IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 30, 2021, accepted August 19, 2021, date of publication August 24, 2021, date of current version September 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3107585

A Dimension Reduction Approach to Player Rankings in European Football

AYSE ELVAN AYDEMIR[1,2], TUGBA TASKAYA TEMIZEL[1], ALPTEKIN TEMIZEL[1], KLIMENT PRESHLENOV[2], AND DANIEL M. STRAHINOV[2]
[1]Graduate School of Informatics, Middle East Technical University (METU), 06800 Ankara, Turkey
[2]Enskai Ltd., 1280 Sofia, Bulgaria

Corresponding author: Ayse Elvan Aydemir (elvan.gunduz@metu.edu.tr)

ABSTRACT Player performance evaluation is a challenging problem with multiple dimensions. Football (soccer) is the largest sports industry in terms of monetary value and it is paramount that teams can assess the performance of players for both financial and operational reasons. However, this is a difficult task, not only because performance differs from position to position, but also it is based on competition, time played and team play-styles. Because of this, raw player statistics are not comparable across players and must be processed to facilitate a fair performance evaluation. Furthermore, teams may have different requirements and a generic player performance evaluation does not directly serve the particular expectations of different clubs. In this study, we provide a generic framework for estimating player performance and performing player-fit-to-criteria assessment, under different objectives, for left and right backs from competitions worldwide. The results show that the players who have ranked high have increased their transfer values and they have moved to suitable teams. Global nature of the proposed methodology expands the analyzed player pool, facilitating the search for outstanding players from all available competitions.

INDEX TERMS Player ranking, player performance, football analytics, sports analytics.

I. INTRODUCTION

Player performance evaluation is a key problem in the sports industry [1], [2]. In team and individual sports alike, it forms the basis of the competition and directly affects the financial aspects such as the transfer market, club investments, betting industry and media rights. Furthermore, by its nature, the goal of every competing entity is to achieve the best possible performance. So, it is important to identify players who fit the requirements with the best potential return on investment, both in financial and performance terms.

Recently there has been an increased focus on evaluating player performance in football using formal analytical and statistical methods [1], [3], [4] assisted by the increased data availability [5]–[7].

Teams achieve their results through collective contribution of the players, and this makes it difficult to separate an individual player's performance from the rest of the team based on results or in-game metrics [8]. In addition, players have certain roles in team sports. These roles differ in their performance requirements in various aspects of the game. For instance, a defence player may be required to be successful in interceptions and duels, whereas a left or right back would be expected to perform accurate shots. Furthermore, players may play different roles in different games, even within a single game [2]. This dynamism requires analysing player performance on multiple dimensions, going beyond simple aggregates.

As a result of being in a team environment, players' performances are impacted by external factors. These factors include, but are not limited to, performance of their teammates, performance of opposing team players, the quality of the competition they play in and how long they play in a game. However, the collected in-game data does not explicitly show the effect of these factors and makes it hard to compare players fairly. A world-class player may have the same statistics as a second league player on these in-game metrics, however, this does not mean they have on-par performance. Furthermore, in ranking player performance there is no established consensus or ground-truth as clubs and roles require different criteria for ranking.

The associate editor coordinating the review of this manuscript and approving it for publication was Dost Muhammad Khan.

In this study, we provide a novel framework for evaluating player performance in football in an unsupervised manner and validate the results under the assumption that players' transfer values and the teams they move to via transfers reflect their perceived performance. We apply the proposed solution to the problem of ranking players for the left or right back positions. The proposed framework is generic in the sense that it is extendable and adaptable to different use-cases and factors. Our objective is to evaluate player performance across the globe and provide a blueprint of analyzing player performance in team sports where there is no ground-truth and complex factors that must be taken into account when analyzing match statistics. In addition to the domain-specific requirements for global analysis in football, general ranking problem specific concerns must also be taken into account. A multi-attribute ranking problem aims to find a mapping from a multivariate dataset to a vector that represents the ordering of the data instances. Under the assumption that the computed features represent a high-dimensional mapping from latent player performance, the problem can be considered as discovering the latent 1D representation of higher-order feature dataset.

This study has the following contributions to the literature:

1. We provide a generic framework to evaluate players across competitions and a larger time frame, thus extending the pool of players. This framework is applicable beyond football, as it formally frames the impact of external factors as a scaling problem. The same framework can be applied to any global team sport.
2. We propose a novel approach to reach per 90 in-game statistics, that approximates the established per 90 scaling averages across games and allows analysis on an individual game level.
3. We adopt an exponential decay approach to extend the time horizon for player performance evaluation.
4. We introduce unsupervised discovery of rankings. While unsupervised ranking techniques are established in information retrieval problems, to the best of our knowledge they have not been applied to player performance rankings.

II. LITERATURE REVIEW

There is a recent focus on evaluating player performance in football with the emergence of new data sources. The most well-known example of player analytics is the Moneyball [9], also adapted into a movie. *The Moneyball concept* is maximizing returns on player investment based on player performance analytics. Since its introduction, Moneyball concept has been applied to various sports fields such as basketball, baseball, rugby and American football [10]–[17]. While lacking the same level of interest for many years, football analytics has gained popularity recently. Table 1 provides a summary of papers related to this study. The table first introduces general references for performance analytics, followed by the economic and financial studies relevant to sports. Then an extensive collection of methods in player performance analysis are provided. In the next two sections references for player valuation and team strategy evaluation are discussed. In addition, ranking methods from other domains are provided in the final section. Literature review demonstrated a clear increase in interest in analytics in football in the recent years. The emergent nature of the field of study and its research potential are also evident from the variety of the problems researchers focus on.

In two recent studies, [2] and [18], authors have developed the most relevant approaches to this study. In [2] authors adopt a supervised machine learning approach to identify the important factors that lead to success, and introduce a methodology to derive the player positions and corresponding ratings based on spatio-temporal data. Gavião *et al.* [18] propose a decision-making framework to probabilistically identify undervalued players who are most likely to perform well. Pantzalis and Tjortjis [19] use match-statistics as well as expected goals metrics and Pezzali scores to predict the performance of a player in next season. While this is not directly a player-ranking approach, it allows the authors to use these predictions to comment on the expected performance of the players. Finally, Liu *et al.* [20] use a deep reinforcement learning technique to assign values based on contribution to scoring to player action values in matches and apply their action-valuation technique to championship matches.

Based on the literature review, it is observed that most of the studies have been conducted in a constrained environment by restricting the region or the dimensions analyzed. In this work, we aim to go beyond such constraints and perform player ranking as a whole, in a manner that is flexible and extendable. Different to the existing literature, the aim is to provide a generic *framework* for player performance evaluation in team sports and not only a *solution* to the problem of ranking players in European football in their isolated competitions.

III. DEFINITIONS AND METHODOLOGY

The proposed approach has four main steps. First, the player-match level scaling coefficients are generated to ensure a comparable match-statistics. Then, in-game player statistics are aligned to represent the player statistics in comparable fashion across different competitions and games. In the third step, aligned player statistics are aggregated to arrive at features on a player-level. In the next step, the players are rated on analysis dimensions. Final player rankings across all analysis dimensions are represented using ranking principle curves [21]. We expand on this further in Section III-C. For clarity, we provide the definitions of the terms used throughout the paper with their notation and dimensions in Table 2. The methodology process flow diagram is given in Figure 1.

A. ALIGNING PLAYER STATISTICS

Aligning player statistics is a crucial step for comparison of players across the globe on the same scale. This alignment is

TABLE 1. Extended literature review.

Authors	Sports	Methods	Summary
			General Analytics
Baumer and Zimbalist [10]	Baseball		Provide a comprehensive analysis of the prominence of analytics in baseball.
Rein and Memmert [24]	Football	Literature Review and Discussion	Give an extensive overview of how big data and advanced analytics techniques are used in various forms to perform tactical analysis in football. They mainly focus on identifying and analysing the tactics of teams while providing insights on the challenges in adoption of analytics in the sports industry, both from technical and cultural stand-points.
Stein et al. [25]	Football	Literature Review and Discussion	Describe the data types and providers. They also give an overview of main analytical problems such as time series analysis, text analysis, video processing and relate those generic problems to sports analytics.
			Economics and Financial
Lucifora and Simmons [26]	Football	Linear Regression	Explore the so-called 'superstar effect' in Italian football. Superstar effect causes the best individual to achieve disproportionate gains no matter how small the difference may be between the best and the rest. The authors propose a model to investigate the factors that affect this phenomenon in Italian football and show that superstar effect exists in football.
Torgler and Schmidt [27]	Football	Panel Analysis	Explore the relationship between player salary and success, and find a relationship between players' relative pays and their performance.
Barros and Leach [3]	Football	Econometric Frontier Model	Employ an Econometric Frontier model to analyze the efficiency of the football clubs in English Premier League. They identify that the football clubs experience diminishing returns on their financial investment.
Gerrard [28]	Football	Win-Cost Analysis	Use standardized win cost to quantify spending efficiency across years in FA Premier League and find that teams are inefficient in various degrees.
			Performance Evaluation
Martinez and Martínez [11]	General	Expert Interviews	Survey the relevant stakeholders to critically assess the value of player evaluation metrics and report that the stakeholders think that the data-driven evaluation metrics are insufficient.
[17]	Football	Linear Regression	Apply the Moneyball concept to player recruitment in Australian football. They apply regression modelling to identify the statistics that are relevant to the match outcomes and propose recruitment strategies based on performance in these identified factors. The data is fairly limited in this application to only 740 games considered for a binary classification problem with over 50 independent variables.
Pettigrew [14]	Hockey	Bayesian Probability	Propose a probability estimation metric for the hockey games and attributes the changes in the probability to player actions, quantifying player impact.
Schulte et al. [16]	Hockey	Markov Decision Process and K-means	Quantify players via a spatial clustering approach and Markov Decision Process to perform scenario planning and impact estimation on player actions.
[13]	Australian Football	Correlation Analysis and Partial Decision Trees	Verify the player ratings based on actual match performance in Australian football.
Pradhan [15]	Basketball	Grey Relational Analysis	Start from player statistics and performance to quantify the quality of each season in NBA across 100 years using Grey Relational Analysis.
[8]	Football	Predictive modelling and ELO	Calculate ELO ratings for teams in isolated competitions and use the ratings to predict the match outcomes.
Duch et al. [4]	Football	Network Analysis	Propose a network based approach to model the probability of each path on the network results in a shot and they use the network metrics to quantify players' match performance based on the flow centrality metric in the network and team performance is the average performance of the players in the same measure. They apply the methodology to the European Cup 2008.
McHale et al. [1]	Football	Linear Regression and Poisson Processes	Propose an approach that builds up from in game activities to interim performance measures, to impact on match outcome heuristically. They propose six separate indices to quantify player performance that additively combine into a final player performance index. They highlight that football is a highly complex game but their client asked for a simple index for general purposes. This approach is the official player performance rating approach used by FA.
Hughes et al. [29]	Football	Expert Opinion and Survey	Identify the key performance indicators and their importance per position to define the important measures in player performance.
Ali et al. [30]	Football	Semi-Markov Decision Process	Use a semi-Markov decision process to predict teams' goal differential and effect of future transfers on team performance. Performance data for EPL between 08/09 and 11/12 seasons are used to perform the analyses.
Cintia et al. [31]	Football	Predictive Modelling	Analyse the passes to arrive at a pass performance index based on 5 passing dimensions, in order to model and predict the team performance in competitions. They show that the results of this simple analysis is correlated with the team performance in competitions and postulate that complexity science can contribute to the game of football.
Brooks et al. [32]	Football	Supervised Learning	Model the value of passes, and build a player ranking model based on this. The supervised model ranks the players on offensive ability, irrespective of the player position. The authors model whether a possession ends in a shot or not and use the learned model weights to quantify player performance.
Power et al. [33]	Football	Supervised Learning	Use event data from EPL and build risk and reward likelihood models for passes. Each pass is quantified in risk and reward terms. Risk is defined as likelihood of completing the pass, whereas reward is defined as the likelihood of creating an opportunity. Risk modelling involves 'micro features' to arrive at more granularity whereas the outcome variable for reward is defined as "likelihood of pass resulting in a shot in 10 seconds".
Barron et al. [34]	Football	Supervised Learning	Use a feed-forward neural network to predict the player movements within English competitions based on the average game statistics. They show that the player performance is predictive of transitioning between 3 pre-identified groups (moving to upper tier, staying in the current tier and moving down). They report an accuracy between 61.5% and 78.8%. However due to the imbalance in their dataset, the models with middle group did not perform well however, the model that compares move down vs. move up showed statistical significance.
Wunderlich and Memmert [35]	Football	Predictive Modelling and ELO	Use the ELO ratings of teams in various competitions to arrive at odds. They extend [8] by incorporating international games and show the predictive power of ELO algorithm.
Pappalardo and Cintia [36]	Football	Supervised Learning	Model the team success based on performance statistics using a supervised learning approach. They demonstrate the complexity of football, and translate this complexity into measuring team success based on how they perform in each dimension. The performance of team is defined as sum of player statistics in a game. They perform the analysis on absolute and relative performance scales and observe that only a combination of features can explain the variance in success. They compare the performance of obtained ratings to actual and ELO ratings, and find that ELO ratings have a higher correlation with actuals as well as low variance.
Palczewski and Sałabun [37]	Football	Multi-criteria decision making	They demonstrate a COMET [38] approach to ranking football teams and show that their results correlate with WhoScored.Com's outputs.
Pereira et al. [39]	Football	Expert Opinion	Identify variables that are important for attacking performance and employ questionnaires to rate the importance of each individual variable. The weights obtained through expert opinion, pass analysis and descriptive statistics are then standardised to arrive at the player attacking index. The approach is applied to player of Atletico Madrid.

TABLE 1. ***(Continued.)*** **Extended literature review.**

Pappalardo et al. [2]	Football	Supervised Learning and K-Means	Propose a role identifying approach using an average position metric and k-means clustering. Furthermore, they perform role-based ratings and rankings for player based on a supervised learning approach and players' performance in variables in the model. They validate their results with expert opinion.
Gavião et al. [18]	Football	Composition of Probabilistic Preferences	Propose a probabilistic model to evaluate the player performance and quality of investment under the Moneyball concept, using CPP-MB methodology.
Ley et al. [40]	Football	Thurstone-Mosteller, Bradley-Terry, Poisson Models	Model the outcome of matches using these models and compare the results in terms of predictive power using Rank Probability Score Epstein [41]. They find that Poisson models outperform other methodologies in predictive power.
Pantzalis and Tjortjis [19]	Football	Supervised Learning	Model several player statistics as a time-series problem and try to predict the performance in the next season. They use the historical results for predictions and report model metrics.
Liu et al. [20]	Football	Reinforcement Learning	Use reinforcement learning to assign values to each player action based on their contribution to reward (i.e. scoring a goal). They aggregate player contributions to scoring to arrive at a ranking of players.
Sałabun et al. [42]	Football	Multi-criteria decision making	Apply COMET [38] MCDM method to player ranking problem. They apply the proposed methodology to 6 players and report ranking metrics compared to the Golden Ball.
Kizielewicz and Dobryakova [43]	Basketball	Multi-criteria decision making	Apply the COMET [38] method to ranking NBA players.
Player Valuation			
Stanojevic and Gyarmati [44]	Football	Supervised Learning	Extract simple aggregate performance features for players as well as demographic information to model player value obtained from Transfermarkt [45].
Müller et al. [46]	Football	Supervised Learning	Model the player value estimates of Transfermarkt [45] performance and the popularity of the player based on social media statistics using multilevel regression models. They find the age-squared, minutes played, goals, assists, dribbles aerial duels, tackles and social media statistics to be significant in predicting Transfermarkt player values.
Team Strategy Evaluation			
Lucey et al. [47]	Football	Supervised Learning	Use occupancy maps, match statistics combined with spatiotemporal data to partially explain so-called 'home advantage'. They show that home teams have more control in final third and there is no significant difference in passing proficiency between home vs away teams. They conclude that the home advantage partially comes from the away teams' defensive behaviour.
Wang et al. [48]	Football	Topic Models	Assume a predefined number of tactics which result in pass sequences. Each pass sequence and the probabilistic modelling of player positions has a distribution of tactics attached to them. The authors define the tactic discovery as inference of the latent tactic that results in the observed sequence. Their model is an extension of LDA Blei et al. [49], applied to football.
Decroos et al. [50]	Football	Agglomerative Clustering	Divide the events data into uninterrupted sequences to automatically group the team tactics. Later, they extract features from these to perform clusters. The clusters are then ranked based on the number of shots they contain. They search for frequent sequences within these clusters using pattern mining algorithms, which are then ranked for user relevance once more. They compare the identified tactics of top EPL teams.
Oskouie et al. [51]	Football	Survey	Provides a detailed survey on football analytics that use video analytics.
Ranking			
Cohen et al. [52]	General	Supervised Learning	Develop the mathematical formulation and notation of the problem of ranking in the presence of various preference vectors. The mathematical formulation is relevant for all problems that require ranking items defined by numerous variables.
Brin and Page [53]	Web and Document Ranking	Graph Analysis	Develop the original PageRank algorithm that powers Google's search engine. The algorithm computes rankings of web-pages based on in-bound hyperlinks into the page from all sources. Original PageRank algorithm is not customized for user preferences however there are many variants that incorporate that functionality. While this is a prominent approach in ranking domain, PageRank and its variants are out of scope for this study.
Klementiev et al. [54]	Rank Fusion	Unsupervised Learning	Formulate the problem of ranking as combining the ranking decisions by multiple rankers. They argue that if a common ranking exists, rankers that capture the high-level ranking information will consistently agree with each other. Therefore, they derive weights for each ranker based on how 'aggreeable' they are in an unsupervised manner to perform rank fusion.
Klementiev et al. [55]	Rank Fusion	Unsupervised Learning	Formulate the problem of ranking as combining the ranking decisions by multiple rankers. They learn the parameters of extended Mallows model in an unsupervised manner.
Li et al. [21]	Multi-attribute Ranking	Unsupervised Learning	Formalize the characteristics of a ranking problem and identify a manifold learning approach to represent a so-called 'ranking skeleton' to represent the ranking result of multi-attribute datasets (such as a dataset that combines country GDP and life expectancy). They derive the principal curves based on Bezier curves which satify all five constraints of a ranking problem.

done through devising scales, which adjust the raw statistics to include the effect of external factors.

The first factor we consider is the game-play duration. This idea is already established in football in terms of per 90 averages. However, we propose a novel approach to this method to adapt the per 90 scaling to our methodology. To reflect the competition quality, we rely on the average player values scraped from Transfermarkt. For game difficulty, we propose a practical application of ELO algorithm [22] that uses a new initialization mechanism. For recency scaling, we adopt an approach from reinforcement learning for infinite horizon tasks [23]. In reinforcement learning, the rewards obtained by actions in continuous tasks are "decayed" using a time-dependent (i.e. recency) factor to facilitate aggregation of an infinite sum and ensure that consequences of recent actions are weighted more heavily. Inspired by this, we adopted the same exponential decay approach. To the best of our knowledge, there is no study in the literature that performs all these alignments for the evaluation of player performance to arrive at a global rating.

Given measurements m and adjustment scale s, alignment is defined as in Eq. (1), arriving at a scaled measure m', through element-wise multiplication. As stated in Table 2, measurements refer to player in-game statistics such as number of passes or shots, whereas scales are designed by us to reflect various factors such as the quality of the competition. Statistics coming from high-quality competitions are assumed to reflect a better overall player performance. The design logic and the assumptions of each scale is detailed in corresponding sections.

$$m' = s \odot m \tag{1}$$

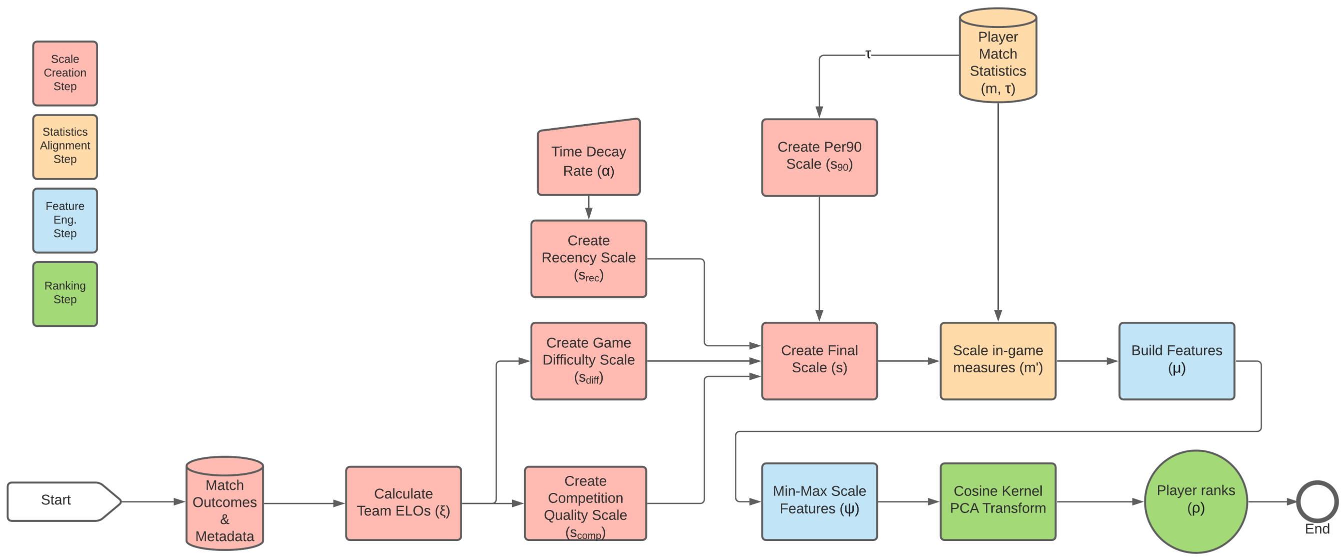

FIGURE 1. Process diagram of methodology.

TABLE 2. Definitions and variables.

Term	Definition	Notation	Dimensions
Match	Event of two games competing against each other, from which player performances are extracted	N	N
Player	Players to be rated	p	p
Team	A football team that is active during the analysis period	λ	λ
Analysis dimension	The player measurements that are relevant to analysis. These are the variables that the user wants to measure the performance on.	d	User defined
Measure	In-game statistics collected by the data providers such as WyScout and OptaStats	m	p x N x d
Minutes played	The duration of player appearance in a match	τ	p x N
Time Decay Rate	Value that specifies the importance given to older matches	α	Scalar
Team Ratings	ELO ratings of teams	ξ	λ x N
Scale	External factors that are relevant to player performance which are not directly visible in measures	s	p x N
Scaling	Process to transform the measures to incorporate the different factors that affect the performance. This is a scalar multiplication of the scale and the measure	fs	
Scaled Measure	In game statistics, adjusted for factors through scales	m'	p x N x d
Aggregated feature	Features computed from scaled in-game statistics m'	μ	p x d
Dimension rating	Player's performance rating on each individual aggregated feature μ	ψ	p x d
Time	Day/Month/Year	t	t

1) LOGARITHMIC PER 90 SCALING

Per 90 minutes scaling is a globally used metric by the industry. Traditionally, the in-game player statistics m for a single player in a single game are transformed into per 90-minute statistics using Eq. (2) [56].

$$m' = \frac{\sum_{i=1}^{N_p} m_i}{\sum_{i=1}^{N_p} \tau_i} 90 \tag{2}$$

where τ_i is the minutes played by the player in game i and N_p is the total number of matches played by player p, m is the in-game statistics of player p and m' is the scaled measures of player p. This is simply the harmonic mean of observations. In this study, we needed to modify the established technique to use it with data for individual games which may have other external factors that need to be considered. To achieve this, we propose an approximation that allows us to scale each game of each player to per 90 individually in Eq. (3).

$$s_{90} = ln(\frac{90}{\tau_i}) + 1 \tag{3}$$

This approximation is empirically shown to have almost perfect linear correlation with outputs of Eq.(2). This logarithmic formulation is closely related to the generalized harmonic mean approximation provided by [57]. The approximation we provide and the approximation provided in [57] are both based on the relationship between the harmonic mean and logarithms. However, their derivation is theoretical whereas we arrived at this representation empirically. For other types of games with fixed durations, such as Basketball, maximum duration of 90 would be replaced by the maximum duration of the corresponding game.

2) GAME DIFFICULTY SCALING

Opponent strength impacts player performance. Performing well against a tougher opponent indicates player potential better than performing well against a weaker one. For a consistent and comparable measure of team strengths, we make use of the existing methods devised for chess [22], [58]. These ratings have also been used in football in isolated competitions [8], [35], [36]. However, these ratings are limited in the sense that they are not able to provide cross-competition ratings, except for minimal cross-competition exposure [35]. To address this limitation, we adopted a novel initialization scheme. In [8], authors show that ELO ratings are correlated with team value and as shown in Figure 2, we find that average player values are log-linearly correlated to the UEFA coefficients. We use these log-transformed average player values to initialize the team ratings, and use the score of each match for each team to update the ratings for the teams via the ELO algorithm [22].

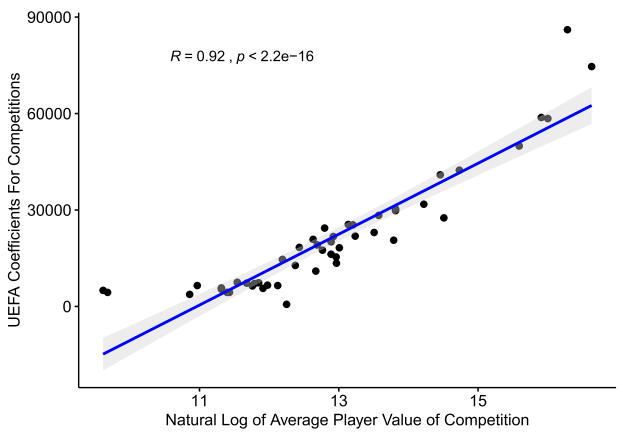

FIGURE 2. **UEFA coefficients vs. player market values.**

Intuitively, the algorithm takes into account the relative ratings of two head-to-head teams and increases the rating of the winning team based on update parameters and decreases the rating of the losing team. The algorithm uses one parameter, K, which determines the expected level of uncertainty in the estimates. The calibration between competitions is further reinforced through the inclusion of international games. Given the team ratings, game difficulty scale for team λ in a single match is computed as in Eq. (4).

$$s_{diff_\lambda} = \frac{\xi_{\neg\lambda}}{\xi_\lambda} \tag{4}$$

where, $\xi_{\neg\lambda}$ is the opponent rating for the most recent match (as the ratings change after every match) and ξ_λ is the rating for the team λ. The parameter values for ELO algorithm are provided in Section IV.

3) COMPETITION QUALITY SCALING

There is a difference in the quality of the football played across different competitions, which are reflected in UEFA Country and club ratings [59]. This makes in-game statistics obtained in different competitions incomparable. However, there is no publicly available global rating of competitions with different divisions, and as a proxy, we first compute the ELO ratings of teams [22] and represent the competition quality as the average ELO of the teams competing in the competition. As opposed to the game difficulty scaling where the statistics are adjusted to account for strength of the opponent, competition scaling boosts the statistics obtained in higher rated competitions to reflect the overall quality of game play. Eq. (5) shows the competition scaling as the average of ELO points of all teams in a given competition where Λ_{comp} is the number of teams in the competition.

$$s_{comp} = \frac{\sum_{\lambda=1}^{\Lambda_{comp}} \xi_\lambda}{\Lambda_{comp}} \tag{5}$$

4) RECENCY SCALING

When evaluating performance, the recent games are more important than the older ones. However, effect size is important in statistical analyses and a sufficient sample size is required for statistical inference. In this study, we adopt a decaying mechanism limiting the impact of older games on the statistical results. Exponential decay mechanism in [2] is based on incremental averages that result in an additional subtraction of higher order decays. In contrast, our approach aims to capture variances in each individual game (like in per 90 scaling), so we rely on being able to scale each game individually as opposed to incremental averages. Recency of the game is encoded using Eq. (6).

$$s_{rec} = \alpha^{\Delta t} \tag{6}$$

where α is a real-valued number $0 < \alpha \leq 1$ and Δt is the time difference in weeks from the latest game to date of analysis.

5) COMBINING THE SCALES

In-game statistics for all players need to be scaled according to different factors that affect the game. We arrive at the final scale by multiplying all the computed scales for the player. For a player p, the final output for an in-game statistic (i.e. measurement) m, would be computed as in Eq. (7).

$$m' = s_{90} s_{comp} s_{diff} s_{rec} m \tag{7}$$

All scales are player and match specific, based on the player's minutes played in a given match, the competition the match is played in, the difficulty of the match, as well as the recency of the match as defined in relevant sections. The scalar production of the scales allows for toggling them on and off as relevant and needed.

B. AGGREGATING PLAYER STATISTICS

To arrive at player-level statistics, in-game statistics for each player must be aggregated into a single value. Aggregation is applied after the scaling step and we employ three basic techniques: scaled averages μ_{avg} (Eq. 8), scaled ratios μ_{rat} (Eq. 9) and strength variable μ_{str} (Eq. 10). Given the definitions of an analysis dimension d, these aggregated features are computed as in respective equations 8, 9, 10.

Eq. (8) reflects the scaled averages of in-game frequencies such as fouls performed or shots taken. This quantifies the activity level of the player in a game. Eq. (9) reflects the scaled performance measures such as recoveries compared to fouls in the form of ratios. This variable quantifies the relative activities between two variables, or when the numerator is the successful subset of the denominator, success rate. Finally, Eq. (10) combined the ratios with the sample size to arrive at a better rounded success measure.

$$\mu_{avg_{pd}} = \frac{\sum_{i=1}^{N_p} m'_{1_{ip}}}{N_p} \tag{8}$$

$$\mu_{rat_{pd}} = \frac{\sum_{i=1}^{N_p} \frac{m'_{1_{ip}}}{m_{2_{ip}}}}{N_p} \tag{9}$$

$$\mu_{str_{pd}} = \frac{\sum_{i=1}^{N_p} F_{binom}(m_{1_{ip}}, m_{2_{ip}}, \overline{\pi})}{N_p} \tag{10}$$

In these equations, $m_{j_{ip}}$ is the j^{th} measure for player p, F_{binom} is the Binomial cumulative distribution function, $\overline{\pi}$ is the average ratio across all players of the two measures in Eq. (9). When calculating μ_{rat}, the measures do not necessarily need to be related. For instance $m_{1_{ip}}$ can refer to fouls and $m_{2_{ip}}$ can refer to recoveries. However when calculated μ_{str}, $m_{1_{ip}}$ must be the successful subset of $m_{2_{ip}}$, because this aggregated feature aims to quantify the amount of evidence towards success in an analysis dimension. Therefore, the strength variable is an extension and improvement over success related ratios in the sense that not only the success of the action is quantified but also the *effect size* [60]. For instance, a ratio of 80% with 5 samples would have less strength than the same ratio of 100 samples, which provides more evidence towards the expected value of the ratio.

There are known issues regarding averaging ratios in μ_{rat}, pointed by [61]. The scaling approach mitigates these issues by transforming them into a non-ratio range.

C. RANKING THE PLAYERS

Aggregated features represent typical player performance across multiple dimensions. These dimensions have different ranges and numerical domains. Therefore, we standardize the aggregated features using min-max standardization F in Eq. (11). We call this standardization *dimension rating* ψ_d for analysis dimension d, which quantifies the performance of the player in this dimension. The result is always in the range of [0, 1].

$$\psi_d = F(\mu_d) \tag{11}$$

As a next step, dimension ratings must be combined to a single metric ρ_p that quantifies the overall player performance. We perform this step by using a dimension reduction approach, Kernel PCA with cosine similarity. In [21], authors specify requirements for a suitable ranking approach and offer PCA as a simple linear solution. However, they also mention that PCA requires orthogonality of the components and therefore it is not suitable for most real-life ranking problems. Kernel PCA is presented as a solution to non-orthogonal multivariate ranking problems. They take the solution a step further and use Bézier curves for one-dimensional embedding. However, Bézier curves require a guarantee for no-ties and in our case we cannot guarantee such topology. Therefore, we opted to use Cosine Kernel-PCA with a cosine kernel which is an established method in document ranking tasks.

IV. EXPERIMENTAL SETUP

To evaluate the methodology, we rated the players who played in the 2016/2017 and 2017/2018 seasons globally. We performed the experiments using the data provided by WyScout, a sample of which is available online [2]. We analyzed the data for 72 competitions and applied the methodology to all players who played in a minimum of 20 games, in which they played as left or right back for at least 90% of the duration of their play-time. These filters result in a total of 3657 players that qualify for the position and the number of player match pairs is 150810, resulting in 1.9 million in-game statistics, for 9 analysis dimensions using 13 variables of in-game player statistics. The data glossary for the individual in-game statistics used can be found in [62].

We also used other publicly available data such as transfer values (for ELO initialization) and UEFA ratings (for referencing domain-specific issues outlined above). We provide a summary of data sources known to us in Table 5. The parameters and their experimental values are provided in Table 3. Out of these parameters, team rating initialization values are based on the best practices coming from chess. In chess, 2500 is roughly assumed to be the grandmaster rating and 1200 is the lowest professional rating level [63].

We set the K parameter of the ELO algorithm to 64 to allow for rapid change in ratings. [64] provide a discussion on the selection of K-factor and explain the effects of different parameter values. Based on their findings, parameter value of 64 represents higher uncertainty than average. Since our aim is to evaluate teams and players on a large scale, this uncertainty is a desired property. For recency scaling, we chose α empirically. The α value provided in Table 3 corresponds to a scaling coefficient of 0.77 for a game that is a year old. The framework can be extended by defining further variables in the same fashion as described in Section III-B.

The only parameter that should be user-defined amongst those listed in Table 3 is the recency scale specifying the importance of older games. The rest are internal parameters to calculate the scaling coefficients. Selection of this parameter is dependent on the importance put on the more recent games.

TABLE 3. Parameter details.

Method Component	Parameter	Details	Exp. Value
Game Difficulty Scale, Competition Quality Scale	K	The factor that determines the amount of impact each game has on the team rating adjustment	64
Game Difficulty Scale, Competition Quality Scale	Min. Rating	The minimum initial rating for the teams	1500
Game Difficulty Scale, Competition Quality Scale	Max. Rating	The maximum initial rating for the teams	2500
Game Difficulty Scale, Competition Quality Scale	Unknown Rating	The initial rating for teams whose average player values are unknown	1200
Recency Scale	α	The decay rate to de-emphasize older matches of the player exponentially	0.995

TABLE 4. Analysis dimensions.

Analysis Dimension	Aggr.	m_1	m_2
Interceptions	μ_{avg}	# Interceptions	
Shot Assists	μ_{avg}	# Shot Assists	
Loose Balls	μ_{str}	# Successful Loose Ball Duels	# Loose Ball Duels
Fouls	μ_{avg}	# Fouls	
Forward Passes	μ_{str}	# Successful Forward Passes	# Forward Passes
Progressive runs	μ_{avg}	# Progressive runs	
Recoveries	μ_{rat}	# Recoveries	# Fouls
Defensive Duels	μ_{str}	# Successful Defensive Duels	# Defensive Duels
Crosses	μ_{str}	# Successful Crosses	# Crosses

V. RESULTS

To showcase the performance of the proposed approach, we apply the methodology and validate results using financial values and team performance in Sections V-A and V-B.

In all the experiments, the data filters (analysis period, position, relevant positions, competitions included, minimum number of games, minimum percentage in relevant roles and minimum duration of game-play) and time decay rate are kept constant. These data filters define the search space for ranking. To apply the methodology to different data filters, we provide a publicly available dashboard.[1] The dashboard also enables the user to apply the methodology to three other target positions.

As stated in [18], there is no established form of validating player performances. The lack of a ground truth makes the validation non-trivial. To mitigate this, we propose a two dimensional validation for player ranks. We first rank the players with the proposed approach, and report the following analyses to assess the rating:

1. We perform financial validation of rankings based on player market valuation after the date of analysis per rank groups. We compare our results to [18] where they use Composition of Probabilistic Preferences (CPP) using triangular distribution [68] to rank players *into groups of ranks (as opposed to assigning an individual rank to each player)* using the published R Package [69]. In addition, we apply COMET [38] method to the rankings and compare outputs of proposed methodology vs. COMET. We also compare both outputs to random rank assignments to establish a baseline for both methods.
2. We perform team-fit validation by analysing the post-analysis ELO ratings of players' teams and show the rank correlation between the player rankings and team ELOs of players' destination teams in case of transfers. We only use the data for players who have been transferred to avoid data leakage. Since the team ELOs are used to calculate scales, the same teams' ELO ratings cannot be used for validation. Therefore, the player should have moved to another team whose ELO rating is de-coupled from the scaled statistics.
3. We provide the standard search ranking evaluation metrics to further validate our results. Player ranking problem is analogous to ordering search results where there is no ground truth or relevance feedback available. We report various information retrieval metrics to quantify the performance of the approach.
4. We provide the of top-20 players who played in the 16/17 and 17/18 seasons, ranked by the proposed methodology. We also provide the market values and ages of the players at the end of 17/18 season for a more comprehensive overview.

As an example of multi-criteria decision making, CPP methodology performs numerical estimations of preference probabilities through combinatorically comparing categories with alternatives, and then calculating overall probabilities of an alternative being better than the rest. CPP-Tri methodology assumes a prior of Triangular distribution when computing initial preferences for CPP. The combinatorial nature of the approach makes it computationally inefficient when there are many alternatives.

When applying CPP-Tri to evaluate player performances, we used standard per 90-minute averages of player statistics that overlap with the analysis dimensions used in the proposed methodology. The choice to use the standard per 90 averages is to ensure consistency with the original paper. Similarly, the specific CPP-Tri algorithm we used assumes a Beta Prior which is kept as is. Class profiles are decentiles which results in 10 distinct player rank groups.

Another MCDM approach is COMET [38]. Unlike CPP-Tri, COMET results in individual rankings as opposed to rank groups. Therefore it is more comparable to the proposed methodology. We compare our outputs to both CPP-Tri and

[1] https://jss-dashboard-aolebn4toq-ew.a.run.app

COMET. These comparisons are kept separate for readability purposes. COMET has already been applied to Football [42] and Basketball [43] however the application to Football is restricted to comparing 6 players, therefore the methodology must be applied to our dataset for comparison purposes.

COMET method relies on having fuzzy membership numbers. To comply with the original application, we divided the standardized analysis dimensions into two distinct parts intervals (0-0.5), (0.5-1). Using these values for all 9 analysis dimensions, 19683 (3^9) characteristic objects were created. This method also relies on having expert feedback on deciding the superiority of one characteristic object over another to create Matrix of Expert Judgement (MEJ). However, in this study no such expert is available, therefore we use an expert decision function f_{exp} given in Eq. (12) to perform judgement where *eps* is an infinitesimally small number to address multiplication with 0. In this equation CO_i represents the i^{th} characteristic object and $\tilde{C}_{ji}$ is the fuzzy number for criteria. For details on the decision process, please refer to [42]. This expert function allows us to parallelize calculation of MEJ and removes the need for hierarchical modelling used in original paper.

$$f_{exp}(CO_i) = \prod_{j=1}^{r}(\tilde{C}_{ji} + eps) \tag{12}$$

A. FINANCIAL VALIDATION

Amongst the numerous factors influencing the valuation in sports, performance is considered to be a significant one [3], [36]. To evaluate the methodology, we have compared the player valuations before and after the analysis period, based on the data from Transfermarkt [45] and cross-referenced this with rank groups. First, we group the player ratings into ten distinct groups using equal-frequency binning. The rank groups are in descending order, where rank-1 holds the highest and rank-10 holds the lowest rated players. For a more comprehensive comparison, we also perform randomization of ratings to establish a baseline to account for effects of inflation of player values. Randomization is done through assigning players into groups randomly using stratified sampling [70]. Table 8 shows the rank group statistics for actual discretization outputs and the randomized case.

In addition, we show the cumulative market value gain of players along the rankings. We define market value gain as the difference between the maximum value after the analysis period ends and the market value of a player immediately after the analysis period. Figure 3 shows the rank-dependent cumulative market value by age group. In other words, if clubs were to invest in a player immediately after ranking, the derivative of this empirical curve dependent on the player's rank and age group would reflect the expected return on player investment. The figure shows diminishing returns after top-25th percentile of rankings for all age groups up to age of 29. For players above age of 29, the expected return

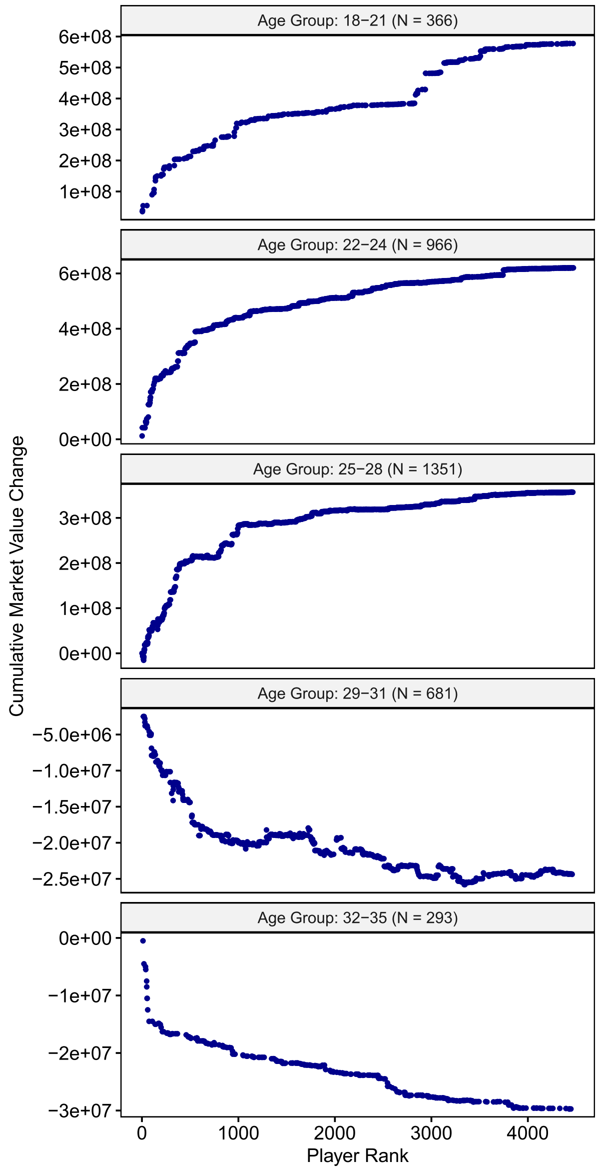

FIGURE 3. Cumulative return on investment per age group and rank.

on investment is always negative. In addition, there are two cut-off points in the youngest age group. This is discussed further in Section VI. Note that while actual transfer values are used for Figure 4, market value estimations from Transfermarkt are used for Figure 3. The same plot cannot be shown for [18] because the methodology presented by the authors only output rank clusters, in our replication 10 groups of player ranks and does not rank players individually.

Figure 4 shows distributions of player values before and after the rating period. For all rank groups, post analysis player values are higher, indicating that player values increase due to inflation. In case of rank groups 3 and onwards, randomized player values are higher than the post-analysis player values. In rank groups 1 and 2, however, the post-analysis player values are higher than the randomized ones. These observations indicate that, our approach identifies the players who have increased their market

TABLE 5. Data sources.

Data Source	Coverage	Collection Method	Data Verification	Match Outcomes	Player In-Game Statistics	Events	Transfers	Financial Information	Club Rankings	Country Rankings
[45]	Major competitions and players	Crowdsourced	Expert Opinion				X	X		
[59]	UEFA Member Association	Expert Opinion	Expert Opinion						X	X
[65]	Major competitions and players	Aggregated from Opta Sports	Unknown	X	X					
[7]	Over 1000 competitions and 200k players	Human experts using bespoke analysis software [66]	Expert Review	X	X	X				
[6]	Over 1000 competitions and 200k players	Human experts using bespoke analysis software [2]	Expert Review	X	X	X	X			
[5]	Unknown	Unknown	Unknown	X	X	X				
[67]	Unknown, however the website claims more granularity than any other provider	Unknown	Unknown	X	X	X				

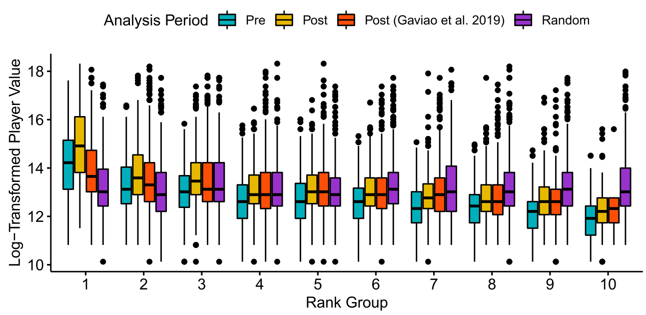

FIGURE 4. Player market value distribution per rank group and analysis period.

TABLE 6. Two-Sample Kolmogorov-Smirnov Test P-Values.

Rank Group	Proposed Method vs. Random	Proposed Method vs. Gaviao et. al	Gaviao et. al vs. Random
1	0.0000	0.0000	0.0000
2	0.0000	0.0357	0.0047
3	0.0003	0.0034	1
4	0.5741	0.4399	1
5	1	0.6010	1
6	1	0.7928	1
7	1	0.7073	1
8	1	0.3848	1
9	1	0.7073	1
10	1	0.7928	1

value beyond inflation, after they have been identified as high-performing players. It has to be noted that the reported values in this figure are log-transformed for visualization, while the effect would be more pronounced in linear scale. For instance, for two players having log-transformed market values of 16 and 13, the corresponding difference in absolute market values would be approximately 8.5 million Euros. For this analysis, only the players who have been transferred after the analysis period have been analyzed by comparing their transfers prior to the analysis and after. As stated in [46], Transfermarkt estimated values could be unreliable. To avoid spurious relationships, we used the actual transfer values as stated in Transfermarkt and not their interim estimates. This figure also contains comparison to a recent paper [18] based on the ranking part of their methodology. While the original paper compared 32 players and used 3 (low-medium-high) performance classes, we applied both methods to 3657 players and used 10 performance classes to account for higher number of the players.

The distribution of pre- and post-analysis values using their methodology is similar to our results, however their results overlap with the distribution of randomly allocated label, showing that our proposed methodology is able to differentiate players who are likely to increase in value beyond the inflation compared to CPP-Tri ranking methodology.

For a more theoretically-based comparison, we also report the results of Two-Sample Kolmogorov-Smirnov Statistical Test (one-tail) for comparing two non-parametric distributions. Formally the hypothesis is defined as:

H_0: Player values of the ranks calculated with the first methodology are drawn from the same distribution as the player values of the ranks calculated with the second methodology.

H_a: Player values of the ranks calculated with the first methodology are drawn from a distribution with a *greater* average than the distribution of the player values of the ranks calculated with the second methodology.

Table 6 shows the 'False Discovery Rate' corrected [71] p-values of the performed test, for ranks obtained via methodologies listed pairwise. The results show that both the proposed methodology and methodology by [18] have significantly greater player values than random allocation for first two rank groups at a significance level of 0.05. The statistical test shows that there is significant evidence that our proposed methodology has greater player values in the third rank group in addition to the first two. Furthermore, the player values in the first rank group identified by our proposed methodology have significantly higher player values than the players identified as rank one by [18]. These results confirm that the proposed methodology is more successful in identifying players who will increase their market value beyond inflation based on their performance statistics. For the low rank groups (8^{th}–10^{th}), statistical non-significance is expected because we perform a one-tailed hypothesis test (i.e. greater average). The statistical non-significance of middle ranking (4^{th}–7^{th}) groups are discussed further in the Discussion section.

B. TEAM-FIT VALIDATION

Team fit validation is based on the idea that high performing teams consist of high performing players. So player and team ranks are expected to have statistically significant correlation. However, team ratings are already utilized in creating of scales and aligning statistics for fair comparison. Therefore,

TABLE 7. Market value change percentages by group.

Market Value Rank Group	Minimum Percentage Change	Maximum Percentage Change
1	1546	1983
2	724	1400
3	456	700
4	280	450
5	168	275
6	105	167
7	60	100
8	12	50
9	-20.6	11.1
10	-93.3	-21.9

TABLE 8. Rank group statistics actual vs. randomized.

	Rank Group						Randomized					
Rank Group	Number of Players	Min. Rating	Max. Rating	Avg. Rating	Median Rating	Std. Rating	Number of Players	Min. Rating	Max. Rating	Avg. Rating	Median Rating	Std. Rating
1	347	1.5	2.17	1.63	1.61	0.11	334	0.7	1.84	1.21	1.22	0.22
2	331	1.39	1.5	1.44	1.44	0.03	325	0.63	1.97	1.18	1.17	0.23
3	347	1.31	1.39	1.35	1.35	0.02	346	0.68	1.88	1.19	1.19	0.22
4	360	1.25	1.30	1.28	1.28	0.02	331	0.63	1.98	1.2	1.19	0.22
5	344	1.19	1.25	1.22	1.22	0.02	328	0.65	1.95	1.19	1.17	0.22
6	370	1.13	1.19	1.16	1.16	0.02	378	0.69	1.83	1.21	1.21	0.22
7	305	1.08	1.13	1.10	1.10	0.01	327	0.49	2.17	1.2	1.18	0.25
8	341	1.01	1.08	1.05	1.05	0.02	361	0.54	1.91	1.19	1.19	0.24
9	366	0.91	1.00	0.96	0.97	0.03	402	0.65	1.91	1.21	1.20	0.23
10	346	0.49	0.91	0.83	0.84	0.07	325	0.58	1.91	1.21	1.19	0.22

to avoid data leakage, the validation can only be performed using the data of the players who have been transferred into a new team after the analysis. This prevents leakage from team ELO points into validation. In other words, by using the data from post-analysis transfers we decouple the scaling from output validation.

Figure 5 shows the relationship between the team player got transferred to (i.e. destination team) post-analysis and the identified ranking of the player. The figure also reports the Spearman rank correlation between two values. The rank correlation is statistically significant in all confidence levels. We also report the linear regression least square estimation of the relationship between rankings and team ratings. Namely, the expected ELO of a player's post-analysis destination, decreases by 0.068 points with every step decrease in their rankings. Most ELO points are distributed between 1500 and 2000, which is in-line with Chess literature [22]. The variance in the ELO ratings given the player ranks is discussed further in Section VI.

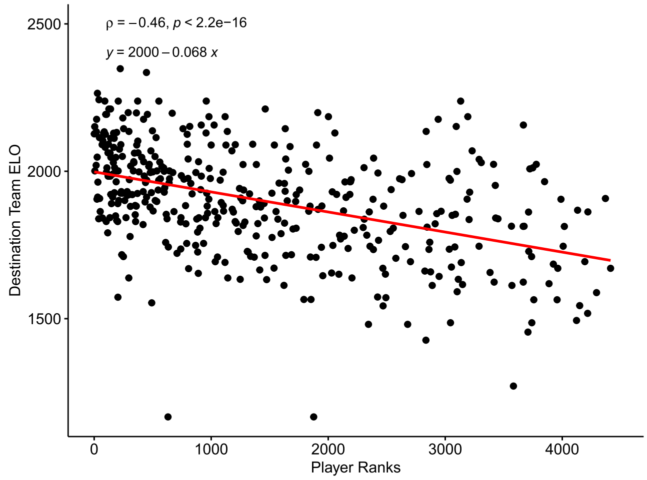

FIGURE 5. Relationship between ranks and post-analysis destination team ratings.

C. PLAYER RANKING EVALUATION USING INFORMATION RETRIEVAL METRICS

Ranking players is an applied problem of ordering a set of items where there is no ground-truth available. Therefore, the correct rankings cannot be learned from feedback. Nor can the final output be validated by comparing to true rankings. This is a well-known problem in retrieval problems where ranking of search results are not customized to user preference (i.e. no feedback). The evaluation metrics developed for such problems are relevant for evaluating our results.

Similar to web search literature, in the absence of ground-truth, comparison of two-rankings is the established method of evaluating ranking outputs. As previously stated, no other methodology outputs such comprehensive and comparable rankings to us, therefore we opt to use post-analysis player values as a proxy to player performance. Note that similar to the return on player investment, these metrics cannot be computed for Gavião *et al.* [18] since their approach does not produce an ordered list of players but rank groups instead.

To validate the results, we provide the following metrics.

1) Kendall's τ [72], which quantifies the overall agreement between two ranked arrays.
2) Average precision [73], which computes performance at several intervals in the rankings and reports the average.

TABLE 9. The results of the proposed solution using information retrieval metrics.

Metric	Proposed Method	COMET score w. Proposed Scaling and Feat. Engineering	COMET score w. Raw In-Game Stats
Kendall's τ	0.364	0.307	-0.026
Average Precision	0.601 ± 0.011	0.549 ± 0.042	0.480 ± 0.070
Normalised Cumulative Discounted Gain (NCDG)	0.707	0.704	0.672
WS Coefficient	0.987	0.845	0.262

TABLE 10. Top 20 Players and Their Metadata.

Rank	Name	Team Prior to Analysis	Team After Analysis	Market Value Before Analysis	Market Value After Analysis	Current League	Age at June, 2018
1	Fabian Delph	Manchester City	Manchester City	10M	15M	Premier League (England)	21
2	Luke Shaw	Manchester United	Manchester United	21M	18M	Premier League (England)	22
3	Joshua Kimmich	Bayern Munich	Bayern Munich	20M	60M	Bundesliga (Germany)	23
4	Ben Chilwell	Leicester City	Leicester City	2.5M	15M	Premier League (England)	21
5	Benjamin Henrichs	Bayer 04 Leverkusen	AS Monaco	1.25M	16M	League 1 (France)	21
6	Alessandro Florenzi	AS Roma	AS Roma	22M	30M	Seria A (Italy)	27
7	Kieran Gibbs	Arsenal FC	West Bromwich Albion	13M	6M	Premier League (England)	28
8	Davide Santon	Inter Milan	AS Roma	6M	5M	Seria A (Italy)	27
9	Christian Fuchs	Leicester City	Leicester City	6M	3.5M	Premier League (England)	32
10	Nico Elvedi	Borussia Mönchengladbach	Borussia Mönchengladbach	5M	20M	Bundesliga (Germany)	21
11	Francisco Femenia Far	Deportivo Alaves	Watford FC	1M	5M	Premier League (England)	27
12	Jonas Hector	1. FC Köln	1. FC Köln	14M	12.5M	Bundesliga (Germany)	28
13	Marcel Schmelzer	Borussia Dortmund	Borussia Dortmund	9M	5M	Bundesliga (Germany)	30
14	Sergio Roberto Carnicer	FC Barcelona	FC Barcelona	20M	55M	La Liga (Spain)	26
15	Joshua Brenet	PSV Eindhoven	TSG 1899 Hoffenheim	3.5M	5M	Bundesliga (Germany)	24
16	Íñigo Lekue Martínez	Athletic Bilbao	Athletic Bilbao	3M	4M	La Liga (Spain)	31
17	Kieran Trippier	Tottenham Hotspur	Tottenham Hotspur	6M	30M	Premier League (England)	27
18	Jeffrey Schlupp	Leicester City	Crystal Palace	8M	10M	Premier League (England)	25
19	Aleksandar Kolarov	Machester City	AC Roma	11M	12M	Seria A (Italy)	32
20	David Olatukunbo Alaba	Bayern Munich	Bayern Munich	45M	55M	Bundesliga (Germany)	26

3) Normalised discounted cumulative gain [74] which quantifies the gain obtained by the ranking system by correctly ranking items.
4) WS Coefficient [75] which is shown to evaluate rankings in a more well-rounded way by giving more emphasis to correctly ranking top items. In addition, unlike NDCG, it requires an input from users regarding the importance of correctly ranking observations.

In information retrieval problems, it is more important to rank the highest performers correctly than ranking lowest performers correctly [76]. Kendall's τ does not take this distinction into account. Conversely, average precision correlation is based on the assumption that correctly ranking the top items is more important. In the case where ranking system's error is distributed randomly, average precision correlation is equivalent to Kendall's τ. If the system has a higher precision for ranking the top-ranking items, then average precision correlation is higher than Kendall's τ, and vice versa. Furthermore, a novel technique WS coefficient [75] also gives more weight to top-ranked items.

Table 9 provides the results of the proposed method using the aforementioned metrics.

All metrics show a positive correlation between the rankings provided by the methodology and the eventual player value. Furthermore, order-dependent metrics (i.e. metrics other than Kendall's τ) have higher values than τ, showing that the proposed methodology is able to capture the rankings of the highest-valued players. However, ranking by player value is an imperfect baseline as player values are impacted by other factors as well. We show this phenomenon in the age-dependent figures and discuss the issue further in Section VI.

In addition, COMET was applied to the proposed scaling and aggregation methodology behaves similarly to the results provided by the proposed methodology, however, the proposed methodology outperforms the results obtained with COMET. This is due to COMET's characteristic object representation dividing the feature space into buckets of preference, and treating all values fall into the same bucket with similar preference values. COMET applied to raw data performs significantly worse than the proposed methodology as well as COMET applied to scaled and aggregated data. This reflects the need for considering factors that affect performance into account while analyzing sports data globally.

The drastic difference between average precision and WS Coefficient should be noted. We hypothesize this is due to large sample size for ranking in this study. More analysis is required to compare AP and WS Coefficient in ranking problems but this is out of scope for this study.

D. REAL-WORLD RESULTS

Current configuration of the methodology outputs the ranks for 3681 players, of which top-20 players according to their ratings are listed in Table 10. This table shows the team and age at the beginning and end of the analysis period to provide information regarding the status of the top-ranked players. It also gives details on the player values obtained from Transfermarkt [45] right before and immediately after the analysis period.

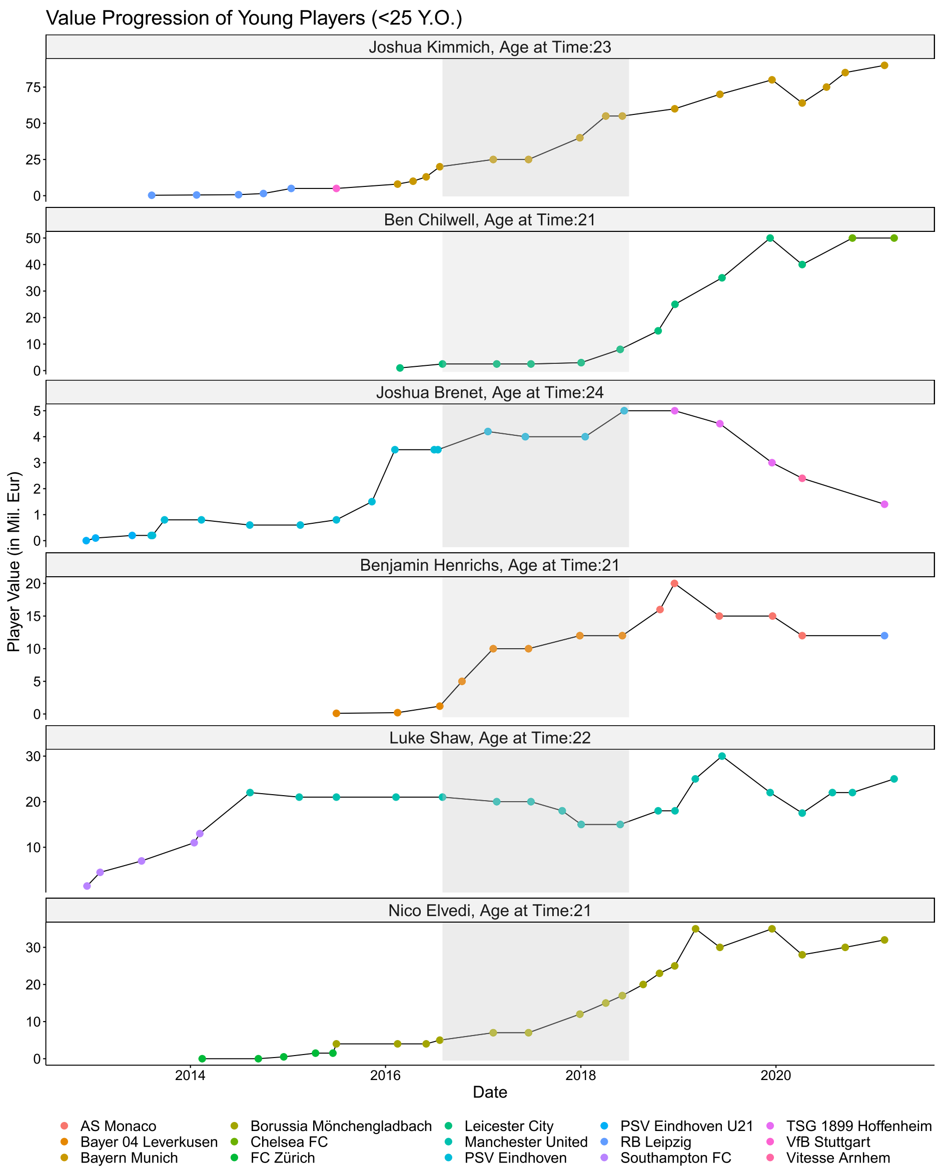

FIGURE 6. Market value changes through time.

Fourteen players in top-20 have increased their market value at the end of analysis period. An exception to this as a young player is Luke Shaw. He sustained an injury which put him on the sidelines for 14 matches in season of 16/17. Shortly after that he sustained another injury that put him on the sidelines for another 11 weeks in the same season. This is potentially one reason why his market valuation has decreased immediately after the end of our analysis period. His current market value is 25M Euros. In other cases, market value decreases are explained by the age of the player which is a general pattern shown in Figure 3.

Figure 6 shows that all but one top-ranking players below age 25 have increased their market values after the period of analysis (i.e. after being identified as top-performers). The period for which the performance data was used in analysis is marked with a grey shade in the figure.

To enable the readers to test the methodology and the configurations, we provide a public API,[2] which uses the publicly available data by [77]. We also provide a graphical user interface to compute results of arbitrary configurations in the form of a dashboard[3] which uses the same API.

VI. DISCUSSION

We have analyzed more than 3500 players across various competitions and shown that the proposed methodology is

[2]https://jss-api-aolebn4toq-ew.a.run.app/redoc
[3]https://jss-dashboard-aolebn4toq-ew.a.run.app

able to identify players who are likely to increase their financial values. Existing literature restricts the pool of the players to a small subset for analysis. In this paper, we proposed a framework that can analyze player performances at a much larger scale and demonstrated statistically significant improvement over the existing methods.

The rating outputs of the methodology correlate to future value increase statistically significantly. These results indicate that the high-rated players are likely to increase their market value beyond inflation. This has implications for teams making player investments. The team-movements are also correlated with the ratings. The high-rated players make more upwards transfers than the low-rated players. Furthermore, the methodology statistically outperforms the existing approaches for the high-rank groups. Therefore it provides an improvement in terms of scalability and performance.

On the other hand, the statistical non-significance of the middle rank groups (4–7) reflects the complexity of the problem. There are many other factors that affect the transfer values, such as player popularity or coach preferences as well as the inflation and even the country of origin of the player [78]. Figures 3 and 6 show the impact of age on market values as well. Additional factors could be included in the methodology to refine the ratings further and the validation can be extended to factor-in contextual information that impacts transfer values.

We also expect similar contextual factors to affect the team ratings shown in Figure V-B. For instance, due to transfer restrictions, a good player may not be eligible to play in top teams before first moving to a lower-rated team, therefore nationality may impact the destination player ratings. We also expect age to be a factor as teams loan out young players to other teams to optimize their rosters while allowing the young player to gain experience.

Table 7 indicates that player values increase for almost all rank groups, mainly due to player value inflation. Average increase in value across competitions is calculated as 42.28% with a standard deviation of 56.18%. The maximum value increase is 264.18% and the highest decrease is 62.29%. However, the magnitude of percentage change in higher rank groups is much larger than lower ones which indicates that rankings capture non-inflation related factors that contribute to market value.

An interesting phenomenon is the observed effect of age in player valuation. Figure 6 shows that the top ranked players under the age of 25 have almost all increased their market values post-analysis regardless of they were transferred or not. This indicates, if combined with age and demographic filters, proposed methodology may help teams optimize their transfer decisions in terms of financial valuation. The same figure also shows that the range of the player values varies. This supports the prior claims that the player values may not always overlap with the player performance and statistics. The additional factors that might impact the player valuation and transfer market are affected by demographic and social factors.

The cumulative gain on market values given in Figure 3 gives a different perspective on player performance. This figure reflects the investment view that could impact player acquisition decisions. After a certain level of performance ranking (around top-25th percentile), return on any potential investment becomes negligible. This cut-off point (i.e. the elbow point on plots) from an investment stand-point varies by age group. The cut-off point decreases with age. The cut-off for players aged below 21, is around 1000^{th} rank. For players below 24 years of age, this point is around 800^{th} rank and for players younger than 28 years of age, the elbow point is around 500^{th} rank. In other words, investing in a top-1000 ranked player of less than 21 years of age might be a financially good decision, however for older players, this is not the case. This shows that age and performance jointly affect player valuation. The same figure also shows two elbow-points in the youngest age group. This is due to scaling approach. Competitions with low player values such as second or third division competitions occasionally produce young super-stars, however, in general these competitions have lower quality. Therefore, young talents in such competitions are ranked low, however they have high future return. An example is Achraf Hakimi who was in Real Madrid Castilla in the beginning of the analysis period and got transferred to Inter in 2020. In other words, ranking methodology is unable to detect future super-star young players, who are outliers in their own competitions.

Under the assumption that market values imperfectly reflect player performance, Figures 3, 4, and Table 6 all show that highly-ranked players also have high performance in financial terms. Same behavior is also evident in Figure 5 from a purely performance stand-point.

VII. CONCLUSION & FUTURE WORK

In this study, we proposed a framework for player performance evaluation across multiple dimensions on a global scale. Naturally, it is impossible to statistically encode every dimension of such a complex game, however, the proposed approach is extendable as a framework, both within football and to other team sports. We aimed to capture the major components of the game and extended the player evaluation application to perform across-competitions, across-time, and across different opponent types. Using this framework, we have shown that the computed ranks reflect both financial and sportive performance of players.

As a future work, we aim to develop solutions to identify gaps in team performance and cross-reference these gaps with our player recommendations by using optimization techniques, as opposed to generating a general ranking for all players that satisfy the criteria. Studies have already been conducted towards automatic inference of team gaps and strategies on match logs [47], [48], [50], [79]. We intend to explore incorporation of such approaches to increase customization of rankings to fit club requirements and queries. We also intend to investigate the transfer market dynamics through various techniques that fall under the domain of

Complexity Economics [80], potentially including important demographic factors such as age.

ACKNOWLEDGMENT

This study has been made possible by a special access provided by [6] for a broader range of dataset. The authors would like to thank WyScout and especially Emanuele Massucuo for their contribution to this work.

REFERENCES

[1] I. G. McHale, P. A. Scarf, and D. E. Folker, "On the development of a soccer player performance rating system for the English premier league," *Interfaces*, vol. 42, no. 4, pp. 339–351, Aug. 2012.

[2] L. Pappalardo, P. Cintia, P. Ferragina, E. Massucco, D. Pedreschi, and F. Giannotti, "PlayeRank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 5, pp. 1–27, Nov. 2019.

[3] C. P. Barros and S. Leach, "Analyzing the performance of the English F.A. premier league with an econometric frontier model," *J. Sports Econ.*, vol. 7, no. 4, pp. 391–407, Aug. 2016.

[4] J. Duch, J. S. Waitzman, and L. A. N. Amaral, "Quantifying the performance of individual players in a team activity," *PLoS ONE*, vol. 5, no. 6, Jun. 2010, Art. no. e10937.

[5] InStat. (2020). *InStat*. [Online]. Available: https://football.instatscout.com/

[6] W. Wyscout. (2020). *US Wyscout*. [Online]. Available: https://wyscout.com/

[7] O. Opta. (2020). *Opta Sports*. [Online]. Available: https://www.optasports.com/sports/football/

[8] L. M. Hvattum and H. Arntzen, "Using ELO ratings for match result prediction in association football," *Int. J. Forecasting*, vol. 26, no. 3, pp. 460–470, Jul. 2010.

[9] M. Lewis, *Moneyball: The Art of Winning an Unfair Game*. New York, NY, USA: WW Norton & Company, 2004.

[10] B. Baumer and A. Zimbalist, *The Sabermetric Revolution*. Philadelphia, PA, USA: Univ. Pennsylvania Press, 2014.

[11] J. A. Martínez and L. Martínez, "A stakeholder assessment of basketball player evaluation metrics," *J. Hum. Sport Exercise*, vol. 6, no. 1, pp. 153–183, Mar. 2011.

[12] D. S. Mason, "Moneyball as a supervening necessity for the adoption of player tracking technology in professional hockey," *Int. J. Sports Marketing Sponsorship*, vol. 8, no. 1, pp. 41–55, Oct. 2006.

[13] S. McIntosh, S. Kovalchik, and S. Robertson, "Validation of the Australian football league player ratings," *Int. J. Sports Sci. Coaching*, vol. 13, no. 6, pp. 1064–1071, Dec. 2018.

[14] S. Pettigrew, "Assessing the offensive productivity of NHL players using in-game win probabilities," in *Proc. 9th Annu. MIT Sloan Sports Anal. Conf.*, vol. 2, 2015, p. 8.

[15] S. Pradhan, "Ranking regular seasons in the NBA's modern era using grey relational analysis," *J. Sports Analytics*, vol. 4, no. 1, pp. 31–63, Feb. 2018.

[16] O. Schulte, Z. Zhao, M. Javan, and P. Desaulniers, "Apples-to-apples: Clustering and ranking NHL players using location information and scoring impact," in *Proc. MIT Sloan Sports Analytics Conf.*, 2017.

[17] M. Stewart, H. Mitchell, and C. Stavros, "Moneyball applied: Econometrics and the identification and recruitment of elite Australian footballers," *Int. J. Sport Finance*, vol. 2, no. 4, pp. 231–248, 2007.

[18] L. O. Gavião, A. P. Sant'Anna, G. B. A. Lima, and P. A. D. A. Garcia, "Evaluation of soccer players under the Moneyball concept," *J. Sports Sci.*, vol. 38, pp. 1221–1247, Dec. 2019.

[19] V. C. Pantzalis and C. Tjortjis, "Sports analytics for football league table and player performance prediction," in *Proc. 11th Int. Conf. Inf., Intell., Syst. Appl. (IISA*, Jul. 2020, pp. 1–8.

[20] G. Liu, Y. Luo, O. Schulte, and T. Kharrat, "Deep soccer analytics: Learning an action-value function for evaluating soccer players," *Data Mining Knowl. Discovery*, vol. 34, no. 5, pp. 1531–1559, Sep. 2020.

[21] C.-G. Li, X. Mei, and B.-G. Hu, "Unsupervised ranking of multi-attribute objects based on principal curves," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 12, pp. 3404–3416, Dec. 2015.

[22] A. E. Elo, *The Rating of Chessplayers, Past and Present*. New York, NY, USA: Arco Pub., 1978.

[23] K. Doya, "Reinforcement learning in continuous time and space," *Neural Comput.*, vol. 12, no. 1, pp. 219–245, 2000.

[24] R. Rein and D. Memmert, "Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science," *SpringerPlus*, vol. 5, no. 1, pp. 1–13, Dec. 2016.

[25] M. Stein, H. Janetzko, D. Seebacher, A. Jäger, M. Nagel, J. Hölsch, S. Kosub, T. Schreck, D. Keim, and M. Grossniklaus, "How to make sense of team sport data: From acquisition to data modeling and research aspects," *Data*, vol. 2, no. 1, p. 2, Jan. 2017.

[26] C. Lucifora and R. Simmons, "Superstar effects in sport: Evidence from Italian soccer," *J. Sports Econ.*, vol. 4, no. 1, pp. 35–55, Feb. 2003.

[27] B. Torgler and S. L. Schmidt, "What shapes player performance in soccer? Empirical findings from a panel analysis," *Appl. Econ.*, vol. 39, no. 18, pp. 2355–2369, Oct. 2007.

[28] B. Gerrard, "Analysing sporting efficiency using standardised win cost: Evidence from the FA premier league, 1995 – 2007," *Int. J. Sports Sci. Coaching*, vol. 5, no. 1, pp. 13–35, Mar. 2010.

[29] M. Hughes, T. Caudrelier, N. James, I. Donnelly, A. Kirkbride, and C. Duschesne, "Moneyball and soccer—An analysis of the key performance indicators of elite male soccer players by position," *J. Hum. Sport Exerc.*, vol. 7, no. 2, pp. 402–412, 2012.

[30] J. Ali, S. Shahram, and M. Thomas, "Modeling team compatibility factors using a semi-Markov decision process: A data-driven approach to player selection in soccer," *J. Quant. Anal. Sports*, vol. 9, no. 4, pp. 347–366, 2013.

[31] P. Cintia, F. Giannotti, L. Pappalardo, D. Pedreschi, and M. Malvaldi, "The harsh rule of the goals: Data-driven performance indicators for football teams," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2015, pp. 1–10.

[32] J. Brooks, M. Kerr, and J. Guttag, "Developing a data-driven player ranking in soccer using predictive model weights," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 49–55.

[33] P. Power, H. Ruiz, X. Wei, and P. Lucey, "Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 1605–1613.

[34] D. Barron, G. Ball, M. Robins, and C. Sunderland, "Artificial neural networks and player recruitment in professional soccer," *PLoS ONE*, vol. 13, no. 10, Oct. 2018, Art. no. e0205818.

[35] F. Wunderlich and D. Memmert, "The betting odds rating system: Using soccer forecasts to forecast soccer," *PLoS ONE*, vol. 13, no. 6, Jun. 2018, Art. no. e0198668.

[36] L. Pappalardo and P. Cintia, "Quantifying the relation between performance and success in soccer," *Adv. Complex Syst.*, vol. 21, nos. 3–4, May 2018, Art. no. 1750014.

[37] K. Palczewski and W. Sałabun, "Identification of the football teams assessment model using the COMET method," *Procedia Comput. Sci.*, vol. 159, pp. 2491–2501, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050919316291

[38] S. Faizi, T. Rashid, W. Sałabun, S. Zafar, and J. Wątróbski, "Decision making with uncertainty using hesitant fuzzy sets," *Int. J. Fuzzy Syst.*, vol. 20, no. 1, pp. 93–103, Jan. 2018, doi: 10.1007/s40815-017-0313-2.

[39] T. Pereira, J. Ribeiro, F. Grilo, and D. Barreira, "The golden index: A classification system for player performance in football attacking plays," *Proc. Inst. Mech. Eng., P, J. Sports Eng. Technol.*, vol. 233, no. 4, pp. 467–477, Dec. 2019.

[40] C. Ley, T. V. D. Wiele, and H. V. Eetvelde, "Ranking soccer teams on the basis of their current strength: A comparison of maximum likelihood approaches," *Stat. Model.*, vol. 19, no. 1, pp. 55–73, Feb. 2019.

[41] E. S. Epstein, "A scoring system for probability forecasts of ranked categories," *J. Appl. Meteorol.*, vol. 8, no. 6, pp. 985–987, Dec. 1969.

[42] W. Sałabun, A. Shekhovtsov, D. Pamučar, J. Wątróbski, B. Kizielewicz, J. Więckowski, D. Bozanić, K. Urbaniak, and B. Nyczaj, "A fuzzy inference system for players evaluation in multi-player sports: The football study case," *Symmetry*, vol. 12, no. 12, p. 2029, Dec. 2020. [Online]. Available: https://www.mdpi.com/2073-8994/12/12/2029

[43] B. Kizielewicz and L. Dobryakova, "MCDA based approach to sports players' evaluation under incomplete knowledge," *Procedia Comput. Sci.*, vol. 176, pp. 3524–3535, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920319293

[44] R. Stanojevic and L. Gyarmati, "Towards data-driven football player assessment," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 167–172.

[45] T. Transfermarkt. (2020). *Transfermarkt*. [Online]. Available: https://www.transfermarkt.com/

[46] O. Müller, A. Simons, and M. Weinmann, "Beyond crowd judgments: Data-driven estimation of market value in association football," *Eur. J. Oper. Res.*, vol. 263, no. 2, pp. 611–624, Dec. 2017.

[47] P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews, "Assessing team strategy using spatiotemporal data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 1366–1374.

[48] Q. Wang, H. Zhu, W. Hu, Z. Shen, and Y. Yao, "Discerning tactical patterns for professional soccer teams: An enhanced topic model with applications," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 2197–2206.

[49] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[50] T. Decroos, J. Van Haaren, and J. Davis, "Automatic discovery of tactics in spatio-temporal soccer match data," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 223–232.

[51] P. Oskouie, S. Alipour, and A.-M. Eftekhari-Moghadam, "Multimodal feature extraction and fusion for semantic mining of soccer video: A survey," *Artif. Intell. Rev.*, vol. 42, no. 2, pp. 173–210, Aug. 2014, doi: 10.1007/s10462-012-9332-4.

[52] W. W. Cohen, R. E. Schapire, and Y. Singer, "Learning to order things," *J. Artif. Intell. Res.*, vol. 10, pp. 243–270, May 1999.

[53] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, Apr. 1998.

[54] A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in *Machine Learning: ECML 2007*, J. N. Kok, J. Koronacki, R. L. D. Mantaras, S. Matwin, D. Mladenič, and A. Skowron, Eds. Berlin, Germany: Springer, 2007, pp. 616–623.

[55] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 472–479.

[56] StatsBomb. (Aug. 2013). *An Introduction to the 'Per 90' Metric*. [Online]. Available: https://statsbomb.com/2013/08/an-introduction-to-the-per-90-metric/

[57] C. R. Rao, X. Shi, and Y. Wu, "Approximation of the expected value of the harmonic mean and some applications," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 44, pp. 15681–15686, Nov. 2014.

[58] P. M. E. Glickman, "Example of the Glicko-2 system," 2003, p. 6.

[59] UEFAcom. (2020). *UEFA Coefficients*. [Online]. Available: https://www.uefa.com/memberassociations/uefarankings/

[60] C. O. Fritz, P. E. Morris, and J. J. Richler, "Effect size estimates: Current use, calculations, and interpretation," *J. Experim. Psychol., Gen.*, vol. 141, no. 1, pp. 2–18, 2012.

[61] V. Larivière and Y. Gingras, "Averages of ratios vs. Ratios of averages: An empirical analysis of four levels of aggregation," *J. Informetrics*, vol. 5, no. 3, pp. 392–399, Jul. 2011.

[62] W. Wyscout. (Oct. 2020). *Wyscout Glossary*. [Online]. Available: https://dataglossary.wyscout.com/

[63] U. USCF. (1996). *About USCF*. [Online]. Available: http://www.uschess.org/index.php/About-USCF/

[64] A. N. Langville and C. D. Meyer, *Who's #1?*. Princeton, NJ, USA: Princeton Univ. Press, 2012.

[65] O. Opta. (May 2020). *WhoScored.com*. [Online]. Available: https://www.whoscored.com/

[66] O. Opta. (2020). *The Inner Beauty of the Beautiful Game: How Opta Analyses Football*. [Online]. Available: https://www.stuff.tv/features/inner-beauty-beautiful-game-how-opta-analyses-%football

[67] StatsBomb. (May 2012). *Football Like Never Before*. [Online]. Available: https://statsbomb.com/

[68] A. P. Sant'Anna, H. G. Costa, and V. Pereira, "CPP-TRI: A sorting method based on the probabilistic composition of preferences," *Int. J. Inf. Decis. Sci.*, vol. 7, no. 3, pp. 193–212, 2015.

[69] L. O. Gavi ao, A. P. Sant'Anna, G. B. A. Lima, and P. Garcia. (2018). *CPP: Composition of Probabilistic Preferences. R Package Version 0.1. 0*. ViennaR Core Team, 2018a. Disponívelem:<. [Online]. Available: https://cran.r-project.org/package=CPP

[70] W. A. Ericson, "Optimum stratified sampling using prior information," *J. Amer. Stat. Assoc.*, vol. 60, no. 311, pp. 750–771, Sep. 1965.

[71] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc., Ser. B*, vol. 57, pp. 289–300, Jan. 1995.

[72] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, pp. 81–93, Jun. 1938.

[73] E. Yilmaz and J. A. Aslam, "Estimating average precision with incomplete and imperfect judgments," in *Proc. 15th ACM Int. Conf. Inf. Knowl. Manage.*, 2006, pp. 102–111.

[74] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.

[75] W. Salabun and K. Urbaniak, "A new coefficient of rankings similarity in decision-making problems," in *Computational Science—ICCS* (Lecture Notes in Computer Science), V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, and J. Teixeira, Eds. Cham, Switzerland: Springer, 2020, pp. 632–645.

[76] E. Yilmaz, J. A. Aslam, and S. Robertson, "A new rank correlation coefficient for information retrieval," in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2008, pp. 587–594.

[77] L. Pappalardo, P. Cintia, A. Rossi, E. Massucco, P. Ferragina, D. Pedreschi, and F. Giannotti, "A public data set of spatio-temporal match events in soccer competitions," *Sci. Data*, vol. 6, no. 1, p. 236, Oct. 2019, doi: 10.1038/s41597-019-0247-7.

[78] E. Nsolo, P. Lambrix, and N. Carlsson, "Player valuation in European football," in *Machine Learning and Data Mining for Sports Analytics*, U. Brefeld, J. Davis, J. Van Haaren, and A. Zimmermann, Eds. Cham, Switzerland: Springer, 2019, pp. 42–54.

[79] H. Anton, "In-game strategy recommendations in association football: A study based on network theory," Tech. Rep., 2019.

[80] W. B. Arthur, "Complexity and the economy," in *Handbook of Research on Complexity* (Chapters), J. B. R. Jr., Ed. Cheltenham, U.K.: Edward Elgar Publishing, Jun. 2009, ch. 2. [Online]. Available: https://ideas.repec.org/h/elg/eechap/3625_2.html

AYSE ELVAN AYDEMIR was born in Kayseri, Turkey, in 1990. She received the B.S. degree in actuarial science from Hacettepe University and the M.Sc. degree in information systems from the Graduate School of Informatics, Middle East Technical University (METU), in 2014. During that time, she worked as a Research Assistant. She continued her academic studies in the Information Systems Ph.D. Program at METU. She is currently writing her Ph.D. dissertation.

In 2018, she took a position as a Lead Data Scientist at Amplify Analytix, Sofia, Bulgaria, in which she led a team of ten data scientists situated in Sofia and Bengaluru. In January 2021, she joined Enskai Ltd., as the Head of Research. Her research interests include sports analytics, explainable and interpretable machine learning, dynamical and complex systems modeling, ethics in data science and machine learning, causal inference, and agent-based economic models. She won the 2nd Best Graduate Thesis Award in computer vision, in 2014.

TUGBA TASKAYA TEMIZEL received the B.Sc. degree from the Computer Engineering Department, Dokuz Eylül University, in 1999, and the Ph.D. degree from the Department of Computing, University of Surrey, U.K., in 2006. She is currently working as an Associate Professor with the Graduate School of Informatics, Middle East Technical University. She is also an NVIDIA Deep Learning Institute Certified Instructor and a University Ambassador. She collaborated with and advised several companies on data mining techniques and social media analysis. She has been an Expert Evaluator of the EU H2020 Research Program and Nationally Funded Programs. She was on sabbatical leave from the University of Birmingham and UCL, U.K., from 2016 to 2017. Her research interests include data mining, mobile and cloud computing, social media analytics, and user behavior analysis.

ALPTEKIN TEMIZEL received the Ph.D. degree from the Centre of Vision, Speech and Signal Processing, University of Surrey, in 2006. He was a Visiting Researcher at Microsoft MLDC-Lisbon in the summers of 2014 and 2015. He was on sabbatical leave from the University of Birmingham, U.K., from 2016 to 2017. He is currently a Professor with the Graduate School of Informatics, Middle East Technical University (METU). He is also the Principal Investigator with the GPU Education and Research Centre and a Deep Learning Institute Certified Instructor and a University Ambassador. He collaborated with and advised several companies on video analytics, video surveillance systems/algorithms, and GPU computing. His research interests include video surveillance, computer vision, machine learning, deep learning, and GPU computing

KLIMENT PRESHLENOV was born in Sofia, Bulgaria, in 1992. He received the B.S. degree in economics and finance from the University of Exeter, U.K., in 2014, and the M.S. degree in real estate economics and finance from the London School of Economics and Political Science, in 2015.

From 2017 to 2019, he was a Data Analytics Consultant and a Data Scientist at Gemseek. Since 2019, he has been the Head of Product at Ensk.ai. His research interests include performance evaluation in sports, marketing impact modeling, and estimating the impact of infrastructure on socio-economic outcomes.

DANIEL M. STRAHINOV was born in Sofia, Bulgaria, in 1989. He received the Bachelor of Economics and Social Studies degree from Cardiff University, in 2012.

From 2013 to 2015, he was a Management Consultant at PricewaterhouseCoopers, London, U.K. After that in 2016, he worked as the Director Data Analytics at GemSeek, and a market research company part of the British Future Thinking Group. In 2018, he became one of the Founders of the Pure Play Data Science Company called Amplify Analytix, whose focus was to develop, deploy and apply practical models in sales and marketing optimization. As part of this role, Daniel and his co-founders grew the company to a scale up stage with 35 employees. Since 2020, he has been leading EnskAI, a sports analytics company with a focus in applying data science to football recruitment. His research interests include mostly in creating practical, but scientifically backed analytical solutions which solve real-world problems with objective and quantifiable results.