

- Panait, L., & Luke, S. (2005). Cooperative multi-agent learning: The state of the art. *Auton. Agents Multi Agent Syst.*, 11(3), 387–434.
- Papandreou, G., Zhu, T., Chen, L.-C., Gidaris, S., Tompson, J., & Murphy, K. (2018). Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In *Proceedings of the european conference on computer vision (ECCV)* (pp. 269–286).
- Papandreou, G., Zhu, T., Kanazawa, N., Toshev, A., Tompson, J., Bregler, C., & Murphy, K. (2017). Towards accurate multi-person pose estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4903–4911).
- Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 10975–10985).
- Pavlakos, G., Zhou, X., Derpanis, K. G., & Daniilidis, K. (2017). Coarse-to-fine volumetric prediction for single-image 3d human pose. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7025–7034).
- Pavllo, D., Feichtenhofer, C., Grangier, D., & Auli, M. (2019). 3d human pose estimation in video with temporal convolutions and semi-supervised training. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7753–7762).
- Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P. V., & Schiele, B. (2016). Deepcut: Joint subset partition and labeling for multi person pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4929–4937).
- Puerzer, R. (2002, 01). From scientific baseball to sabermetrics: Professional baseball as a reflection of engineering and management in society. *Nine: A Journal of Baseball History and Culture*, 11, 34-48.
- Quiroga, J., Carrillo, H., Maldonado, E., Ruiz, J., & Zapata, L. M. (2020, June). As seen on tv: Automatic basketball video production using gaussian-based actionness and game states recognition. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (CVPR) workshops*.
- Ramos, G., Vaz, J., Mendonca, G., Pezarat-Correia, P., Rodrigues, J., Alfaras, M., & Gamboa, H. (2020, 01). Fatigue evaluation through machine learning and a global fatigue descriptor. *Journal of Healthcare Engineering*, 2020, 1-18.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems 28* (pp. 91–99). Curran Associates, Inc. Retrieved from <http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf>
- RoboCup project*. (2020). Retrieved 2020-09-09, from <https://www.robocup.org>

- Rossi, A., Pappalardo, L., Cintia, P., Iaia, F., Fernández, J., & Medina, D. (2018, 06). *Machine learning approach to injury prediction*.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117. (Published online 2014; based on TR arXiv:1404.7828 [cs.NE]) doi: 10.1016/j.neunet.2014.09.003
- Shih, H.-C. (2017). A survey of content-aware video analysis for sports. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(5), 1212–1231.
- Shoham, Y., Powers, R., & Grenager, T. (2007). If multi-agent learning is the answer, what is the question? *Artif. Intell.*, 171(7), 365–377.
- Sindik, J., & Vidak, N. (2008, 06). Application of game theory in describing efficacy of decision making in sportsman's tactical performance in team sports. *Interdisciplinary Description of Complex Systems - scientific journal*, 6, 53-66.
- Skinner, B. (2010). The price of anarchy in basketball. *Journal of Quantitative Analysis in Sports*, 6(1).
- Song, A., Severini, T., & Allada, R. (2017). How jet lag impairs major league baseball performance. *Proceedings of the National Academy of Sciences*, 114(6), 1407–1412.
- Spearman, W. (2016, 02). *Quantifying pitch control*. doi: 10.13140/RG.2.2.22551.93603
- Spearman, W. (2018). Beyond expected goals. In *Proceedings of the 12th mit sloan sports analytics conference* (pp. 1–17).
- Statsbomb*. (2020). Retrieved 2020-09-09, from <https://statsbomb.com/>
- Stats perform*. (2020). Retrieved 2020-09-17, from <https://www.statsperform.com/resource/stats-playing-styles-introduction/>
- Stone, P., Sutton, R. S., & Kuhlmann, G. (2005). Reinforcement learning for RoboCup-soccer keepaway. *Adaptive Behavior*, 13(3), 165–188.
- StriVR Immersive Sports Analytics*. (2020). Retrieved 2020-09-09, from <https://www.strivr.com/use-cases/sports/>
- Su, S.-Y., Hajimirsadeghi, H., & Mori, G. (2019). Graph generation with variational recurrent neural network. *arXiv preprint arXiv:1910.01743*.
- Sun, C., Karlsson, P., Wu, J., Tenenbaum, J. B., & Murphy, K. (2019). Stochastic prediction of multi-agent interactions from partial observations. *arXiv preprint arXiv:1902.09641*.
- Sun, J., Xie, J., Hu, J.-F., Lin, Z., Lai, J., Zeng, W., & Zheng, W.-s. (2019). Predicting future instance segmentation with contextual pyramid convlstm. In *Proceedings of the 27th ACM international conference on multimedia* (pp. 2043–2051).
- Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5693–5703).
- Sun, X., Davis, J., Schulte, O., & Liu, G. (2020). Cracking the black box: Distilling deep sports analytics. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1–10).

- ence on knowledge discovery & data mining* (p. 3154–3162). New York, NY, USA: Association for Computing Machinery. Retrieved from <https://doi.org/10.1145/3394486.3403367> doi: 10.1145/3394486.3403367
- Sypetkowski, M., Kurzejamski, G., & Sarwas, G. (2019). Football players pose estimation. In *Image processing and communications challenges 10* (pp. 63–70). Cham: Springer International Publishing.
- Sypetkowski, M., Sarwas, G., & Trzcinski, T. (2019). Synthetic image translation for football players pose estimation. *J. UCS*, 25(6), 683–700.
- Szymanski, S. (2020). Sport analytics: Science or alchemy? *Kinesiology Review*, 9(1), 57–63.
- Tuyls, K., Pérrolat, J., Lanctot, M., Hughes, E., Everett, R., Leibo, J. Z., ... Graepel, T. (2020). Bounds and dynamics for empirical game theoretic analysis. *Auton. Agents Multi Agent Syst.*, 34(1), 7.
- Tuyls, K., & Weiss, G. (2012). Multiagent learning: Basics, challenges, and prospects. *AI Mag.*, 33(3), 41–52.
- Unterthiner, T., van Steenkiste, S., Kurach, K., Marinier, R., Michalski, M., & Gelly, S. (2018). Towards accurate generative models of video: A new metric & challenges. *arXiv preprint arXiv:1812.01717*.
- Urieli, D., MacAlpine, P., Kalyanakrishnan, S., Bentor, Y., & Stone, P. (2011, May). On optimizing interdependent skills: A case study in simulated 3d humanoid robot soccer. In *Proc. of 10th int. conf. on autonomous agents and multiagent systems (AAMAS)* (Vol. 2, pp. 769–776). IFAAMAS.
- Villegas, R., Yang, J., Zou, Y., Sohn, S., Lin, X., & Lee, H. (2017). Learning to generate long-term future via hierarchical prediction. *arXiv preprint arXiv:1704.05831*.
- Visser, U., & Burkhard, H.-D. (2007, 06). Robocup: 10 years of achievements and future challenges. *AI Magazine*, 28, 115–132.
- Vondrick, C., Pirsiavash, H., & Torralba, A. (2016). Anticipating visual representations from unlabeled video. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 98–106).
- Vu, T.-H., Choi, W., Schulter, S., & Chandraker, M. (2018). Memory warps for learning long-term online video representations. *arXiv preprint arXiv:1803.10861*.
- Wang, J. M., Fleet, D. J., & Hertzmann, A. (2007). Gaussian process dynamical models for human motion. *IEEE transactions on pattern analysis and machine intelligence*, 30(2), 283–298.
- Weissenborn, D., Täckström, O., & Uszkoreit, J. (2019). Scaling autoregressive video models. *arXiv preprint arXiv:1906.02634*.
- Wellman, M. P. (2006). Methods for empirical game-theoretic analysis. In *Aaaai* (pp. 1552–1556).
- Wyscout*. (2020). Retrieved 2020-09-09, from <https://wyscout.com/>

- Xiong, B., Kalantidis, Y., Ghadiyaram, D., & Grauman, K. (2019, June). Less is more: Learning highlight detection from video duration. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (CVPR)*.
- Xu, J., Ni, B., Li, Z., Cheng, S., & Yang, X. (2018). Structure preserving video prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1460–1469).
- Yang, H., Wang, B., Lin, S., Wipf, D., Guo, M., & Guo, B. (2015). Unsupervised extraction of video highlights via robust recurrent auto-encoders. In *Proceedings of the IEEE international conference on computer vision* (pp. 4633–4641).
- Yee, T., Lisì, V., & Bowling, M. H. (2016). Monte carlo tree search in continuous action spaces with execution uncertainty. In *Ijcai* (pp. 690–697).
- Yeh, R. A., Schwing, A. G., Huang, J., & Murphy, K. (2019). Diverse generation for multi-agent sports games. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4610–4619).
- Zhang, F., Zhu, X., & Ye, M. (2019, June). Fast human pose estimation. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (CVPR)*.
- Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019). Self-attention generative adversarial networks. In *International conference on machine learning* (pp. 7354–7363).
- Zhang, H., Sciutto, C., Agrawala, M., & Fatahalian, K. (2020). Vid2player: Controllable video sprites that behave and appear like professional tennis players. *arXiv preprint arXiv:2008.04524*.
- Zhang, K., Chao, W.-L., Sha, F., & Grauman, K. (2016). Video summarization with long short-term memory. In *ECCV* (7).