Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Niğde Ömer Halisdemir University Journal of Engineering Sciences

Araştırma makalesi / Research article

www.dergipark.org.tr/tr/pub/ngumuh / www.dergipark.org.tr/en/pub/ngumuh

Machine learning-based identification of the strongest predictive features of scoring penalty kick in football

Futbolda başarılı penaltı atışı için en güçlü belirleyici özniteliklerin makine öğrenimi tabanlı tespiti

Ural Akıncıoğlu[1,*], Önder Aydemir[2], Ahmet Çil[3]

[1,3] *Karadeniz Technical University, Faculty of Technology, Electronics and Communication Engineering, 61830, Trabzon, Türkiye*
[2] *Karadeniz Technical University, Electrical and Electronics Engineering, 61080, Trabzon, Türkiye*

Abstract

In football, the penalty is the situation that has one of the highest chances of scoring a goal. However, the success of a penalty kick highly depends on many kinds of attributes, including the penalty-takers' abilities, the amount of fan pressure, the minute of the match, and the current score. In this paper, 16 features were extracted from penalty kick positions, penalty-takers' information, and match-day preferences, and machine learning was used to predict penalty kick outcomes. Moreover, we revealed the most important feature combination that significantly affected the success of a penalty kick. The proposed method was trained with 120 and tested with 50 penalty kicks from the Turkish Super League in terms of classification accuracy and polygon area metric. We concluded that the result of a penalty kick can be predicted with an average classification accuracy and average polygon area metric rates of 79.80% and 0.60 using the *k*-nearest neighbor classifier.

Keywords: Football, machine learning, penalty kick, polygon area metric, classification

Öz

Penaltı, futbolda gol atma şansının en yüksek olduğu durumlardan bir tanesidir. Bir penaltı vuruşunun başarısı penaltıyı kullananların yetenekleri, taraftar baskısının seviyesi, maçın dakikası ve mevcut skor dahil olmak üzere birçok etkene bağlı değişkenlik gösterir. Bu makalede, penaltı pozisyonlarından, penaltıyı kullananların bilgilerinden ve maç günü tercihlerinden 16 öznitelik çıkarılmıştır. Çıkarılan öznitelikler, makine öğrenimi aracılığıyla penaltı vuruşu sonucunu tahmin etmek için kullanılmıştır. Ayrıca bir penaltı vuruşunun başarısını büyük ölçüde etkileyen en önemli öznitelik kombinasyonu elde edilmiştir. Önerilen yöntem, Türkiye Süper Ligi'nden 120 penaltı vuruşu ile eğitilirken 50 penaltı vuruşu ile sınıflandırma doğruluğu ve poligon alanı metriği açısından test edilmiştir. Penaltı vuruşu sonucunun, *k*-en yakın komşu sınıflandırıcısı kullanılarak %79.80 ortalama sınıflandırma doğruluğu ve 0.60 ortalama poligon alanı metriği oranlarıyla tahmin edilebileceği sonucuna varılmıştır.

Anahtar kelimeler: Futbol, makine öğrenmesi, penaltı atışı, poligon alan metriği, sınıflandırma

1 Introduction

Technology is advancing day by day in the world. The developments in hardware and software are also increasing with the advancement of technology. These developments allow different algorithms to be used in various fields. Pattern recognition and machine learning algorithms are especially used in many areas, from medicine to finance, engineering to agriculture, in daily life. Moreover, the speed of these algorithms increases with advanced hardware. By the help of these algorithms, optimum solutions can be offered, and quick decisions can be taken. These are also used more and more in sports. Machine learning has been used in various studies of different sports branches. It is used to increase the training variety in chess, which is a strategy game [1]. In martial arts such as karate and taekwondo, it is used for predicting winners [2], classifying two taekwondo kick techniques [3], and accuracy analysis of movements [4]. In sports such as basketball, volleyball, rugby and football which are played with ball, it is used for predicting winners [5], classification of tactical behaviors [6], recommending a relational-learning-based approach to identify different strategies based on optical tracking data [7], increasing training efficiency [8-11], athlete-specific injury status monitoring [12], wonderkid prediction [13], determining the factors of affecting success [14], referee performance analysis [15], match result prediction [16-18], understanding and estimating the risk of injury [19, 20], improving the offensive game [21], determining the position of the football player [22, 23], increasing the game advantage by predicting the opponent's substitutions [24], estimating the value of football players in the transfer market [25], player selection and team building [26], estimating the match result from player performances [27], predicting the recovery time of professional football players after an undiagnosed injury [28] and classification of passes [29].

Today, football has started to take its place among the most critical sectors expressed in huge numbers economically and commercially [30]. Success is necessary for teams to get more shares from this ever-growing industry. As the success increases, the number of supporters will

* Sorumlu yazar / Corresponding author, e-posta / e-mail: uralakincioglu@ktu.edu.tr (U. Akıncıoğlu))
Geliş / Recieved: 17.05.2024 Kabul / Accepted: 20.08.2024 Yayımlanma / Published: xx.xx.20xx
doi: 10.28948/ngumuh.1485962

increase, so economic and commercial development will also be achieved. To achieve sportive success, clubs make investments with large budgets such as player transfers, manager/coaching team agreements, youth team expenditures, training facility/stadium constructions or maintenance costs of existing ones, technological training equipment purchases, etc. The primary purpose of these investments, expenses, and training is to win matches. To reach this aim, achieving more goal(s) than the opponent is necessary. One of the closest positions to the goal is the penalty kick. In short, making a difference by increasing the percentage of goal scoring in penalty kicks is very effective on the way to success, which is the target.

Penalty kicks have been researched from various perspectives in the literature. Timmis et al. [31] recorded the data from 12 participants who wore an eye-tracking device while taking penalty kicks. They found significant differences between goals and missed penalty kicks in the quiet eye (QE). As a result of their research, the QE period was effective for the penalty kick to be scored. Missed penalty kicks had shorter QE periods and earlier QE offsets. In addition, missed penalty kicks were taken more centrally, and the goalkeeper saved the ball more easily. In another study, Ellis and Ward [32] researched the effect of high and low pressure of supporters on the penalty kicker's performance, psychology, and psychophysiological response. Data were recorded from 20 professional football players under high-pressure and low-pressure situations. A questionnaire was applied to measure the stressful situations, and the respiratory and heart rates of the football players were acquired via sensors. In addition, the players were asked to hit the targets for accuracy detection. In the case of high-pressure situations, the penalty kickers listened to the sounds from the headphones. These were the supporters' sounds that were recorded in major football tournaments. On the contrary, the sound was not played under low-pressure situations. Quantitative results showed that increased cognitive anxiety, respiratory rate, bivariate variable error, and low self-confidence under high pressure. Qualitative results showed that high pressure was predominantly distracting. In another penalty kick-based study, Ferraresi and Gucciardi [33] collected penalty kick data, including scored or missed penalty kick information from the played football matches. They aimed to analyze social environment effects on individual performances. Data consist of 545 penalties from the France, Germany, Italy, Spain, and United Kingdom major leagues (from the season 2019-2020), and 148 of them were related to games played behind closed doors (295 of them awarded to home teams). In closed-door matches, the probability of missing a penalty increased, so the fans' support especially led to improved home teams' performance. The result from the analysis was that the fans affect individual performances. In another approach, the penalty kick result was estimated through the accelerometer sensor and camera data. Data were recorded from four bachelor students. They took 10-12 penalties on each part of the goal, which was divided into six parts. Thus, a total of 268 penalty kicks were taken. Camera footage was divided into four sections: running, preparing for the shot, hitting, and post-hit. In the running phase, statistical calculations such as mean, variance, standard deviation, average amplitude, and average power were made from the data obtained from the accelerometer, and a better data set was created. 5-fold cross-validation was performed on the acquired data set. Random forest, decision tree (DT), support vector machines (SVM), and multiple linear perceptron algorithms were applied. SVM reached the highest accuracy among the algorithms at 47.98%. The accuracy reached 70.98% when convolutional neural networks were used [34]. Bloechle et al. [35] realized an augmented reality penalty simulator with a holographic goalkeeper to train sensorimotor kicking skills. The simulator was tested with 13 young elite soccer players. Using the simulator, the threshold time required to re-determine the direction of the kick according to the goalkeeper's movement before kicking the ball was reduced by 120 ms in ten training sessions of twenty kicks. In other words, the probability of successfully redirecting the kick to score the penalty was increased. The individual performance of each player was modeled with the Bayesian network. A 35% increase in the success rate was achieved with training and feedback. Based on the aforementioned studies reported in the literature, it can be said that the studies focused on topics such as analyzing the eye movements during penalty kick using eye tracking, in which area of the goal the probability of missing the penalty increases, the effect of stress caused by social environment during penalty kick, penalty kick outcome prediction through data recorded using camera and accelerometer, and the realization of a virtual reality simulator for penalty kick training. Although penalty kicks have been analyzed from various perspectives, there is no study that classifies the penalty kick outcome as a goal scored or missed using 16 features that have a high potential to affect the result by machine learning. In this paper, different from the previous studies, penalty kicks are analyzed from various perspectives, and 16 different features with high potential to affect the penalty kick outcome are extracted from match videos and some sports websites. These features are used together in penalty kick prediction by machine learning. This study aims to determine the combination of features that predict the penalty kick outcome with high accuracy among various features and thus contribute to the literature on penalty kick outcome prediction. In addition, this study is also important in the field of soccer academy as the penalty kick outcome can be improved positively by training on the features in the determined combination.

To determine the most important features, all possible feature subsets of the 16 features were tested one by one with different classification methods, including *k*-nearest neighbor (*k*-NN), DT, SVM, and linear discriminant analysis (LDA). To evaluate the performance of the classifiers, classification accuracy (CA), sensitivity (SE), specificity (SP), area under curve (AUC), Jaccard index (J), and F-measure (FM) metrics were obtained via polygon area metric (PAM) [36]. Thus, classifier performance can be evaluated with the PAM without needing various metrics. As a result of this paper, it is aimed to automatically determine the outcome of a penalty kick by using machine learning algorithms.

In section 2, the data set description is introduced, classification and feature selection procedures are mathematically explained. Number of feature selection results for different classifiers are provided in figures, and various metrics results are provided in tables in section 3. section 4 presents the conclusion and discusses the findings.

2 Material and methods

2.1 Data set description

The data set consists of 170 penalties, 85 of which were scored and 85 of which were missed. All penalties were selected from the Turkish Super League, including the 2016-2017, 2018-2019, 2019-2020, 2020-2021, 2021-2022, and 2022-2023 seasons. Penalty kicks' and penalty-takers' information was taken from the Transfermarkt and Turkish Football Federation websites [37-38]. Analysis of penalty kicks were compiled with data from penalty positions for each season. As a result of the analysis of penalty kick videos, 16 attributes were determined as features. The features and their groups are shown in Table 1.

Penalty-taker's position, nationality, preferred foot, age, number of goals, success rate, game times, and match experience were obtained from Transfermarkt. Penalty-takers were divided into three different groups (defender, midfielder, and forward) according to their positions, three different groups (European, African, American) according to their nationalities, two different groups (right, left) according to their preferred foot while taking a penalty kick, four different groups (1stGroup<=23, 23<2ndGroup<=26, 26<3rdGroup<=31, 4thGroup>31) according to their ages. Penalty-taker and goalkeeper age difference feature was obtained by subtracting the goalkeeper's age from the penalty-taker's age. This feature was divided into seven groups (1stGroup<-9, -9<=2ndGroup<=-5, -5<3rdGroup<0, 4thGroup=0, 0<5thGroup<5, 5<=6thGroup<=9, 7thGroup>9) according to age difference. Penalty-taker's success rate, game times, and match experience were determined by looking at their career statistics.

League week, home/away status, league position of the penalty-taking team, penalty minute, match starting time, and match score before taking penalty kick information was obtained from the Turkish Football Federation official website. Due to the COVID pandemic, different numbers of clubs have competed in the Turkish Super League in the last years, so the number of league weeks has changed. The league week feature was obtained by dividing the number of week in which the penalty kick was taken by the total number of weeks of the season. The league position feature of the team that took the penalty kick was determined by the previous week's league table. This feature was set to zero for the first week's penalty kicks. Six groups were created for the penalty minute feature. These groups were determined by dividing 90 minutes into six equal parts (0<1stGroup<=15, 15<2ndGroup<=30, 30<3rdGroup<=45, 45<4thGroup<=60, 60< 5thGroup<=75, 75<6thGroup<=90). Additional times for both halves were grouped as the 45th and 90th minutes. Two different groups (before and after 6 pm) were created for the match starting time feature. Goal difference was used as the match score before taking the penalty kick feature. If the team that took a penalty kick was winning, the goal difference was set to plus goal difference. If it was losing, the goal difference was set to the minus goal difference. In the case of a draw, the goal difference was set to zero. Home/away status was determined on the team that took the penalty kick.

Table 1. Features and groups

Feature	Groups
Penalty-taker's position (F1)	1-Defence, 2-Midfielder, 3-Forward
Penalty-taker's nationality (F2)	1-European, 2-African, 3-American
League week (F3)	2022-2023 Season 38 weeks in total 2021-2022 Season 38 weeks in total 2020-2021 Season 42 weeks in total 2019-2020 Season 34 weeks in total 2018-2019 Season 34 weeks in total 2016-2017 Season 34 weeks in total
Home/Away Status (F4)	1-Home, 2-Away
League position of the penalty-taking team (F5)	Penalty-taking teams league position in the previous week
Penalty-taker's preferred foot (F6)	1-Right, 2-Left
Penalty minute (F7)	(0<1stGroup<=15, 15<2ndGroup<=30, 30<3rdGroup<=45, 45<4thGroup<=60, 60< 5thGroup<=75, 75<6thGroup<=90)
Match starting time (F8)	1-Before 6 pm, 2-After 6 pm
Match score before taking penalty kick (F9)	Goal difference between the two teams
Potential target point of penalty kick (F10)	1-right lower, 2-right medium, 3-right upper, 4-center lower, 5-center medium, 6-center upper, 7-left lower, 8-left medium, 9-left upper
Penalty-taker's age (F11)	1stGroup<=23, 23<2ndGroup<=26, 26<3rdGroup<=31, 4thGroup>31
Penalty-taker and goalkeeper age difference (F12)	1stGroup<-9, -9<=2ndGroup<=-5, -5<3rdGroup<0, 4thGroup=0, 0< 5thGroup<5, 5<=6thGroup<=9, 7thGroup>9
Number of goals scored by penalty-taker (F13)	Total number of goals in the penalty-taker's career until the penalty
Penalty-taker's success rate (F14)	Success rate on penalties in their career
Penalty-taker's game times (F15)	Total minutes on the pitch in career
Penalty-taker's match experience (F16)	Number of matches played in their career

As shown in Figure 1, 9 different groups (according to the goalkeeper, 1-right lower, 2-right medium, 3-right upper, 4-center lower, 5-center medium, 6-center upper, 7-left lower, 8-left medium, 9-left upper) were determined for the potential target point feature. If the goalkeeper was able to catch the ball in the middle of the goal without moving left and right and without opening their arms to the sides, this point was defined as the center of the goal, if they moved to the right or left for catching the ball, the direction was defined as the goal’s right or left, respectively. The section from the ground to the goalkeeper's knee was defined as lower, the section from the goalkeeper's knee to the middle of the goal (122 cm) was defined as medium, and the middle (122 cm) to upper (244 cm) part of the goal was defined as upper.

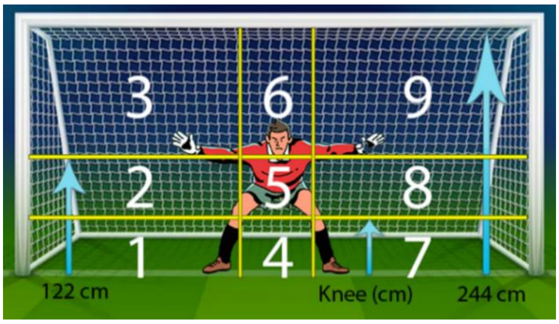

Figure 1. Potential target points of a penalty kick

The data set was randomly divided into training and test sets. The training data set consisted of a total of 120 data, 60 of which were goals, and 60 were missed. The test data set included a total of 50 data, 25 of which were goals, and 25 were missed. The test data set was used to evaluate the proposed method after the classifier model was trained with the training data set.

2.2 Classification and feature selection procedures

The *k*-NN is a simple and popular algorithm that classifies two or multi-class problems. The *k*-NN algorithm tries to find the nearest neighbors by calculating the distance between the test trial and all training trials. The class label of the test trial is defined by a majority vote on the *k*-NN. The Euclidean distance function is usually used to calculate the distance between two points.

SVM is a supervised algorithm used in classification and regression problems. This algorithm aims to separate the classes on a boundary line by finding a hyperplane in the feature space. We tried to minimize the classification error by increasing the distance between the hyperplane and the nearest data point from each class. Class determination can be calculated with the following equation:

$$C = \{Class\ 1\ if\ w^T.\ i + d < 0 \quad Class\ 2\ if\ w^T.\ i + d\ \geq 0\} \tag{1}$$

Where C is class information, i is the input vector, w is the weight vector, and d is the deviation.

The decision tree algorithm is also a supervised learning algorithm used in regression and classification problems. This algorithm is based on the top-down learning approach [39]. Each node of the decision tree is labeled with an input feature. This algorithm aims to obtain nodes that are easier to decide by applying a set of rules to the nodes.

LDA searches for a linear transformation by reducing the original data to much lower-dimensional space. In short, it aims to find the transform that maximizes class separability and the linear transform that provides the best separation of classes in reduced space [40]. The default settings of MATLAB were used for all classifiers.

Generally, CA, SE, SP, J, FM, and AUC are calculated to evaluate the performance of machine learning systems. The mathematical definitions are respectively given as follows:

$$\mathrm{CA} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}) \tag{2}$$

$$\mathrm{SE}=\mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \tag{3}$$

$$\mathrm{SP}=\mathrm{TN}/(\mathrm{TN}+\mathrm{FP}) \tag{4}$$

$$\mathrm{J}=\mathrm{TP}/(\mathrm{TP}+\mathrm{FP}+\mathrm{FN}) \tag{5}$$

$$\mathrm{FM}=2\mathrm{TP}/(2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}) \tag{6}$$

$$\mathrm{AUC} = \left(\int_0^1 f(a)da\right) \tag{7}$$

TP, TN, FP, and FN are defined as the number of goal trials correctly predicted, the number of missed trials correctly predicted, the number of goal trials incorrectly predicted, and the number of missed trials incorrectly predicted, respectively. f(a) is a receiver operating characteristic curve in which the TP rate (SE) is plotted as a function of the FP rate (1-SP) for different cut-off points.

PAM is calculated using the area of the polygon formed by the points CA, SE, SP, AUC, J, and FM in a regular hexagon. A polygon formed in a regular hexagon is shown in Figure 2.

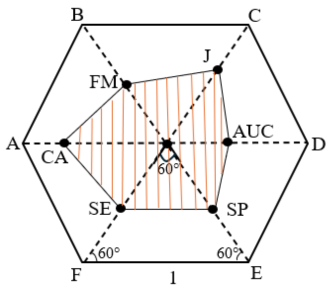

Figure 2. A Polygon created in a regular hexagon

It should be noted that the length (*L*) of each side of the regular hexagon is equal to 1, and it consists of six equilateral triangles. The Area of a regular hexagon is calculated from formula 8 as 2.59807.

$$Area\ of\ a\ regular\ hexagon = \frac{3\sqrt{3}}{2}\ x\ L \tag{8}$$

The PAM is calculated using the following formula:

$$\mathrm{PAM} = \mathrm{PA}/2.59807 \tag{9}$$

The area of the polygon is PA. The PA is obtained by summing the areas of the six triangles formed. PA value is divided by 2.59807, and PAM is normalized to the [0,1] interval. Although it is a single value in PAM, it contains more information than the CA, SE SP, AUC, J, and FM metrics whose formulas are given above. The polygon area is created using the values of the six metrics, and the PAM metric is obtained from the created area. In other words, the PAM value reflects the values of the six metrics, and obtaining a single metric from the six makes the comparison easier. In this paper, PAM was used to evaluate the performance of the four classifiers.

The CAs of 65536 possible subset combinations obtained from 16 features were calculated separately. It is worth mentioning that the subset combination providing the highest CA in 65536 possible subsets was considered. Due to the limited amount of data set and to verify the robustness of the method and to avoid randomization problems in the training and test sets, the classification result was run ten times through a random data selection. The highest CAs of each classifier in 10 runs were averaged, and the classifier with the highest average CA was determined. The features in the feature combination that provided the highest CA of the determined classifier were set as selected features. The average PAM, CA, SE, SP, J, FM, AUC, and their standard deviations (Stds) were calculated for each classification model. In the calculation of stds, the data providing the highest CA in 10 runs of each classifier and other metrics obtained from these data were used. The most effective features were determined based on the selection frequencies of features in the combinations that provided the highest CAs. The effective features were used only for evaluation separately from the selected features. The flowchart of the proposed method is shown in Figure 3. All runs were performed in the MATLAB R2022a environment on a 2.2 GHz Intel Core i7 processor-powered computer with 16GB, 2933 MHz DDR4 memory. The CAs for all classifiers of possible subsets of 16 features in one run performed in approximately 30 hours.

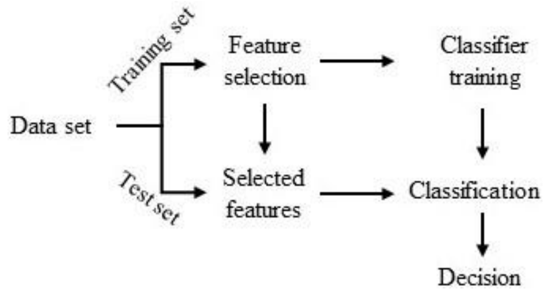

Figure 3. Flowchart of the proposed method

3 Results

In this paper, the outcome of a penalty kick was predicted with 16 features determined through penalty kick videos, penalty-takers' information, and match day information. Figure 4 shows the *k*-NN, DT, SVM, and LDA classifiers' feature selection frequencies of 10 runs in the training sets.

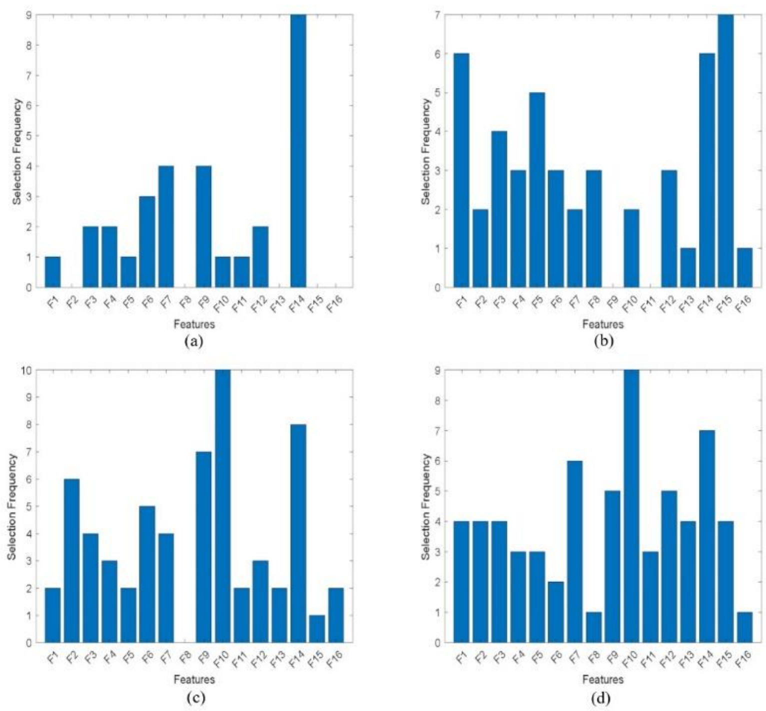

Figure 4. Selection frequencies of features (a) *k*-NN, (b) DT, (c) SVM, (d) LDA

The highest CAs of 10 runs of each classifier are shown in Table 2. According to these results, the 2nd run with 84% in *k*-NN, the 8th run with 82% in DT and SVM, and the 5th, 7th, 8th, and 10th runs with 78% in LDA provided the highest CAs. The highest CAs were achieved with the combinations of F1-F3-F4-F9-F10 in *k*-NN, F1-F6-F10-F14-F15 in DT, F1-F2-F6-F10-F11-F13-F14-F16 in SVM, and F1-F3-F5-F10-F11-F13-F14-F15, F3-F9-F10-F12-F13-F14-F16, and F7-F8-F10-F15 in LDA.

Table 2. The highest CAs of 10 runs

Run	k-NN CAs (%)	DT CAs (%)	SVM CAs (%)	LDA CAs (%)
1	80	76	76	76
2	84	78	74	72
3	82	70	72	70
4	80	76	76	76
5	78	76	76	78
6	80	76	76	76
7	78	74	76	78
8	80	82	82	78
9	78	78	72	72
10	78	76	76	78

The average PAM, CA, SE, SP, AUC, J, and FM values of each classifier and their stds were calculated. These results are given in Table 3. According to the results in Table 3, it is seen that the *k*-NN classification method provided the highest average CA and PAM values. A Small value of the std supports the robustness of the proposed algorithm. Therefore, the F1, F3, F4, F9, and F10 features in the combination that provided the highest CA of 84% in the *k*-NN classifier were determined as the selected features. The PAM and various metric results of the data using the *k*-NN classifier in 10 runs are shown in Table 4. In Table 3, and Table 4 all values except average CA, average SE, and average SP were normalized to the [0,1] interval, and average CA, average SE, and average SP values are given as percentages. Table 5 shows the confusion matrices of the run of each classifier that provided the highest CA. It can also be seen in the confusion matrix that the *k*-NN classifier provided balanced and high accuracy results in both classes.

Figure 4(a) shows that the feature with the highest frequency of selection in the *k*-NN classifier is the penalty taker's success rate corresponding to F14. This is also considered to be consistent. In addition, it is concluded that F7 and F9 features are also effective based on the selection frequencies. Figure 4(b), Figure 4(c), and Figure 4(d) show the effective features of DT, SVM and LDA. These effective features are F1, F14, F15 in DT, F9, F10, F14 in SVM and F7, F10, F14 in LDA.

Table 3. Average CA, PAM, SE, SP, AUC, J, FM results and stds

Classifier	*k*-NN	DT	SVM	LDA
Average CA ± std (%)	79.80 ± 1.99	76.20 ± 3.05	75.60 ± 2.80	75.40 ± 2.99
Average PAM ± std	0.60 ± 0.03	0.54 ± 0.04	0.52 ± 0.04	0.53 ± 0.05
Average SE ± std (%)	81.20 ± 6.27	73.20 ± 7.32	68.80 ± 4.13	72.80 ± 7.25
Average SP ± std (%)	78.40 ± 7.5	79.20 ± 8.80	82.40 ± 5.06	78.00 ± 7.83
Average AUC ± std	0.80 ± 0.02	0.76 ± 0.03	0.76 ± 0.03	0.75 ± 0.03
Average J ± std	0.67 ± 0.03	0.61 ± 0.04	0.59 ± 0.04	0.60 ± 0.04
Average FM ± std	0.80 ± 0.02	0.75 ± 0.03	0.74 ± 0.03	0.75 ± 0.03

Table 4. PAM and various metrics results of data using *k*-NN classifier in 10 runs

Run	PAM	CA(%)	SE(%)	SP(%)	AUC	J	FM
1	0.60	80.00	76.00	84.00	0.80	0.66	0.79
2	0.67	84.00	84.00	84.00	0.84	0.72	0.84
3	0.64	82.00	88.00	76.00	0.82	0.71	0.83
4	0.60	80.00	76.00	84.00	0.80	0.66	0.79
5	0.58	78.00	84.00	72.00	0.78	0.66	0.79
6	0.60	80.00	76.00	84.00	0.80	0.66	0.79
7	0.58	78.00	92.00	64.00	0.78	0.68	0.81
8	0.59	80.00	72.00	88.00	0.80	0.64	0.78
9	0.57	78.00	80.00	76.00	0.78	0.65	0.78
10	0.58	78.00	84.00	72.00	0.78	0.66	0.79

Table 5. Confusion matrices of the highest CAs

Classifier	Run	TP	FN	TN	FP
k-NN	2	21	4	21	4
DT	8	19	6	22	3
SVM	8	18	7	23	2
LDA	7 and 8	19	6	20	5
LDA	5 and 10	17	8	22	3

4 Conclusion and discussion

In this paper, the penalty kick outcome was predicted with 16 features, including statistical and match-day features. The proposed algorithm was tested using *k*-NN, DT, SVM, and LDA classifiers. We ran the proposed method 10 times to demonstrate the proposed algorithm's robustness and avoid random selection in testing and training sets. Average PAM values of *k*-NN, DT, SVM, and LDA classifiers were obtained as 0.60, 0.54, 0.52, and 0.53, respectively. We obtained significant results not only for PAM but also for other metrics by using the *k*-NN classifier. We calculated the average CA, SE, SP, AUC, J, and FM values as 79.80%, 81.20%, 78.40%, 0.80, 0.67, and 0.80, respectively. The PAM graph displays the results of six metrics on a single graph. It allows the results of six metrics to be compared and evaluated simultaneously. The PAM graph of the *k*-NN's 2nd run is shown in Figure 5.

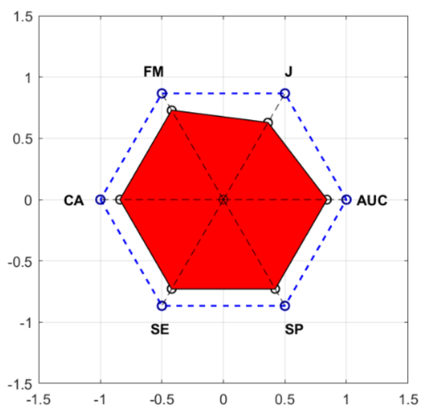

Figure 5. PAM graph

According to the results, it can be said that *k*-NN is the most suitable classifier for predicting the outcome of a penalty kick. In addition, small standard deviation values indicated the robustness of the system. The *k*-NN classifier selection frequencies results show that the most effective features were F7, F9, and F14. The selected features were F1, F3, F4, F9, and F10 which combination provided the highest CA in 10 runs.

The selected F3, F4, and F9 features are related to stress and pressure. Studies in the literature have also shown the effect of stress and pressure on penalty kicks [32,33]. Another feature selected by the study is the F10, which is related to the target point. In the literature, the penalty target point has also been proven to be effective [31]. The fact that the proposed method selects these features in the feature combination that provides the highest CA in predicting the penalty kick outcome can be considered as evidence that it works accurately and effectively.

In the most similar penalty prediction study in the literature, accelerometer sensor and camera data were obtained from 268 penalty kicks. The penalty kick outcomes were predicted with 70.98% CA using convolutional neural networks [34]. Different from the most similar study in the literature, the conducted study analyzed penalty kicks by combining different features with each other without focusing only on a few features and achieved higher CA. This study contributed to the literature by determining the feature combination that predicts the penalty kick outcome with the highest CA.

In the proposed method, among the five features in the combination that predicts the outcome of the penalty kick with the highest CA, F3, F4, and F9 features are known before the penalty kick, and F1 and F10 features can be dynamically adjusted. By inputting data into the system, trained using five features determined with the proposed method, the success rate of the penalty kick can be calculated, and the ideal target point of the penalty kick and the most suitable penalty-taker can also be determined. This real-time penalty kick success determination system can be used in matches and training as it provides fast results.

Deep learning was not used in this paper because the data structure was not suitable. In future studies, performance will be increased with deep learning by adding data such as electromyography and eye tracking. In addition, it is aimed to improve the average PAM performance by adding not only the penalty-taker's related features but also the goalkeeper's related features.

Acknowledgements

Ural AKINCIOĞLU and Ahmet ÇİL was supported by a scholarship from The Scientific and Technological Research Council of Turkey.

Conflict of interest

The authors declare no potential conflict of interest.

Similarity rate (iThenticate): 15%

References

[1] J. A. Brown, A. Cuzzocrea, M. Kresta, K. D. Kristjanson, C. K. Leung and T. W. A. Tebinka, Machine learning tool for supporting advanced knowledge discovery from chess game data. 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 649-654, 2017. https://doi.org/10.1109/ICMLA.2017.00-87.

[2] M. R. Albuquerque, P. H. C. Mesquita, T. Herrera-Valenzuela, D. Detanico and E. Franchini, Predicting taekwondo winners in high-level competition using ranking scores and country performance scores: an analysis of the 2019 World Taekwondo Championship. Ido Movement for Culture. Journal of Martial Arts Anthropology, 21(2), 2021. https://doi.org/10.14589/ido.21.2.4.

[3] M. H. Zhong, J. C. Hung, Y. C. Yang and C. P. Huang, GA-SVM classifying method applied to dynamic evaluation of taekwondo. International Conference on Advanced Materials for Science and Engineering, 534-537, 2016. https://doi.org/10.1109/ICAMSE.2016.7840191.

[4] B. Emad, O. Atef, Y. Shams, A. El-Kerdany, N. Shorim, A. Nabil and A. Atia, Ikarate: Improving karate kata. Procedia Computer Science. 170, 466-473, 2020. https://doi.org/10.1016/j.procs.2020.03.090.

[5] T. Chellatamilan, M. M. Ravichandran and K. Kamalakkannan, Modern Machine Learning Approach for Volleyball Winning Outcome prediction. Global Journal of Multidisciplinary Studies, 4(12), 63-71, 2015.

[6] S. Wenninger, D. Link and M. Lames, Performance of machine learning models in application to beach volleyball data. International Journal of Computer Science in Sport, 19(1), 24-36, 2020. https://doi.org/10.2478/ijcss-2020-0002.

[7] J. Van Haaren, H. B. Shitrit, J. Davis and P. Fua, Analyzing volleyball match data from the 2014 world championships using machine learning techniques. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 627-634, 2016. https://doi.org/10.1145/2939672.2939725.

[8] X. Dai and S. Li, Volleyball data analysis system and method based on machine learning. Wireless Communications and Mobile Computing, 1-11, 2021. https://doi.org/10.1155/2021/9943067.

[9] A.K. Holatka, H. Suwa and K. Yasumoto, Volleyball setting technique assessment using a single point sensor. IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 567-572, 2019. https://doi.org/10.1109/PERCOMW.2019.8730811.

[10] V. Y. Kulkarni and P. K. Sinha, Pruning of random forest classifiers: A survey and future directions. International Conference on Data Science & Engineering (ICDSE), 64-68, 2012. https://doi.org/10.1109/ICDSE.2012.6282329.

[11] T. Xu and L. Tang, Adoption of machine learning algorithm-based intelligent basketball training robot in athlete injury prevention. Frontiers in Neurorobotics, 14, 620378, 2021. https://doi.org/10.3389/fnbot.2020.620378.

[12] A. W. de Leeuw, S. V. der Zwaard, R. V. Baar and A. Knobbe, Personalized machine learning approach to injury monitoring in elite volleyball players. European journal of sport science, 22(4), 511-520, 2022. https://doi.org/10.1080/17461391.2021.1887369.

[13] Z. Mahmood, A. Daud, R. A. Abbasi, Using machine learning techniques for rising star prediction in basketball. Knowledge-Based Systems, 211. 106506, 2021. https://doi.org/10.1016/j.knosys.2020.106506.

[14] M. Migliorati, Detecting drivers of basketball successful games: an exploratory study with machine learning algorithms. Electronic Journal of Applied Statistical Analysis, 13(2), 454-473, 2020. https://doi.org/10.1285/i20705948v13n2p454.

[15] S. W. Wang and W. W. Hsieh, Performance analysis of basketball referees by machine learning techniques. In International Congress on Sport Sciences Research and Technology Support, 2, 165-170, 2016.

[16] R. Baboota and H. Kaur, Predictive analysis and modelling football results using machine learning approach for English Premier League. International Journal of Forecasting, 35(2), 741-755, 2019. https://doi.org/10.1016/j.ijforecast.2018.01.003.

[17] A. Joseph, N. E. Fenton and M. Neil, Predicting football results using Bayesian nets and other machine learning techniques. Knowledge-Based Systems, 19(7), 544-553, 2006. https://doi.org/10.1016/j.knosys.2006.04.011.

[18] F. Thabtah, L. Zhang and N. Abdelhamid, NBA game result prediction using feature analysis and machine learning. Annals of Data Science, 6(1), 103-116, 2019. https://doi.org/ 10.1007/s40745-018-00189-x.

[19] J.L. Oliver, F. Ayala, M. B. D. S. Croix, R. S. Lloyd, G. D. Myer and P. J. Read, Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players. Journal of science and medicine in sport, 23(11), 1044-1048, 2020. https://doi.org/10.1016/j.jsams.2020.04.021.

[20] Y. Liang, Sports Injury Prediction Model based on Machine Learning. Economic Management and Big Data Application, 969-981, 2024. https://doi.org/10.1142/9789811270277_0086.

[21] M. Herold, F. Goes, S. Nopp, P. Bauer, C. Thompson and T. Meyer, Machine learning in men's professional football: Current applications and future directions for improving attacking play. International Journal of Sports Science & Coaching, 14(6), 798-817, 2019. https://doi.org/10.1177/1747954119879350.

[22] M. Frey, E. Murina, J. Rohrbach, M. Walser, P. Haas and M. Dettling, Machine learning for position detection in football. 6th Swiss Conference on Data Science (SDS), 111-112, 2019. https://doi.org/10.1109/SDS.2019.00009.

[23] A. Garcia-Aliaga, M. Marquina, J. Coteron, A. R. Gonzales and S. L. Sanchez, In-game behaviour analysis of football players using machine learning techniques based on player statistics. International Journal of Sports Science & Coaching, 16(1), 148-157, 2021. https://doi.org/10.1177/1747954120959762.

[24] A. Mohandas, M. Ahsan and J. Haider, Tactically Maximize Game Advantage by Predicting Football Substitutions Using Machine Learning. Big Data Cogn. Comput., 7, 117, 2023. DOI:10.3390/bdcc7020117.

[25] I. Behravan and S. M. Razavi, A novel machine learning method for estimating football players' value in the transfer market. Soft Computing, 25(3), 2499-2511, 2021. https://doi.org/10.1007/s00500-020-05319-3.

[26] D. Abidin, A case study on player selection and team formation in football with machine learning. Turkish Journal of Electrical Engineering and Computer Sciences, 29(3), 1672-169, 2021. https://doi.org/10.3906/elk-2005-27.

[27] J. Almulla and T. Alam, Machine learning models reveal key performance metrics of football players to win matches in qatar stars league. IEEE Access, 8,

213695-213705, 2020. https://doi.org/10.1109/ACCESS.2020.3038601.

[28] S. Kampakis, Comparison of Machine Learning Methods for Predicting the Recovery Time of Professional Football Players After an Undiagnosed Injury. MLSA@ PKDD/ECML. 1969, 58-68, 2013.

[29] S. Chawla, J. Estephan, J. Gudmundsson and M. Horton, Classification of passes in football matches using spatiotemporal data. ACM Transactions on Spatial Algorithms and Systems (TSAS). 3(2), 1-30, 2017. https://doi.org/10.1145/3105576.

[30] S. Terekli and H. O. Çobanoğlu, Developing economic values in football: Example of Turkish Football Federation. Open Access Library Journal, 5(2), 1-14, 2018. https://doi.org/10.4236/oalib.1104263.

[31] M. A. Timmis, A. Piras and K. N. Van Paridon, Keep your eye on the ball; the impact of an anticipatory fixation during successful and unsuccessful soccer penalty kicks. Frontiers in psychology, 9, 2058, 2018. https://doi.org/10.3389/fpsyg.2018.02058.

[32] L. Ellis and P. Ward, The effect of a high-pressure protocol on penalty shooting performance, psychological, and psychophysiological response in professional football: A mixed methods study. Journal of Sports Sciences, 40(1), 3-15, 2022. https://doi.org/10.1080/02640414.2021.1957344.

[33] M. Ferraresi and G. Gucciardi, Who chokes on a penalty kick? Social environment and individual performance during Covid-19 times. Economics Letters, 203, 109868, 2021. https://doi.org/10.1016/j.econlet.2021.109868.

[34] A. Chakma, A. Z. M. Faridee, N. Roy and H. S. Hossain, Shoot like ronaldo: Predict soccer penalty outcome with wearables. IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 1-6, 2020. https://doi.org/10.1109/PerComWorkshops48775.2020.9156244.

[35] J. L. Bloeche, J. Audiffren, T. Le Naour, A. Alli, D. Simoni, G. Wüthrich and J. P. Bresciani, It's not all in your feet: Improving penalty kick performance with human-avatar interaction and machine learning. The Innovation, 5(2), 100584, 2024. https://doi.org/10.1016/j.xinn.2024.100584.

[36] O. Aydemir, A new performance evaluation metric for classifiers: polygon area metric. Journal of Classification, 38, 16-26, 2021. https://doi.org/10.1007/s00357-020-09362-5.

[37] Transfermarkt, Football player and match statistics, https://www.transfermarkt.com/, Accessed 16 May 2024.

[38] Turkish football federation official website, Match statistics, https://www.tff.org/, Accessed 16 May 2024.

[39] B. R. Patel and K. K. Rana, A survey on decision tree algorithm for classification. International Journal of Engineering Development and Research, 2(1), 1-5, 2014.

[40] C. H. Park and H. Park, A comparison of generalized linear discriminant analysis algorithms. Pattern Recognition, 41(3). 1083-1097, 2008. https://doi.org/10.1016/j.patcog.2007.07.022.