AÇIK ERİŞİM
OPEN-ACCESS

Gazi Beden Eğitimi ve Spor Bilimleri Dergisi
Gazi Journal of Physical Education and Sports Sciences
2024, 29(3), 202-209

Serving Up Success: Unveiling the Power of Machine Learning for Volleyball League Prediction

Başarıya Hizmet Etmek: Voleybol Ligi Tahmini için Makine Öğreniminin Gücünün Ortaya Çıkarılması

Research Article / Araştırma Makalesi

Emre ALTUNDAG [1]
Hasan AKA [2]
Çağlar SOYLU [3]
Pervin DEMIR [4]

[1] Faculty of Sport Sciences, Kütahya Dumlupınar University, Kütahya, TÜRKİYE

[2] Faculty of Sport Sciences, Coaching Education Department, Niğde Ömer Halisdemir University, Nigde, TÜRKİYE

[3] Gülhane Faculty of Physical Therapy and Rehabilitation, University of Health Sciences, Ankara, TÜRKİYE

[4] Faculty of Medicine, Basic Medical Sciences, Biostatistics and Medical Informatics, Ankara Yıldırım Beyazıt University, Ankara, TÜRKİYE

Corresponding Author/ Sorumlu Yazar
Asst. Prof. Çağlar SOYLU
caglar.soylu@sbu.edu.tr

Received / Geliş Tarihi : 04.05.2024
Accepted / Kabul Tarihi : 31.07.2024
Published / Yayın Tarihi : 31.07.2024

Ethical Statement / Etik Bilgilendirme
This study was ethically approved by the decision of University, Social and Human Sciences Scientific Research and Publication Ethics Committee dated 27.07.2023 and numbered 214071.

DOI: 10.53434/gbesbd.1478533

Abstract

This study investigates the efficacy of Artificial Neural Networks (ANN) in predicting volleyball league standings, focusing on the Turkish Volleyball Federation's Sultanlar and Efeler leagues over five seasons (2018-19 to 2022-23). Given the complexity and volume of performance data in volleyball, traditional analysis methods often face challenges such as data overload and high operational costs. ANN models, known for their ability to learn from and generalize data, present a promising solution to these challenges. By analyzing 23 input variables related to match performance, including points scored, services, attacks, and blocks, this study aims to identify the most influential factors on final league standings and provide a more objective, rapid, and economical analysis method. The results indicate significant potential for ANN in sports analytics, demonstrating high accuracy rates in predictions, especially for the Sultanlar League. However, the study also acknowledges limitations such as data quality and model complexity, suggesting areas for future research to enhance predictive accuracy and applicability of ANN in volleyball and other sports analytics.

Keywords: Artificial Neural Networks, Volleyball League Standings, Machine Learning, Sports Analytics, Performance Prediction

Öz

Bu çalışma, Türkiye Voleybol Federasyonu'nun Sultanlar ve Efeler liglerine odaklanarak, beş sezon boyunca (2018-19- 2022-23) voleybol lig sıralamalarını tahmin etmede Yapay Sinir Ağlarının (YSA) etkinliğini araştırmaktadır. Voleybolda performans verilerinin karmaşıklığı ve büyüklüğü göz önüne alındığında, geleneksel analiz yöntemleri genellikle aşırı veri yükü ve yüksek operasyonel maliyetler gibi zorluklarla karşılaşmaktadır. Verilerden öğrenme ve genelleme yetenekleriyle bilinen YSA modelleri, bu zorluklara umut verici bir çözüm sunmaktadır. Bu çalışma, atılan sayılar, servisler, ataklar ve bloklar dahil olmak üzere maç performansıyla ilgili 23 girdi değişkenini analiz ederek, nihai lig sıralaması üzerinde en etkili faktörleri belirlemeyi ve daha objektif, hızlı ve ekonomik bir analiz yöntemi sağlamayı amaçlamaktadır. Sonuçlar, özellikle Sultanlar Ligi için tahminlerde yüksek doğruluk oranları göstererek spor analitiğinde YSA için önemli bir potansiyel olduğunu göstermektedir. Bununla birlikte, çalışma aynı zamanda veri kalitesi ve model karmaşıklığı gibi sınırlamaları da kabul etmekte ve YSA'nın voleybol ve diğer spor analitiklerinde tahmin doğruluğunu ve uygulanabilirliğini artırmak için gelecekteki araştırmalar için alanlar önermektedir.

Anahtar Kelimeler: Yapay Sinir Ağları, Voleybol Lig Sıralamaları, Makine Öğrenimi, Spor Analitiği, Performans Tahmini

Citation / Atıf: Gazi, Ü. (2024). Altundağ, E., Aka, H., Soylu, Ç. and Demir, P. Serving Up Success: Unveiling the Power of Machine Learning for Volleyball League Prediction. _Gazi Journal of Physical Education and Sports Sciences, 29_(3), 202-209.

Introduction

As in all sport branches, the recent increase in the use of digital technologies for different purposes such as statistics and analysis depending on the improvement in the components special to volleyball is remarkable. The purpose of the use of digital technologies in volleyball is mainly to analyze the matches by an expert through analysis programs (Aka, Akarçeşme, Aktuğ, Özden, 2021a; Aka, Aktuğ, Kılıç, 2021b). According to the results of a study evaluating the efficiency of analysis programs, trainers find these programs useful (Aka et al., 2021a; Aka et al., 2021b; Akarçeşme, Aka, Özden, Aktuğ, 2020). These analysis programs, developed special to volleyball, enable trainers to reach a vast amount of data about athletes. These data help trainers in a lot of issues such as planning the training process, techniques and tactics that influence the team performance directly (Bai & Bai, 2021). The fact that the data obtained from the analysis programs used for volleyball are quite many may be evaluated as a restrictive factor for data analysis, accurate inference and the decision processes of trainers. This requires a high cost since expensive analysis programs, a long period of time and labor force are needed for the teams playing at least one match per week.

Machine learning is a method that estimates according to the inferences from the present data cluster by using the sciences of statistics and mathematics and that is named as artificial intelligence. This method is a field that studies for computers to ensure learning structure, as in the learning process of human, and aims at developing algorithms and methods in this respect (Beck, 2018). The methods used under machine learning can be classified under two main titles as Supervised and Unsupervised learning (Beck, 2018). Together with these learnings, Semi-Supervised Learning, Awarded Learning, Genetic algorithms and Artificial Neural Network methods also enable machine learning. Artificial Neural Networks (ANN) can be defined as an information operating system working similarly to biological neural systems and performing skills such as new knowledge generation and exploration through learning (Cossich, Carlgren, Holash, Katz, 2023).

Since ANN can learn simple structures special to problems, generalize, perform parallel processing and create solutions for complicated problems that are difficult to be modeled and nonlinear through its skills such as tolerating errors, it has found a wide range of execution area in modeling and supervision of complicated systems (Cortsen & Rascher, 2018). One of the execution areas is sports. ANN is used for different purposes such as data analysis and evaluation special to sport branches and performance tracking (Gorriz, Alvarez- Illan, Alvarez- Marquina., 2023).

It is believed that inferences to be obtained by an analysis performed through developed ANN models may be more economical since they will be more objective and rapid. Thus, ANN can provide quite significant inferences for trainers in evaluating the performances of athletes and team in terms of technic and tactics. The recent use of ANN models for different purposes in team sports having a vast number of data clusters and athlete numbers is remarkable (Joao, Vaz, Mota, 2019). It is seen in the literature that the evaluations for volleyball through ANN model are limited (Fernandez-Echeverria, Mesquita, Gonzalez-Silva, Ehcerria., 2017; Komar, Egrıoglu, Semiz, 2023). In addition, no study on the prediction of league standing by analyzing both women and men volleyball leagues through machine learning methods was found in the literature. Given the burgeoning use of digital technologies for statistics and match analysis in sports, including volleyball, and the potential limitations associated with traditional analysis tools, this study proposes a novel approach leveraging ANN for predicting league standings. Despite the invaluable insights provided by existing volleyball-specific analysis programs, challenges such as data overload, high costs, and the extensive time and labor required for effective use can impede the optimization of training and tactical planning. Furthermore, the literature review underscores a gap in utilizing machine learning methods, particularly ANN, for comprehensive analysis across both women's and men's volleyball leagues.

It is believed that inferences to be obtained by an analysis performed through developed ANN models may be more economical since they will be more objective and rapid. Thus, ANN can provide quite significant inferences for trainers in evaluating the performances of athletes and team in terms of technic and tactics. The recent use of ANN models for different purposes in team sports having a vast number of data clusters and athlete numbers is remarkable (Aka et, al. 2021; Aktuğ et al., 2022). It is seen in the literature that the evaluations for volleyball through ANN model are limited (Kautz et al, 2017, Koch and Tilp, 2009, Jörg et al. 2017, Tümer and Koçer, 2017; Akarçeşme et al., 2021; Aka et al., 2020). In addition, it is noteworthy that there are limited number of studies conducted on predicting the league ranking by means of input variables specific to volleyball competitions (Tümer and Koçer, 2017; Akarçeşme et al., 2021). On the other hand, in the current literature there are no studies on the prediction of league available by analyzing both women and men volleyball leagues through machine learning methods. In this respect, the aim of this study is to predict the league, which is the output variable, through 23 input variables related to the Played Matches, Points Scored, numbers of Service Performed, Service Received, Attack and Block in the Turkish Volleyball Federation (TVF) Sultanlar League and Efeler League matches over the past 5 seasons (2018-19, 2019-20, 2020-21, 2021-22, 2022-23) by using developed Artificial Neural Network (ANN) models.

Method

Data Collection

A total of 23 variables were determined as input variables. These were the main variables of the number of Matches (played matches and sets), Scored Points (Total, Service break, total point on reception, Win-Lost), Serving (Total, Ace, Error, Ace per set, Efficiency rate), Reception (Total, Error, Negative, Perfect, Perfect%, Efficiency), Attack (Total, Error, Blocked balls, Perfect (point), Perfect%, Efficiency rate) and Block (Net Contact, Block point, point per set) and the related sub-variables. League standing was determined as the output variable. The data used in the study were obtained from TVF official website, which is open access.

Table 1. *Descriptive statistics for variables*

	Sultanlar		Efeler			Sultanlar		Efeler	
	Median	Mean±SD	Median	Mean±SD		Median	Mean±SD	Median	Mean±SD
Match	29	28±4	29	28±4	rec_exc	554	555±112	528	511±102
Set	107	103±16	109	106±18	rec_exc%	0	0±0	0	0±0
t_pt	1681	1655±340	1729	1677±347	rec_eff	0	0±0	0	0±0
sk_pt	699	701±177	605	608±152	att_tot	3225	3155±521	2706	2633±457
k_pt	982	954±183	1093	1078±199	att_err	245	242±38	208	209±37
w_l	756	774±291	678	690±233	att_blc	238	234±44	237	239±45
ser_tot	2303	2227±403	2448	2342±450	att_exc	1280	1270±264	1323	1306±257
ser_ace	148	149±39	138	139±38	att_exc%	0	0±0	0	0±0
ser_err	237	253±48	414	410±79	att_eff	0	0±0	0	0±0
Ser_aceps	1	1±0	1	1±0	blc_nt	304	301±129	52	77±78
ser_eff	0	0±0	0	0±0	blc_pt	241	237±59	239	242±58
rec_tot	2010	1950±295	1982	1929±301	blc_pps	2	2±0	2	2±0
rec_err	150	154±38	143	139±26	l_ran	7	7±4	7	7±4
rec_neg	577	580±108	590	580±107					

Table 2. *Variable names*

Played	M	Match	Match
	S	Set	Set
Scored Points	M	Match	Match
	S	Set	Set
	Tot.	Total Points	t_pt
	SK	Break Points	sk_pt
Services	K	Total Points on Reception	k_pt
	Win-Lost	Win-Lost	w_l
	Tot.	Total	ser_tot
	Ace	Ace	ser_ace
	Err.	Error	ser_err
Reception	A.P.S.	Ace per set	ser_aceps
	Eff.	Efficiency	ser_eff
	Tot.	Total	rec_tot
	Err.	Error	rec_err
	Neg.	Negative	rec_neg
	Exc.	Excellent	rec_exc
Attack	Exc. %	Excellent %	rec_exc%
	Ver	Efficiency	rec_eff
	Tot.	Total	att_tot
	Err.	Error	att_err
	Pt	Blocked	att_blc
	Exc..	Excellent	att_exc
BLC	Exc.%	Excellent %	att_exc%
	Eff	Efficiency	att_eff
	Net	Net contact	blc_nt
	Pt	Points	blc_pt

Statistical Analysis

The Matches (played matches and sets), Scored Points (Total, Service break, total point on reception, Win-Lost), Serving (Total, Ace, Error, Ace per set, Efficiency rate), Reception (Total, Error, Negative, Perfect, Perfect%, Efficiency), Attack (Total, Error, Blocked balls, Perfect (point), Perfect%, Efficiency rate) and Block (Net Contact, Block point, point per set) over the past 5 seasons (2018-19, 2019-20, 2020-21, 2021-22, 2022-23) by TVF Sultanlar League and Efeler League and the sub-variables related to these main variables were determined as input variables and the league standing as output variable.

Before ANN model was generated, it was analyzed whether there was multicollinearity between the independent variables via correlation coefficient. The fact that the coefficient was 0.90 and above (in other words, the fact that the variance inflation factor was above 10) revealed that there was a strong correlation (VIF exceeding 10 or corr coefficient higher than 0.90 indicates high multicollinearity between independent variables). Since particularly the total scores had a high correlation with most of the other variables, the total scores in question (match, set, t_pt) were not included in the model. Consequently, the feedforward regression model that best estimated (with the lowest error rates: RMSE, MAD and the highest accuracy values) the league standing variable through sigmoid activation function (preferred due to its ability to output values within the 0-1 range) was determined from the input layer that was generated by using the total 23 input variables and the output variable. The training set included the results of the first 4 seasons and the test set includes that of the last season. Values were defined in the range of 0-1 because algorithms in neural networks perform better with standardized data. In the first step, 3 different ANN models were generated as one layered and in different neuron numbers [sultanlar: 23*1-8*1-12*1 (SL Model 3); efeler: 23*1-7*1-10*1(EL Model 3)} and they were compared with each other. To evaluate model performance, a Welch two-sample t-test was used to check for statistically significant differences between the models' RMSE (root mean squared error) and MAD (median absolute deviation) metrics. These metrics were derived from k-fold cross-validation, where each fold's predictions were compared to actual values. The t-test then assessed whether the observed differences in performance across models were significant, ensuring a robust evaluation across various data subsets. K-fold cross-validation was used to ensure that the model's accuracy and generalizability were assessed across different subsets of the data. Finally, the model having the lowest error, and the highest rate of accuracy was reported. Relative significance rates were obtained for each input variable through Garson's algorithm and their orders of significance on the output variable were determined. Statistical analyses were performed by using "corrplot", "neuralnet", "NeuralNetTools" and "ggplot2" packages in open-source coded R programming language (Bai & Bai,2021), (Kufel et., 2023), (Millington & Millington, 2015).

Ethical Statement

This study was ethically approved by the decision of University, Social and Human Sciences Scientific Research and Publication Ethics Committee dated 27.07.2023 and numbered 214071.

Results

Firstly, a model was generated through 23 neurons and one hidden layered structure (23*1) including all variables for both Sultanlar and Efeler Leagues. Then, the second one-layered model was created with the variables having the variable significance level of 5% and above that were determined in the first model, and the third one-layered model was created from the variables that were determined to have a correlation of above 0.50 with the league standing. The obtained models and the results related to them were given in Table 1. It was observed that the model results created with the variables having a correlation of above 50% with the league standing had the best performance criteria for both leagues.

Table 3. The performance results of ANN models

LEAGUE	Model	Train MSE	Test MSE	Accuracy (%)	RMSE ($\sqrt{MSE}$)	MAD
Sultanlar	**23*1**	0.012	2.956	93.54	1.719 [a]	1.110 [a]
	8*1	0.514	1.761	91.69	1.327 [b]	1.514 [b]
	12*1	**0.347**	**2.405**	**92.27**	**1.551** [b]	**0.751** [b]
Efeler	**23*1**	0.078	13.158	36.89	3.627 [a]	3.270 [a]
	7*1	1.254	n/a	n/a	n/a	4.131 [b]
	10*1	**0.209**	**6.603**	**79.12**	**2.570** [b]	**3.070** [c]

MSE: Mean square error, RMSE: root mean square error, MAD: median absolute deviation, n/a: non-applicable
[a,b,c] The models, a significant difference was determined of which RMSE or MAD statistical values as a result of the paired comparison performed through Welch two sample t-test were indicated with different letters.
Variables were presented in significance order.
Sultanlar league models:

Sultanlar Ligi Model 1 (SL Model 1):23*1:l_ran~0.066*rec_err+ 0.057*rec_exc%+ 0.054*ser_err+ 0.053*att_exc%+ 0.046*rec_neg+ 0.046*w_l+ 0.046*blc_nt+ 0.046*rec_exc+ 0.044*att_blc+ 0.044*att_tot+ 0.044*ser_tot+0.043*blc_pps+ 0.042*att_err+0.042*rec_exc+ 0.039*att_eff+ 0.038*ser_ace+ 0.038*ser_aps+ 0.037*k_pt+0.037*att_exc+ 0.036*blc_pt+0.035*ser_eff+ 0.033*rec_tot+ 0.033*sk_pt.

Sultanlar Ligi Model 2 (SL Model 2):8*1: l_ran~0.160*blc_nt+ 0.146*ser_err+ 0.143*w_l+ 0.128*rec_exc+ 0.125*rec_neg+ 0.119*rec_exc%+0.105*rec_err+ 0.074*att_exc%.

Sultanlar Ligi Model 3 (SL Model 3)12*1:l_ran~ 0.119*att_exc%+0.115*ser_ace+0.099*rec_err+0.096*ser_err+0.082*blc_pt+ 0.082*ser_ace + 0.078*w_l+ 0.073*sk_pt+ 0.072*ser_tot+ 0.067*att_exc+ 0.063*att_eff+0.054*blc_pps.

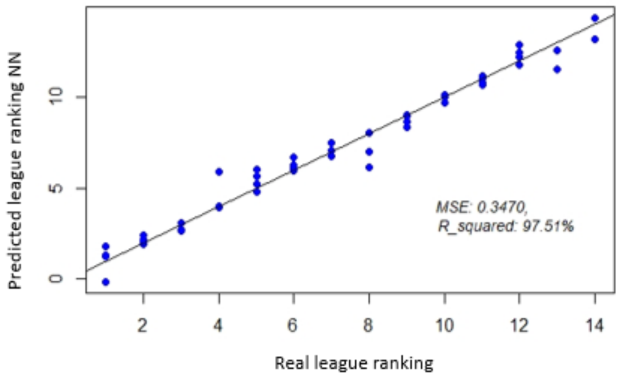
Efeler league models:

Efeler Ligi Model 1 (SL Model 1): 23*1: l_ran~ 0.074*blc_nt+ 0.065*rec_eff+ 0.058*rec_exc%+ 0.053*ser_ace+ 0.052*ser_err+0.051*ser_neg+ 0.046*ser_aps+ 0.045*att_blc+ 0.045*att_eff+ 0.044*w_l+ 0.044*rec_err+ 0.042*blc_pt+ 0.041*k_pt+0.038*att_exc%+ 0.038*ser_tot+ 0.037*att_tot+ 0.036*ser_eff+ 0.035*rec_exc+ 0.035*sk_pt+0.034* blc_pps+ 0.033*att_err+ 0.029*att_exc+ 0.027*rec_tot.

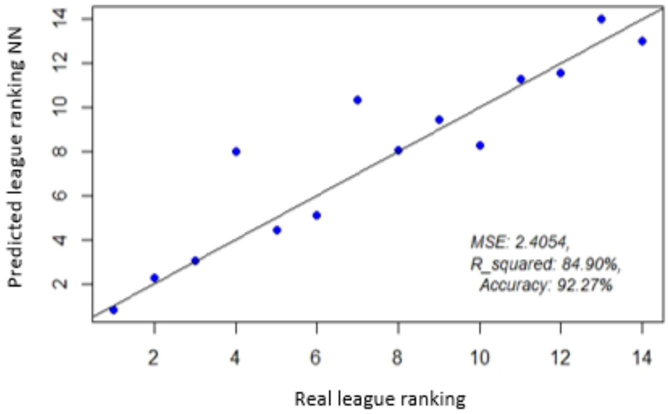
Efeler Ligi Model 2 (SL Model 2):7*1: l_ran~ 0.191*rec_exc%+ 0.191*ser_ace+0.143*ser_err+ 0.141*blc_nt+ 0.133*rec_neg+0.105rec_eff+0.096*ser_aps.

Efeler Ligi Model 3 (SL Model 3):10*1: l_ran~ 0.142*att_eff+0.125*att_exc%+ 0.111*sk_pt+ 0.110*blc_pt+0.107*ser_ace+0.105*w_l+ 0.102*ser_aps+0.007*ser_tot+0.066*ser_eff+ 0.062*att_exc.

While a significant difference was determined between the first model and the other two models in terms of RMSE and MAD values in Sultanlar League standing (p<0.001), second and third model results were similar (for RMSE p=0.659 and for MAD p=0.400). The results of the SL Model 3 were reported to have higher accuracy rates compared to the SL Model 2. The coefficient of determination of the model training set consisting of SL Model 3 variables was determined as 97.51%. The RMSE value of 1.0803, obtained through cross-validation, indicates that, on average, the model's predictions deviate from the actual values by approximately 1.0803 units. The coefficient of determination determined for the test set was 84.90% and the accuracy rate was 92.27% (Figure 1a-b). The first 3 variables having the highest effect on the league standing in Sultanlar League were respectively att_exc%, set_aps and rec_err.



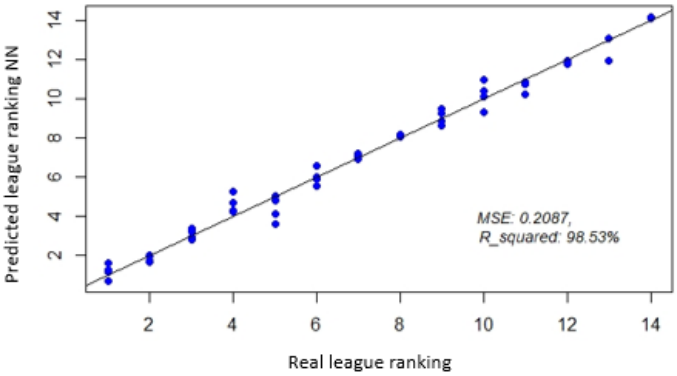
(a) Training set



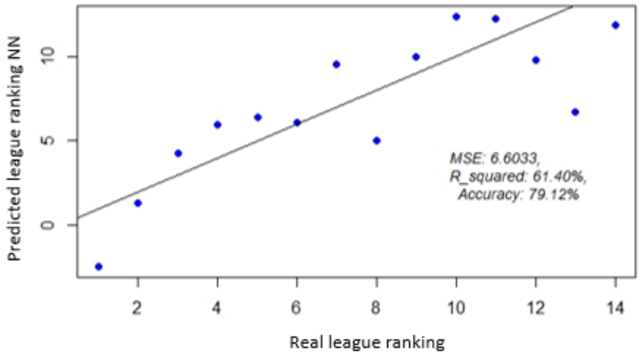
(b) Test set

Figure 1. The Training Set and Test Set results of the 12*1 model (Sultanlar League)

A significant difference was determined between the three models generated over the league standing of Efeler League in terms of RMSE and MAD values (p<0.05). SL Model 3 results having higher accuracy rates than others were reported. The coefficient of determination of the model training set consisting of SL Model 3 variables were determined as 98.53%. The RMSE value of 2.0075, obtained through cross-validation, indicates that, on average, the model's predictions deviate from the actual values by approximately 2.0075 units. The coefficient of determination determined for the test set was 61.40% and the accuracy rate was 79.12% (Figure 2a-b). The first 3 variables having the highest effect on the league standing in Efeler League were respectively Att_eff, att_exc% and sk_pt.



(a) Training set



(b) Test set

Figure 2. The Training and Test Set results of 10*1 model (Efeler League)

Sultanlar and Efeler League 2022/2023 league standing, and the test set prediction results obtained as a result of the models generated from the suggested variables are presented in Table 3. Considering the accuracy rates of the models, it is seen that the predictions determined in Sultanlar League gives closer results to the real league standing (p-value=0.596). It is observed that there are differences especially between Efeler League final league standing and the predicted league standing, but this difference was not statistically significant (p=0.812). It is seen that an adequate prediction has not been made with the variables used for the prediction of Efeler League season standing.

Table 4. Sultanlar and Efeler 2022-2023 league standing and model predictions

Sultanlar League			Efeler League		
Team	2022/2023 League Standing	12*1 Model Prediction	Team	2022/2023 League Standing	10*1 Model Prediction
Eczacıbaşı	1	1	Halkbank	1	-2
Vakıfbank	2	2	Ziraat Bankası	2	1
Fenerbahçe	3	3	Fenerbahçe	3	4
Türk Hava Yolları	4	8	Arkas Spor	4	6
Galatasaray	5	4	Bursa B.Şehir Bld.	5	6
Nilüfer Bld.	6	5	Galatasaray	6	6
Sarıyer Bld.	7	10	Türşad	7	10
Aydın B.Şehir Bld.	8	8	Cizre Bld.	8	5
Çukurova Bld. Adana Demirspor	9	9	Spor Toto	9	10
Kuzeyboru	10	8	Hekimoğlu Global Connect Travel Bvi	10	12
Ptt	11	11	Develi Bld.	11	12
Sigorta Shop	12	12	Altekma	12	10
Bolu Bld.	13	14	Hatay B.Şehir Bld.	13	7
İlbank	14	13	Tokat Belediye Plevne	14	12

Discussion

The meticulous records of athletes' performances amassed during matches yield a vast dataset, and technological advancements have streamlined the analysis of these data sets (Palao & Hernández-Hernández, 2014). Such analyses are pivotal for the enhancement of athletes' performance (Taye, 2023). Presently, technological innovations serve varied purposes in sports clubs, including the enhancement of athletes' performance, analysis of opposing teams, and tactical strategies for securing victories (Tümer &Koçer, 2017). Owing to these advancements and the significant volume of sports-specific data accrued, the significance of data mining in the sports arena has seen an uptick (Yang, 2021). Despite limitations in data mining, volleyball has also witnessed the application of algorithms for technical and tactical analyses (Wei & Simko, 2021). This study aimed to identify the most influential variables on the final standings of the TVF Efeler and Sultanlar leagues by conducting a machine learning analysis of their data, and to predict the league standings accordingly.

In the Sultanlar League, the refinement of variables to those most significantly correlated with league standings (above 50%) resulted in the most efficacious performance criteria. This is evidenced by the notable accuracy rates, with the SL Model 3 demonstrating a compelling balance of explanatoriness and accuracy (92.27%), underlining the potential of ANN in sports analytics. The critical variables identified—attack excellence percentage (att_exc%), ace per set (ser_aceps), and reception error (rec_err) —underscore the multifaceted nature of performance in volleyball, highlighting the importance of both offensive and defensive play in determining league standings. The findings for the Efeler League further emphasize the complexity of predictive modeling in sports, where the SL Model 3 model, despite a lower accuracy rate (79.12%) than its Sultanlar counterpart, sheds light on the crucial aspects of volleyball performance, such as attack efficiency (att_eff), attack excellence percentage (att_exc%), and service break (sk_pt) (Figure 2a-b). This divergence in model performance between leagues suggests the nuanced differences in competitive dynamics and perhaps the variability in data quality or the relevance of selected variables to each league's specific context.

No previous studies have been found on the impact of specific factors on final volleyball league standings through ANN models, indicating a gap in the literature. Komar et al. (2023) employed ANN models to analyze variables influencing match outcomes in Turkey's and Italy's premier leagues (2013-2020 seasons). Their findings revealed that for men's matches, guest teams' attack blocks and direct service points, along with the host teams' attack and block points, were the most impactful variables (Wicham et al., 2016). Similarly, for women's matches, guest teams' attack and direct service points, alongside host teams' attack and block points, emerged as the key determinants of match results.

Prior research on volleyball performance prediction using ANNs demonstrates promising accuracy. Komar et al. (2023) achieved 89.1% accuracy in match outcome prediction (Wicham et al., 2016), while Aka et al. (2021a) successfully predicted set scores using points before technical timeouts (Górriz et al., 2023). Akarçeşme et al. (2020) even achieved over 98% accuracy in predicting team standings for the Rio Olympics (men's and women's) (Fernandez-Echeverria et al., 2017). Similarly, Tümer & Koçer (2017) reported 98% accuracy in predicting Turkey's volleyball league standings based on home/away win-loss records (Komar et al., 2023). These findings, alongside successful applications in other sports like football and basketball highlight the potential of machine learning for volleyball performance analysis (Schumaker, Solieman, Chen, 2010). It is remarkable that predictions at high accuracy rates were obtained through the developed ANN models in the studies mentioned above although the analyses of the volleyball matches were conducted by using different input and output variables. Although this situation limits to compare our results, it is believed that volleyball match analyses can be conducted also by machine learning. These findings underscore the high accuracy rates in determining the three most influential variables on the final league standings of the Sultanlar and Efeler Leagues over the last five seasons (2018-19 to 2022-23). The analysis highlights the importance of attack efficiency, attack excellence percentage, and service break in the Efeler League, and attack excellence percentage, ace per set, and reception error in the Sultanlar League as determinants of final league standings. It's noteworthy that predictions for the Sultanlar League standings closely match the actual team standings, although discrepancies exist between the predicted and actual standings in the Efeler League.

This study leveraged machine learning to identify key factors influencing volleyball league standings, achieving promising prediction accuracy, especially in the Sultanlar League. However, limitations exist. Data quality, model complexity, and unforeseen external factors (e.g., injuries) can impact results. Future research should explore data augmentation, comparing machine learning algorithms, incorporating external variables, and developing league-specific models to improve prediction accuracy for both leagues. By addressing these limitations, machine learning can be further optimized to analyze volleyball performance and predict league standings with even greater precision, providing valuable insights for athletes, coaches, and sports analysts to elevate strategic decision-making within the sport.

Conclusion

This study has demonstrated the robust potential of Artificial Neural Networks (ANN) in predicting the league standings within the context of volleyball, specifically analyzing data across the Sultanlar and Efeler leagues over a five-season span. Our findings reveal that ANN models not only manage large data sets with greater efficiency but also provide predictive insights with considerable accuracy, particularly in the Sultanlar League. The success of these models emphasizes the critical role of specific performance metrics such as points scored, services, and blocks, which were pivotal in achieving high prediction accuracy. However, the study also recognizes the challenges related to data quality and model complexity that could affect the predictive outcomes. These findings suggest that while ANN can streamline sports analytics and potentially enhance coaching and performance strategies, there remains a clear need for further research. Future studies should focus on refining data collection processes, exploring more complex ANN architectures, and possibly integrating other machine learning techniques to extend the robustness and applicability of predictive models in sports analytics.

In conclusion, leveraging ANN in sports not only furthers our understanding of athletic performance dynamics but also opens new pathways for technological integration in sports management and strategy formulation. This study contributes to the burgeoning field of sports analytics by highlighting the effectiveness of machine learning tools in real-world applications and sets the stage for future innovations in the domain.

Financial Support

No financial support was received from institutions and/or institutions during the preparation and writing of this study.

Conflict of Interest

There is no conflict of interest between the authors regarding the publication of this article.

Authors Contributions

Research Idea: EA, CS, HA; Research Design: EA, CS, HA; Analysis of Data: PD; Writing: EA, CS, HA, PD; Critical Review: EA, CS, HA, PD.

References

1. **Aka, H., Akarçeşme, C., Aktuğ, Z. B., & Ozden, S.** (2021a). The estimation of the set results in 2016/2017 &stel &nus sultans league games by artificial neural network. *European journal of human moment*, 47: 32-39.
2. **Aka, H., Aktuğ, Z. B., & Kılıç, F.** (2021b). Estimating the England premier league ranking with artificial neural network. *Applied artificial intelligence*, 35: 393-402.
3. **Akarçeşme C, Aka H, Özden S, & Aktug, Z.B.** (2020). Estimating the volleyball team ranking in the 2016 Rio Olympics by artificial neural network and linear model: Yapay sinir ağları & doğrusal model ile 2016 Rio Olimpiyatlarındaki voleybol takım sıralamasının tahmin edilmesi. *Journal of human sciences,* 17: 1069-1078.
4. **Bai, Z., & Bai, X.** (2021). Sports big data: Management, Analysis, Applications, and Challenges. *Complexity*;6676297.
5. **Beck, M. W.** (2018). Neural NetTools: Visualization and analysis tools for neural networks. *Journal of statistical software*; 85: 1.
6. **Cossich, V. R. A., Carlgren, D., Holash, R. J., & Katz, L.** (2023). Technological Breakthroughs in Sport: Current Practice and Future Potential of Artificial Intelligence, Virtual Reality, Augmented Reality, and Modern Data Visualization in Performance Analysis. *Applied sciences* 13: 12965.

7. **Cortsen, K., & Rascher, D. A.** (2018). The application of sports technology and sports data for commercial purposes. *The use of technology in sport: Emerging challenges*: 47-84.
8. **Górriz, J.M., Álvarez-Illán, I., Álvarez-Marquina, A., et al** (2023). Computational approaches to explainable artificial intelligence: Advances in theory, applications and trends. *Information Fusion* 100: 101945.
9. **João, P. V., Vaz, L., & Mota, M. P.** (2019). The statistics which qualified Portugal for the European Volleyball Championship 2019. *Motricidade* 15: 139-139.
10. **Fernandez-Echerria, C., Mesquita, I., González-Silva, J., & Eche&rría, F. C.** (2017). Match analysis within the coaching process: a critical tool to impro& coach efficacy. *International Journal of Performance Analysis in Sport*, 17: 149-163.
11. **Jörg, M., Perl, J., & Schöllhorn, W.** (2017). Analysis of players' configuration by means of artifical neural networks. *International Journal of performance analysis in sport*, 7: 90-105.
12. **Kautz, T., Groh, B. H., Hannink, J., Jensen, U.** (2017). Activity recognition in beach volleyball using a Deep Convolutional Neural Network: Le&raging the potential of Deep Learning in sports. *Data mining and knowledge disco&ry* 31: 1678-1705.
13. **Komar, E., Egrioglu, E., & Semiz, K.** (2023). Türkiye & İtalya Voleybol Süper Ligleri 2013-2020 İstatistik &rilerinin &ri Madenciliği Yöntemleriyle Analizi. *Eurasian research in sport science*, 8: 54-66.
14. **Kufel, J., Bargieł-Łączek, K., Kocot, S., et al.** (2023). What Is Machine Learning, Artificial Neural Networks and Deep Learning?- Examples of Practical Applications in Medicine, *Diagnostics (Basel, Switzerland)* 13 2023/08/12. DOI: 10.3390/diagnostics13152582.
15. **Millington, B., & Millington, R.** (2015). 'The datafication of e&rything': Toward a sociology of sport and big data. *Sociology of sport journal*, 32: 140-160.
16. **Palao, J. M., & Hernández-Hernández, E.** (2014). Game statistical system and criteria used by Spanish volleyball coaches. *International journal of performance analysis in sport,* 14: 564-573.
17. **Taye, M. M.** (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*; 12: 91.
18. **Tümer, A. E., & Koçer, S.** (2017). Prediction of team league's rankings in volleyball by artificial neural network method. *International Journal of performance analysis in sport*, 17: 202-211.
19. **Yang, Y.** (2021). Data Mining Algorithm in Volleyball Match Technical and Tactical Analysis. In: *Application of Big Data, Blockchain, and Internet of Things for Education Informatization: First EAI International Conference, BigIoT-EDU 2021, Virtual E&nt, August 1–3, 2021, Proceedings, Part II 1* 2021, pp.204-213. Springer.
20. **Wei, T., & Simko, V.**(2021). package "corrplot": Visualization of a Correlation Matrix. 2017. *&rsion 084* 2021.
21. **Wickham, H., Chang, W., & Wickham, M. H.** (2016). Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics &rsion* 2016; 2: 1-189.
22. **Schumaker, R. P., Solieman, O. K., Chen, H.** (2010). Sports data mining: The field. *Sports data mining* 1-13.