Yugoslav Journal of Operations Research
xx (20xx), Number xx, xxx–xxx
DOI: https://doi.org/xxxx

Team-bounded efficiency scores of UEFA Champions League players

Luka Ivanović
University of Belgrade, Faculty of Organizational Sciences
luksivanovic19@gmail.com

Sandro Radovanović
University of Belgrade, Faculty of Organizational Sciences
sandro.radovanovic@fon.bg.ac.rs

Gordana Savić
University of Belgrade, Faculty of Organizational Sciences
gordana.savic@fon.bg.ac.rs

Boris Delibašić
University of Belgrade, Faculty of Organizational Sciences
boris.delibasic@fon.bg.ac.rs

Milena Popović
University of Belgrade, Faculty of Organizational Sciences
milena.popovic@fon.bg.ac.rs

Received: / Accepted:

Abstract: Measuring the efficiency of a football player is an interesting task. Every player invests time and energy to produce an outcome during the game that can lead to a better score for the team. However, football is a team sport, thus the synergy of the entire team is an *invisible* factor that influences the outcomes a player produces. In this paper, we introduce a novel efficiency estimation model based on the data envelopment analysis that incorporates the team effect in the efficiency score. To achieve this effect, we presented data envelopment analysis models using a single mathematical model. This allows us to define lower and upper bounds on the team efficiency score. More specifically, we introduce a novel type of assurance region constraint around the team performance such that the efficiency score cannot exceed γ times the average team efficiency score. The results of the proposed method on the UEFA Champions League 2021/22 season show that efficiency scores can be heavily affected by inefficient teammates. However, if the entire team performed well, the drop in efficiency scores is insignificant.

Keywords: Efficiency, DEA, Linear Programming, Football, Sports Analytics.

MSC: 90C05, 90C90.

1. INTRODUCTION

The sports industry has become one of the most interesting professions in recent years and one of the highest-paid. Because of this, a big number of sponsors have started to get involved with sports. Expected profit and big investments are more and more frequent. Consequently, results are put in the foreground [1]. The results can be achieved only with detailed analyzed reports and keeping track of statistical parameters of both players and clubs [2]. Exploring the advantages and disadvantages of players and investing in which direction a player can improve his skills helps every interested side in football to predict results better and to influence more on player's and team's efficiency and every other organization directly connected with football and sport. On the other hand, all these analyses are giving applicable statistics, which can improve even the economic football side [3].

It is noted in the literature how important is to collect, store, transform, and analyze data to see any improvement in the sport performance, or as in this paper football (or soccer). One approach to calculating individual athlete or team performance is data envelopment analysis (DEA). There are a lot of papers and research on this topic, both on the theoretical and practical aspects [4, 5, 6]. However, most of the analyses consider single-player performance independent of the team. In more technical terms, the process of transforming the inputs of a single player into outputs is the sole effect of a player's knowledge and skills. While the independence assumption is easier to implement, it can lead to false conclusions and consequently to false decisions and policies, especially if the assumption is that the purpose of football is to be a collective sport.

The paper's main goal is to improve the DEA mathematical model by introducing team performance into a process of calculation of efficiency scores. The novel DEA model called *Team DEA* introduces the influence of the team on every player's efficiency. This is done by extending the basic DEA model into a single mathematical model that solves n decision making-units simultaneously. Having a single linear programming mathematical model at hand, we introduce a novel type of assurance region constraint around the team performance such that the efficiency score cannot exceed γ times the average team efficiency score. It is worth noting that, in contrast to other papers on this topic, we focus solely on events on the field, thus players' efficiency is measured without the economic segment.

The dataset used contains football players who played at the UEFA Champions League in 2021/22, associated with their club, and teammates, players who play in the same club. The motivating idea is that teammates affect every individual, but the team's performance, as well. This method aims to show an *invisible* statistic, which is caused by excellent cohesion of the team and players focusing on the success of the team and improvement of the team's spirit.

The remained of the paper is structured as follows. In Section 2 we describe the literature review based on which we identified the research gap. The proposed Team DEA mathematical model is derived in Section 3, while Section 4 presents the results and discussion of results. Finally, we conclude the paper in Section 5.

2. LITERATURE REVIEW

DEA has been applied in many fields to assess the performance of decision-making units. One can find applications in both non-profit and profit institutions, assessment of employees within the companies, entire branches of a company, or companies between each other [7, 8, 9]. Similarly, one can evaluate efficiency in sports, where one can evaluate individual or team performance and compare them among other players or teams within the league they play in.

A comprehensive review of DEA in sports can be found in [10]. They summarize in which sports the DEA method was used and for what purposes. Therefore, one can find that DEA was used in football (soccer), basketball, baseball, cricket, cycling, golf, handball, and tennis. As can be observed, most of the sports listed above are team sports. Thus, it is very hard to find a function that will explain the performance of a DMU, although it is easy to measure both inputs and outcomes. In addition, most of the efficiency assessments are referred to individual players, which means that performances are considered independent of the performance of another player. Regarding football, the most studied league was English Premier League, mainly due to its tradition, worldwide reach, precise statistics, and due to the fact that some of the best players play in clubs that compete in this league. The reason why DEA is used is the interpretability of the results. Not only the decision-maker gets information about the efficiency of a player, but also a piece of additional information on how and where a player can and should improve.

In paper [11], one can find a specific application we used as a baseline for the experimental setup of the TeamDEA method. More specifically, the paper assesses the efficiency of teams that share the same defensive and attacking characteristics that are considered inputs to the DEA method. The output variable is the ratio of scored and conceded goals of the team. The analysis was performed on the Spanish La Liga for eight consecutive seasons. As for the DEA design choices, they performed the CCR window DEA model. The authors emphasized that better-ranked teams have a higher number of goals scored and that their game is more attacking, while teams struggling to stay in the league should emphasize the defense. The inputs they used served us as a guideline for the experimental part of our paper. More specifically, the number of shots on target, crosses, and passes represent attacking inputs, while the number of touches, clearances, headers, and interceptions represent defensive inputs.

One can find an application of a two-stage relational network DEA model [12] that assumes inputs and outputs are not directly connected but connected through intermediate concepts that mediate the outcome. Therefore, instead of modeling the outputs depending on the inputs, they state that the financial resources of the clubs are influencing two abstract concepts, namely social and sporting dimensions. Further, those concepts contribute to results, which are the business performance of a club. It turns out that in Spanish La Liga the most efficient distribution of their input resources is leaned toward the social dimension (Percentage of attendance at the stadiums and Number of followers on principal social networks), rather than toward the sporting dimension.

An interesting analysis can be found in [13] where the role of the manager is inspected as their influence on the performances of the team in the Italian Serie A league. As inputs they used age, a set of signals such that a manager is Italian, the manager has international experience, a manager is a former professional football player, had a manager managed a lower league club, and if a manager played for the national team, while the outputs are average points per match, sports performance, and financial performance. Their findings are suggesting that the biggest influence on efficiency has both sports and financial performance. In other words, average points per match and sport performance are very correlated.

We can observe that there is a gap in the literature where one can provide dependence of efficiency score of DMU that are somewhat dependent, i.e. two players playing for the same team. Therefore, the aim of this paper is to propose an extension of the DEA method such that the efficiency score of an individual is regulated by the performance of the team. In other words, the efficiency score of an individual should be upper-bounded by the performance of the worst-performing player within the team.

3. METHODOLOGY

The methodology section briefly explains the DEA method and provides the process of derivation of the proposed Team DEA method. After the explanation of the DEA method and the proposed Team DEA method, we provide a description of the UEFA Champions League dataset.

3.1. Data Envelopment Analysis

DEA is a method originating from operational research that calculates the relative efficiency of decision-making units (DMU). Each DMU is characterized by inputs and outputs, each associated with its weight. The task of the DEA method is to find such weights so that DMU achieves the best possible efficiency. [14]

The term decision-making unit is a flexible one. Over the past years, as evaluated by the DEA method itself, the efficiency of different decision-making units has been studied [15, 16, 9, 17]. In the beginning, the application of the DEA method was aimed at non-profit organizations, such as schools, universities, the army, cities, and hospitals, but later its application was significantly expanded and recently it has been used in profit-oriented companies, but in the sports field, as well [18, 19].

In general, DEA is a non-parametric method that estimates the efficiency frontier by considering the best values of DMUs. A non-parametric method means that no assumptions are made about the dataset's characteristics and that the parameters' number and nature are flexible rather than fixed in advance. Inefficient units can become efficient by increasing the volume of output while maintaining the same level of input or by maintaining the same level of output but simultaneously reducing input resources [15]. There is also a combination of the previous two scenarios, in which case an inefficient unit can become efficient [15]. The mentioned method is very suitable for calculating the efficiency of athletes, because by comparing input and output resources between athletes, a relatively efficient unit is obtained, which potentially represents a model for others [20]. Later, based on the model, one can see the segments of the game where the improvement and advancement of the athletes are possible, and this is achieved through the variables that are generated in the dual model.

There are many types of DEA models. The one used in this paper is CCR (Charnes, Cooper, and Rhodes), named after the scientists who first constructed the method. The mentioned model is based on *constant return-to-scale* [20]. Constant returns to scale represent a situation when a change in input resources causes a proportional change in output resources. It is also worth noting that the DEA model we implemented is an *output oriented* DEA model. More specifically, the model aims at generating the largest amount of output resources, given a level of input resources.

For each of the decision-making units, the mathematical model shown by expression (1) is solved, to obtain its efficiency [14, 15]. The decision unit is relatively efficient only if the solution of the objective function (that is, the efficiency index f_k) is equal to 1. Otherwise, the unit is considered inefficient and its inputs or outputs must be altered, depending on the orientation of the model. In the output-orient CCR model, each inefficient decision unit must increase its outputs while keeping the input resources at the same level to become relatively efficient. In addition, it is important to note that the smallest value that can be obtained in the objective function of the output-oriented CCR model is one and that any higher value shows that the decision-making unit is relatively inefficient. The reciprocal value of the outcome of the objective function is used to calculate the efficiency coefficient for the decision unit [14].

The CCR model's dimensions are equal to the sum of its variables $(m+s)$, that is, the number of control variables is equal to the sum of the number of inputs and outputs. Using the formula (1) one CCR model has a single constraint regarding the weighted sum of outputs, n constraints regarding the difference between the weighted sum of outputs and the weighted sum of inputs, and additional $r+i$ constraints. The last constraint represents a hyper-parameter of the DEA model as ϵ is provided before the optimization.

$$
\begin{aligned}
& min f_k = \sum_{i=1}^{m} v_i x_{ik} \\
& s.t. \\
& \sum_{r=1}^{s} u_r y_{rk} = 1 \\
& \sum_{r=1}^{s} u_r y_{rp} - \sum_{i=1}^{m} v_i x_{ip} \leq 0, \forall p = 1, ..., n \\
& u_r, v_i \geq \epsilon, \forall r = 1, ..., s, \forall i = 1, ..., m \\
& \epsilon \geq 0
\end{aligned}
\tag{1}
$$

3.2. Team-based Assurance Region Data Envelopment Analysis

As colloquially emphasized, football is a team sport. A single individual, no matter how good it is, cannot outperform the synergistic effect of multiple individuals (i.e., team). This paper aims at addressing the issue of independence assumption, which is present in many DEA models. This is done by providing an assurance region around every DMU, such that a single player cannot be much better or much worse than the average efficiency score of a team that the player plays for.

Team-bounded assurance region is hard to implement in the CCR model. The simple solution would be to create an iterative procedure where one would solve the CCR model to obtain efficiency scores, followed by another CCR model with constraints regarding the efficiency score. This would result in solving $n + n$ mathematical models. First, n models obtain unbounded efficiency scores and the latter n models with an additional team-based constraint on the efficiency score. One should be aware that between the two parts, one needs to calculate average efficiency scores and introduce them into the mathematical model.

The solution that we adopted aims at mitigating the two-step procedure as it introduces additional complexity in the process of obtaining efficiency scores. More specifically, instead of solving $n+n$ mathematical models, we aim at creating and solving a single mathematical model. Since every DMU is represented in the goal function and in constraints as independent of each other, one can merge n mathematical models into a single one with guarantees that the optimal solution of a single mathematical model is the same as the optimal solution of n independent ones [21]. The benefit of using a single mathematical model is the ability to construct more complex constraints than the ones available in the CCR model. One can limit the input and output weights according to the other, joint criterion, as we did with the team of the player. The Team DEA mathematical model is presented in the formula (2).

$$
\begin{aligned}
&minf = \sum_{p=1}^{n}\sum_{i=1}^{m} v_{ip}x_{ip}\\
&s.t.\\
&\sum_{r=1}^{s} u_{rp}y_{rp} = 1, \forall p = 1, ..., n\\
&\sum_{r=1}^{s} u_{rp}y_{rk} - \sum_{i=1}^{m} v_{ip}x_{ik} \leq 0, \forall p = 1, ..., n, \forall k = 1, ..., n\\
&\sum_{i=1}^{m} v_{ip}x_{ip} \geq \frac{\gamma}{|1_{team(k)=team(p)}|}\sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}x_{ik}1_{team(k)=team(p)}, \forall p = 1, ..., n\\
&\sum_{i=1}^{m} v_{ip}x_{ip} \leq \frac{1}{\gamma|1_{team(k)=team(p)}|}\sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}x_{ik}1_{team(k)=team(p)}, \forall p = 1, ..., n\\
&u_{rp}, v_{ip} \geq \epsilon, \forall rp, r = 1, ..., s, p = 1, ..., n, \forall ip, i = 1, ..., m, p = 1, ..., n\\
&\epsilon \geq 0
\end{aligned}
\tag{2}
$$

The goal function represents the sum (over players) of weighted sums of inputs. As a consequence of this adjustment of the CCR model, the goal function no longer represents the efficiency score, but it represents the sum of efficiency scores. However, by simple calculation, one can obtain each individual efficiency score.

The first constraint is the same as in the CCR model. It represents the equality constraint where the weighted sum of outputs should be equal to one. This constraint transforms the non-linear nature of efficiency into a linear formulation that can be solved efficiently using linear programming solvers such as the Simplex, or Interior points method. The difference compared to the CCR model is that the proposed Team DEA model has p of these constraints, one for each player in the dataset.

The second constraint represents the efficiency constraint. More specifically, one must ensure that not a single player is over-efficient (having an efficiency score over one). Thus, for weights of a player p, a player k should have at most efficiency score equal to 1. Although this type of constraint is the same as in the CCR model, the proposed model it is more complex. The reason for this is the fact that a single mathematical model requires to have n^2 of these constraints since both p and k represent players.

A lower bound of a team assurance region is provided as the third constraint in equation (2). It states that the weighted sum of inputs should be at a most average of the weighted sum of inputs of the players in the same team. To calculate the average of the team's weighted sum we introduce the function $team(p)$ and the indicator function 1. More specifically, the indicator function ($1_{team(k)=team(p)}$) provides us the information on whether players play in the same team or not. The width of the assurance region is controlled with hyper-parameter γ, which can take values between zero and one. Similarly, we provide an upper bound of a team assurance region as the fourth constraint in equation (2).

The proposed model includes the dependency between DMUs into a single mathematical model, which means that the model should be defined and solved only once. In addition to lowering the number of mathematical models, one can introduce more complex constraints, as we did with team-based assurance regions. However, the downside of the proposed model is the number of variables and constraints. The number of variables increases to $(s+m)n$ compared to $(s+m)$ in the CCR model. Although the optimal solution is the same, some solvers like Simplex would take significantly longer time to solve the Team DEA model, as they construct and iterates over the vertices (extreme points) of a given polytope [22]. Therefore, in the implementation of the proposed method we utilized the interior point method as a solving procedure as a more efficient linear programming solver. More specifically, instead of iterating over the vertices of the polytope, the optimization procedure iterates towards the lowest cost vertex without regard for vertices [23]. Another point worth noting is that the proposed model has $(3n + n^2 + (s+m)n)$ constraints, while a single CCR model has $(1 + n + s + m)$ constraints. This increase in the number of constraints can cause an issue during the efficiency calculation as the constraint matrix can be extremely large and not fit the memory of the computer. However, this issue is avoided by using sparse matrix representation techniques and using solvers specialized for computation with such matrices [24, 25].

3.3. Data Description

The dataset consists of football players in the 2021-22 season. Players' goals, pass distribution, attacking, defending, goalkeeping, and other statistics were tracked in separate data files. All these datasets were combined into a single dataset that had all attributes and observations. Players that didn't have values for certain attributes in the primary dataset have their missing values replaced with 0 in the dataset. The replacement of missing values with zero is justified as those players haven't had a pass, goal, or assist.

The processed dataset contains 747 football players playing for 32 clubs and 43 attributes, of which 40 are quantitative, while three attributes are qualitative. More specifically, the football player's name, the team he plays for, and the position he plays on the field are categorical data.

The first filter criterion was the selection of outfield players. This is done because goalkeepers have different behaviour in terms of inputs and outputs. Their task is solely to stop the opposing player from scoring. They seldom score goals and have passes and interceptions. Thus, they are incomparable to the remainder of the dataset. Since there are 54 goalkeepers, the dataset was reduced to 693 players.

Also, the dataset contains 693 players. However, many of them have played only a couple of minutes or just one or two games. These observations are excluded

The dataset is available at the following link.

from the final dataset because they are outliers and can distort the efficiency score. The filter criteria for players to stay in the dataset are that they played three or more games and that they played more than 200 minutes. After implementing this criterion, the final dataset contains 441 observations.

The DEA model we find the most suitable for the data at hand is the CCR model. The CCR model is output-oriented, thus in order for the players to be relatively efficient, the values inputs must be low for fixed outputs.

After observing the literature [26, 27, 10], we selected the following outputs:

- Goals ($O1$) - A number of goals scored by the player. A crucial thing in football is to score every game in as many as possible. It is most often associated with strikers, who are often the best scorers on the team.
- Assists ($O2$) - A number of passes that led to a goal. An essential item of any good midfielder in football and often their primary task is to facilitate and enable the striker to score.
- Balls recovered ($O3$) - A number of recovered balls. An event in which a player gains possession after control of the ball has been lost by the opposition. In most cases, this is the primary task of defensive players.

Since there are 40 attributes, we performed attribute selection as a part of data preprocessing. First, we excluded attributes that were uninformative due to low variance, as well as those that were carrying the same information as some other attribute (i.e., goals scored with the left foot), or being a consequence of an event of interest. After the initial attribute selection, the dataset consists of 25 attributes.

The selection of the input attributes is done using a supervised filter attribute selection approach [28]. More specifically, we aimed at filtering the 10 most important attributes regarding each output attribute, where the importance score was done using the F statistic. Table (1) shows the essential input variables for each of the output attributes. A value *Yes* in the table signals if an attribute is in the top 10 most important ones, while an empty value represents that an attribute is not within the top 10 most important ones.

Table 1: A signal if an attribute is in the top 10 most important ones

Attribute	Goals	Assists	Balls Recovered
Minutes Played	Yes		Yes
Match Played	Yes	Yes	
Distance Covered			Yes
Pass Accuracy			Yes
Pass Attempted			Yes
Pass Completed			Yes
Cross Accuracy			
Cross Attempted		Yes	
Cross Completed		Yes	
Free kicks Taken			Yes
Fouls Committed			
Fouls Suffered	Yes		
Red Cards			
Yellow Cards			
Tackles			Yes
Tackles Won			Yes
Tackles Lost			Yes
Clearance Attempted			Yes
Total Shot Attempts	Yes	Yes	
Shots on Target	Yes	Yes	
Shots off Target	Yes	Yes	
Shot Blocked	Yes	Yes	
Corners Taken		Yes	
Offsides	Yes	Yes	
Dribbles	Yes	Yes	

Four attributes are not significant for any of the output attributes; thus, they are excluded from further analysis. These attributes are *Cross Accuracy*, *Fouls Commited*, *Red Cards*, and *Yellow Cards*. On the other hand, there are nine attributes that are crucial for two of the three target attributes and 12 attributes that are crucial for one of the three outcomes. As a result, all the attributes that have an impact on two of the three outputs as well as a few other attributes that the authors selected based on the domain expertise and the literature constitute potential inputs. Attributes that are chosen because of domain knowledge are *Pass Attempted*, *Pass Completed*, *Tackles*, *Tackles Won*, and *Clearance Attempted*. The number of potential inputs is reduced from 25 to only 14.

To further reduce the number of attributes, we calculated the correlation coefficient between potential input attributes and output attributes. Attributes with absolute values higher than 0.5 were chosen for at least one targeted attribute after a thorough analysis of the correlation matrix of correlation between inputs and outputs, as well as between inputs themselves as we try to mitigate input multi-colinearity. For instance, there was a significant correlation between the attributes *Total Attempts*, that is the number of total shots, and the *Shots on Target*, the number of total shots in the goal. These features have a causal connection, and one is a subset of the other. In such cases, we select attributes with greater

Table 2: Correlation between input and output attributes

Attribute	$I1$	$I2$	$I3$	$I4$	$I5$	$I6$	$O1$	$O2$	$O3$
$I1$	1.00	0.34	0.34	0.73	0.35	0.43	0.30	0.33	0.62
$I2$		1.00	0.49	0.03	-0.21	-0.25	0.87	0.39	-0.25
$I3$			1.00	0.10	-0.09	-0.17	0.38	0.64	-0.08
$I4$				1.00	0.38	0.36	0.00	0.17	0.72
$I5$					1.00	0.33	-0.18	-0.06	0.55
$I6$						1.00	-0.20	-0.13	0.67

interpretability for decision-making.

Based on the above-mentioned analysis, we choose the following set of input attributes:

- Minutes Played ($I1$) - The number of minutes each player has played during the season.
- Shots on Target ($I2$) - The total number of shots in a goal that each player has had.
- Dribbles ($I3$) - Total number of dribbling that the player has made.
- Pass Completed ($I4$) - Total number of accurate passes that the player has had to his teammates.
- Tackles Won ($I5$) - Total number of successful tackles against opponents by the player.
- Clearance Attempted ($I6$) - The total number of defensive players' attempts to prevent a goal from being scored in the last moment.

The final dataset also includes two additional attributes—the player's club and position—as well as the player's name and last name, which serve as each player's index values. Table (2) shows the correlation between each input and output.

3.4. Experimental Setup

The output CCR model and the proposed TeamDEA model are developed in Python programming language using scipy [29] adjusted for sparse matrix calculations [25]. A total of 441 players are examined, each having six inputs and three outputs. The *Club* attribute is used during the TeamDEA model optimization.

The main goal of the paper is to inspect what players are efficient and what characterizes them as efficient players. For inefficient players, we would like to provide an answer what are their downsides and how can they improve. However, we would like to provide an answer on how teammates influence the efficiency score. For this part, we must employ the TeamDEA method and inspect at what value of parameter γ efficiency score starts to drop.

For the CCR model, we solve 441 CCR DEA models, one for each player in the dataset. Each model consists of 451 constraints (one for the sum of virtual outputs, 441 for each player in the dataset regarding efficiency scores, 3 for outputs, and 6 for inputs). As a result, we get efficiency scores, slacks, and dual variables that are analyzed.

For the TeamDEA model, the objective function minimizes all the values of the input variables (due to the output orientation of the DEA model). There are 3,969 variables in the objective function vector (for each player six input values and three output values). Input values are filled with values from the dataset, whereas output resource values are set to 0. In addition, the TeamDEA method has 199,773 constraints. There are two sets of constraints. The first set of constraints requires that the virtual output sum is equal to one. As a result, the first constraint matrix is constructed, with each row containing the values of the output variable, of the specific player. The supplied values are placed next to the virtual output's unknown variables, which are three for each player. All other values are 0. Because there is a value of one on the other side of the equation, the vector of the first constraint is also created. There are 441 ones in the vector, one for each player. The second constraint is that the difference between virtual output and virtual input must be less than zero. A linear mathematical model with 441 (number of players) inequalities is solved for each player. This happens because each player is compared to other players and thus their relative efficiency is evaluated. In the case of a single-player inequality, input and output values are taken from the dataset and placed next to the virtual input and virtual output variables. The above is done for 441 players, which means there are 194,481 rows of the second constraint matrix. The second constraint's vector is the upper bound of the inequality (all are zeros in the CCR model), and the number of rows is the same as for the matrix. The remainder of the constraints is related to the positive values of input and output weights. We inspect and discuss the results based on the value of γ. The value of γ varies from zero to one with an increment of 0.1.

4. RESULTS AND DISCUSSION

This section consists of two parts. The first part provides an analysis of the CCR DEA model, while the second part provides an analysis of the proposed TeamDEA model.

4.1. Results of the CCR DEA model

After running the CCR DEA model, there are 94 efficient players in total. The rest of them, 347 players, need to improve their input resources to become efficient. The most relatively inefficient player is Depay, from Barcelona, who has an efficiency index with a value of 0.096. Also, Evanilson (Porto), Munir (Sevilla), and Marlos (Shakhtar Donetsk) have an index of efficiency less than 0.25. Therefore, one can consider them the most inefficient players.

With six of them, Liverpool is the team with the most efficient players. Atalanta, Benfica, Chelsea, Dortmund, Salzburg, Sporting SP, and Villarreal have five players who generate efficient output based on their input. There is just one efficient player on teams like Club Brugge, Inter, Malmo, PSG, and Young Boys.

Teams such as Dynamo Kyiv, LOSC, and Shakhtar Donetsk don't have an efficient player in the squad at all. Football club Atletico has the highest number of inefficient players, 16 of them. After Atletico, LOSC has 15, and Manchester City has 14 players that are inefficient. The situation is as follows when we look at the percentage of which teams have the most and which teams have the least efficient players: the percentage of efficient players for Atalanta, Benfica, Dortmund, and Salzburg is greater than 0.35 and these are the best-ranked clubs. However, the teams with the lowest percentage of efficient players include Inter, Malmo, PSG, and Young Boys, all of which have percentages of less than 0.1.

The efficiency of input/output resources is displayed in the coefficients of virtual variables, which highlight the attributes that the DEA approach focused on to provide each player with their best shot. Below is the analysis of the three selected interesting players. These players are selected because they are considered one of the best football players in the World, and one of them is inefficient according to the DEA method.

For instance, the Real Madrid player Luka Modrić, who is inefficient with an efficiency score of 0.594, is better in assists and balls recovered, but poor in goals when compared to other players. This is because he had 39 balls stolen from opponents and four assists during the competition, but no goals. For this football player, the DEA model concentrated on the number of dribbles and minutes played as input attributes.

Timo Werner, a Chelsea player who was efficient, did a great job in both goals scored and assists made for output variables. With only five shots on goal but four goals in the tournament, he was quite good in this input attribute, and DEA gave it a virtual coefficient of 1 for it. Because he has fewer shots on target than other players, but a higher output as a result, all his input is thus concentrated on this. In the meantime, the output is presented sequentially with coefficients of 0.8, 0.2, and 0 for goals, assists, and balls recovered. Since he had two assists, but only 8 balls recovered, the basic DEA model gave it a coefficient of zero for the last output attribute.

Lewandowski (Bayern Munich), on the other hand, was the second-best scorer in the tournament with 13 goals, thus the DEA model only considered his output in terms of goals. While his minutes played and shots on target were only significant for input resources.

An overview of virtual coefficients for inputs and outputs of the above-mentioned players is presented in Table 3.

Table 3: Virtual coefficients for inputs and outputs for Luka Modrić, Timo Werner, and Robert Lewandowski

Player Name	Luka Modrić (Real Madrid)	Timo Werner (Chelsea)	Robert Lewandowski (Bayern Munich)
$I1$	0.950	0.000	0.125
$I2$	0.051	1.000	0.871
$I3$	0.532	0.000	0.004
$I4$	0.000	0.000	0.000
$I5$	0.071	0.000	0.000
$I6$	0.079	0.000	0.000
$O1$	0.000	0.800	1.000
$O2$	0.527	0.200	0.000
$O3$	0.473	0.000	0.000
Eff.	0.594	1.000	1.000

Players that are inefficient have their role models. The values of the output attributes would be equal to or identical to those of their role models if the inefficient players maintained their inputs while raising their outputs and went to the efficiency frontier. This analysis is crucial because it indicates which players they should imitate and how they should allocate their resources to produce efficient results. The authors see that a player could have multiple role models. Players who are role models to others can be found in the dual model or in the second constraint of the primal model (the difference between the virtual output and the virtual input must be less than zero). When comparing the inefficient player to the best role model player, λ value in the dual model is maximized. While assessing the player's efficiency in the second constraint, the value of the slack variable is the least.

The most frequent role models are Dahoud, Djimsiti, Fernando, Solet, Uribe, and Werner. They have all performed as role models more than 75 times, and as may be assumed, they are all relatively efficient.

4.2. TeamDEA Results

The analysis of the TeamDEA method depends heavily on the γ hyper-parameter. More specifically, its value directly influences whether players on the same team are exposed to looser or tighter restrictions, allowing for varying player efficiencies to be reached. It was established that the influence of the team on the individual begins with a value of 0.3 and that the influence of the modified DEA model stops at the value of 0.8. This was concluded after the model of the modified DEA method was run repeatedly with gamma values ranging from 0.1 to 1 with a step of 0.1. Only one player's efficiency changes for a value of 0.3, namely Pique (Barcelona) is no longer efficient but has an efficiency score of 0.944.

As the value of γ increases, the constraints imposed in the TeamDEA mathematical model become increasingly tight and stricter for evaluating player efficiency. Therefore, there are only 20 efficient players in the dataset when γ equals 0.7, and no efficient player with a γ value of 0.8. The efficiency indices of players at the same club with γ values of 1 are the same.

Players Modrić, Werner, and Lewandowski are examined for the TeamDEA model as well. Modrić was inefficient in the basic DEA model and remained so until the γ value of 0.8, at which point he was inefficient with an index of 0.589 and the most inefficient with a γ value of 1 (index was 0.43). Lewandowski has been able to maintain efficiency up to 0.8 γ, while Werner's efficiency index has decreased to a γ value of 0.7. All values of players' indices can be seen in Table 4.

Table 4: Efficiency scores for Luka Modrić, Timo Werner, and Robert Lewandowski for different values of γ

Player Name	Luka Modrić	Timo Werner	Robert Lewandowski
	(Real Madrid)	(Chelsea)	(Bayern Munich)
$\gamma = 0.3$	0.594	1.000	1.000
$\gamma = 0.4$	0.594	1.000	1.000
$\gamma = 0.5$	0.594	1.000	1.000
$\gamma = 0.6$	0.594	1.000	1.000
$\gamma = 0.7$	0.594	0.793	1.000
$\gamma = 0.8$	0.589	0.607	0.880
$\gamma = 0.9$	0.531	0.480	0.695
$\gamma = 1.0$	0.430	0.389	0.563

Due to the restrictive nature of the hyper-parameter γ and the ease of interpretation, we examine results for the γ value of 0.7. There are few efficient players, but even under these restrictions, each efficient player produces exceptional output based on input.

Therefore, there are 20 efficient players in the TeamDEA model. Depay, Gavi, and F. De Jong, all from Barcelona, are the least efficient. They all have the same, extremely low-efficiency index of 0.096. Only Depay had this index in the basic DEA model, but now his teammates are included in this inefficient group of players. Additionally, there were only 4 inefficient players with index values below 0.25 in the CCR DEA model. Now, there are 27 of them in the TeamDEA one.

There are slight changes in the efficiency scores within the teams. There are only six clubs with players who have an efficiency score of one. Namely, Ajax and Bayern with two, Juventus and Wolfsburg with three efficient players, and Dortmund and Salzburg with six players. In addition, with 19 players, Liverpool became the team with the least number of efficient players. Atletico, with 18 players, is the next club after Liverpool, but based on the CCR DEA model, it was also a very inefficient team. With a value of 0.5, Dortmund has the highest percentage of efficient players.

For the analysis of virtual coefficients, we analyze the same three players. More specifically, Modrić, Werner, and Lewandowski. The distribution of resources has changed slightly because of the new constraints, and they now concentrate on different input and output variables. Values are presented in Table 5. The red-colored values in the table represent a decrease in the value compared with the CCR DEA model, while the green color represents an increase in value.

Table 5: Virtual coefficients for inputs and outputs of TeamDEA for Luka Modrić, Timo Werner, and Robert Lewandowski

Player Name	Luka Modrić	Timo Werner	Robert Lewandowski
	(Real Madrid)	(Chelsea)	(Bayern Munich)
$I1$	0.950	0.000	0.997
$I2$	0.051	0.000	0.871
$I3$	0.532	0.162	0.003
$I4$	0.000	1.098	0.000
$I5$	0.071	0.000	0.000
$I6$	0.079	0.000	0.000
$O1$	0.000	0.000	1.000
$O2$	0.527	1.000	0.000
$O3$	0.473	0.000	0.000
Eff.	0.594	0.793	1.000

The input and output values of Modrić are identical to those in the basic DEA model. The main reason for the consistency of the efficiency score over the γ values is the efficiency scores of other players within Real Madrid. Therefore, additional constraints did not influence the virtual inputs and virtual outputs of Luka Modrić. Team efficiency scores per value of γ can be inspected in Figure 1.

Werner, on the other side, can't perform any more efficiently and had an index score of 0.793. So, in relation to the basic DEA model, his efficiency went down by just over 20%. The TeamDEA approach concentrates on balls recovered in the output resource while putting this player's dribbles and pass-completed attributes in the input resource as the most significant. One can observe that he had an efficiency score equal to one in the CCR model, with a virtual coefficient of 1 for the total number of shots on target, and a high value of an output virtual coefficient for the total number of goals scored.

Lewandowski has continued to be efficient, although his resource allocation has changed significantly. He no longer stands out in his team in shots on target, obviously since that input attribute has decreased substantially in Bayern with new restrictions. In this situation and with this γ, TeamDEA adopts his resource allocation, and he has the highest virtual coefficient for input variable minutes played. He continues to be the best at scoring goals, and his virtual coefficient for this output is 1.

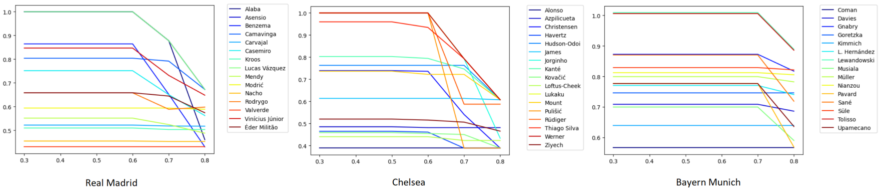

Figure 1: Efficiency scores of players of Real Madrid (left), Chelsea (middle), and Bayen Munich (right) for different values of γ

The most frequent role models mostly remained the same. There were six of these in the CCR DEA model, but in TeamDEA there are four of them. Although Werner and Solet are no longer included in this group, Dahoud, Djimistri, Fernando, and Uribe continue to be role models more than 75 times. There are also many inefficient athletes that serve as role models for others. In terms of terminology, dominating decision-making units are inefficient decision-making units that serve as examples of inefficient players. In this dataset with γ value of 0.7, dominating role models are Djimistri, Fernando, and Uribe.

Analysis of active constraints identifies the players who are blocked to be more efficient by a teammate. As an illustration, three players on the Real Madrid squad have active restrictions, that are Alaba, Asensio, and Lucas Vazquez. Their efficiency coefficients are 0.878. This indicates that they had the capacity to be more efficient, but one of the players blocked and limited them. After the inspection of the constraints, one can observe that Valverde is the team's least efficient player and thus limits others from being more efficient than they currently are.

An additional level of analysis is per position analysis. We select the top 15 players per position and observe their efficiency scores given the γ parameter equal to 0.8. One can observe efficiency scores in Figure 2.

One can observe that many defenders have dropped in efficiency score after $\gamma = 0.7$. At that point, five players are efficient. Those are, namely, Solet (Salzburg), Meunier (Dortmund), De Sciglio, Alex Sandro (Juventus), and Fernando Costanza (Sheriff). However, for most of them, efficiency hinders them as they play in a team with highly inefficient players. More specifically, as we increase the demand that a player is efficient if the team is efficient, their efficiency scores drop. One can also note players who are consistent regardless of the γ like Akanji and Süle, who have efficiency scores over 0.8. This is a strong indicator that their performance influences a team to play better.

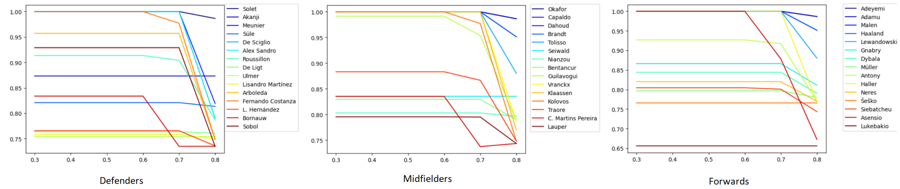

Figure 2: Efficiency scores of top 15 defenders (left), midfielders (middle), and forwards (right) for different values of γ

A similar conclusion can be drawn for midfielders. Several players are efficient until $\gamma = 0.7$, but their efficiency score drops at $\gamma = 0.8$. It is worth emphasizing Okafor and Capaldo (Salzburg), as well as Dahoud and Brandt (Dortmund) as their efficiency scores drop slightly at $\gamma = 0.8$. Salzburg indeed had a very good UEFA Champions League season and has reached the round of 16 with dominant performances in a group stage. Dortmund failed to advance to the group stage of the competition. However, these players were one of the most important ones in games where Dortmund won. Finally, forwards from Salzburg, Adeyemi, and Adamu were efficient or close to efficient regardless of γ. Other forwards, such as Asensio from Real Madrid suffered from having inefficient players in their team.

5. CONCLUSION

The major objective of the paper is to calculate the team's effect on an individual and apply this invisible influence to every player on the same team. For many players, the efficiency coefficient decreased because of their teammates' inefficient performance, which contributed to the team's overall lower efficiency. The newly proposed data envelopment analysis mathematical model TeamDEA added upper and lower assurance regions that limit player efficiency regarding the performance of other players within the team. The proposed assurance region successfully captured team efficiency and cohesion, as an essential component of football. After using the basic DEA approach, 94 players were efficient. However, when new constraints were added and the hyper-parameter $\gamma = 0.7$, only 20 players remained efficient. As a result, it is possible to see how the TeamDEA model affects players, specifically how inefficient teammates affect the rest of the team.

Further, by analysis of active constraints, it is found that certain players made their teammates inefficient. In other words, some players couldn't make the most use of the input resources due to their teammates. This type of analysis is crucial because it makes it possible to identify the *culprits* for the team's slight decline and investigate the causes and potential solutions further.

Hyper-parameter γ, which directly impacted how restrictive the assurance regions are, had a significant influence in determining how efficient the player is. By decreasing γ, the player's influence decreased, allowing him to better utilize input resources and thus increase the efficiency score. The team's effect can be observed at a minimum γ value of 0.3. Below this threshold, very inefficient teammates were no longer able to participate in evaluating the efficiency of other players and reducing their indices. By doing the contrary, i.e., increasing the value of γ, the team's influence grew, and the players' efficiency decreased.

This paper provides a better introduction to the weak and inefficient individuals of each team. These inefficient players would contribute to improving the general efficiency of the team by reducing their input resources, which is one of the necessary conditions for good results for the team.

References

[1] V. De Bosscher, S. Shibli, and A. C. Weber, "Is prioritisation of funding in elite sport effective? an analysis of the investment strategies in 16 countries," *European Sport Management Quarterly*, vol. 19, no. 2, pp. 221–243, 2019.

[2] M. Du and X. Yuan, "A survey of competitive sports data visualization and visual analysis," *Journal of Visualization*, vol. 24, pp. 47–67, 2021.

[3] I. Guzmán-Raja and M. Guzmán-Raja, "Measuring the efficiency of football clubs using data envelopment analysis: Empirical evidence from spanish professional football," *SAGE Open*, vol. 11, no. 1, p. 2158244021989257, 2021.

[4] G. Rossi, D. Goossens, G. L. Di Tanna, and F. Addesa, "Football team performance efficiency and effectiveness in a corruptive context: the calciopoli case," *European Sport Management Quarterly*, vol. 19, no. 5, pp. 583–604, 2019.

[5] M. Terrien and W. Andreff, "Organisational efficiency of national football leagues in europe," *European Sport Management Quarterly*, vol. 20, no. 2, pp. 205–224, 2020.

[6] M. Espitia-Escuer and L. I. Garcia-Cebrian, "Efficiency of football teams from an organisation management perspective," *Managerial and Decision Economics*, vol. 41, no. 3, pp. 321–338, 2020.

[7] S. Kaffash, R. Azizi, Y. Huang, and J. Zhu, "A survey of data envelopment analysis applications in the insurance industry 1993–2018," *European journal of operational research*, vol. 284, no. 3, pp. 801–813, 2020.

[8] P. Peykani, E. Mohammadi, R. F. Saen, S. J. Sadjadi, and M. Rostamy-Malkhalifeh, "Data envelopment analysis and robust optimization: A review," *Expert systems*, vol. 37, no. 4, p. e12534, 2020.

[9] V. Cvetkoska and G. Savic, "Dea in banking: Analysis and visualization of bibliometric data," *Data Envelopment Analysis Journal*, 2021.

[10] Z. U. H. Bhat, D. Sultana, and Q. F. Dar, "A comprehensive review of data envelopment analysis (dea). approach in sports," *Journal of Sports Economics & Management*, vol. 9, no. 2, pp. 82–109, 2019.

[11] R. Sala-Garrido, V. L. Carrión, A. M. Esteve, and J. E. Boscá, "Analysis and evolution of efficiency in the spanish soccer league (2000/01-2007/08)," *Journal of Quantitative Analysis in Sports*, vol. 5, no. 1, 2009.

[12] A. Pérez-González, P. de Carlos, and E. Alén, "An analysis of the efficiency of football clubs in the spanish first division through a two-stage relational network dea model: a simulation study," *Operational Research*, vol. 22, no. 3, pp. 3089–3112, 2022.

[13] L. Buzzacchi, F. Caviggioli, F. L. Milone, and D. Scotti, "Impact and efficiency ranking of football managers in the italian serie a: Sport and financial performance," *Journal of Sports Economics*, vol. 22, no. 7, pp. 744–776, 2021.

[14] W. W. Cooper, L. M. Seiford, and K. Tone, *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*. Springer, 2007, vol. 2.

[15] W. W. Cooper, L. M. Seiford, and J. Zhu, "Handbook on data envelopment analysis," 2011.

[16] M. Radojicic, V. Jeremic, and G. Savic, "Going beyond health efficiency: what really matters?" *The International journal of health planning and management*, vol. 35, no. 1, pp. 318–338, 2020.

[17] S. Radovanović, G. Savić, B. Delibašić, and M. Suknović, "Fairdea—removing disparate impact from efficiency scores," *European Journal of Operational Research*, vol. 301, no. 3, pp. 1088–1098, 2022.

[18] S. Radovanović, "Two-phased dea-mla approach for predicting efficiency of nba players," *Yugoslav Journal of Operations Research*, vol. 24, no. 3, 2016.

[19] A. de Cássio Rodrigues, C. A. Gonçalves, and T. S. Gontijo, "A two-stage dea model to evaluate the efficiency of countries at the rio 2016 olympic games," *Economics Bulletin*, 2019.

[20] Y.-b. Ji and C. Lee, "Data envelopment analysis," *The Stata Journal*, vol. 10, no. 2, pp. 267–280, 2010.

[21] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[22] S. Smale, "On the average number of steps of the simplex method of linear programming," *Mathematical programming*, vol. 27, no. 3, pp. 241–262, 1983.

[23] I. J. Lustig, R. E. Marsten, and D. F. Shanno, "Interior point methods for linear programming: Computational state of the art," *ORSA Journal on Computing*, vol. 6, no. 1, pp. 1–14, 1994.

[24] R. Hassani, A. Fazely, P. Luksch *et al.*, "Analysis of sparse matrix-vector multiplication using iterative method in cuda," in *2013 IEEE Eighth International Conference on Networking, Architecture and Storage*. IEEE, 2013, pp. 262–266.

[25] M. Aslam, O. Riaz, S. Mumtaz, and A. D. Asif, "Performance comparison of gpu-based jacobi solvers using cuda provided synchronization methods," *IEEE Access*, vol. 8, pp.

31 792–31 812, 2020.

[26] T. Tiedemann, T. Francksen, and U. Latacz-Lohmann, “Assessing the performance of german bundesliga football players: a non-parametric metafrontier approach,” *Central European Journal of Operations Research*, vol. 19, pp. 571–587, 2011.

[27] D. Santín, “Measuring the technical efficiency of football legends: who were real madrid’s all-time most efficient players?” *International Transactions in Operational Research*, vol. 21, no. 3, pp. 439–452, 2014.

[28] Z.-H. Zhou, *Machine learning.* Springer Nature, 2021.

[29] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, “Scipy 1.0: fundamental algorithms for scientific computing in python,” *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.