applied sciences

MDPI

Article

Aggregated Rankings of Top Leagues' Football Teams: Application and Comparison of Different Ranking Methods

László Gyarmati [1], Éva Orbán-Mihálykó [1,*], Csaba Mihálykó [1] and Ágnes Vathy-Fogarassy [2,*]

1 Department of Mathematics, University of Pannonia, Egyetem u. 10, 8200 Veszprém, Hungary; gyarmati.laszlo@phd.mik.uni-pannon.hu (L.G.); mihalyko.csaba@mik.uni-pannon.hu (C.M.)
2 Department of Computer Science and Systems Technology, University of Pannonia, Egyetem u. 10, 8200 Veszprém, Hungary
* Correspondence: orban.eva@mik.uni-pannon.hu (É.O.-M.); vathy.agnes@mik.uni-pannon.hu (Á.V.-F.); Tel.: +36-88-624-000 (ext. 6109) (É.O.-M.)

Abstract: In this study, the effectiveness and characteristics of three ranking methods were investigated based on their performance in ranking European football teams. The investigated methods were the Thurstone method with ties, the analytic hierarchy process with logarithmic least squares method, and the RankNet neural network. The methods were analyzed in both complete and incomplete comparison tasks. The ranking based on complete comparison was performed on match results of national leagues, where each team had match results against all the other teams. In the incomplete comparison case, in addition to the national league results, only a few match results from international cups were available to determine the aggregated ranking of the teams playing in the top five European leagues. The rankings produced by the ranking methods were compared with each other, with the official national rankings, and with the UEFA club coefficient rankings. In addition, the correlation between the aggregated rankings and the Transfermarkt financial ranking was also examined for the sake of interest.

Keywords: ranking aggregation; analytic hierarchy process (AHP); Thurstone method; RankNet neural network; evaluation of sports results

Citation: Gyarmati, L.; Orbán-Mihálykó, É.; Mihálykó, C.; Vathy-Fogarassy, Á. Aggregated Rankings of Top Leagues' Football Teams: Application and Comparison of Different Ranking Methods. *Appl. Sci.* **2023**, *13*, 4556. https://doi.org/10.3390/app13074556

Academic Editor: Luis Hernández-Callejo

Received: 28 February 2023
Revised: 17 March 2023
Accepted: 29 March 2023
Published: 3 April 2023

1. Introduction

Sports competitions promise hard effort for competitors and exciting entertainment for supporters. However, sports and sporting competitions are not just about entertainment; they are also important from an economic point of view. Organizing and running them usually costs a lot of money; furthermore, sports betting, advertising, and player transfer fees also involve vast amounts. Perhaps this is why the importance of sports can be observed not only in the economic and entertainment sectors, but also in scientific research, which deals with predicting the results of matches and competitions.

Evaluating matches' results plays a crucial role in setting out the rankings and predicting the results of subsequent matches. The most frequently used method for ranking competitors is the point-based method. When two competitors face each other, the winner receives one, two, or three points (in chess, handball, and soccer, respectively, as examples) according to the scoring system used in the sport. At the same time, the loser does not receive any points. In the case of a tie, both competitors receive one or half a point. This method is called the row sum method [1]. If the number of matches played differs between competitors, the number of points collected does not reflect the strength of the competitors. The row sum method does not take into consideration the strength of the opponent: a win guarantees the same number of points whoever the opponent is, either weak or strong. This drawback can be partly eliminated by applying the generalized row sum method elaborated by Chebotarev [2]. However, the generalized row sum method requires a parameter that can be changed arbitrarily, and its value significantly impacts the evaluation result.

Another branch of evaluation methodologies is the set of Elo evaluation methods originally elaborated by Árpád Élő [3]. An Elo method is applied, for example, in chess, where players are ranked with their Elo points, but it can also be applied in evaluating horse sports [4]. The method changes the strengths of competitors in pairs. If two players play a match, then their Elo points change as a function of the result of the match played and the difference in the players' strengths. As this can be achieved in various ways, we can speak about different Elo methods. The Elo methods can be considered local methods, as only the strengths of the competitors participating in the current competition are changed. It is worth mentioning that, besides chess players' rankings, an Elo-type method has been used for the FIFA ranking of women's national football teams and the FIFA ranking of men's national teams since 2018 [5]. For further details about the Elo methods, see [6]; additionally, their predicting properties are detailed in [7].

The most popular method of evaluating through comparisons in pairs is the analytic hierarchy process (AHP) method elaborated by Saaty [8,9]. This method starts with a reciprocal symmetric matrix, with the elements expressing the priorities of the objects relative to each other. There are several possible ways to evaluate the matrix; the most frequently used is the eigenvector method. The result is a priority vector, which expresses the weights of the objects to be evaluated. The method has thousands of applications in every field of life, for example, in education [10], psychology [11], security [12], technology [13], business [14], and decision-making [15]. Its application for evaluating sports results is presented, for example, in [16,17]. The latter paper used the pairwise comparison method for ranking the best tennis players in the world, but "comparison in pairs" methods can also be used for racing sports [18], handball [19], chess [20], e-sports [21], and football [22].

An alternative means of comparison in pairs is the stochastic Thurstone-type methods. In this case, the performances of the competitors are considered random variables. This idea is not far from reality. Think of the surprising changes in the performances of teams. The model was presented by Thurstone [23] and was primarily elaborated for psychological questions. Since then, these method have often been used for evaluating subjective opinions when scaling is haphazard: in cases when it is difficult to compare the entity to a scale, but it is easy to compare the entities to each other. Thurstone applied Gauss distribution and allowed two options: better and worse. Bradley and Terry [24] used logistic distribution, but other distributions could also be applied [25]. Later, the model was generalized, allowing ties [26]. In [27,28], further generalizations were presented for an arbitrary number of options and a wide class of distributions. In these papers, the authors proved a sufficient condition for the existence and uniqueness of the parameter estimation, which was suitable for setting up both the national and international rankings of the teams investigated in this paper.

Despite the widespread use of neural networks, only a few papers have been published on their application to ranking problems. Although we can find some examples of their application in information retrieval [29,30], image classification [31], and developing recommendation systems [32], these implementations are mainly unique due to the specific nature of the tasks. For example, while mainly convolutional neural networks are used for object ranking, in text-processing-based ranking tasks, recurrent networks or BERT [33] models are mainly used. Nevertheless, for paired comparison tasks, the RankNet [34] neural network can provide a solution. This kind of neural network implements the learn-to-rank approach and uses gradient descent to obtain a ranking function.

When we examined the studies that have used these methods to rank sports teams, we found that the published articles typically deal with the results of different championships in isolation and do not deal with aggregating the rankings of different championships. Therefore, the aim of the current paper was to create a unified ranking of the teams playing in the strongest European football leagues, namely in the Premier League (English Championship), Ligue 1 (French Championship), the Bundesliga (German Championship), Serie A (Italian Championship), and La Liga (Spanish Championship). This unified ranking contained not only those teams that play international matches but also those that do not attend international cups. We have to note that UEFA also provides a ranking called the club coefficient ranking, but it contains only those teams that play international matches. Therefore, UEFA's ranking is only a restricted list of football teams. Aggregating the rankings of teams in national championships is appropriate to compare the performances of these teams that are not included in the cups. Moreover, aggregated ranking is suitable for evaluating and comparing national championships, as it provides a broad picture of all teams. It can also act as a basis for determining the number of participating teams from different countries in the international cups during the following season. In our opinion, aggregated rankings provide more sophisticated evaluations than rankings that consist of a selected subset of teams and work with a restricted dataset.

We have to emphasize that when the UEFA club coefficients are determined, only the international cups' matches are taken into account. In this paper, for the evaluations, a larger dataset was used. All match results from the investigated national championships were used, and the results of matches played in international cups served as links for aggregating the national championships. The considered international cups were the UEFA Champions League (UCL), the UEFA European League (UEL), and the UEFA Super Cup (USC).

The difficulty of creating an aggregated ranking of teams playing in different national championships was that only a small amount of linking information was available to rank the teams. Therefore, the problem could be seen as a ranking problem based on the incomplete comparison of teams. In our study, the evaluations were performed using the Thurstone method with ties (TH), the analytic hierarchy process with logarithmic least squares method (AL) and the RankNet (RN) neural network. During our analysis, we also investigated the effect of the weighting factor of international matches. The resulting rankings were compared to each other and also to the UEFA coefficient rankings. Additionally, we also made comparisons with data from Transfermarkt [35] that contained the financial values of the teams. Furthermore, we also compared the entire national leagues to each other and determined the rankings of the national championships.

Summarizing the methods and results related to this research, the contributions of this article are as follows: (1) The unified ranking of the teams playing in the five strongest European football leagues was determined based on *incomplete comparisons*. For this purpose, the results of all matches played in the national championships and the results of matches played between the same teams in international leagues were considered. (2) The methods used allowed teams that did not play international matches to be included in the rankings. (3) The ranking capabilities of the Thurstone method, the analytic hierarchy process with logarithmic least squares method, and RankNet were compared in a ranking task based on incomplete comparisons. (4) Finally, based on the experimental results, the main characteristics of the investigated methods were also highlighted.

The structure of the paper is as follows: In Section 2, we briefly present the research papers connected to the evaluation of football championships. Then, in Section 3, the methods applied for evaluation in the current study are briefly presented, namely the Thurstone method with maximum likelihood estimation, the analytic hierarchy process with logarithmic least squares method, and the RankNet method. Section 4 presents the research methodology. In Section 5, the evaluation results are presented and discussed. Section 5.1 contains the evaluation of match results in the Premier League by all three methods. Section 5.2 comprises the unified rankings of the teams in the five above-mentioned championships. Section 5.3 presents the correlations between the resulting rankings and the financial value of the teams, while Section 5.4 contains the rank correlations between the presented rankings and the UEFA club coefficient rankings in the case of those teams that played matches in international cups during the 2020/2021 season. In Section 5.5, the national leagues are compared to each other. Then, in Section 6, we summarize the properties of the applied evaluation methods. Finally, some conclusions are drawn in Section 7.

2. Literature Review

Football is a popular sport in Europe, and it attracts special attention. In this section, we briefly present some research papers dealing with the possibilities of football team evaluation. These works mainly concentrate on predicting the results of matches.

In [36], the English Premier League was investigated by a refined score-driven method, and this method was also extended to the best national leagues (English, German, Spanish, French, Italian, and Dutch) in [37]. The authors elaborated a bivariate Poisson model for predicting the number of goals scored by the teams. The success of the method was shown by the fact that the predictions concerning the differences in the number of goals coincided well with the odds provided by some bet offices.

In [38], the results of the UEFA Champions League were forecast using a cumulative probit model. The teams' strengths were supposed to be Gauss distributed, and the forecasting was expressed by linear regressions for the UEFA club rankings.

In [7], seven evaluation methods applying a gradient descent algorithm were analyzed from a predictive point of view. In this paper, the authors also considered the time effects. The starting point was the Elo method [3], which was modified according to the authors' aims. The predictive abilities of the algorithms were presented for the results of the "Big Six" in the Premier League (Arsenal, Chelsea, Liverpool, Manchester City, Manchester United, and Tottenham).

The authors of [39] measured the predictive abilities of different methods. In this paper, the authors compared ten scoring methods for predicting teams' rank probabilities.

Similarly to [39], in [40], the comparison of different methods was again the focus of the research. The authors investigated the Thurstone model, the Bradley–Terry model, the Bradley–Terry–Davidson model, and Poisson models. The article concluded that the Poisson models outperformed the Thurstone and Bradley–Terry models from a predictive point of view. However, we have to note that the Poisson models use more information than the Thurstone and Bradley–Terry models, as they also consider the number of goals scored by teams in addition to the relations (better, equal, worse) between them. The predictive power of fuzzy logic was analyzed in [41] through a football case study.

Nowadays, machine learning provides new possibilities for predicting sports results. Although neural-network-based methods are popular, their disadvantage is that they require lots of data to provide reliable results. In [42], the authors built a neural network for predicting football match results, and they demonstrated its effectiveness using data from the Chinese Football Super League. In [43], Premier League match results were also analyzed by machine learning methods. In [44], a similarity learning approach was proposed to predict football match rankings in the English Premier League. The study utilized transfer learning, and the performances of standard neural networks, Siamese networks, the RankNet method, the traditional sports tally ranking method, and graph-based PageRank methods were evaluated. Besides using the English Premier League seasonal team statistics, the data were also augmented with financial and transfer data. The authors concluded that no model consistently performed the best from all perspectives. However, machine learning methods have been applied to not only evaluate the performances of football teams but also determine the values of football players [45]. This information can also be used to estimate the results of matches [46].

Hybrid methods for evaluating football results have also been developed. For example, in [47], the random forest machine learning method was combined with the Poisson ranking method and the bookmaker consensus. In this article, the FIFA Women's World Cup 2011 and 2015 were used to train the proposed method, and then the FIFA Women's World Cup 2019 was simulated 100,000 times to determine the winning probabilities of the 24 participating teams.

As can be seen, the main direction of research is to predict match results, typically focusing on a single league or sporting event. Unfortunately, only a few articles have been published on the combined analysis of the results of different championships. In [48], the authors analyzed the results of the top football leagues and international cup matches.

The applied approach was a dynamic extension of a method based on logistic regression. To be able to estimate a large number of parameters, the authors considered a long time period, namely from 1996 to 2001. Therefore, they included about 9000 match results for estimating the parameters. As the international cups' results comprised a small portion of all match results, they used weights to highlight the importance of international matches. The study also covered the analysis of the effects of the weights. The authors concluded that the rankings changed depending on whether the international matches were weighted or not. Although they established that it was useful to use weights, they provided no suggestion for the values of weights; in their study, they used ad hoc values.

3. Methods Applied

This section presents the methods used for ranking the teams and merging them. The Thurstone method and analytic hierarchy process are among the most frequently used methods in the case of comparisons in pairs, while the RankNet method uses an increasingly popular neural network model to rank competing pairs.

3.1. Thurstone Method with Ties

The Thurstone method [23] is a stochastic method in the sense that it supposes that the performances of players (in this case, teams) are random variables. This assumption can be reflective of reality. If we keep a close watch on matches, we can also observe some fluctuations in the performances of the individuals and teams. The strengths of the teams are the expectations for the random variables. These random variables are denoted by ξ_i, $i = 1, 2, 3, \ldots, n$, where n is the number of objects (teams) to evaluate, and the expectations are denoted by m_i, $i = 1, 2, 3, \ldots, n$, respectively.

The actual result of a match depends on the actual values of the random variables, or, more precisely, on their differences. Since in football there are three possible results, win, tie, and defeat, we also allowed three options for the result of a comparison: better (win), draw (tie), and worse (defeat). The differences in performances are $\xi_i - \xi_j = m_i - m_j + \eta_{i,j}$, where $\eta_{i,j}$ are supposed to be independent, identically distributed random variables with the cumulative distribution function F. If F is the standard normal (Gauss) cumulative distribution function ($F = \Phi$), then the model is the generalization of the Thurstone model. If F is the logistic cumulative distribution function, then the Bradley–Terry model with ties is used. In this paper, we used Gauss distribution.

If the result of a match between teams i and j is a tie, then the difference $\xi_i - \xi_j$ is close to zero. The bounds of the tie are represented by $-d$ and d, where d is a positive parameter estimated from the data. If the result of the match is a win, then the difference is above the parameter d. If team i suffers a defeat to team j, the difference is below $-d$ (see Figure 1). Of course, if i beats j, then j suffers a defeat to i.

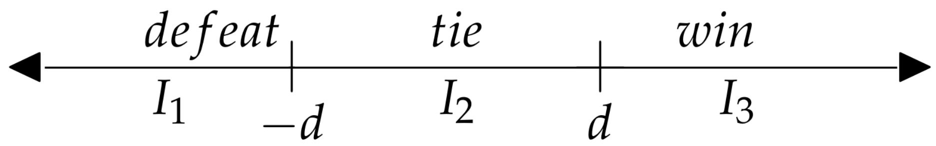

Figure 1. The possible results and the intervals belonging to them.

The probabilities of the defeat/tie/win results for the matches between teams i and j can be written as follows:

$$p_{i,j,1} = P(\text{team } i \text{ is defeated by team } j) = P(\xi_i - \xi_j \in I_1) = \Phi(-d - (m_i - m_j)) \quad (1)$$

$$p_{i,j,2} = P(\text{the result is a tie}) = P(\xi_i - \xi_j \in I_2) = \Phi(d - (m_i - m_j)) - \Phi(-d - (m_i - m_j)) \quad (2)$$

$$p_{i,j,3} = P(\text{team } i \text{ wins over team } j) = P(\xi_i - \xi_j \in I_3) = 1 - \Phi(d - (m_i - m_j)). \quad (3)$$

Let **A** be a three-dimensional data matrix with size $n \times n \times 3$ and with elements $A_{i,j,k}$ $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, $k = 1, 2, 3$. $A_{i,j,k}$ is the number of matches when team i has result k against team $j, i < j$; $k = 1$ stands for a defeat, $k = 2$ for a draw, and $k = 3$ for a win. If $i \geq j$, then let $A_{i,j,k} = 0$ for all $k = 1, 2, 3$. The probability of the results provided by data matrix **A** in the function $\underline{m} = (m_1, m_2, \ldots, m_n)$ and $0 < d$, supposing the independence of the sample elements, is

$$L(\mathbf{A}|\underline{m}, d) = \prod_{k=1}^{3} \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} p_{i,j,k}^{A_{i,j,k}}. \tag{4}$$

where L is the likelihood function. The maximum likelihood estimation of parameters $\underline{m}$ and $0 < d$, denoted by $\underline{\widehat{m}}$ and $\hat{d}$, is the $n + 1$ dimensional argument, where the function L reaches its maximal value [49].

For Gauss distribution and for the case of more than two options, sufficient conditions for the existence and uniqueness of the maximal value of (4) are provided in [27], which was generalized in [28]. In the following, we formulate the statement for the three options. Let us define the graph G_{TH} as follows: let the vertices be the teams and let the nodes i and j $(i < j)$ be connected if $0 < A_{i,j,2}$ or $0 < A_{i,j,1} \cdot A_{i,j,3}$ (i.e., teams i and j have either drawn at least once, or both teams have beaten each other at least once).

Theorem 1 ([27])**.** *Suppose that there is a pair (i_1, j_1), $i_1 < j_1$, for which*

$$0 < A_{i_1,j_1,2} \tag{5}$$

and a pair (i_2, j_2), $i_2 < j_2$, for which

$$0 < A_{i_2,j_2,1} \cdot A_{i_2,j_2,3}. \tag{6}$$

If the graph G_{TH} is connected, then, fixing $m_1 = 0$, the likelihood function attains its maximum, and the maximizer is unique.

This statement can be applied even to the national leagues of all five investigated nations, as well as to the whole set of matches. Therefore, the evaluations by the Thurstone method could be uniquely performed. The estimated expectations could be transformed into weights as follows [27]:

$$\underline{\widehat{w}} = \left(\frac{\exp(\widehat{m}_1)}{\sum_{i=1}^{n} \exp(\widehat{m}_i)}, \ldots, \frac{\exp(\widehat{m}_n)}{\sum_{i=1}^{n} \exp(\widehat{m}_i)} \right). \tag{7}$$

The coordinates w_i are called weights in the TH model and are used for comparisons between different methods.

3.2. Analytic Hierarchy Process with Logarithmic Least Squares Method

In the case of the analytic hierarchy process (AHP) method [8,9], the starting point is a reciprocal symmetric matrix called a pairwise comparison (PC) matrix. Its elements express how much stronger a team is relative to another. Let the PC matrix be $\mathbf{B} = (b_{i,j})$, $i = 1, \ldots, n$, $j = 1, \ldots, n$. In AHP, the evaluation "better" is expressed by the number 3, "much better" by 5, and so on. Applying the Thurstone method, we distinguished three options (win, tie, loss), and we did not differentiate between wins based on how many more goals were scored by the winning team. Therefore, we also used only three categories in the case of AHP. We constructed the PC matrix as follows: during the mth match between teams i and j, $b_{i,j}^{(m)} = 3$ if team i wins over team j; $b_{i,j}^{(m)} = 1$ if the result is a draw; and $b_{i,j}^{(m)} = 1/3$ if team i suffers a defeat to team j. In the case of k matches between the teams i and j, we took the geometric mean as follows [50]:

$$b_{ij} = \sqrt[k]{\prod_{m=1}^{k} b_{i,j}^{(m)}}. \tag{8}$$

If team i does not play any matches with team j, then the element $b_{i,j}$ is not defined, and its place remains empty in the matrix. If there is at least one match between the teams i and j, it can be easily seen that $b_{i,j} = \frac{1}{b_{j,i}}$, letting $b_{i,i} = 1$ for each $i = 1, 2, \ldots, n$. If every team plays with every other team, i.e., the comparison is complete, then the most frequently applied evaluation method is the eigenvector method. Completeness holds for all national leagues but, unfortunately, is not satisfied when we concatenate these national leagues. That is, lots of teams did not qualify for international cups, but we intended to locate them in the international rankings all the same. In the case of an incomplete comparison, the most frequently applied method is the logarithmic least squares method (LLSM) [51]. That is, minimize the function

$$H(\underline{w}) = \sum_{(i,j)\in I} \left(\log(b_{i,j}) - (\log(w_i) - \log(w_j))\right)^2 \tag{9}$$

under the conditions

$$0 < w_i, i = 1, 2, \ldots, n \tag{10}$$

and

$$\sum_{i=1}^{n} w_i = 1. \tag{11}$$

In [51], the authors proved the necessary and sufficient conditions for the existence and uniqueness of the optimization problem. Let I be the set that contains all the pairs of teams for which there is at least one comparison. Let us define the graph of comparisons G_I as follows: the vertices are the teams, and there is an edge between two vertices if the pair is an element of I.

Theorem 2 ([51])**.** *The optimization problem minimizing (9) under conditions (10) and (11) can be solved uniquely if and only if the graph G_I is connected.*

We note that $G_{TH} \subseteq G_I$; therefore, if the Thurstone method with ties works, so does the analytic hierarchy process with logarithmic least squares method. One can easily see that the requirements of Theorem 2 are less strong than those of Theorem 1.

3.3. RankNet

The RankNet model is a feed-forward deep learning neural network developed to implement the learn-to-rank approach [34]. Although RankNet is mainly used for ranking items (documents or images) based on their relevance to a given query, the network structure also allows its application to rank sports teams.

RankNet uses a paired comparison between instances (in our case, teams) to learn the ranking function. The base network receives two inputs that pass through the same hidden layers and calculates a score for each input, expressing the strengths of the instances. Following this, the difference between these scores is calculated, and the difference is passed through an activation function, which provides the output of the RankNet network. Finally, the ranking of instances is performed based on the scoring values learned by the Base Network.

If we use this architecture to rank football teams, then the inputs are provided by the teams playing against each other, while the output is the result of the match, i.e., win, draw, or defeat from the perspective of the first team. In our study, the match outcomes were modeled with 0 point for a defeat, 0.5 point for a tie, and 1 point for a win; therefore, we placed a single neuron with sigmoid activation in the output layer of the RankNet network. During the learning process, the base network learned the scoring function, which correctly represented the teams' strengths according to their paired match results. For this purpose, we used two hidden layers containing 32 and 16 neurons with Leaky ReLU activation. To avoid overfitting, dropout regularization was applied for the first hidden layer with a dropout rate of 0.1. The input layer contained two neurons (one for the teamID and one for the points gained by the team) and used Leaky ReLU activation, and dropout regularization with a dropout rate of 0.1 was applied. The output layer of the base network was implemented as a single neuron with linear activation. It is important to emphasize that the output of this layer provided the strength of each team. In this layer, the linear activation function was used; therefore, the value of the output was not constrained, and the strengths of the teams could be expressed with an arbitrary value. A block diagram of the network applied in our study is illustrated in Figure 2, and Table 1 contains detailed information about the layers.

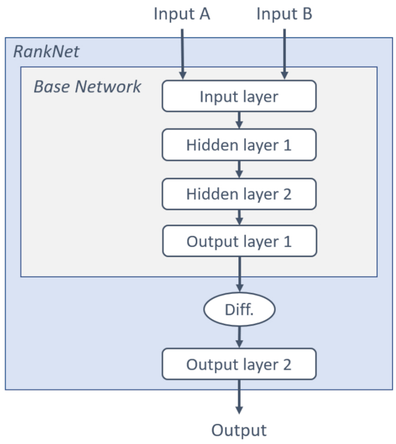

Figure 2. The sequential block diagram of the RankNet network used in the study.

Table 1. The number of neurons and the applied activation functions in the layers of the RankNet network designed for the study.

Layer	Neurons	Activation	Dropout Regularization
Input layer	2	Leaky ReLU	Yes, dropout rate = 0.1
Hidden layer 1	32	Leaky ReLU	Yes, dropout rate = 0.1
Hidden layer 2	16	Leaky ReLU	No
Output layer 1	1	Linear	No
Output layer 2	1	Sigmoid	No

During the training, the error of the RankNet network was measured by the mean squared error, abbrevated by MSE, provided by Equation (12). For the optimization, the "adam" optimizer was applied with a learning rate equal to 0.001.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{12}$$

where y_i denotes the expected value for the ith input (1 for a win for team A, 0.5 for a draw, and 0 for a win for team B), and $\hat{y}_i$ is the value computed by the RankNet for the ith input.

In our study, all parameters and hyper-parameters of the network were tuned with the grid search method to find the best values. As the neural-network-based scoring of the teams was not deterministic, the training of the networks was, in each case, performed 100 times. The final scores of the teams were calculated as the average of the resulting scores. The scores were transformed into weights via a process similar to that described in (7). In each training iteration, the RankNet network was trained for 200 epochs with a batch size of one. To stop the training in time (to avoid overfitting), early stopping regularization was applied with a patience value of 20.

4. Data and Research Methodology

In our study, the top five European national leagues in the 2020/21 season were analyzed, and this dataset was augmented with the results of the UEFA Champions League (UCL), the UEFA Europa League (UEL), and the UEFA Super Cup (USC) from the same season.

In the study, the following national championships were enrolled: the English Premier League (ENG), the French Ligue 1 (FRA), the German Bundesliga (GER), the Italian Serie A (ITA) and the Spanish La Liga (ESP). We chose these leagues because they were the top 5 leagues in the season 2020/21 (and in some other seasons, too); furthermore, the numbers of teams playing in these domestic leagues and the systems for running these championships were similar. The ENG, FRA, ITA, and ESP national leagues comprise 20–20 teams, and every team plays with the others twice; therefore, these national leagues have 380 matches. In the Bundesliga, there are only 18 teams; hence, the number of matches is only 306. Altogether, we considered in this study 1826 matches from national leagues.

In the season considered, teams from the mentioned five leagues played 53 international matches against each other in the UCL, 19 matches in the UEL, and 1 match in the USC. This resulted in 73 international matches that provided the basis for merging the individual rankings arising from the match results achieved in the national leagues. The number of matches for concatenating the national leagues' teams was not small (73), but if we compare it to the total number of matches, it did not represent a significant portion. The ratio was $73/(1826 + 73) = 0.0384$, which was under 4%. We note that the comparisons in the national leagues were always complete, but the unified ranking was based on incomplete comparisons. This explains the necessity of using the LLSM method instead of the eigenvector method. We checked that in the case of each national league and also in the case of all matches, the assumptions of Theorem 1 were satisfied. Therefore, the evaluations by the Thurstone and AHP with LLSM methods were mathematically established in the study. It is also worth noting that, although many teams did not play any international matches, we could fit them into the international ranking with the help of their results against teams who played in international cups.

In [48], the authors argued for considering international matches with multiple weights. Following this recommendation, the merging was performed both without weights and with weights equal to 4. The reason for weighting, as explained by Held and Vollnhals, was the following: as international matches are usually more important, they are given priority attention by teams. The ranking and weights of the teams for international matches could also serve as a basis for comparing the national championships themselves, in addition to the comparison of the teams.

For our study, all the match data mentioned above were downloaded with the help of [52]. Additionally, Transfermarkt data [35] containing the financial values of the teams were also downloaded and were used to evaluate the results more extensively. The UEFA club coefficients were downloaded from [53].

5. Results and Discussion

This section presents the results of the evaluations and related discussions. First, the ranking of the teams was carried out for each league separately by excluding the international matches. Section 5.1 presents the ranking results for the English Premier League as an example. Following this, in Section 5.2, the aggregated rankings are shown. The rank correlations are presented in Section 5.3. In Section 5.4, the aggregated rankings are compared with the UEFA club coefficient rankings. Finally, Section 5.5 presents the comparison of the different national leagues' strengths according to different indicators.

5.1. Evaluation of the Rankings of Premier League Teams

In this subsection, the rankings of the Premier League teams are presented and evaluated. For ranking the teams, only the match results in the English Championship were considered. Table 2 shows the orders of the teams based on the Thurstone method with ties (TH), the analytic hierarchy process with LLSM method (AL), and the RankNet network (RN). Additionally, the table also includes the weights assigned to the teams by these methods. For the sake of interest, the last columns contain the financial values of the teams based on Transfermarkt data (TM) and the official ranking (OR) based on the points collected during the whole season.

Table 2. Evaluations of the Premier League teams based on their matches in the national championship by different methods (Thurstone (TH), analytic hierarchy process with LLSM method (AL), and RankNet (RN)); the financial values of the teams based on Transfermarkt (TM) data; and the official ranking (OR) based on the number of points received in the Premier League.

	TH		AL		RN		TM		OR	
Rank	**Name**	**Weight**	**Name**	**Weight**	**Name**	**Weight**	**Name**	**Value (€)**	**Name**	**Point**
1	Manch City	0.10637	Manch City	0.08523	Manch City	0.34021	Manch City	1040.00 m	Manch City	86
2	Manch Utd	0.07886	Manch Utd	0.07228	Manch Utd	0.14567	Liverpool	969.65 m	Manch Utd	74
3	Liverpool	0.06772	Liverpool	0.06476	Liverpool	0.08614	Chelsea	889.20 m	Liverpool	69
4	Chelsea	0.06558	Chelsea	0.06301	Chelsea	0.07099	Manch Utd	770.05 m	Chelsea	67
5	Leicester	0.0619	West Ham	0.05964	West Ham	0.06401	Tottenham	703.50 m	Leicester	66
6	West Ham	0.06162	Leicester	0.05964	Leicester	0.05341	Arsenal	619.25 m	West Ham	65
7	Tottenham	0.05732	Tottenham	0.05645	Tottenham	0.04191	Everton	511.20 m	Tottenham	62
8	Arsenal	0.05521	Arsenal	0.05492	Arsenal	0.03644	Leicester	510.70 m	Arsenal	61
9	Everton	0.05320	Everton	0.05343	Everton	0.03279	Wolverhampton	450.30 m	Leeds	59
10	Leeds	0.05314	Leeds	0.05199	Leeds	0.02987	Aston Villa	430.75 m	Everton	59
11	Aston Villa	0.04717	Aston Villa	0.04921	Aston Villa	0.02275	West Ham	334.90 m	Aston Villa	55
12	Newcastle	0.03852	Wolverhampton	0.04173	Newcastle	0.01485	Brighton	290.40 m	Newcastle	45
13	Wolverhampton	0.03844	Newcastle	0.04173	Wolverhampton	0.01182	Southampton	282.80 m	Wolverhampton	45
14	Crystal Palace	0.03765	Crystal Palace	0.04060	Brighton	0.01073	Newcastle	255.35 m	Crystal Palace	44
15	Brighton	0.03707	Brighton	0.04060	Crystal Palace	0.01048	Fulham	239.75 m	Southampton	43
16	Southampton	0.03607	Southampton	0.03950	Southampton	0.00945	Leeds	238.10 m	Brighton	41
17	Burnley	0.03290	Burnley	0.03739	Burnley	0.00781	Crystal Palace	193.85 m	Burnley	39
18	Fulham	0.02721	Fulham	0.03171	Fulham	0.00467	Sheffield Utd	149.85 m	Fulham	28
19	West Brom	0.02541	West Brom	0.03001	West Brom	0.00362	West Brom	141.15 m	West Brom	26
20	Sheffield Utd	0.01865	Sheffield Utd	0.02616	Sheffield Utd	0.00239	Burnley	132.30 m	Sheffield Utd	23

As we can see, Manchester City was the best team in the English Championship. Moreover, the ranking of the first four teams was the same for all methods except for the data from Tranfermarkt. The same could be stated for the last four teams, too.

Concerning the different evaluation methods (TH, AL, and RN), although the weights of the teams were different, in the rankings, one can see only small differences: if we compare the results of the TH and AL methods, Leicester and West Ham, and Newcastle and Wolverhampton swapped places with each other. Comparing the Thurstone and RankNet methods, we found differences in the 5th and 6th, as well as the 14th and 15th positions. Considering the AL and RN methods, there were differences in the cases of the 12th and 13th, as well as the 14th and 15th places. Overall, we found that all three methods produced very similar results, and the rankings were not far from the official rankings of the Premier League.

The Kendall's tau rank correlations between the rankings can be seen in the upper-left sub-matrix of Table 6. It is interesting that all the rank correlations were the same and very high (0.979). However, this was the right result due to the equal number of changeovers. A similar phenomenon could also be recognized in the case of each national championship (GER, FRA, ITA, and ESP). At first sight, we realized that the ranking of the Transfermarkt values showed substantial differences from the rankings of the TH, AL, and RN methods. This pointed to the fact that the return on investment was different for all teams. Manchester Utd was less valuable than Liverpool and Chelsea, while Wolverhampton was more valuable financially than Leeds, Aston Villa, and Newcastle. In addition, Transfermarkt's ranking differed significantly from the official ranking, although the correlation was moderate.

Summarizing the findings, we can state that all three ranking methods performed equally well in relation to the official ranking result. However, we have to note that the weights resulting from the RankNet network differed significantly from those of the TH and AL methods: the top teams had much larger weights, and at the bottom of the ranking, the teams had much lower weights in the case of RN than in the case of the TH and AL methods.

5.2. Evaluation of the Overall International Rankings

This subsection contains the aggregated rankings of the teams playing in the top national championships. The aggregated rankings were based on the joint results of national championships and international cup matches. Tables 3 and 4 contain the first 20 and the last 10 teams, as evaluated by the TH, AL, and RN methods, using W = 1 and W = 4 weights for the international matches. In detail, Table 3 shows the outcome when the results of international cup matches were taken into account in the same manner as the national championship matches, while Table 4 shows the outcome when international matches were considered with four-times larger weights than matches played in national championships. We could observe that teams who won lots of matches in international cups advanced in the rankings. A typical example is Chelsea: this team was the fourth team in the national championship in every evaluation (Table 2). In the unified ranking (W = 1), it reached second place among the English teams but the sixth among all teams according to the TH method, second according to the AL method, and tenth according to the RN method (see Table 3). When we considered the international matches with a weight of four, Chelsea reached second place in the international ranking according to the TH and AL methods and third place according to the RN method (see Table 4). As this team was the winner of both the UCL and the UEFA Super Cup that year, it deserved these positions.

When we considered at the overall picture based on the three evaluation methods and two weights applied for international matches, we could establish that the best six teams were the following (in this order): Manchester City, Chelsea, Real Madrid, Bayern Munich, Manchester United, and Liverpool. Furthermore, it could be seen that Manchester City was the top team in all rankings, regardless of whether international matches were weighted or not.

Table 3. Rankings after aggregation, taking into account the international matches with W = 1 weights.

	TH		AL		RN		TM	
Rank	**Name**	**Weight**	**Name**	**Weight**	**Name**	**Weight**	**Name**	**Value (€)**
1	Manchester City	0.02874	Manchester City	0.02616	Manchester City	0.10103	Manchester City	1040.00 m
2	Bayern Munich	0.02767	Chelsea	0.02113	Lille	0.07739	Liverpool	969.65 m
3	Real Madrid	0.02351	Manchester Utd	0.01967	Inter	0.06707	Chelsea	889.20 m
4	Inter	0.02158	Liverpool	0.01902	Bayern Munich	0.06283	Bayern Munich	858.23 m
5	Atl. Madrid	0.02045	Bayern Munich	0.01831	Real Madrid	0.05445	Barcelona	817.5 0m
6	Chelsea	0.01993	Real Madrid	0.01826	Monaco	0.03338	Paris SG	805.00 m
7	Manchester Utd	0.01894	West Ham	0.01776	Lyon	0.03338	Real Madrid	787.30 m
8	Barcelona	0.01718	Leicester	0.01776	Atl. Madrid	0.03275	Atl. Madrid	773.10 m
9	RB Leipzig	0.01717	Tottenham	0.01681	Juventus	0.03186	Manchester Utd	770.05 m
10	Liverpool	0.01679	Atl. Madrid	0.01619	Chelsea	0.02920	Tottenham	703.50 m
11	Wolfsburg	0.01673	Everton	0.01592	Manchester Utd	0.02875	Inter	663.90 m
12	Sevilla	0.01650	Arsenal	0.01566	Napoli	0.02399	Juventus	633.30 m
13	Eintracht Frankfurt	0.01643	Leeds	0.01548	Barcelona	0.02239	Dortmund	628.40 m
14	Dortmund	0.01591	Aston Villa	0.01466	Paris SG	0.02209	Arsenal	619.25 m
15	Lille	0.01578	Inter	0.01465	AC Milan	0.02141	RB Leipzig	574.95 m
16	Juventus	0.01577	Barcelona	0.01444	Atalanta	0.02118	AC Milan	547.78 m
17	Atalanta	0.01541	Sevilla	0.01406	Liverpool	0.01869	Napoli	530.20 m
18	Paris SG	0.01512	Lille	0.01387	Sevilla	0.01823	Everton	511.20 m
19	Napoli	0.01508	RB Leipzig	0.01357	Eintracht Frankfurt	0.01512	Leicester	510.70 m
20	AC Milan	0.01501	Paris SG	0.01316	Wolfsburg	0.01470	Wolverhampton	450.30 m
...	...	...	...	...	...	...	...	...
89	Strasbourg	0.00530	Udinese	0.00613	Benevento	0.00097	Spezia	71.68 m
90	Nantes	0.00527	Torino	0.00613	Valladolid	0.00097	Dijon	65.63 m
91	Brest	0.00523	Spezia	0.00613	Fulham	0.00094	Huesca	64.10 m
92	Benevento	0.00520	Cagliari	0.00580	Eibar	0.00076	Valladolid	63.05 m
93	Schalke	0.00442	Nimes	0.00549	West Brom	0.00072	Nimes	60.4 m
94	Nimes	0.00441	Benevento	0.00549	Sheffield Utd	0.00045	Arminia Bielefeld	56.73 m
95	Sheffield Utd	0.00420	Schalke	0.00507	Dijon	0.00043	Crotone	52.75 m
96	Parma	0.00371	Dijon	0.00429	Crotone	0.00038	Benevento	50.13 m
97	Crotone	0.00345	Crotone	0.00429	Parma	0.00037	Cadiz CF	47.65 m
98	Dijon	0.00303	Parma	0.00429	Schalke	0.00028	Elche	45.65 m

Table 4. Rankings after aggregation, taking into account the international matches with W = 4 weights.

	TH		AL		RN		TM	
Rank	**Name**	**Weight**	**Name**	**Weight**	**Name**	**Weight**	**Name**	**Value (€)**
1	Manchester City	0.03985	Manchester City	0.02575	Manchester City	0.14878	Manchester City	1040.00 m
2	Chelsea	0.03276	Chelsea	0.02145	Lille	0.08444	Liverpool	969.65 m
3	Real Madrid	0.02816	Manchester Utd	0.01967	Chelsea	0.07968	Chelsea	889.20 m
4	Bayern Munich	0.02791	Liverpool	0.01902	Bayern Munich	0.06167	Bayern Munich	858.23 m
5	Liverpool	0.02036	Real Madrid	0.01830	Real Madrid	0.05888	Barcelona	817.50 m
6	Manchester Utd	0.01998	Bayern Munich	0.01830	Lyon	0.03717	Paris SG	805.00 m
7	Atl. Madrid	0.01736	Leicester	0.01776	Monaco	0.03552	Real Madrid	787.30 m
8	Inter	0.01728	West Ham	0.01776	Inter	0.02765	Atl. Madrid	773.10 m
9	Leicester	0.01702	Tottenham	0.01682	Manchester Utd	0.02702	Manchester Utd	770.05 m
10	Sevilla	0.01603	Atl. Madrid	0.01623	Juventus	0.02581	Tottenham	703.50 m
11	Wolfsburg	0.01597	Everton	0.01592	Napoli	0.02260	Inter	663.9 0m
12	Tottenham	0.01592	Arsenal	0.01566	Liverpool	0.01896	Juventus	633.30 m
13	Eintracht Frankfurt	0.01553	Leeds	0.01549	Eintracht Frankfurt	0.01752	Dortmund	628.40 m
14	Paris SG	0.01535	Inter	0.01466	West Ham	0.01722	Arsenal	619.25 m
15	Barcelona	0.01507	Aston Villa	0.01466	Wolfsburg	0.01659	RB Leipzig	574.9 5m
16	RB Leipzig	0.01477	Barcelona	0.01446	Bayer Leverkusen	0.01525	AC Milan	547.78 m
17	Leeds	0.01473	Sevilla	0.01409	Leicester	0.01447	Napoli	530.20 m
18	Lille	0.01469	Lille	0.01384	Sevilla	0.01232	Everton	511.20 m
19	Everton	0.01463	RB Leipzig	0.01356	Barcelona	0.01217	Leicester	510.7 0m
20	Juventus	0.01447	Paris SG	0.01313	Tottenham	0.01114	Wolverhampton	450.30 m
…	…	…	…	…	…	…	…	…
89	Angers	0.00478	Udinese	0.00614	Werder Bremen	0.00108	Spezia	71.68 m
90	Strasbourg	0.00463	Torino	0.00614	West Brom	0.00107	Dijon	65.63 m
91	Lorient	0.00461	Brest	0.00612	Valladolid	0.00105	Huesca	64.10 m
92	Nantes	0.00459	Cagliari	0.00581	Benevento	0.00103	Valladolid	63.05 m
93	Brest	0.00457	Benevento	0.00550	Eibar	0.00081	Nimes	60.40 m
94	Schalke	0.00421	Nimes	0.00548	Sheffield Utd	0.00061	Arminia Bielefeld	56.73 m
95	Nimes	0.00383	Schalke	0.00507	Dijon	0.00048	Crotone	52.75 m
96	Parma	0.00334	Parma	0.00429	Parma	0.00039	Benevento	50.13 m
97	Crotone	0.00318	Crotone	0.00429	Crotone	0.00038	Cadiz CF	47.65 m
98	Dijon	0.00262	Dijon	0.00428	Schalke	0.00028	Elche	45.65 m

The correlations of the rankings applying different weights for international matches are summarized in Table 5. The use of different weights for each method also led to significant variations in the rankings, and this was clearly shown in the rank correlation indices. The only exception was the AL method, where there was only a slight change in the ranking when using different weights for international matches. The reason was the following: the pairwise comparison matrix **B** hardly changed if we used W = 4 instead of W = 1 due to Equation (8). The only values that changed in the matrix were those belonging to team pairs that played with each other in both national leagues and international cups. The only such pair was Chelsea and Manchester City. Therefore, only two elements of the pairwise comparison matrices were different when we used W = 1 and W = 4 weights, and the results of the evaluations were very close to each other.

Table 5. Rank correlations between all teams' aggregated rankings computed by different methods and applying different weights for international matches.

		International W = 1			International W = 4			
		TH	**AL**	**RN**	**TH**	**AL**	**RN**	**TM**
Int. W = 1	TH	1	0.763	0.716	0.884	0.772	0.685	0.622
	AL	0.763	1	0.638	0.805	0.987	0.640	0.631
	RN	0.716	0.638	1	0.637	0.628	0.880	0.592
Int. W = 4	TH	0.884	0.805	0.637	1	0.814	0.641	0.621
	AL	0.772	0.987	0.628	0.814	1	0.637	0.633
	RN	0.685	0.640	0.880	0.641	0.637	1	0.552
	TM	0.622	0.631	0.592	0.621	0.633	0.552	1

5.3. Rank Correlations between the Aggregated Rankings

This subsection provides a comparative analysis of the aggregated rankings obtained using different methods. The analysis examined the effects of the application of different methods (TH, AL, and RN) and different weightings for international matches (W = 1 or W = 4). In the evaluation, the similarity between the rankings was measured by calculating the Kendall's tau rank correlation. Table 5 summarizes the correlation coefficients of the rankings that contained all teams from the five leagues.

Table 5 shows that the correlations were higher between the rankings of the same method, even if the weights differed, than between the rankings generated by different methods using the same weights. By applying the same method and different weights, as previously mentioned, we observed hardly any differences in the case of the AL method (the correlation was 0.987). In the case of the TH method, this measurement was 0.884, while in the case of RN, it equaled 0.880. Therefore, the method itself had a stronger effect on the rankings than the choice of weights. The high values of the rank correlation coefficients could be easily understood by comparing the rankings produced using the same method but different weights (Tables 3 and 4). For example, when we compared the rankings generated by the TH method using different weights, we saw that 8 of the top 10 teams were the same. This observation was true for the weakest 10 teams, too. Concerning the AL and RN methods, the number of equally placed teams was at least 9 among both the best and worst 10 teams. Based on our observations, we could say that modifying the weights (W) caused only local changes in the rankings.

Comparing the different methods, the largest difference from the other methods was in the case of the RN method: The correlation coefficients with TH were 0.716 and 0.641, and those with AL were 0.638 and 0.637, while in the case of TH and AL, the correlation coefficients were 0.763 and 0.814. Therefore, the TH method resembled the AL method more than the RN method, and the same could be stated for the AL method. Finally, we noted that the TM ranking had moderate correlations with each ranking, but the lowest correlation coefficients (0.592 and 0.552) were seen again with the RN method.

Table 6 contains the correlation coefficients between the different rankings of English teams. Additionally, the last two columns of the table contain the Transfermarkt ranking and the official ranking. Concerning the correlation coefficients of the TM ranking, we observed values that were moderate (around 0.7) but still higher than in the case of the TM ranking for all teams (Table 5). Concerning the official ranking, we realized that the national rankings had larger correlations with the official rankings than the international ones. The larger the weight of the international matches, the lower the correlation with the official ranking, which was an easily explicable phenomenon. We observed that the international match results principally affected the rankings of the TH method and least affected the rankings of the AL method. The TM ranking had a medium correlation (around 0.7) with both national and international rankings.

Table 6. Kendall's tau rank correlations between the different rankings of English teams.

		ENG National			International W = 1			International W = 4				
		TH	**AL**	**RN**	**TH**	**AL**	**RN**	**TH**	**AL**	**RN**	**TM**	**OR**
ENG Nat.	TH	1	0.979	0.979	0.895	0.958	0.968	0.8	0.958	0.916	0.716	0.979
	AL	0.979	1	0.979	0.895	0.958	0.968	0.8	0.958	0.916	0.716	0.958
	RN	0.979	0.979	1	0.874	0.958	0.968	0.8	0.937	0.916	0.716	0.958
Int. W = 1	TH	0.895	0.895	0.874	1	0.916	0.905	0.905	0.937	0.895	0.737	0.874
	AL	0.958	0.958	0.958	0.916	1	0.989	0.821	0.979	0.958	0.695	0.937
	RN	0.968	0.968	0.968	0.905	0.989	1	0.811	0.968	0.947	0.705	0.947
Int. W = 4	TH	0.800	0.800	0.800	0.905	0.821	0.811	1	0.842	0.821	0.705	0.800
	AL	0.958	0.958	0.937	0.937	0.979	0.968	0.842	1	0.937	0.716	0.937
	RN	0.916	0.916	0.916	0.895	0.958	0.947	0.821	0.937	1	0.653	0.937
	TM	0.716	0.716	0.716	0.737	0.695	0.705	0.705	0.716	0.653	1	0.695
	OR	0.979	0.958	0.958	0.874	0.937	0.947	0.8	0.937	0.937	0.695	1

5.4. Comparison of the Aggregated Rankings and the UEFA Club Coefficient Rankings

In this subsection, we compare the evaluation results generated by the TH, AL, and RN methods to the UEFA club coefficient rankings. These rankings and ratings are computed every season for the teams playing in international cups. Teams not included in international cups do not receive such coefficients for that season. The method of the calculation is a point-based algorithm, which is detailed on the website [54]. The five-year club coefficient is the sum of the annual club coefficient in the year of interest and the previous four years. Financial support and the granting of spots in international cups are calculated based on these club coefficients.

In the 2020/21 season, 34 teams from the investigated leagues played matches in the international cups. The list of these teams and their UEFA club coefficients for the 2020/21 season are contained in Table 7.

If we consider Table 4, we can observe that Manchester City and Chelsea were first and second, respectively, in the TH and AL rankings, as is also shown in Table 7; moreover, Real Madrid, Bayern Munich, Manchester United, and Liverpool were among the best teams according to the TH, AL, and UEFA club coefficient models. Surprisingly, Villareal was placed third by the UEFA coefficients, but it was not among the best 20 teams as evaluated by the TH, AL, and RN methods. The reason was that it was the champion of the UEL, and it achieved many wins; therefore, it gained many points. However, these victories were against "weak" teams. Villareal hardly played any international matches against the strongest teams. In ESP, it played two matches against Real Madrid, was defeated once, and had one tie. Although Real Madrid was among the best teams according to all evaluations and was the winner of ESP, the position of Villareal was only fifth in ESP. That was why it reached the group stage of the UEL. Therefore, the position of Villareal seemed to have been overestimated in the UEFA club coefficient rankings. Unfortunately, point-based methods do not take into consideration the strength of the opponent, which may cause bias. The inconsistency of aggregation based on point-based methods is detailed in [55].

Table 7. UEFA club coefficient rankings.

Rank	Name	Score	Rank	Name	Score
1	Manchester City	35	18	Atl. Madrid	16
2	Chelsea	33	19	Tottenham	15
3	Villarreal	30	20	B. Monchengladbach	15
4	Bayern Munich	27	21	Granada CF	13
5	Real Madrid	26	22	AC Milan	12
6	Manchester Utd	26	23	Hoffenheim	12
7	Paris SG	24	24	Napoli	10
8	Liverpool	24	25	Bayer Leverkusen	10
9	AS Roma	24	26	Leicester	10
10	Arsenal	23	27	Inter	9
11	Dortmund	22	28	Real Sociedad	8
12	Juventus	21	29	Lille	8
13	Barcelona	20	30	Marseille	6
14	Sevilla	19	31	Rennes	5
15	RB Leipzig	17	32	Nice	3
16	Atalanta	17	33	Wolfsburg	2.5
17	Lazio	17	34	Reims	2

As the calculation of the UEFA club coefficients considered only the international matches and not the national ones, its ranking significantly differed from the rankings of the TH, AL, and RN methods. The Kendall's tau rank correlations between the UEFA coefficient rankings and the TH, AL, RN, and TM rankings applying W = 1, W = 4, and W = 16 weights for international matches are included in Table 8.

Table 8. Kendall's tau rank correlations between the different rankings.

	International W = 1			International W = 4			International W = 16			
	TH	AL	RN	TH	AL	RN	TH	AL	RN	TM
UEFA	0.405	0.474	0.321	0.434	0.471	0.307	0.522	0.471	0.339	0.540
TM	0.643	0.697	0.597	0.665	0.693	0.515	0.547	0.693	0.340	1

First, we can see that the correlation between the UEFA coefficient rankings and TM was 0.54, which was a moderate correlation. The correlations were moderate in the case of TH, too, with an increasing tendency for W. Certainly, the higher the value of W, the more significant the results of the international matches. In the case of AL, the correlations with the UEFA coefficient rankings were almost the same. The reason was that the elements of the AHP matrix (defined by Equation (8)) were not sensitive to the number of matches. The smallest correlations were between the results of RankNet and the UEFA club coefficient rankings. These correlations were weak.

We could establish that the correlations between TM and TH, AL, and RN were stronger than those between the UEFA coefficient rankings and TH, AL, and RN; they were mostly between 0.6 and 0.7. The smallest values appeared again between TM and RN.

Summarizing the observations, we could conclude that the best teams identified by TH and AL were the same as those identified by the UEFA club coefficient rankings. The only difference in the top teams (Villareal) could be clearly explained. It seemed to be the consequence of the point-based aggregation of the UCL and UEL, which is inappropriate in some cases. While TH and AL are evaluation methods that take into consideration the strength of the opponent, the UEFA club coefficient rankings do not. On the other hand, the TH and AL methods were based on a much larger set of information than that used for calculating the UEFA coefficient rankings. Not only international matches but also the results of the matches in the national leagues were taken into account when calculating the TH and AL rankings. The results of the national championships provided a more detailed picture of the performances of the teams playing in the international cups, and the evaluation results seemed to be more realistic.

5.5. Comparison of Different National Championships

In this section, the results of the above-presented evaluations for each team are aggregated for the ranking of the national championships. For this purpose, we used both the rankings and weights provided by the international evaluations. Table 9 contains the average of the estimated weights, presented in Tables 3 and 4. In Table 9, we can see that generally, England topped the list. In the case of the Thurstone method using W = 1 for the international matches, the German teams' average was slightly larger than the average weight of the English teams, but in the case of W = 4, the balance became positive for the English teams. This was in line with the fact that in the international cups, English teams performed better than German teams. The final of the UCL was played between two English teams, and there was an English team in the final of the UEL. On top of all this, the winner of the Super Cup was also an English team, Chelsea. According to the TH and AL methods, regarding the ranking of nations, second was GER, third was ESP, fourth was ITA, and last was FRA.

Table 9. Average weight values of the teams by nation.

	International W = 1			International W = 4		
	TH	**AL**	**RN**	**TH**	**AL**	**RN**
ENG	0.01182	0.01491	0.01281	0.01442	0.01490	0.01837
FRA	0.00756	0.00800	0.01063	0.00668	0.00799	0.01047
GER	0.01206	0.01007	0.00825	0.01143	0.01006	0.00824
ITA	0.00938	0.00813	0.01053	0.00837	0.00814	0.00692
ESP	0.01040	0.00989	0.00860	0.01024	0.00991	0.00682

The RN method ranked the French championship in second position. It was interesting to observe that in the international rankings, French teams were ranked ahead in the case of the RankNet method. See, for example, Lille, which was ranked second according to the RN method and eighteenth according to the TH and AL methods. Its international experience in cups did not support this position. Lille was 29th on the list of the UEFA club coefficient rankings. Similarly, we can mention Lyon and Monaco, which were not even in the best twenty teams according to the TH and AL methods and were not contained in the UEFA club coefficient rankings. The explanation for this phenomenon could be the following: there were only two international matches between Lille and teams from other nations investigated. The results of these two matches were good: one win and one tie for Lille. Although Lille played several matches with other teams, these teams were not in the investigated groups; therefore, these results were not contained in the present evaluation. Lille achieved only good results internationally against teams from ENG, ESP, GER, and ITA, with no negative results to drag them down in the rankings. This seemed to be decisive in the case of RN. Some other French teams (such as Lyon and Monaco) did not play any international matches, but they were linked to Lille in the standings through the national championship; therefore, the average weight became high. We did not consider these rankings to be realistic.

If we focus on the positions of the teams in the international rankings, the situation was somewhat similar and somewhat different. Ranking the best by a value of 1 and the worst by a value of 98, the average positions by nation are presented in Table 10. In the case of the Thurstone method using W = 1, the leader was Germany, followed by ENG, ESP, ITA, and FRA. In the case of the AL and RN methods, the best nation was England, but the rankings of the other teams differed. If we consider W = 4, according to TH and AL, the rankings were identical: ENG, GER, ESP, ITA, and FRA. The RankNet method identified England as the best, with France in second position. Moreover, the values of FRA, ITA, GER, and ESP were surprisingly close.

Table 10. Average ranking positions of the teams by nation.

	International W = 1			International W = 4		
	TH	**AL**	**RN**	**TH**	**AL**	**RN**
ENG	39.000	20.800	43.700	31.600	20.850	38.400
FRA	67.850	65.200	47.900	71.750	66.100	50.150
GER	36.389	46.167	50.778	38.722	46.278	52.556
ITA	55.500	65.800	51.600	58.150	65.000	52.250
ESP	47.450	49.200	53.650	46.200	48.950	54.450

Finally, we considered the best 20 teams in the international rankings. According to the Thurstone method using W = 4, we found eight English, four German, four Spanish, two Italian, and two French teams among them. According to the AL method, the top 20 teams included eleven English, four Spanish, two German, two French, and one Italian team. The RankNet method distributed the places uniformly: seven English, four German, three Spanish, three French, and three Italian teams were contained in the best 20 teams in the international rankings. Summarizing the observations, we could unequivocally declare that, concerning the performances, the English championship was the strongest. If we look at the financial ranking of the teams according to Transfermarkt data, the English teams were the most expensive, followed by ITA, GER, ESP, and FRA, respectively (see Table 11).

Table 11. Average values of the teams by nation.

	TM (€)
ENG	457.653 m
FRA	201.006 m
GER	266.406 m
ITA	272.250 m
ESP	261.403 m

Figure 3 shows an example for evaluating the correlation between the teams' real and estimated financial values. The horizontal axis presents the teams' values estimated by their performances. On the vertical axis, the TM values are shown. In this example, the evaluation was performed for the English teams, and the estimated values were computed by the AL method. For the calculations, only the match results in the national championship were taken into account. Figure 3 clearly demonstrates that in the case of poor performances, the teams were rather undervalued, and in the case of good performances, the teams were frequently overvalued in England. This phenomenon could also be observed in the case of other national championships.

Let us consider the financial values of the teams averaged by country. Figure 4 shows that the Transfermarkt data presented higher values than those computed on the basis of the performance weights produced by the TH and AL methods in the case of ENG and ITA teams. In the case of GER and ESP, the TH and AL methods estimated larger values than the Transfermarkt values. Conversely, the RankNet method overestimated the English and French teams while underestimating the other countries compared to the TM values. This phenomenon aligned with the observation that RN allocated multiple weights for the top teams and only fractional weights for the bottom teams. Overall, the average values computed by the TH and AL methods were closer to the average values of the Transfermarkt data than those computed by the RN method.

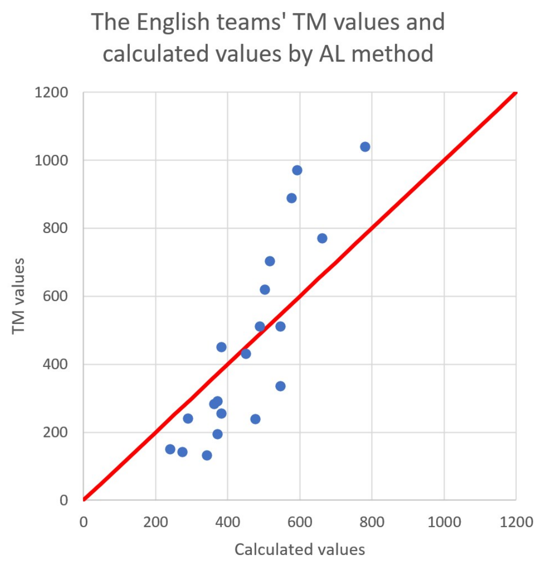

Figure 3. Correlation between estimated values and real TM values of the English teams.

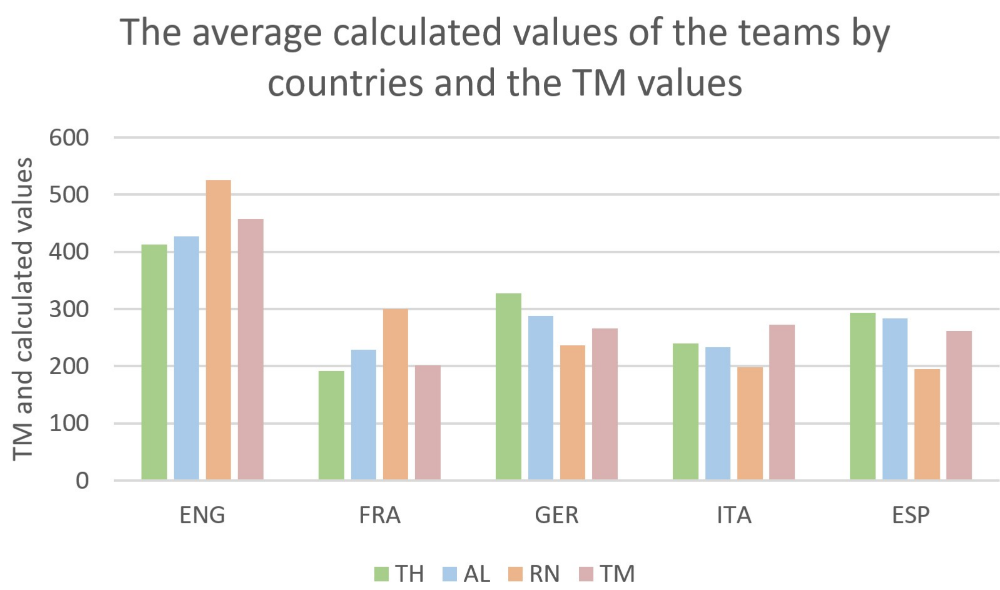

Figure 4. Estimated values and real average Transfermarkt (TM) values of the teams by nation.

Finally, for the sake of completeness, we also evaluated the UEFA country coefficients in the 2020/21 season, downloaded from the website [56]. According to this ranking, the best championship, based on international matches, in the 2020/21 season was ENG, the second was ESP, the third was ITA, the fourth was GER, and the last was FRA among the investigated championships. The first and last countries were the same as in our evaluations. Furthermore, FRA was definitely lower, being far behind GER in terms of points.

6. Comparison of the Different Methods

In this section, the characteristics of the three applied evaluation methods are compared.

All three methods (TH, AL, and RN) are based on optimization. In one case, the search for the maximum of a likelihood function, and in another case, a minimum search based on a least squares method solves the problem. The learning process of the RankNet neural network also requires optimization, which is generally carried out using variants of the stochastic gradient descent method. Each method can provide rankings and weights, i.e., in addition to the relative positions of the objects to each other, the difference in their strengths is also obtained. The TH method is an additive model, and AL is a multiplicative model. RN also works as an additive model. Due to the complexity of the functions to be maximized, the TH method requires numerical optimization, whereas the AL method does not. The RN method varies the scores representing the strength of the objects step-by-step through iterations. The scale of the weights resulting from the RankNet network is larger than the usual scale of the Thurstone method. This may be due to the fact that the last layer in the base network in RankNet has a linear activation function. As a result, the RankNet method attributed significantly larger weights to the top-ranked teams and extraordinarily small weights to the bottom-ranked teams.

A fundamental feature of the methods is that TH and AL treat the comparisons as a global system, which is not true for RN.

In the previous chapters, each of these methods demonstrated its excellent ranking ability in the case of national leagues. In the case of the rankings based on the results of national league matches, the comparison was complete, as each team played every other team. However, if teams from different leagues were aggregated into a single ranking, the comparison was not complete, as there were only a few links between teams from different nations. TH and AL performed adequately in this incomplete comparison case, but the RankNet method did not. RankNet seemed to require more connections among the national championships to provide an appropriate unified ranking. However, the result was not surprising. In the training phase, neural networks generally need a lot of training data to learn the optimal values for a large number of parameters.

Focusing on the running time, due to the step-wise training, RN required the most computational time. This was followed by the Thurstone method, due to the complexity of multivariate numerical optimization, and AL required the shortest computational time.

Comparing the evaluation results to the UEFA club coefficient rankings, we observed moderate correlations in the case of the TH and AL methods, similar to TM, but the RN method showed a weak correlation.

Regarding the applicability of the methods, the most complex set of conditions belonged to the Thurstone method, where only sufficient conditions have been proven so far, and not necessary and sufficient ones. Condition checking requires the investigation of the connectivity of a graph for which an effective algorithm exists. For the AL method, a necessary and sufficient condition for applicability is known, and this condition is also related to graph connectivity, but the condition itself is more often satisfied than the graph connectivity required for the Thurstone method. Although both the TH and AL methods were applicable in the analyzed case, mathematically, the AL method might also produce results in cases where TH cannot. The reason is that when constructing the AHP matrix, a match result is converted into a number, whereas in the case of the Thurstone method, the result is just a relation (worse/tie/better). In the case of the RankNet network, step-wise optimization is applicable in all cases, but a large amount of learning data is required to obtain a satisfactory result. In its absence, the final result may be biased, as in this paper.

Among the three methods, only the TH method was directly capable of providing probabilistic forecasts. Thus, it would be very useful for betting, for example, when deciding whether it is worth betting on the outcome of a match between teams. For this to be worthwhile, of course, it would require an objective function that already used other information (e.g., odds from bookmakers). Unfortunately, neither the AL nor RN method would be suitable for making predictions based on probabilities.

During our study, we established that the TH and RN methods were sensitive to the choice of the multiplier W, but the AL method was not. The reason was hidden in the construction of the elements of the pairwise comparison matrix $\mathbf{B}$, as we explained previously.

For all three methods, there were some characteristics whose choice was arbitrary. In the case of the TH method, the type of distribution was arbitrarily chosen as standard normal distribution. In the case of the RN method, the parameters (e.g., the number of hidden layers and the number of neurons placed in the hidden layers) and hyper-parameters (e.g., the learning rate for the optimization) were arbitrarily chosen characteristics. In the case of the AL method, the value of three corresponding to the outcome "better"was an arbitrary value.

7. Conclusions

This paper investigated the performance of different ranking methods in the case of complete and incomplete comparisons. The investigated methods were the Thurstone method with ties, the analytic hierarchy process with logarithmic least squares method, and the RankNet neural network. Since one of the possible application areas of the investigated methods is the ranking of football teams, the performance of the methods and their characteristics were evaluated by ranking the teams of the top five European football leagues and constructing a composite ranking of the teams playing in these leagues. In addition, we ranked the national championships for the five nations under study based on their estimated strengths and compared the results obtained by the different methods. A further part of the research was to investigate the weighting factor for international matches in the case of aggregated ranking based on incomplete comparisons. The methods were compared through the rankings and ratings resulting from the different methods. During the evaluations, both the positions of the teams in the rankings and the correlation of the rankings were considered.

Looking at the rankings of the teams and leagues, it was clear that in the 2020/2021 season, Manchester City was the best team in Europe. Regarding the national championships, the Premier League was the strongest among the five investigated championships (English, German, Italian, French, and Spanish).

As a result of our analysis, we established that in the case of national championships, all three methods performed equally well. In the case of rank aggregation based on incomplete comparisons, the RankNet network performed worse than the other two methods tested. The RankNet method preferred the French teams to the German, Italian, and Spanish teams. We concluded that the RankNet method did not provide a trustworthy international ranking due to the few connections between the championships. A further characteristic of the RankNet network was that it tended to overweight the best teams compared to the other two methods.

Furthermore, we found in our study that in the case of the aggregation of the rankings, the ranking method affected the ranking results more than the choice of weights for the linking comparisons. Larger weights caused noticeable modifications in the rankings in the case of the Thurstone method, but the result of the analytic hierarchy process with logarithmic least squares method was almost unaffected by the applied weights.

In addition, we also compared the methods' properties (e.g., running time and stochastic nature). We concluded that RankNet was the least recommended method of the three investigated methods for combining rankings based on sparse data. Its application problems arose mainly from the small amount of data available for ranking. As it is a neural network, it would need more data for aggregation.

This research had some obvious limitations. The results were instantaneous exposures, as they were based on the matches of a single (2020/2021) season and confined to the teams in the five investigated leagues. No temporal changes were taken into account.

It would be interesting to analyze the changes from one season to another. Further important aspects (such as home-field advantage and differences in the number of goals) could also be considered. These could all be the topic of further research.

The methods examined in this article have a wide range of applications. Ranking based on pairwise comparisons and the aggregation of rankings based on incomplete comparisons can be applied to the results of arbitrary sports competitions and championships, provided that there are some (in our study 3.84%) linking results between the participating competitors/teams. The extent to which the amount of comparisons affects the performance of the RankNet algorithm that produced the weakest results will be the subject of a subsequent study.

A practical application of this research could be the following: by extending the evaluations of TH and/or AL to the teams of other countries, conducting aggregation, and performing measurements, one could form clusters of countries. These clusters could be used to determine the number of spots for each UEFA competition for a given national league.

Author Contributions: Conceptualization, L.G. and É.O.-M.; methodology, L.G., Á.V.-F. and É.O.-M.; software, L.G.; validation, L.G., C.M. and Á.V.-F.; data curation, L.G.; writing, L.G., É.O.-M. and Á.V.-F.; review and editing, L.G., É.O.-M. and Á.V.-F.; visualization, L.G.; supervision, É.O.-M., C.M. and Á.V.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Only public data were used. They are available on the public websites listed in the References.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. González-Díaz, J.; Hendrickx, R.; Lohmann, E. Paired comparisons analysis: An axiomatic approach to ranking methods. *Soc. Choice Welf.* **2014**, *42*, 139–169. [CrossRef]
2. Chebotarev, P.Y. Aggregation of preferences by the generalized row sum method. *Math. Soc. Sci.* **1994**, *27*, 293–320. [CrossRef]
3. Elo, A.E. *The Rating of Chessplayers, Past and Present*; BT Batsford Limited: London, UK, 1978.
4. Giontella, A.; Sarti, F.M.; Biggio, G.P.; Giovannini, S.; Cherchi, R.; Silvestrelli, M.; Pieramati, C. Elo method and race traits: A new integrated system for sport horse genetic evaluation. *Animals* **2020**, *10*, 1145. [CrossRef] [PubMed]
5. FIFA. Revision of the FIFA / Coca Cola World Ranking. 2018. Available online: https://img.fifa.com/image/upload/edbm045h0udbwkqew35a.pdf (accessed on 9 November 2022).
6. Aldous, D. Elo ratings and the sports model: A neglected topic in applied probability? *Stat. Sci.* **2017**, *32*, 616–629. [CrossRef]
7. Lasek, J.; Gagolewski, M. Interpretable sports team rating models based on the gradient descent algorithm. *Int. J. Forecast.* **2021**, *37*, 1061–1071. [CrossRef]
8. Saaty, T.L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource, Allocation*; McGraw-Hill: New-York, NY, USA, 1980.
9. Saaty, T.L. Decision making—The analytic hierarchy and network processes (AHP/ANP). *J. Syst. Sci. Syst. Eng.* **2004**, *13*, 1–35. [CrossRef]
10. Qin, Y.; Hashim, S.R.M.; Sulaiman, J. An Interval AHP Technique for Classroom Teaching Quality Evaluation. *Educ. Sci.* **2022**, *12*, 736. [CrossRef]
11. Xu, W.; Zhao, S. The influence of entrepreneurs' psychological capital on their deviant innovation behavior. *Front. Psychol.* **2020**, *11*, 1606. [CrossRef]
12. Hwang, J.W.; Kim, J.Y. A study on industrial security psychology. In Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; pp. 500–506.
13. Improta, G.; Converso, G.; Murino, T.; Gallo, M.; Perrone, A.; Romano, M. Analytic hierarchy process (AHP) in dynamic configuration as a tool for health technology assessment (HTA): The case of biosensing optoelectronics in oncology. *Int. J. Inf. Technol. Decis. Mak.* **2019**, *18*, 1533–1550. [CrossRef]
14. Canco, I.; Kruja, D.; Iancu, T. AHP, a Reliable Method for Quality Decision Making: A Case Study in Business. *Sustainability* **2021**, *13*, 13932. [CrossRef]
15. Katarina, D.; Nurrohman, A.; Putra, A.S.; Wibisono; Priatno. Decision Support System for the Best Student Selection Recommendation Using Ahp (Analytic Hierarchy Process) Method. *Int. J. Educ. Res. Soc. Sci.* **2021**, *2*, 1210–1217.
16. Csató, L. Ranking by pairwise comparisons for Swiss-system tournaments. *Cent. Eur. J. Oper. Res.* **2013**, *21*, 783–803. [CrossRef]
17. Bozóki, S.; Csató, L.; Temesi, J. An application of incomplete pairwise comparison matrices for ranking top tennis players. *Eur. J. Oper. Res.* **2016**, *248*, 211–218. [CrossRef]
18. Anderson, A. A Monte Carlo comparison of alternative methods of maximum likelihood ranking in racing sports. *J. Appl. Stat.* **2015**, *42*, 1740–1756. [CrossRef]
19. Orbán-Mihálykó, É.; Mihálykó, C.; Gyarmati, L. Application of the Generalized Thurstone Method for Evaluations of Sports Tournaments' Results. *Knowledge* **2022**, *2*, 157–166. [CrossRef]
20. Csató, L. On the ranking of a Swiss system chess team tournament. *Ann. Oper. Res.* **2017**, *254*, 17–36. [CrossRef]
21. Urbaniak, K.; Wątróbski, J.; Sałabun, W. Identification of players ranking in e-sport. *Appl. Sci.* **2020**, *10*, 6768. [CrossRef]
22. Kiani Mavi, R.; Kiani Mavi, N.; Kiani, L. Ranking football teams with AHP and TOPSIS methods. *Int. J. Decis. Sci. Risk Manag.* **2012**, *4*, 108–126.
23. Thurstone, L. A law of comparative judgment. *Psychol. Rev.* **1927**, *34*, 273–286. [CrossRef]

24. Bradley, R.A.; Terry, M.E. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* **1952**, *39*, 324–345. [CrossRef]
25. Stern, H. A continuum of paired comparisons models. *Biometrika* **1990**, *77*, 265–273. [CrossRef]
26. Rao, P.; Kupper, L.L. Ties in paired-comparison experiments: A generalization of the Bradley–Terry model. *J. Am. Stat. Assoc.* **1967**, *62*, 194–204. [CrossRef]
27. Orbán-Mihálykó, É.; Mihálykó, Cs.; Koltay, L. A generalization of the Thurstone method for multiple choice and incomplete paired comparisons. *Cent. Eur. J. Oper. Res.* **2019**, *27*, 133–159. [CrossRef]
28. Orbán-Mihálykó, É.; Mihálykó, Cs.; Koltay, L. Incomplete paired comparisons in case of multiple choice and general log-concave probability density functions. *Cent. Eur. J. Oper. Res.* **2019**, *27*, 515–532. [CrossRef]
29. Shen, Y.; Deng, Y.; Yang, M.; Li, Y.; Du, N.; Fan, W.; Lei, K. Knowledge-aware attentive neural network for ranking question answer pairs. In Proceedings of the The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 901–904.
30. Guo, J.; Fan, Y.; Pang, L.; Yang, L.; Ai, Q.; Zamani, H.; Wu, C.; Croft, W.B.; Cheng, X. A deep look into neural ranking models for information retrieval. *Inf. Process. Manag.* **2020**, *57*, 102067. [CrossRef]
31. Li, Y.; Song, Y.; Luo, J. Improving pairwise ranking for multi-label image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3617–3625.
32. Song, B.; Yang, X.; Cao, Y.; Xu, C. Neural collaborative ranking. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Turin, Italy, 22–26 October 2018; pp. 1353–1362.
33. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
34. Burges, C.; Shaked, T.; Renshaw, E.; Lazier, A.; Deeds, M.; Hamilton, N.; Hullender, G. Learning to rank using gradient descent. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 89–96.
35. FIFA. Transfermarkt. 2021. Available online: https://www.transfermarkt.com (accessed on 28 February 2023).
36. Koopman, S.J.; Lit, R. A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *J. R. Stat. Soc. Ser. A (Statistics Soc.)* **2015**, *178*, 167–186. [CrossRef]
37. Koopman, S.J.; Lit, R. Forecasting football match results in national league competitions using score-driven time series models. *Int. J. Forecast.* **2019**, *35*, 797–809. [CrossRef]
38. Foulley, J.L. A simple Bayesian procedure for forecasting the outcomes of the UEFA Champions League matches. *J. Soc. Fr. Stat.* **2015**, *156*, 38–50.
39. Constantinou, A.C.; Fenton, N.E. Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models. *J. Quant. Anal. Sport.* **2012**, *8*, 1–14. [CrossRef]
40. Ley, C.; Wiele, T.V.d.; Eetvelde, H.V. Ranking soccer teams on the basis of their current strength: A comparison of maximum likelihood approaches. *Stat. Model.* **2019**, *19*, 55–73. [CrossRef]
41. Sałabun, W.; Shekhovtsov, A.; Pamučar, D.; Wątróbski, J.; Kizielewicz, B.; Więckowski, J.; Bozanić, D.; Urbaniak, K.; Nyczaj, B. A fuzzy inference system for players evaluation in multi-player sports: The football study case. *Symmetry* **2020**, *12*, 2029. [CrossRef]
42. Guan, S.; Wang, X. Optimization analysis of football match prediction model based on neural network. *Neural Comput. Appl.* **2022**, *34*, 2525–2541. [CrossRef]
43. Baboota, R.; Kaur, H. Predictive analysis and modelling football results using machine learning approach for English Premier League. *Int. J. Forecast.* **2019**, *35*, 741–755. [CrossRef]
44. Manack, H.; Van Zyl, T.L. Deep similarity learning for soccer team ranking. In Proceedings of the 2020 IEEE 23rd International Conference on Information Fusion (FUSION), Rustenburg, South Africa, 6–9 July 2020; pp. 1–7.
45. Behravan, I.; Razavi, S.M. A novel machine learning method for estimating football players' value in the transfer market. *Soft Comput.* **2021**, *25*, 2499–2511. [CrossRef]
46. Stübinger, J.; Mangold, B.; Knoll, J. Machine learning in football betting: Prediction of match results based on player characteristics. *Appl. Sci.* **2019**, *10*, 46. [CrossRef]
47. Groll, A.; Ley, C.; Schauberger, G.; Van Eetvelde, H. A hybrid random forest to predict soccer matches in international tournaments. *J. Quant. Anal. Sport.* **2019**, *15*, 271–287. [CrossRef]
48. Held, L.; Vollnhals, R. Dynamic rating of European football teams. *IMA J. Manag. Math.* **2005**, *16*, 121–130. [CrossRef]
49. Eliason, S.R. *Maximum Likelihood Estimation: Logic and Practice*; Number 96; Sage: Thausand Oaks, CA, USA, 1993.
50. Bozóki, S.; Tsyganok, V. The (logarithmic) least squares optimality of the arithmetic (geometric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. *Int. J. Gen. Syst.* **2019**, *48*, 362–381. [CrossRef]
51. Bozóki, S.; Fülöp, J.; Rónyai, L. On optimal completion of incomplete pairwise comparison matrices. *Math. Comput. Model.* **2010**, *52*, 318–333. [CrossRef]
52. Data. 2021. Available online: https://allsportsapi.com/ (accessed on 28 Februray 2023).
53. UEFA Club Coefficients. 2020–2021. Available online: https://www.uefa.com/nationalassociations/uefarankings/club/#/yr/2021 (accessed on 28 February 2023).

54. Method of Caluculation of UEFA Club Coefficients. Available online: https://www.uefa.com/nationalassociations/uefarankings/country/about/ (accessed on 28 February 2023).
55. Orbán-Mihálykó, É.; Mihálykó, C.; Gyarmati, L. Evaluating the capacity of paired comparison methods to aggregate rankings of separate groups. *Cent. Eur. J. Oper. Res.* **2023**, 1–21. [CrossRef] [PubMed]
56. UEFA Country Coefficients. Available online: /www.uefa.com/nationalassociations/uefarankings/country/#/yr/2021 (accessed on 28 February 2023).