



# Stacked-Based Ensemble Machine Learning Model for Positioning Footballer

Selim Buyrukoğlu<sup>1</sup> · Serkan Savaş<sup>1</sup>

Received: 29 December 2021 / Accepted: 29 March 2022 / Published online: 24 April 2022  
© King Fahd University of Petroleum & Minerals 2022

## Abstract

Due to the high performance of machine learning (ML) methods in different disciplines, these methods have been used frequently in various sports fields, especially in the last decade. Researchers have used ML algorithms in football on various subjects such as match result prediction, estimation of factors affecting match results, prediction of league standings, and analysis of the performances of football players. However, there has not been enough work on determining the position of the football player, which is one of the leading problems for coaches in football. Therefore, this study aims to classify footballer positions employing a stacked ensemble ML model using the FIFA'19 game dataset. To achieve this aim, a two-stage application is followed. In the first stage, 10 features are selected using four different feature selection algorithms. In the second stage, Deep Neural Networks, Random Forest, and Gradient Boosting were used as single-based algorithms, and Logistic Regression was employed as a meta-learner in the stacked model. The results show that the combination of the Chi-square feature selection technique and the stacked-based ensemble learning model yielded the best accuracy (83.9%). The findings emphasize the validity and robustness of our stacked ensemble learning model to determine the positions of footballers.

**Keywords** Footballer position classification · Meta-learner · Ensemble learning · Feature selection · Chi-square

## 1 Introduction

The use of artificial intelligence (AI) has significantly increased in sports science, especially in football, volleyball, basketball, etc. Employing AI algorithms in sport science has become important for managers in terms of analyzing data about players and many more features. Moreover, the success of machine learning (ML) algorithms in sports science has caused to an increase in its interest for use in different sports fields. Thus, many sports clubs have focused on using ML algorithms to obtain robust and reliable statistical information about the players, tactical struggles, predictions, and more.

Today, analysis and evaluation studies are carried out with AI and ML in almost all types of sports. There are various

types of studies in sports science that have been proposed for different purposes. Tümer and Koçer [1] use Artificial Neural Networks (ANNs) to predict team rankings in a volleyball league. A different study proposed by [2] employed k-nearest neighbor (kNN), Logistic Regression (LR), Multi-layer Perceptron (MP), Naive Bayes (NB), j48, and Voting ML algorithms to predict basketball league match results. The Decision Tree (DT) algorithm is also employed to predict passing and attacking games in American football in the National Football League [3]. However, it should be noted that football is one of the most popular sports in the world. Therefore, this study aims to analyze the data relating to football.

To illustrate the theoretical framework of our study, the rest of this section provides information about the research problem, aim & objectives. It also gives criticism on the theories developed by other researchers. Footballers mostly play in certain positions on the football pitch based on their abilities, but their positions can be changed by the coach if necessary. Moreover, many people start playing football as amateur players. If the coaches determine the positions of the amateur players appropriately based on their talents,

✉ Serkan Savaş  
serkansavas@karatekin.edu.tr

Selim Buyrukoğlu  
sbuyrukoglu@karatekin.edu.tr

<sup>1</sup> Department of Computer Engineering, Çankırı Karatekin University, Çankırı 18100, Turkey



these players can be professionals in their positions. However, determining the appropriate position of a football player is not an easy task in football. In general, coaches determine football players' positions using their observations and experiences [4]. However, it cannot be considered as an effective way to determine football players' positions which is the research problem of our study. Instead, ML algorithms can be effectively employed in the determination of footballers' positions. Thus, the study aims to classify footballers based on ML algorithms. In the literature, limited studies have been proposed in the determination of footballers' positions using ML algorithms.

In a study focusing on player position estimation, case-based reasoning (CBR) was used together with the kNN, and an accuracy rate of 97% was obtained using 36 case data [5]. In another study, a framework that performs position estimation on 100 football players' data was created by grouping the physical, mental, and technical abilities of the football players. Results from the classification experiments using Bayesian Networks, DT, and kNN have shown an average of 98% accuracy [4]. Apart from these, different football player position estimation study apart from these could not be found in the literature, to the best of the authors' knowledge. These studies have some limitations. The first limitation observed is that these studies are conducted on a limited amount of data. For this reason, we aim to classify football players' positions using The Federation Internationale de Football Association (FIFA) dataset. It is not an easy task to obtain actual data that will be used to categorize their positions on a football pitch, and so the FIFA'19 video game dataset is used in this study. FIFA is one of the most important football games in the world and constantly updates football player data. Additionally, the use of the FIFA dataset is important in terms of showing the potentials of ML in the classification of footballer positions. The second limitation is that the studies presented above employed uncommon ML algorithms in the classification process. In this sense, a case study could be carried out to determine the appropriate ML algorithms to be used in the studies. In our study, the employed ML algorithms are determined as a result of a case study, and the background of these algorithms is presented in Sect. 2.3 in detail. The final limitation of these studies is that feature selection techniques could be used with ML algorithms to obtain a better statistical performance in the classification problem. The objectives of this study are presented below with bullet points based on the limitations of the previous studies [4,5] and the aim of our study.

1. To determine the optimal subsets obtained through the employed different feature selection methods.
2. To predict player positions using single-based models including Deep Neural Network (DNN), Random Forest

(RF), Gradient Boosting (GB) (the obtained subsets are used separately).

3. To predict player positions using a heterogeneous stacked ensemble model (SEM) (the obtained subsets are used separately).
4. To compare the performance of the base models and the proposed heterogeneous SEM in terms of accuracy, confusion matrix, and receiver operating characteristic (ROC).

As can be understood from the objectives presented above, the main aim of this study is to determine footballers' positions, creating a new hybrid stacked ensemble learning model. The combination of feature selection and stacked model creates the hybrid model. Feature selection techniques reduce overfitting, improve the model accuracy, and reduce analysis time in the classification/prediction problems. Four different statistical feature selection techniques (Information Gain, Gain Ratio, ReliFF, Chi-square) have been used to determine the most appropriate technique. Detailed information on the used feature selection approach is presented in Sect. 2.2. On the other hand, a SEM combines the predictions from more than one single-based machine learning algorithm on the same dataset. The combination makes it stronger compared to single-based algorithms. Thus, in our study, the stacked approach was preferred in the classification of footballers' positions. Random Forest, Gradient Boosting, and Deep Neural Networks algorithms were employed in this stacked model as single-based algorithms. The creation of a new stacked model is the main contribution to this study. A detailed description of the SEM and the employed single-based algorithms in the creation of the proposed stacked model is given in Sect. 2.3. Additionally, the performance comparison of the employed Random Forest, Gradient Boosting, and Deep Neural Networks in the footballer positioning based on statistical metrics is another contribution of this study. Moreover, determining the optimal feature selection technique to improve the performance of the employed algorithms in the positioning footballer is the final contribution in this study.

The structure of this paper is organized as follows. In Sect. 2, the proposed models are explained. In Sect. 3, the results obtained from these models are explained and discussed. In Sect. 4, the contribution of the study and future studies are presented.

## 2 Proposed Stacked Ensemble Machine Learning Model

In recent years, the increasing complexity and difficulty of real-world problems have led to the need for more reliable models, algorithms, optimization techniques, and meta-heuristic optimization algorithms [6]. Thus, various

meta-heuristic optimization algorithms [7–10], meta-learner models (ensemble), transfer learning studies [11], feature selection studies [12–14], and fine-tuning studies [15] have been carried out to produce reliable models. Additionally, a new meta-learner model (stacked ensemble learning model) is created for positioning footballers in this study. SEMs have been used in different fields, especially in recent years, and have achieved high performance. The analysis of football data with these models also offers various opportunities to both researchers, sports managers, and sports fans. Researchers can achieve higher performance by researching ensemble models, which may or may not have been carried out before. Football managers can achieve successful results in different lanes with the results of these models. For example, the football player position estimation performed in this study can predict the possible different positions of a football player and allow the setup in the game to be changed dynamically. Similarly, higher-performance predictions can be made by applying these models to predict league rankings. Managers can build their transfer policies on this. Betting companies can set the betting rates with the high performance of these models, or game companies can make the characters in the games learn with these models. As huge amounts of data flow in football, it has become a necessity to test the performances of ensemble models on these data. The model proposed in this study is presented as an answer to one of these needs. Therefore, as highlighted in Sect. 1, this study proposed a new hybrid stacked ensemble learning approach for the classification of footballers' positions.

The general methodology of our study is illustrated in Fig. 1 including three main stages. The first stage is data pre-processing which includes two parts. They are data cleaning (see Sect. 2.1) and feature selection (see Sect. 2.2). Data are split as the training (80%) and test set (20%) after the data cleaning and feature selection parts are completed. The second stage is related to the creation of the employed models. The single-based algorithms and the SEM are created for positioning footballers (see Sect. 2.3). 10-fold cross-validation is used for the evaluation of the proposed models. Finally, in the third stage, scoring metrics are presented to evaluate the performances of the employed algorithms (see Sect. 2.3.5). A detailed presentation about these three stages is given in the rest of this section.

## 2.1 Dataset and Data Pre-processing

The FIFA 2019 data<sup>1</sup> was used in the study. This dataset consists of 89 columns and 18,207 rows. In the data pre-processing process, initially, redundant features were removed from the dataset such as “id,” “name,” “photo,” “flag,” “club logo,” etc. Secondly, the Goalkeeper (GK) fea-

<sup>1</sup> <https://www.kaggle.com/karangadiya/fifa19>.

**Table 1** Position-class transformation

Position	Class	Position	Class	Position	Class
ST	3	CAM	2	R2	2
CF	3	RM	2	RB	1
LW	3	CM	2	CB	1
RW	3	LM	2	LB	1
RS	3	CDM	2	LWB	1
LS	3	LAM	2	RWB	1
RF	3	LDM	2	L1	1
LF	3	RDM	2	R1	1
L2	2	RAM	2	3B	1

ture was also removed from the dataset. The reason behind not using of GK feature is that this study aims to classify the defender, midfielder, and forward player features. Another reason for removing the GKs from the data set is to increase the reliability of the study. In other words, the GK characteristics are considerably different from the characteristics of the defender, midfielder, and striker features. This makes it easier for the model to learn this position and affects the reliability of the model. Thirdly, rows related to GKs were removed from the dataset. Fourthly, rows that include missing/blank values were also excluded from the dataset. The last step of the data pre-processing is normalization. In this step, the position data (which is the output data) were reduced to 3 classes according to the in-field layout as defender, midfielder, and striker. Figure 2 illustrates the sample on-site layout and positions. The meanings of the positions are presented in Fig. 2 are available at <https://fifauteam.com/fifa-21-positions/>. In this figure, there are 3 main positions except GK. These positions are “B” for “back,” “M” for “midfielder,” and “F” for forward. Also, there is another information inside the acronym of the word “R,” for “right,” “L” for “left,” and “C” for “center.”

There are differences in the data set for all three positions, including right, left, middle, and more different positions. The positions and class conversion in the FIFA 19 dataset are also presented in Table 1. After data normalization, the dataset consists of 60 columns and 16,097 rows.

## 2.2 Feature Selection

Feature selection methods are embedded, wrapper, and filter based. Filter-based selection methods are less time-consuming compared to the others which is one of its important advantages. Thus, the following filter-based feature selection methods are used in this study, namely Information Gain (IG), Gain Ratio, RelifF, and Chi-square. Information Gain can be named as mutual information. It is used to measure which feature in the dataset provides maximum

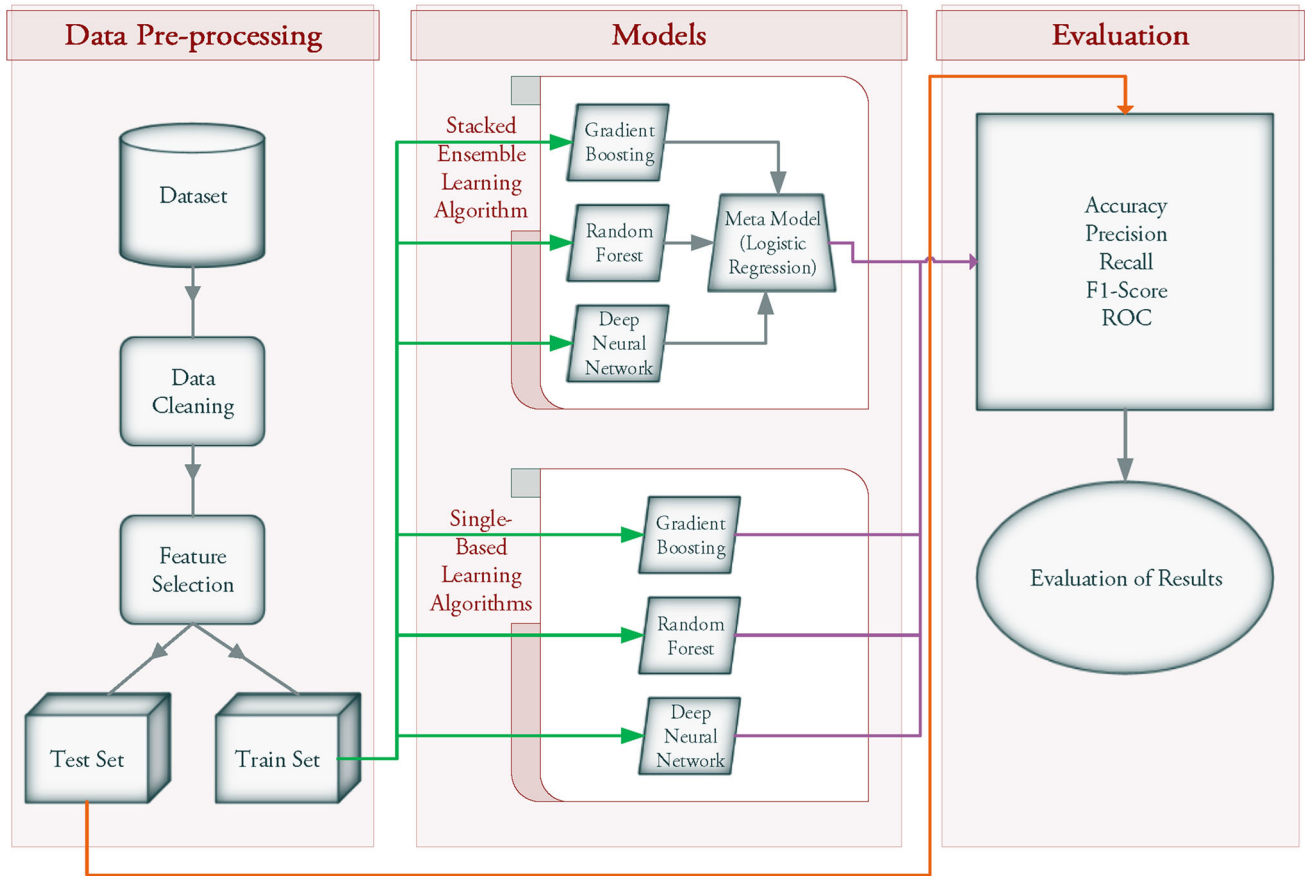


Fig. 1 Proposed footballer position classification approach

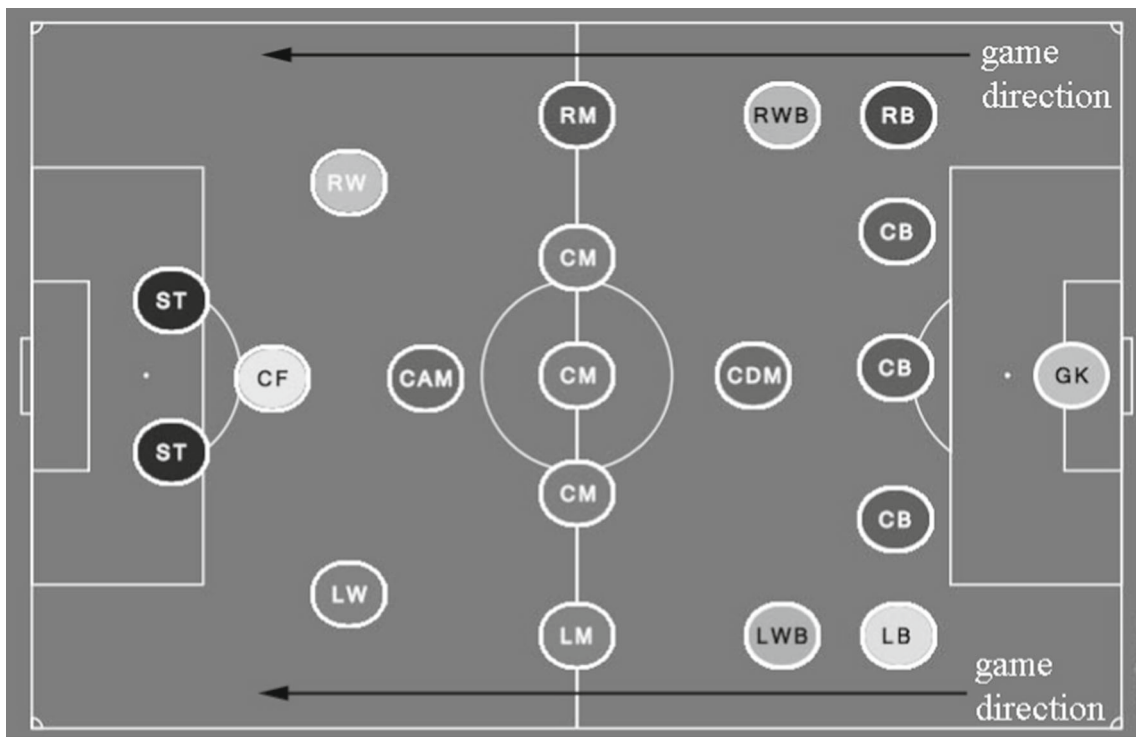


Fig. 2 Sample on-site layout and positions

information according to entropy [16]. Entropy plays an important role in the calculation of Information Gain. For each attribute in the data set, the Information Gain is calculated separately. However, it may not be an effective method for features having a large number of distinct values. Thus, the Gain Ratio is the modified version of the Information Gain [16]. The Gain Ratio is measured for class to evaluate a feature value. Also, it aims to reduce the bias of Information Gain [17]. ReliefF is the modified version of Relief. It finds one or more neighboring samples relating to each class [18]. Chi-square is one of the well-known filter-based feature selection methods for categorical features. The Chi-square value is measured between the target and each feature in terms of obtaining the top features [17]. It works on the statistical significance of the difference between the root node and the child nodes. It is calculated by summing the squares of the standardized differences between the observed and expected frequencies of the target variable.

## 2.3 Employed Machine Learning Algorithms

In this section, the algorithms used in the study are presented. The parameters used in the implementation of the algorithms and the structures of the models are explained. In addition, explanations regarding the performance indicators used are also given in this section. Python Programming Language in Jupyter Platform is used in this study.

### 2.3.1 Deep Neural Network

In this study, the structure of the proposed DNN consists of an input, three hidden layers, and an output layer. According to [19], a neural network can be accepted as a DNN if it includes more than one hidden layer. Thus, more extensive relationships are established from simple to complex data in DNN. Each layer tries to establish a relationship between itself and the previous layer. Thus, a more detailed examination of the inputs is made and more accurate decisions are made [20,21]. Different activation functions can be used when forming the structure of DNN. These functions may vary according to the type, structure, size, and model of the data. The activation function determines the output that the cell will produce in response to the input itself. A nonlinear function is usually selected [22]. Thus, the proposed approach can be considered as a DNN because of the used number of hidden layers. Figure 3 illustrates the structure of the proposed DNN. The number of neurons for the hidden layer is 40, 20, and 10, respectively. Also, the Sigmoid activation function has been used in these hidden layers. Different numbers of epochs and batch sizes have been tried to evaluate the DNN model for the classification of footballer positions. The batch size is determined as five in this model, and since the best accuracy

value without overfitting was obtained with 100 epochs, the number of epochs was set as 100.

### 2.3.2 Random Forest

RF is a tree structure that is created using a large number of DT and has branches according to randomly selected parameters [23]. It is also an effective bagging ensemble learning algorithm. In the RF algorithm, each DT makes a class prediction, and the results are aggregated by voting to determine the class taking the most votes. The results of these unrelated DTs can produce a prediction that is more accurate than any of the individual predictions. In the employed RF algorithm, the number of trees is set to 100, and the subsets are not split when they are less than five.

### 2.3.3 Gradient Boosting

One of the effective boosting ensemble learning algorithms is GB, and it can be applicable for regression and classification problems [24]. The GB ensemble algorithm is constructed using DTs. Ensemble GB algorithm consisting of sequentially added DTs. Then, it trains weak learners by aiming to minimize the loss function. In the employed GB ensemble algorithm, the number of estimators is set to 100 while the learning rate of the model is 0.1.

### 2.3.4 Stacked Ensemble Algorithm

Stacking is an ensemble learning approach created using more than one ML algorithm. Two levels (level 0 and 1) are used in any SEM. Single-based models are used in Level-0. On the other hand, the meta-model is used in Level-1. In level 1, predictions of based models obtained from Level-0 are combined through the meta-model. In other words, the outputs of the based models are used as an input value for the meta-model (Level-1). Then, the final classification or regression score is obtained through the stacked ensemble learning algorithm [25].

Stacking is based on the idea that instead of using trivial functions (such as hard voting) to aggregate the predictions of all predictors in an ensemble, it uses training a model to perform this aggregation [26]. Figure 4 shows an example of an ensemble performing a regression task on a new instance. Each of the (level-0) three predictors predicts a different value (8.6, 8.2, and 8.0), and then, the final predictor (called a meta-learner) takes these predictions as inputs and makes the final prediction (8.4).

To train the stacking, a common approach is to use a hold-out set. In this approach, first, the training set is split into two subsets. The first subset is used to train the predictors in the first layer. Next, the first layer's predictors are used to make predictions on the second (held-out) set. This ensures

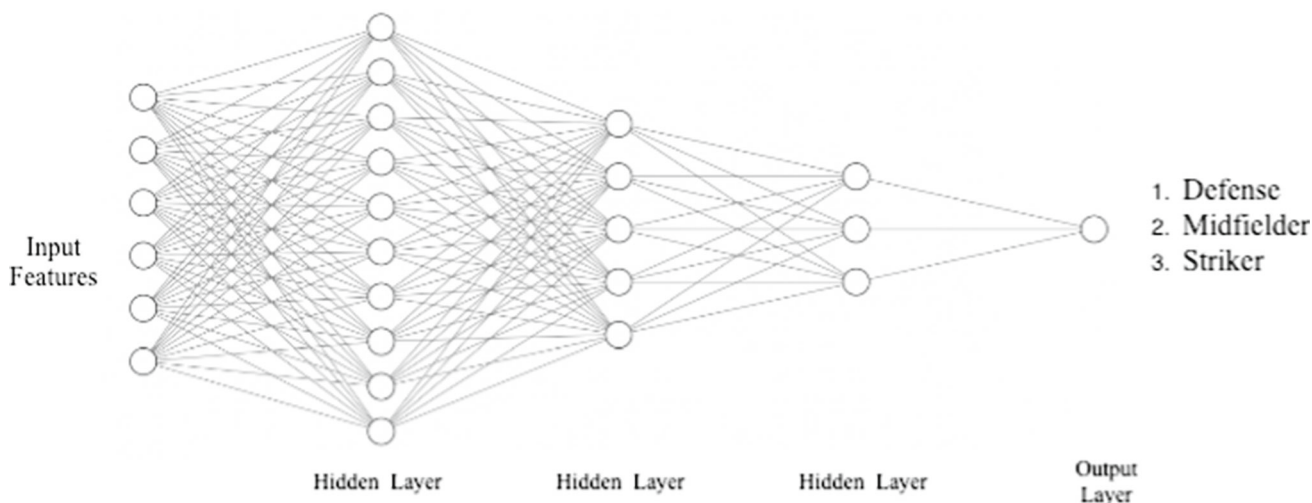


Fig. 3 Structure of the proposed DNN

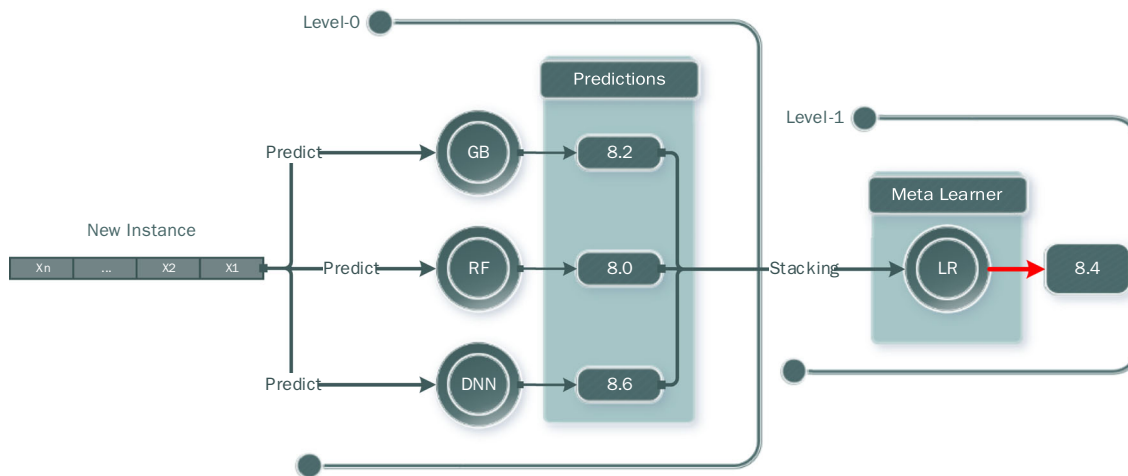


Fig. 4 Generation of predictions using a stacking predictor

that the predictions are “clean,” since the predictors never saw these instances during training. For each instance in the hold-out set, there are three predicted values. Then, a new training set using these predicted values can be created as input features (which makes this new training set 3D) while keeping the target values. The model is trained on this new training set, so it learns to predict the target value, given the first layer’s predictions. It is possible to train several different meta-learner this way (e.g., one using Linear Regression, another using Random Forest Regression), to get a whole layer of staking [26].

In this study, DNN, RF, and GB algorithms are used as based models in Level-0. Then, the LR algorithm is used as a meta-model in Level-1. It should be noted that a variety of ML algorithms (AdaBoost, DNN) were initially used in Level-1 to improve the performance of the stacked model as much as possible. However, the best accuracy performance

was obtained through the LR, and so it is used in Level-1. Moreover, Table 2 shows the mathematical notations of the employed ML models.

### 2.3.5 Scoring Metrics

Accuracy, Precision, Sensitivity, F1-Score scoring metrics are used in the evaluation of the employed algorithms. Formulas of the used scoring parameters are presented below [27].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Sensitivity(Recall)} = \frac{TP}{TP + FN} \tag{3}$$

**Table 2** Mathematical notations of the employed DNN, RF, GB, and Stacked Learner

Model name	Parameters
Deep Neural Network (DNN)	Hidden layer number: 3 Activation function in hidden layers: Sigmoid Batch size: 5 Number of epochs: 100
Random Forest (RF)	Number of trees: 100 The subsets were not split when they are less than 5
Gradient Boosting (GB)	The number of estimators: 100 The learning rate of the model: 0.1
Stacked Learner	Level 0 The parameters used in DNN, (DNN, RF, GB) RF and GB are used in level 0. Level 1 Penalty: L2 (ridge regression) (Logistic regression) Random state: 0 Maximum number of iterations: 100

$$F1 - Score = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

where TN: True Negative, TP: True Positive, FN: False Negative, FP: False Positive.

Accuracy measure refers to the ratio of the number of correct predictions to the number of input samples. Precision is a measure of how accurately all classes are predicted. It is also known as positive predictive value. Sensitivity (Recall) can be defined as the ratio of the total number of correctly classified positive examples divided by the total number of positive examples. F1-score is the harmonic mean of Precision and Recall. It is a measure of how well the classifier performs and is often used to compare classifiers [22,28].

### 3 Results and Discussion

In this section, firstly, the feature selection results based on the same dataset are presented in Sect. 3.1. Then, the statistical performances of the employed models are compared in Sect. 3.2. In addition to these, the ROC curve results of the models are given in Sect. 3.3. Summary of findings is presented to reveal the efficiency of the proposed hybrid stacked ensemble approach in Sect. 3.4. Later, in Sect. 3.5, previous studies and the proposed stacked model are compared in terms of the structure of the studies. Finally, the limitations of the study are stated in Sect. 3.6.

#### 3.1 Feature Selection Results

As presented in Sect. 2.3, four different feature selection methods (Information Gain, Gain Ratio, ReliefF, Chi-square) were used to measure their efficiency in the footballer position classification. Table 3 presents the selected top ten

features based on the used feature selection methods. The meaning of the selected features is available at <https://www.kaggle.com/karangadiya/fifa19>. Optimal subsets are determined based on these four-feature selection methods, thus fulfilling objective 1 of this study presented in Introduction of this study.

#### 3.2 Comparison of Model Performances

Table 4 shows the model performances based on the feature selection approaches. The stacked-based ensemble learning model provides better accuracy performance in the footballer position classifier using ReliefF and Chi-square features selection techniques than the others. On the other hand, the GB algorithm yields the best accuracy classification performance as a result of using the Information Gain and Gain Ratio feature selection techniques. Overall, the best classification accuracy performance (83.9%) is obtained through the combination of the Chi-square feature selection technique and the stacked-based ensemble learning model. It means that the Chi-square feature selection technique can be considered a more effective feature selection technique compared to the others in the classification of footballer position. Moreover, the second-best accuracy performance (82.9%) is obtained with the combination of the ReliefF feature selection technique and the stacked-based ensemble learning model. On the other hand, the worst accuracy performance (82.2%) of the stacked ensemble learning model is obtained using the features selected based on the Gain Ratio.

Many researchers highlight that DNN generally provides more beneficial statistical performances in classification and regression problems than bagging and boosting algorithms [29,30]. The use of neurons and the number of iterations in the neural network make it more effective than the others. However, the proposed DNN did not provide a convincing

**Table 3** Selected top ten features based on Information Gain, Gain Ratio, ReliefF, Chi-square

Feature No	Chi_Square	ReliefF	Information Gain	Gain Ratio
1	LB	RB	SlidingTackle	Skill Moves
2	RB	LB	StandingTackle	SlidingTackle
3	LDM	SlidingTackle	Finishing	StandingTackle
4	CDM	StandingTackle	Marking	Finishing
5	RDM	Interceptions	Interceptions	Marking
6	LWB	RWB	Skill Moves	Interceptions
7	RWB	LWB	Positioning	Positioning
8	LS	Marking	RB	Dribbling
9	ST	RDM	LB	Volleys
10	RS	CDM	Dribbling	LongShots

**Table 4** Model performances based on the feature selection approaches

Feature selection methods	Algorithm	Accuracy	Precision	Recall	F1-Score
Information Gain	DNN	74.2	74.2	74.2	73.7
	RF	73.9	73.9	73.9	73.9
	Gradient Boosting	82.8	82.8	82.8	82.8
	Stacked Learner	82.6	83.6	83.1	81.8
Gain Ratio	DNN	75.2	75.3	75.2	75.1
	RF	74.1	74.2	74.1	74.1
	Gradient Boosting	82.6	82.5	82.6	82.5
	Stacked Learner	82.2	83.1	82.8	82.7
ReliefF	DNN	76.1	76.6	76.1	75.7
	RF	74.8	74.8	74.8	74.8
	Gradient Boosting	82.5	82.5	82.5	82.5
	Stacked Learner	82.9	83	82.4	81.2
Chi-square	DNN	73	74.6	73	72.9
	RF	77	77	77	77
	Gradient Boosting	82	82.1	82	82
	Stacked Learner	83.9	84.3	83.2	83.4

accuracy performance in the footballer position classification. For example, the best accuracy performance of DNN was obtained with the combination of the ReliefF feature selection technique (76.1%). Also, this accuracy classification score of the DNN is 7.8% less than the classification score of the employed stacked ensemble learning model. From this result, it can be inferred that feature selection techniques play an important role in classification problems. Additionally, a stacked-based ensemble algorithm cannot always provide better accuracy performance compared to the others even if the best accuracy performance is obtained through the proposed heterogeneous stacked ensemble learning model. As shown in Table 4, footballer positions are predicted applying single-based models including Deep Neural Network, Random Forest, Gradient Boosting, and a heterogeneous SEM, thus fulfilling objectives 2 and 3 of this study. The third objective refers to the main contribution to this study.

Generally, stacked ensemble learning algorithms do not provide the best performance in the positioning footballer based on the results given in Table 4. However, in the creation of the stacked ensemble learning model, we carried out a case study to determine the algorithms used in level 0 of the stacked model. In this case, Decision Trees, Random Forest, Support Vector Machines, Deep Neural Networks, Gradient Boosting, and Random Forest Algorithms are used individually for positioning footballers. Then, the Deep Neural Network, Gradient Boosting, and Random Forest Algorithms are selected based on the adjusted  $R^2$  value. In other words, the adjusted  $R^2$  value is used to select appropriate algorithms used in level 0 of the stacked model. The use of the appropriate models with the feature selection methods made the proposed new stacked model superior to the others used in this study.

We fit four different models with different features obtained through the feature selection methods. The accu-



racies of the stacked model fit are between 82.2% and 83.9% as shown in Table 4. Therefore, this proposed new stacked model is considerably robust. On the other hand, the other employed algorithms (DNN, RF, GB) can be also considered robust because of the small differences among evaluation scores as given in Table 4.

The rest of this section gives information about the strengths and weaknesses of the proposed new SEM. The proposed SEM is created using the single-based DNN, RF, and GB algorithms which have well-adjusted  $R^2$  performance. These single-based algorithms provided convincing classification performance for positioning footballers. However, the proposed new SEM has better classification performance than any single-based model used in the stacked model with the Chi-square and ReliF feature selection techniques, which is its strength. On the other hand, the training time of the proposed stacked model is longer than the single-based models and requires more memory than the others, which is the weakness of the proposed SEM. For instance, the training time of GB, RF, DNN, and the new SEM in the use of Chi-square feature selection is 8, 10, 10, and 188 seconds, respectively. In other words, the training time of the proposed stacked model is approximately 19 times more than the single-based algorithms. Moreover, it should be highlighted that the Orange Data Mining platform (<https://orangedatamining.com>) is used in the feature selection process. Thus, this process took the same time for all feature selection methods used in this study (1 second).

### 3.3 ROC Curve Results

Figure 5 presents the ROC results of the proposed SEM and the single-based learners based on the used feature selection approaches. It should be noted that the use of the ROC value for the evaluation of any model is useful because it does not depend on the class distribution. The stacked ensemble learning approaches yielded better ROC results compared to the others based on each feature selection approach separately. Also, the stacked ensemble learning approach was achieved to provide the best ROC result using the features selected through the Chi-square (95.1%). In Table 4 and Fig. 5, performance scores of the employed algorithms are presented in terms of accuracy, confusion matrix, and ROC, thus fulfilling the fourth objective of this study.

As shown in Fig. 5, the GB algorithm obtained the second-best result after the SEM. Although the DNN algorithm has achieved convincing performances in different studies, especially in the last decades, it could not achieve the same success in this research. Another interesting result is that the RF algorithm has the lowest performing algorithm in this study when used as a single-based algorithm.

### 3.4 Summary of Findings

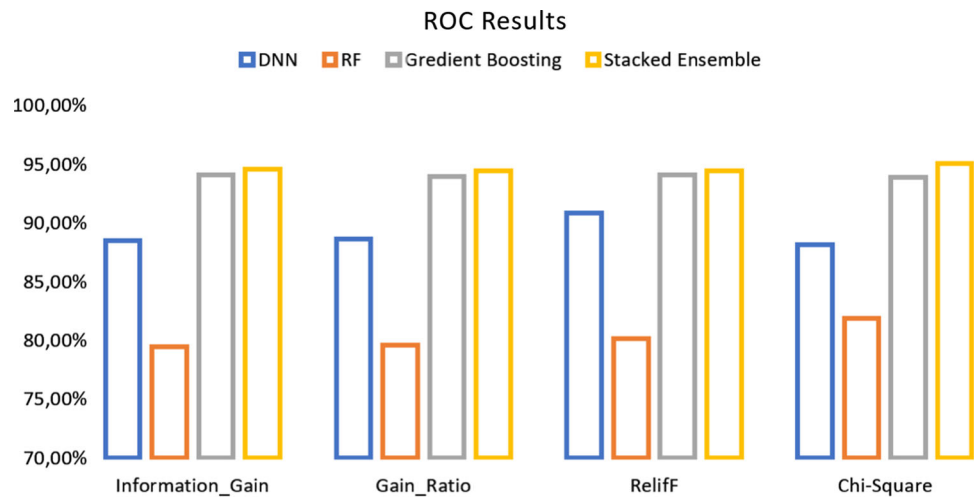
Chi-square can be considered the most effective statistical feature selection technique when it is used with the proposed new stacked ensemble learning model. On the other hand, the GB ensemble learning algorithm achieved better classification performance compared to the employed Stacked Learner in the use features obtained through the Gain Ratio and Information Gain. This finding emphasizes the efficiency of the features selection techniques in the classification problems. The stacked ensemble learning model has the best performance as specified in Sect. 3.2 based on the results given in Table 4. On the other hand, GB ensemble learning is the second most effective and robust model. This finding shows that boosting ensemble learning is better than the bagging and DNN algorithms in the positioning of footballers.

### 3.5 Comparison of Previous Studies and Current Models

As highlighted in Sect. 1, studies on the footballer position prediction are limited (see Table 5) and most of the studies in the literature have focused on different aims including prediction of match results, prediction of league ranking, creation of game setup, analysis of player performances (see Table 6).

As shown in Table 5, even if the study (proposed by [4]) succeeded to obtain 99% accuracy for predicting player position using the Bayesian Networks, the sample size is small (100). Also, the test size is accepted as 1% while the rest of this data is accepted as a training set (99%). In this sense, the obtained accuracy result can be considered as bias. Moreover, the performance of the employed models in this study could be examined using different test sizes to justify the efficiency of the models for predicting player position. In a different study [5], case-based reasoning and nearest-neighbor algorithms are used to predict footballer positions. According to this study, the similarity rate between the new problem and the problems saved in the database should not be less than 80% in the reuse process of case-based reasoning. This means that if the similarity rate is more than 80% between the new problem and any problem in the database, the solution of the saved problem in the database is used to provide a solution to the new problem. In this sense, the proposed approach obtained 97.22% accuracy in the classification of footballer position. However, this study has two limitations. The first one is that the sample size of the data is small (36). The second limitation is that the used threshold value (80%) may be a reason to assign the footballer positions incorrectly. [31] carried out a case study to determine the optimal threshold value to be used in studies related to the CBR. The case study revealed that the optimal threshold value can be accepted as 90% to obtain convincing, reliable, and consistent classification results. In this sense, the threshold value could be

**Fig. 5** ROC results of the proposed approach and the single-based learners



increased at least 90%, or a case study could be carried out to obtain the optimum threshold value. Thus, prediction of the footballer position can be more consistent and reliable.

As specified in the above-mentioned studies, small samples are a limitation in the ML models. Due to this limitation, FIFA game data are used in this study because it has 16,097 samples. On the other hand, the above-mentioned studies used single-based algorithms. However, the efficiencies of the ensemble learning algorithms (bagging, boosting, and stacked) have never been measured in the prediction of footballer positions, to the best of the authors' knowledge. Thus, a new heterogeneous stacked ensemble learning model is proposed to predict footballer positions which is the main contribution of this study.

Even if the proposed studies by [4] and [5] succeed to obtain better accuracy results compared to the proposed SEM, the accuracy results of these studies can be biased due to the limitations presented above. Moreover, the study proposed by [5] used the threshold value of 80%. However, if the similarity value is less than the threshold, the manager has to manually assign the footballer positions which is an extra workload for the manager. Also, a point that should not be forgotten is that the manager (human classifier) may perform less efficiently than ML models. Overall, we believe that this study solved the limitations of these studies and proposed a novel approach for this field using stacked-based ensemble learning.

Apart from the studies on positioning footballers shown in Table 5, ML methods have been employed for different purposes. As given in Table 6, these studies can be examined under topics such as match result prediction, estimation of factors affecting match results, prediction of league standings, and analysis of the performances of football players.

As provided in Table 6, different machine learning algorithms have been proposed for different purposes. For instance, a Lasso penalized regression algorithm provided

the best statistical performance (75%) compared to the other employed algorithms (Ridge, Elastic, Neural Network, Random Forest, etc.) in the football match result prediction [32]. In a different study, an Extreme Gradient Boosting (EGB) learning algorithm is employed for the estimation of factors affecting match results, and it achieved 89.6% accuracy score [33]. While the algorithm used in league ranking prediction is ANN, the highest performance in this algorithm was obtained with 99% accuracy [34–36]. With the ML algorithms used in the analysis of the performances of the football players, the accuracy rate was 80% by [37], 82% by [38], and 94% by [39].

As shown in Table 6, the stacked ensemble learning approach has not yet been employed for any topic. The proposed stacked model in this study could be used to improve the performances of the currently employed ML algorithms for different purposes. Ensemble learning models have become frequently used in recent years. The most important reason behind this is that ensemble learning models are more likely to achieve higher performance than single-based machine learning algorithms. Also, ensemble learning models combine more than one single-based algorithm which makes them more effective compared to the single-based ML algorithms in classification and regression problems. Thus, they can show higher success rates. The use of ensemble learning models on football data can bring higher success to the ML studies made for different purposes in Table 6.

### 3.6 Limitations of this Study

- The first limitation is the use of FIFA game data instead of real data. Generally, footballer data consist of parameters such as training status, match frequency, injury, etc. Also, these data are recorded by technical teams with various technological equipment. However, important information about the players is not shared by the technical teams

**Table 5** Previous studies and the proposed approach for prediction player position

Author(s)	Method	Result
Razali et al. [4]	Bayesian networks decision tree nearest neighbor	Accuracy: Bayesian networks: 99% Decision tree: 98% Nearest Neighbor: 97%
Maulany et al. [5]	Case base reasoning, nearest neighbor	Accuracy: 97.22%
The proposed stacked model	Hybrid: two steps step 1: Chi-square feature selection step 2: stacked ensemble level 0: DNN, RF, Gradient Boosting level 1: logistic regression	Accuracy: 83.9%

**Table 6** Studies for different purposes using ML methods

Author(s)	Method	Result
<i>Match result prediction</i>		
South and Egros [32]	Ridge Regression, Lasso, Elastic Network, NN, RF, kNN, Stochastic Gradient Boosting (SGB), Bayesian Regression	Lasso: 75% BR: 72.2 % SGB: 71.7% RF: 78.9%
Hubáček et al. [40]	Gradient Asst. Tree	Ranked Prob. Score (RPS): 0.2063
Baboota and Kaur [41]	GB	RPS: 0.2012
<i>Estimation of factors affecting match results</i>		
Baboota and Kaur [41]	GB	RPS: 0.2156
Karabiyik and Zaim [42]	Discriminant analysis (DA), Mathematics models (MM)	DA: 65.4%, MM: 71%, 74%, 77%, 79%
Geurkink et al. [33]	Extreme Gradient Boosting (EGB)	Accuracy: 89.6%
Bilek and Ulas [43]	One-way ANOVA, Tukey HSD, k-Means, DT	DT (Best): 64.8%
<i>Prediction of league standings</i>		
Kılıçet al. [36]	ANN	Accuracy: 94%
Hasan et al. [34]	ANN	Accuracy: 99%
Hasan et al. [35]	ANN	Accuracy: 99%
<i>Analysis of the performances of football players</i>		
Almulla and Alam [37]	LR	Accuracy: 80%
Ćwiklinski [38]	RF, NB, AdaBoost	Accuracy: 82%
Ćwiklinski et al. [39]	Adaptive back propagation NN	Accuracy: 94%

with people outside the club. Thus, publicly available football players’ data through platforms such as the game industry (FIFA game dataset) or transfer exchange are used in this study.

- The second limitation is that the GK position was not included in the data. The reason is that GK features (parameters) are distinguishable from the other positions because GK can easily affect the reliability of the models compared to the other positions.
- The third limitation is that a case study has not been carried out to determine the optimal machine learning model

as a meta-model yet. In this study, Logistic Regression is used as a meta-model in the employed SEM. Due to the use of Logistic regression in most of the classification problems, Logistic Regression was applied as a meta-model.

### 4 Conclusion and Future Directions

Determining the football players’ position is a challenge for coaches, and they mostly decide the footballer positions as

a result of their observations. However, the determination of footballer positions based on coaches' observation is not an effective way. In this sense, ML algorithms can be used to help the coaches in the determination of the footballer positions. Thus, this study proposes a stacked ensemble learning model to predict footballer positions using FIFA'19 game data. Initially, we used four filter-based feature selection techniques (Information Gain, Gain Ratio, Relief, and Chi-square) to select optimal feature subsets and also measure the efficiency of these techniques. Then, a stacked ensemble learning model is used in this study. Deep Neural Network, Random Forest, and Gradient Boosting algorithms are used as based models in Level-0. Then, the Logistic Regression algorithm is used as a meta-model in Level-1. The proposed stacked model achieved 83.9% accuracy using feature subset obtained through the Chi-square feature selection technique. Also, the proposed model can be considered as a hybrid approach due to the combination of the feature selection technique and the SEM, which is the main contribution of this study. Also, the classification performance of bagging (Random Forest), boosting (Gradient Boosting), and DNNs is compared in the determination of footballer positions, which is another contribution of our study. Four different statistical feature selection methods are used to improve the performance of the proposed SEM. Then, Chi-square is determined as an optimal feature selection method, which is the final contribution of our study.

Logistic Regression is used as a meta-model in the creation of the SEM due to the use of most classification problems. However, a case study could be carried out to determine an optimal algorithm to be used as a meta-model which is the limitation of this study. The findings reflect the importance of the Chi-square feature selection technique and the stacked ensemble learning model in the prediction of footballer positions. As a further study, different stacked ensemble learning models with a variety of meta-models can be proposed to predict football players' marketing value and league rankings of football teams. Moreover, Deep Belief Networks (DBNs) can be used in level 1 of the SEMs as a future study.

**Funding** This work was not funded.

## Declarations

**Conflict of interest** No potential conflict of interest was reported by the authors.

## References

- Tümer, A.E.; Koçer, S.: Prediction of team league's rankings in volleyball by artificial neural network method. *Int. J. Perform. Anal. Sport* **17**(3), 202–211 (2017). <https://doi.org/10.1080/24748668.2017.1331570>
- Özdemir, E.; Ballı, S.: Türkiye erkekler basketbol ligi maç sonuçlarının makine Öğrenmesi yöntemleri ile tahmini. *Mühendislik Bilimleri ve Tasarım Dergisi* **8**(3), 740–752 (2020)
- Fernandes, C.J.; Yakubov, R.; Li, Y.; Prasad, A.K.; Chan, T.C.Y.: Predicting plays in the national football league. *J. Sports Anal.* **6**(1), 35–43 (2020). <https://doi.org/10.3233/JSA-190348>
- Razali, N.; Mustapha, A.; Yatim, F.A.; Aziz, R.A.: Predicting player position for talent identification in association football. In: *IOP Conference Series: Materials Science and Engineering*, vol. 226, p. 012087. IOP Publishing (2017)
- Maulany, G.J.; et al.: Application of case based reasoning and nearest neighbor algorithm for positioning football players. *Int. J. Mech. Eng. Technol. (IJMET)* **9**(13), 258–265 (2018)
- Abualigah, L.; Diabat, A.; Mirjalili, S.; Elaziz, M.A.; Gandomi, A.H.: The arithmetic optimization algorithm. *Comput. Methods Appl. Mech. Eng.* **376**, 113609 (2021a)
- Abualigah, L.; Yousri, D.; Elaziz, M.A.; Ewees, A.A.; Al-qaness, M.A.A.; Gandomi, A.H.: Aquila optimizer: a novel meta-heuristic optimization algorithm. *Comput. Ind. Eng.* **157**, 107250 (2021b)
- Abualigah, L.; Elaziz, M.A.; Sumari, P.; Geem, Z.W.; Gandomi, A.H.: Reptile search algorithm (RSA): a nature-inspired meta-heuristic optimizer. *Expert Syst. Appl.* **191**, 116158 (2022)
- Agushaka, J.O.; Ezugwu, A.E.; Abualigah, L.: Dwarf mongoose optimization algorithm. *Comput. Methods Appl. Mech. Eng.* **391**, 114570 (2022)
- Oyelade, O.N.; Ezugwu, A.E.-S.; Mohamed, T.I.A.; Abualigah, L.: Ebola optimization search algorithm: a new nature-inspired meta-heuristic optimization algorithm. *IEEE Access* **10**, 16150–16177 (2022). <https://doi.org/10.1109/ACCESS.2022.3147821>
- Savaş, S.: Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures. *Arab. J. Sci. Eng.* **47**, 2201–2218 (2022). <https://doi.org/10.1007/s13369-021-06131-3>
- Karakış, R.; Güler, İ.; Işık, A.H.: Feature selection in pulmonary function test data with machine learning methods. In: *2013 21st Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4 (2013). <https://doi.org/10.1109/SIU.2013.6531578>
- Buyrukoğlu, S.: Early detection of Alzheimer's disease using data mining: comparison of ensemble feature selection approaches. *Konya Mühendislik Bilimleri Dergisi* **9**(1), 50–61 (2021)
- Abualigah, L.M.Q.; et al.: Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering. Springer (2019)
- Too, E.C.; Yujian, L.; Njuki, S.; Yingchun, L.: A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* **161**, 272–279 (2019)
- Porkodi, R.: Comparison of filter based feature selection algorithms: an overview. *Int. J. Innov. Res. Technol. Sci.* **2**(2), 108–113 (2014)
- Vanaja, S.; Kumar, K.R.: Analysis of feature selection algorithms on classification: a survey. *Int. J. Comput. Appl.* **96**(17), 28–35 (2014)
- Tang, J.; Alelyani, S.; Liu, H.: Feature Selection for Classification: A Review, pp. 37–64. CRC Press (2014). ISBN 9781466586741. <https://doi.org/10.1201/b17320>
- Canziani, A.; Paszke, A.; Cukurciello, E.: An analysis of deep neural network models for practical applications. *CoRR* (2016). [arxiv:1605.07678](https://arxiv.org/abs/1605.07678)
- Calp, M.H.: Use of Deep Learning Approaches in Cancer Diagnosis, pp. 249–267. Springer, Singapore, 2021. ISBN 978-981-15-6321-8. [https://doi.org/10.1007/978-981-15-6321-8\\_15](https://doi.org/10.1007/978-981-15-6321-8_15)
- Büttner, R.; Calp, M.H.: Covid-19 detection from lung tomography images using deep learning and machine learning methods. Technical report, EasyChair (2020)
- Savaş, S.; Topaloğlu, N.; Kazıcı, Ö.; Koşar, P.N.: Classification of carotid artery intima media thickness ultrasound images with deep



- learning. *J. Med. Syst.* **43**(8), 1–12 (2019). <https://doi.org/10.1007/s10916-019-1406-2>
23. Liu, Y.; Wang, Y.; Zhang, J.: New machine learning algorithm: random forest. In: Liu, B., Ma, M., Chang, J. (eds.) *Information Computing and Applications*, pp. 246–252. Springer, Berlin (2012). ISBN 978-3-642-34062-8
  24. Natekin, A.; Knoll, A.: Gradient boosting machines, a tutorial. *Front. Neurobotics* **7**, 21 (2013). ISSN 1662-5218. <https://doi.org/10.3389/fnbot.2013.00021>
  25. Buyrukoğlu, S.: Improvement of machine learning models' performances based on ensemble learning for the detection of Alzheimer disease. In: 2021 6th International Conference on Computer Science and Engineering (UBMK), pp. 102–106 (2021). <https://doi.org/10.1109/UBMK52708.2021.9558994>
  26. Aurélien, G.: *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media (2019)
  27. Hossin, M.; Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **5**(2), 1 (2015)
  28. Awadh, K.; Akbaş, A.: Intrusion detection model based on TFIDF and C4.5 algorithms. *Politeknik Dergisi* **24**(4), 1691–1698 (2021). <https://doi.org/10.2339/politeknik.693221>
  29. Hung, C.-Y.; Chen, W.-C.; Lai, P.-T.; Lin, C.-H.; Lee, C.-C.: Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3110–3113. IEEE (2017)
  30. Ul Haq, A., Li, J., Memon, M.H., Khan, J., Ud Din, S., Ahad, I., Sun, R., Lai, Z.: Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for prediction of Parkinson disease. In: 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pp. 101–106. IEEE (2018)
  31. Buyrukoglu, S.; Batmaz, F.; Lock, R.: Improving marking efficiency for longer programming solutions based on a semi-automated assessment approach. *Comput. Appl. Eng. Educ.* **27**(3), 733–743 (2019). <https://doi.org/10.1002/cae.22094>
  32. South, C.; Egos, E.: Forecasting college football game outcomes using modern modeling techniques. *J. Sports Anal.* **6**(1), 25–33 (2020)
  33. Geurkink, Y.; Boone, J.; Verstockt, S.; Bourgois, J.G.: Machine learning-based identification of the strongest predictive variables of winning and losing in Belgian professional soccer. *Appl. Sci.* **11**(5) (2021). ISSN 2076-3417. <https://doi.org/10.3390/app11052378>
  34. Hasan, A.; Aktuğ, Z.B.; Kılıç, F.: Türkiye süper lig sezon sonu takım sıralamasının geliştirilen yapay sinir ağları modeli ile tahmin edilmesi. *Spor ve Performans Araştırmaları Dergisi* **11**(3), 258–268 (2020)
  35. Hasan, A.; Serkan, İ.; Aktuğ, Z.B.; Kılıç, F.: Yapay sinir ağları modeli ile İspanya futbol ligi (la lıga) sezon sonu takım sıralamasının tahmin edilmesi. *SPORMETRE Beden Eğitimi ve Spor Bilimleri Dergisi* **19**(1), 147–155 (2021)
  36. Kılıç, F.; Hasan, A.; Aktuğ, Z.B.: Futbolda yapay sinir ağları modeli ile lig sıralaması tahmini. *Uluslararası Güncel Eğitim Araştırmaları Dergisi* **6**(2), 379–391 (2020)
  37. Almulla, J.; Alam, T.: 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>
  38. Ćwiklinski, B.; Gietczyk, A.; Choraś, M.: Who will score? A machine learning approach to supporting football team building and transfers. *Entropy* **23**(1), 90 (2021)
  39. Khan, M.A.; Habib, M.; Saqib, S.; Alyas, T.; Khan, K.M.; Ghamdi, M.A.A.; Almotiri, S.H.: Analysis of the smart player's impact on the success of a team empowered with machine learning. *Comput. Mater. Continua* **66**(1), 691–706 (2021). <https://doi.org/10.32604/cmc.2020.012542>. ISSN 1546-2226.
  40. Hubáček, O.; Šourek, G.; Železný, F.: Learning to predict soccer results from relational data with gradient boosted trees. *Mach. Learn.* **108**(1), 29–47 (2019)
  41. Baboota, R.; Kaur, H.: Predictive analysis and modelling football results using machine learning approach for English premier league. *Int. J. Forecast.* **35**(2), 741–755 (2019). ISSN 0169-2070. <https://doi.org/10.1016/j.ijforecast.2018.01.003>
  42. Karabıyık, A., Zaim, A.H.: Futbol maç sonuçlarına etki eden faktörlerin tespiti ve faktörlerin etki paylarının hesaplanarak maç sonuçlarının tahmin edilmesi. II. Mühç Araştırmacılar Kongresi (2005)
  43. Bilek, G.; Ulas, E.: Predicting match outcome according to the quality of opponent in the English premier league using situational variables and team performance indicators. *Int. J. Perform. Anal. Sport* **19**(6), 930–941 (2019). <https://doi.org/10.1080/24748668.2019.1684773>

