



Exploring polynomial classifier to predict match results in football championships



Rodrigo G. Martins^{a,c}, Alessandro S. Martins^a, Leandro A. Neves^b, Luciano V. Lima^c,
Edna L. Flores^c, Marcelo Z. do Nascimento^{d,*}

^aIFTM, r. Belarmino Vilela Junqueira S/N, 38305-200, Ituiutaba, MG, Brazil

^bUNESP - DCSS, r. Cristóvão Colombo 2265, S. J. Rio Preto, 15054-000, SP, Brazil

^cUFU - FEELT, av. João Neves de Ávila 2121, BLX, Uberlândia, 38400-902, MG, Brazil

^dUFU - FACOM, av. João Neves de Ávila 2121, BLB, Uberlândia, 38400-902, MG, Brazil

ARTICLE INFO

Article history:

Received 8 October 2016

Revised 17 April 2017

Accepted 18 April 2017

Available online 19 April 2017

Keywords:

Polynomial classifier
Football championship
Prediction
Feature selection
Machine learning

ABSTRACT

Football is the team sport that mostly attracts great mass audience. Because of the detailed information about all football matches of championships over almost a century, matches build a huge and valuable database to test prediction of matches results. The problem of modeling football data has become increasingly popular in the last years and learning machine have been used to predict football matches results in many studies. Our present work brings a new approach to predict matches results of championships. This approach investigates data of matches in order to predict the results, which are win, draw and defeat. The investigated groups were different type of combinations of two by two pairs, win-draw, win-defeat and draw-defeat, of the possible matches results of each championship. In this study we employed the features obtained by scouts during a football match. The proposed system applies a polynomial algorithm to analyse and define matches results. Some machine-learning algorithms were compared with our approach, which includes experiments with information obtained from the football championships. The association between polynomial algorithm and machine learning techniques allowed a significant increase of the accuracy values. Our polynomial algorithm provided an accuracy superior to 96%, selecting the relevant features from the training and testing set.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Football is a game played by two teams with eleven players each and any scoring corresponds to kicking a ball into the opponent's goal. There are three possible results of a match: win, draw and defeat. One of the teams wins when it achieves the highest score within a period of 90 min (Sfeir, 2011). Namely, it is a simple and inexpensive game, which explains the reason why it became so popular world-wide. In 2014 Brazil hosted the last World Cup, whose audience reached more than 3.5 billion TV viewers (FIFA, 2014). The 2012 Olympic Games in London had an audience of 4 billion people (Olympic, 2012). The difference between the two events is that the World Cup was played among 32 countries for a

single sport whereas in the Olympic Games 204 countries participated among 26 types of sports.

The first World Cup happened in 1930, whereas the history of Olympic Games started much earlier, namely in 1896 (Cardinal, Boonchaaythanasit, Cheung, Lee, & Si, 2016; Moore, 2016). From 1936 on they included football definitely. Hence, this long-lived and widespread interest in football yielded a huge dataset with detailed information about each single match in championships until now (Perin, Vuillemot, & Fekete, 2013).

The result of a football match has been considered as subject of numerous scientific efforts in the endeavour to improve the game tactics and team features. In literature there are many studies which focus on football matches forecasts (Constantinou, Fenton, & Neil, 2013). Forecasting football is comprised of result of a match (win, draw and defeat) and score, which can be used by the bookmakers on the bets. Much less effort has been devoted, to the understanding of football from the perspective of the predicting results. Predicting the results is a difficult problem because of the number of factors which must be taken into account may not represent quantitatively (Hucaljuk & Rakipović, 2011). For instance, a

* Corresponding author.

E-mail addresses: rodrigo.grassimartins@gmail.com (R.G. Martins), alessandrosm2006@yahoo.com.br (A.S. Martins), neves.leandro@gmail.com (L.A. Neves), vieira@ufu.br (L.V. Lima), edna@ufu.br (E.L. Flores), marcelo.zanchetta@gmail.com (M.Z. do Nascimento).

team may completely dominate statistical measures such as number of good finish, number of successful pass or possessions in the offensive area, and fail to score one goal more than the opposing team to win a match (Brooks, Kerr, & Guttig, 2016).

In the literature we can find studies that involve predict matches results in a championship and of other team sports. Ulmer and Fernandez (2013) studied such techniques (baseline, Gaussian naïve Bayes, hidden Markov model, multimodal naïve Bayes, support vector machine (SVM), random forest (RF), One vs All SGD) to predict results by using the goals scored by each team in 10 seasons (from the 2002–03 season to the 2011–12 season) of the English Championship. Hucaljuk and Rakipović (2011) have surveyed the prediction of results to the UEFA Champions League also through scored goals with the following algorithms: naïve Bayes (NB), Bayesian networks, logitboost, k-nearest neighbours (KNN), RF and artificial neural networks. With SVM, Igiri (2015) studied the data related to scores in the English Championship.

With a much greater amount of information, Parinaz and Sadat (2013) employed data related to physiology and football strategies in order to analyse the Futbol Club Barcelona in the Spanish Championship. In that work the authors described a Bayesian network approach for football results prediction with the NETICA software. The model considered only one team to predict the result of football matches. Tax and Joustra (2015) used the following classification algorithms: CHIRP, logitboost, DTNB, FURIA, hyperpipes, J48, naïve Bayes, multilayer perceptron and RF. For selection they applied ReliefF, CfsSubsetEval and principle components to 65 features from the Dutch Championship, but they could not determine the relevant features in order to raise the accuracy of score predictions. Duarte, Soares, and Teixeira (2015) surveyed the C5.0, JRip, RF, KNN, SVM and NB algorithms to predict matches of the Portuguese Championship with the following pieces of information: previously scored goals, number of goals, lapse between matches and features of the competition. New features and also information about the psychology information of the football players were applied in Duarte et al. (2015), which however did not improve the performance of the model. For this reason, it is a challenge to investigate information and strategy of classification which facilitate the prediction of matches results. Thus, the focus of this research is to propose a new approach to predict football matches results of championships.

Others studies have been developed to predict matches results exploring machine learning algorithms. These algorithms are tools that receives as inputs the features and provides as output the prediction of matches results (win, draw and defeat). There are algorithms that can provide the most appropriate response to the problem (Pendharkar, Khosrowpour, & Rodger, 2000; Ramirez-Villegas & Ramirez-Moreno, 2012). Our decision for the polynomial classification was conditioned by its capacity of learning complex patterns that could be linearly inseparable and the success reached in other applications (Park, Oh, & Kim, 2008). A polynomial classification is a parameterized nonlinear map which non-linearly expands a sequence of input vectors to a higher dimension and maps them to a desired output sequence. This expansion can improve the separation of the different classes in an expanded vector space. Moreover, this strategy presents the advantages of providing only one model for optimal separation of classes and to consider this as the solution of the problem, which does not occur with the models presented in Campbell, Assaleh, and Broun (2002) and Ajmera and Holambe (2010).

Our present study introduces a novel approach to predict football matches results of championships. This approach investigate data of matches in order to predict the results, which are win, draw and defeat. The proposed system employs a polynomial classification algorithm that analyses and defines the matches results. We apply features related to information obtained by scouts. The

investigated groups were win-draw, win-defeat and draw-defeat. In this paper, we used the 10-fold cross-validation method and the sliding window technique, in which the training part and the group test used 4 and 2 rounds, respectively, to evaluate the ability of the model to predict football matches results of championships. The accuracy rates were calculated and evaluated considering datasets from real tournaments. Comparisons with machine learning classification algorithms were drawn and their new accuracy values showed a significant increase. Furthermore, the robustness of the proposed method was also investigated with addition of different noise levels on the features: the results were important even in adverse conditions. Here are the main contributions of this work:

- It presents a novel approach to identify football matches results based on the polynomial classifier with features obtained by scouts.
- It shows the usefulness of our approach when applied to datasets commonly investigated in the literature.
- It proposes the use of the polynomial classifier as an algorithm for feature selection and to improve the performance of machine learning techniques.
- It investigates the impact of noise at different attributes and the robustness of the proposed method given by the accuracy rates.

This paper is organised as follows. Section 2.1 describes details about how the datasets of the football championships were obtained. Sections 2.2 and 2.3 provide a brief description of each machine learning technique used in this survey and they also explain our polynomial algorithm. Experimental results, together with a discussion, are presented in Section 3. Section 3.4 describes some related to researches, and finally we draw our conclusions in Section 4.

2. Methodology

The proposed approach was organised into three stages, data collection, which is based on procedures related to the events that occur during a football match, feature selection for polynomial classification, and the classification through the polynomial algorithm and machine learning techniques. Fig. 1 presents the proposed flowchart of approach and in the following sections the main details are described.

The polynomial algorithm was implemented in MATLAB R2012b, the machine learning classifiers and feature selection were implemented applying the Waikato Environment for Knowledge Analysis (WEKA) software (Hall et al., 2009).

2.1. Data collection

The data used for the proposed approach were the football matches results, which were obtained from the different Championships: English Premier League (EPL), season 2014/2015; La Liga Primera Division (LLPD), season 2014/2015; and Brazilian League Championships, seasons 2010 (BLC 2010) and 2012 (BLC 2012). These different datasets can provide the predictive power of the proposed method. The championships were played on the point system accrued by 20 teams. Each team faced the other both home and away, playing in their own stadium as well as that of the visiting team. In this manner, there were a total of 38 rounds, and in each 10 matches were played, thus producing a total of 380 matches per championship.

EPL and LLPD: The data were the results (win, draw or defeat) of all English Premier League matches from season 2014/15. Similar number of instances were obtained of Liga Primera Division matches from season 2014/2015. This information is available online at <http://www.football-data.co.uk/>. These datasets were used

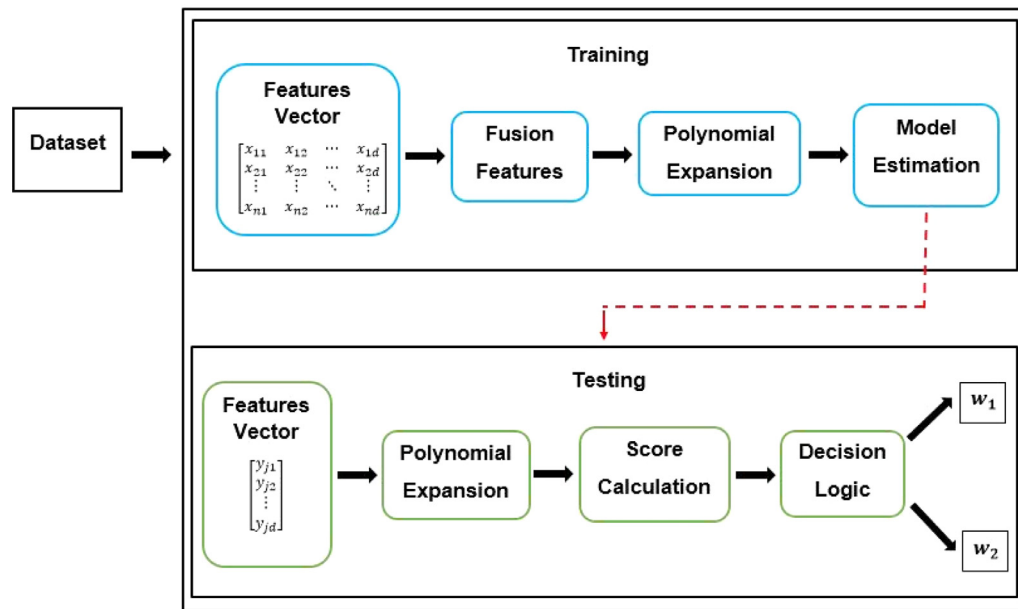


Fig. 1. Flowchart of proposed approach to predict football match results of championships.

Table 1

Features of the EPL and LLPD datasets of the home team and away team employed by the proposed approach and the machine learning algorithms.

Feature	Description
Goal	Is the act of scoring in a game
Goal shot	Is the player taking a shot in an attempt to score in a game
Shot on target	Is the player taking a shot towards the goal assuming no interception from players
Hit woodwork	Is the ball hitting the frame of the goal
Corner	Is awarded to the attacking team when the ball leaves the field of play by crossing the goal line (byline)
Foul committed	Is the act committed by players which are judged by the referee to be unfair
Offside	Is a football rule that occurs when the ball is touched by a teammate interfering with play
Yellow card	Indicates that a player was officially cautioned
Red card	Indicates that a player was dismissal from the game (sent off)

in other studies (see Tax & Joutsa, 2015; Ulmer & Fernandez, 2013). Each example of dataset consists of a set of fundamental features of a match for each team (scalar data type): (x_1) Home Team Goals, (x_2) Away Team Goals, (x_3) Home Team Goals Shots, (x_4) Away Team Goals Shots, (x_5) Home Team Shots on Target, (x_6) Away Team Shots on Target, (x_7) Home Team Hit woodwork, (x_8) Away Team Hit woodwork, (x_9) Home Team Corner, (x_{10}) Away Team Corner, (x_{11}) Home Team Foul Committed, (x_{12}) Away Team Foul Committed, (x_{13}) Home Team Offside, (x_{14}) Away Team Offside, (x_{15}) Home Team Yellow Card, (x_{16}) Away Team Yellow Card, (x_{17}) Home Team Red Card and (x_{18}) Away Team Red Card. The features are details in Table 1 and these information were stored in a vector with 18 features.

BLC: In this study the collected data were consolidated in reference to the Brazilian League Championships from 2010 to 2012. In this dataset, each example have a set of features of a match for each team (scalar data type): (x_1) Home Team Assistance, (x_2) Away Team Assistance, (x_3) Home Team Received Ball, (x_4) Away Team Received Ball, (x_5) Home Team Recovered Ball, (x_6) Away Team Recovered Ball, (x_7) Home Team Lost Ball, (x_8) Away Team Lost Ball, (x_9) Home Team Yellow Card, (x_{10}) Away Team Yellow Card, (x_{11}) Home Team Red Card, (x_{12}) Away Team Red Card, (x_{13}) Home Team Received Cross, (x_{14}) Away Team Received Cross, (x_{15}) Home Team Missed Cross, (x_{16}) Away Team Missed Cross, (x_{17}) Home Team Defense, (x_{18}) Away Team Defense, (x_{19}) Home Team

Successful Tackle, (x_{20}) Away Team Successful Tackle, (x_{21}) Home Team Unsuccessful Tackle, (x_{22}) Away Team Unsuccessful Tackle, (x_{23}) Home Team Successful Dribble, (x_{24}) Away Team Successful Dribble, (x_{25}) Home Team Unsuccessful Dribble, (x_{26}) Away Team Unsuccessful Dribble, (x_{27}) Home Team Given Corner, (x_{28}) Away Team Given Corner, (x_{29}) Home Team Received Corner, (x_{30}) Away Team Received Corner, (x_{31}) Home Team Received Foul, (x_{32}) Away Team Received Foul, (x_{33}) Home Team Committed Foul, (x_{34}) Away Team Committed Foul, (x_{35}) Home Team Good Finish, (x_{36}) Away Team Good Finish, (x_{37}) Home Team Bad Finish, (x_{38}) Away Team Bad Finish, (x_{39}) Home Team Finish, (x_{40}) Away Team Finish, (x_{41}) Home Team Goal, (x_{42}) Away Team Goal, (x_{43}) Home Team Own Goal, (x_{44}) Away Team Own Goal, (x_{45}) Home Team Offside, (x_{46}) Away Team Offside, (x_{47}) Home Team Good Throw-in, (x_{48}) Away Team Good Throw-in, (x_{49}) Home Team Bad Throw-in, (x_{50}) Away Team Bad Throw-in, (x_{51}) Home Team Successful Pass, (x_{52}) Away Team Successful Pass, (x_{53}) Home Team Unsuccessful Pass and (x_{54}) Away Team Unsuccessful Pass. The features are detailed in Table 2. This information was stored as a set composed by 54 features.

2.2. Polynomial classifier

The polynomial classifier (POL) is a supervised classification method that has presented relevant results in data analysis of

Table 2

Features used of the BLC championships from 2010 and 2012 for investigation of the algorithms.

Feature	Description
Assistance	Indicates that a player passed or crossed the ball to a player from his team
Received ball	Indicates that a player received the ball from a cross or good throw-in
Recovered ball	Indicates that a player recovered possession of the ball for his team
Lost ball	Indicates that a player lost the ball to the opposing team
Yellow card	Indicates that a player was cautioned
Red card	Indicates that a player was dismissed from the game (sent off)
Received cross	Indicates that a player passed the ball from an area of the field towards the centre of the field near the opponent's goal
Missed cross	Indicates that a player was not able to pass the ball from an area of the field towards the centre of the field near the opponent's goal
Defense	Indicates that an outfield player whose primary role was prevent the opposing team from scoring a goal
Successful tackle	Indicates that a player stopped an opponent team player from taking control of the ball
Unsuccessful tackle	Indicates that a player was not able to stop an opponent team player from taking control of the ball
Successful dribble	Indicates that a player attempted to drive the ball toward the opposing goal through individual control of the ball
Unsuccessful dribble	Indicates that a player was not able to drive the ball towards their opponents' goal through individual control of the ball
Given corner	Indicates that the ball completely crossed the goal-line (byline) after last being touched by a player of the defending team
Received corner	Indicates that a player of the opposing team took a corner
Received foul	Indicates that a player received foul committed by a player from the opposing team
Committed foul	Indicates that a player committed a foul on a player from the opposing team
Good finish	Indicates that a player kicked the ball towards the goal and it hits the beam and the goalkeeper defended or there was a goal
Bad finish	Indicates that a player kicked the ball towards the goal and the ball went out of the field of play
Finish	Indicates that a player kicked the ball towards the goal
Goal	Indicates that a player scored the a goal for his team
Own goal	Indicates that a player scored in the goal of his own team during the game
Offside	Is a football rule that occurs when the ball is touched or played by a teammate, and is not considered as a legitimate move in the game
Good throw-in	Indicates that a player was able to throw-in the ball to teammate
Bad throw-in	Indicates that a player was not able to throw-in the ball to teammate
Successful pass	Indicates that a player passed the ball to the player of the same team
Unsuccessful pass	Indicates that a player was not able to pass the ball to the player of the same team

medical images, mainly for problems in which the data are not linearly separable (Do Nascimento et al., 2013).

This algorithm looks to expanding the space of input features into a greater space dimension, in a manner that allows for a more adequate separation between the analysed classes. Therefore, one defines Eq. (1):

$$g(\mathbf{x}) = \mathbf{a}^T p_n(\mathbf{x}), \quad (1)$$

where \mathbf{a} is the function of the vector coefficients of the polynomial base, $p_n(\mathbf{x})$ is the function of the polynomial base and n the order or degree of the polynomial function.

For example, given a two-dimensional input vector $\mathbf{x} = [x_1 \ x_2]^T$, the elements of $\mathbf{p}_2(\mathbf{x})$ result in parameters similar to those shown in Campbell et al. (2002):

$$\mathbf{p}_2(\mathbf{x}) = [1 \ x_1 \ x_2 \ x_1^2 \ x_1x_2 \ x_2^2]^T. \quad (2)$$

In the first step, one should transform the feature set d -dimensional into polynomial base vectors L -dimensional. The coefficients of the polynomial function are calculated using the least squared method. Therefore, an expression given by $y_i = g(\mathbf{x})$ is obtained after a linear combination of the expanded terms $p_n(\mathbf{x})$ through the polynomial coefficient vector defined by Eq. (1). Following this, a test vector should be expanded in terms of the polynomial base. This process is performed by means of the scaling product from the expanded vector with the polynomial coefficient vector obtained from the feature set d -dimensional. The algebraic symbol of $g(\mathbf{x})$ allows one to determine which class the analysed vector belongs to.

The rule of decision for only two classes ω_1 and ω_2 , is defined by Eq. (3).

$$\text{Decide} \begin{cases} \omega_1, & \text{if } g(\mathbf{x}) > 0 \\ \omega_2, & \text{if } g(\mathbf{x}) < 0 \end{cases}. \quad (3)$$

The array \mathbf{X} represents the input data set, where N is the number of standards used in the construction phase of the polynomial base vectors L -dimensional. Furthermore, the robustness of the method was also investigated with addition of different noise levels on the features: the results were important even in adverse conditions.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nd} \end{bmatrix}. \quad (4)$$

The expansion of the polynomial base for each sample, array \mathbf{M} of dimension $N \times L$, is determined by (5).

$$\mathbf{M} = \begin{bmatrix} p_n(\mathbf{x}_1)^T \\ p_n(\mathbf{x}_2)^T \\ \vdots \\ p_n(\mathbf{x}_N)^T \end{bmatrix}. \quad (5)$$

The solution to the simultaneous equations is simplified with a matrix notation, as in Eq. (6).

$$\mathbf{M}\mathbf{a} = \mathbf{b}, \quad (6)$$

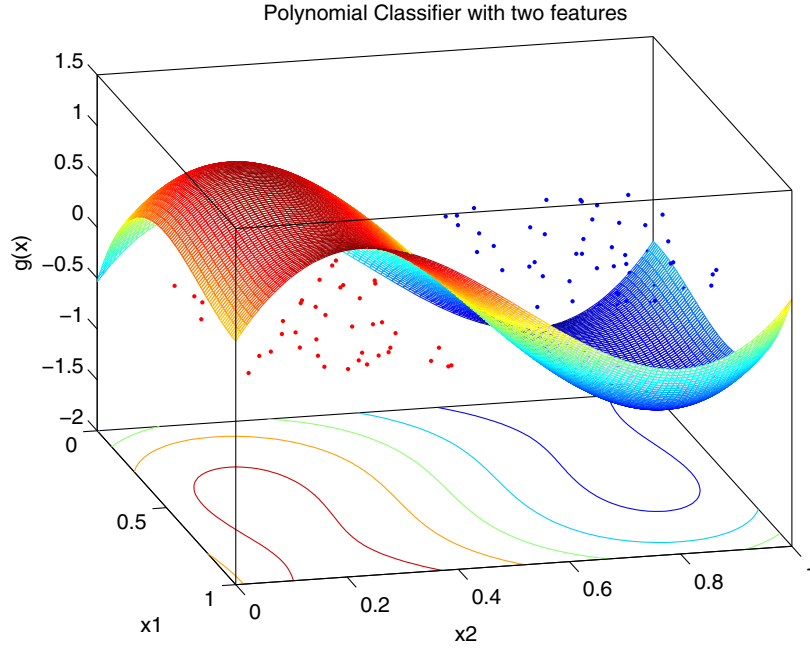


Fig. 2. Example of a polynomial function calculated for separating the data into two different classes.

where \mathbf{b} is a vector for which the elements are all random constants of value 1 or -1 , according to the input standard class. When there exist more equations than incognitos, there does not exist a single solution. Therefore, one should look for a vector that minimizes the error between \mathbf{Ma} and \mathbf{b} . This vector is given by Eq. (7):

$$\mathbf{e} = \mathbf{Ma} - \mathbf{b}. \quad (7)$$

As the error minimisation problem is a classic problem, known in the literature as least squares, the problem for $\mathbf{Ma} = \mathbf{b}$ is dealt with by Eq. (8).

$$\mathbf{M}^T \mathbf{Ma} = \mathbf{M}^T \mathbf{b}. \quad (8)$$

As the array $\mathbf{M}^T \mathbf{M}$ is a squared array of dimension $L \times L$ and not singular, one can find a single solution \mathbf{a} as in Eq. (9).

$$\mathbf{a} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{b} = \mathbf{M}^\dagger \mathbf{b}, \quad (9)$$

where the array \mathbf{M}^\dagger of dimension $L \times N$ as shown in Eq. (10)

$$\mathbf{M}^\dagger = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T, \quad (10)$$

is called *pseudoinverse* of \mathbf{M} (Duda, Hart, & Stork, 2000).

In this study, the feature set for 3-dimensions was used to calculate the polynomial bases of the 4th order. In this procedure different combinations of features were concatenated for 3-dimensions in order to obtain the final result. Fig. 2 gives an example of separation with features analysed by the polynomial algorithm.

2.3. Machine learning

In order to investigate the prediction of the football matches results, some of the main machine learning algorithms were used in this work. We chose NaiveBayes module, J48 (WEKA's own version of C4.5) module, MultilayerPerceptron module, RBF module and SMO (WEKA's own version of support vector machine) module for implementation of the classifiers naïve Bayes (NB), decision tree (DT), multilayer perceptron (MLP), radial basis function (RBF) and support vector machine (SVM), respectively, in WEKA (Constantinou et al., 2013; Igiri, 2015; Tax & Joutstra, 2015; Ulmer & Fernandez, 2013). The parameters for each machine learning algorithm are summarised in Table 3.

Table 3

The machine learning parameters used for each of the different algorithms in WEKA.

Algorithms	Parameters
NB	weka.classifiers.bayes.NaiveBayes
DT	weka.classifiers.trees.J48 -C 0.25 -M 2
MLP	weka.classifiers.functions.MultilayerPerceptron -L 0.1 -M 0.05 -N 3000 -V 0 -S 0 -E 40 -H a
RBF	weka.classifiers.functions.RBFNetwork -B 2 -S 1 -R 1.0E-8 -M -1 -W 0.1
SVM	weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V 1 -W 1 -K

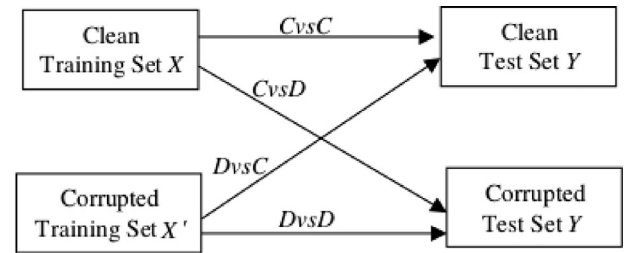


Fig. 3. Exploring the effects of feature noise with classification accuracy Zhu and Wu (2004, Fig. 2).

NB is the implementation of a simple probabilistic classifier based on applying Bayes' theorem that calculates a set of probabilities by counting the frequency and combinations of values in a given data set (Karabatak, 2015). The feature values for each class are then treated independently. This factor makes it deal efficiently with large amounts of data avoiding dimensionality problems.

DT is a machine learning algorithm that uses classification rules and a representation based on the data structure denominated as tree (Quinlan, 1986). A decision tree is made up of a set of nodes, where each of the nodes is composed of two basic types: decision nodes, which fragment the decision making and then constructs a path through the tree, and the terminal nodes which allows for the determination of which class the object under evaluation belongs to.

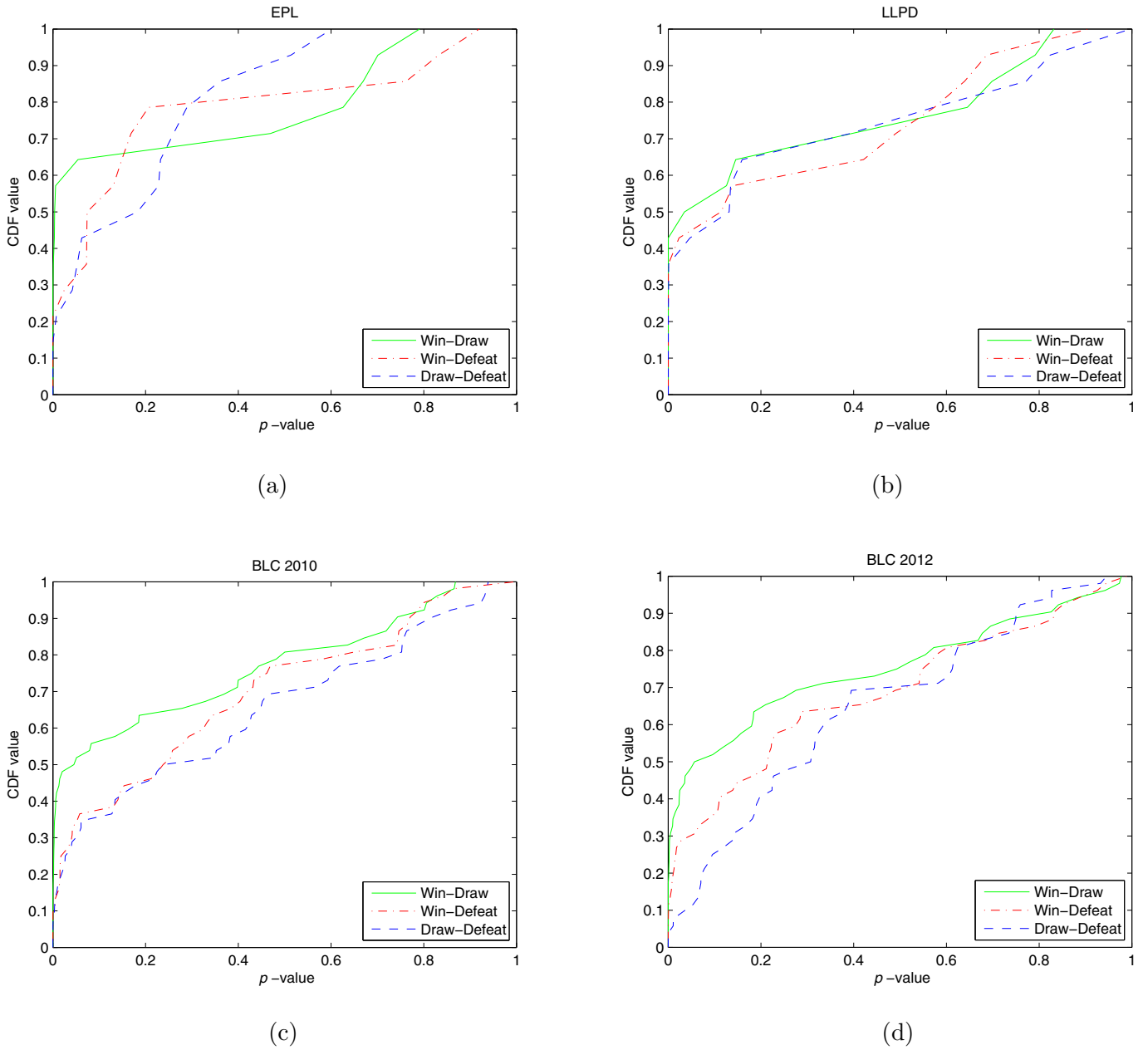


Fig. 4. Empirical cumulative distribution function of the p -values of the datasets: (a) EPL, (b) LLPD, (c) BLC 2010 and (d) BLC 2012.

MLP is a feed forward neural network with one or more hidden layer and an output layer of perceptrons (Cybenko, 1989). The input signals are propagated in a forward direction on a layer-by-layer basis. This algorithm utilises a supervised learning technique called back propagation for training the model that is able to classify non-linearly separable data.

RBF is a typical forward neural network, consisted of three layers and nodes in the input layer is to transfer signals to hidden level, in which nodes are composed of radius base functions (Broomhead, 1988). In this algorithm, the conversion from input layer to the hidden is nonlinear and the hidden layer performs nonlinear transformation of input vectors. The output implement a weighted sum of hidden unit outputs. The input into an RBF network is nonlinear while the output is linear.

SVM is a machine learning model used in classification problems of two classes. The input features are mapped to construct a dimensional space. This dimensional space is used to elaborate

a decision surface. Originally it has an excellent classification of linearly separable classes (Vapnik, 1995). Cortes and Vapnik (1995) presented a new implementation in which it was possible to classify non-linear classes by means of a larger dimensional space for data classification. Thus, the tests were performed with the SVM using the polynomial kernel.

2.4. Feature selection

We implemented the principal component analysis (PCA) and Relief methods, which are techniques for space dimensionality reduction of the data (Tax & Joutstra, 2015; Tüfekci, 2016). Moreover, the POL algorithm was also applied as a method for feature selection and associated with the classifiers based on machine learning methods.

PCA is a statistical approach can be applied to analyse large multivariate datasets, mainly when it is often desirable to reduce

Table 4

Accuracy rates obtained from the different championships.

Classifier	Cross-validation technique											
	EPL			LLPD			BLC 2010			BLC 2012		
	Win-draw	Win-defeat	Draw-defeat	Win-draw	Win-defeat	Draw-defeat	Win-draw	Win-defeat	Draw-defeat	Win-draw	Win-defeat	Draw-defeat
POL	99.6 ± 1.0	100.0 ± 0.0	99.6 ± 1.2	99.6 ± 1.0	99.6 ± 1.1	99.4 ± 1.6	99.5 ± 1.4	100.0 ± 0.0	99.3 ± 1.9	98.7 ± 3.0	99.1 ± 1.8	99.0 ± 3.1
NB	75.1 ± 8.7	95.7 ± 3.5	71.2 ± 9.3	80.7 ± 8.6	95.9 ± 2.9	72.7 ± 10.3	69.7 ± 12.0	92.6 ± 7.9	74.7 ± 5.6	75.0 ± 12.3	94.2 ± 7.1	77.0 ± 5.7
DT	98.8 ± 1.7	99.7 ± 0.9	99.6 ± 1.1	99.6 ± 1.1	99.2 ± 1.6	99.5 ± 1.4	97.2 ± 5.6	97.4 ± 5.9	96.7 ± 3.7	96.8 ± 8.9	96.1 ± 9.9	98.0 ± 4.7
MLP	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	94.4 ± 8.7	96.2 ± 3.5	94.0 ± 5.0	95.0 ± 7.3	95.6 ± 3.7	95.0 ± 5.0
RBF	96.7 ± 4.8	98.6 ± 2.4	96.8 ± 4.3	96.6 ± 4.8	100.0 ± 0.0	95.3 ± 5.4	76.5 ± 9.2	93.1 ± 4.6	78.4 ± 7.1	81.1 ± 11.7	92.9 ± 5.2	78.9 ± 9.9
SVM	97.2 ± 2.9	100.0 ± 0.0	98.6 ± 3.2	92.3 ± 7.4	100.0 ± 0.0	88.0 ± 7.6	96.1 ± 8.4	99.5 ± 1.4	97.2 ± 4.1	94.8 ± 7.3	98.9 ± 3.3	98.1 ± 3.3

Table 5

Match classification based on championships with sliding window technique.

Classifier	Sliding window technique											
	EPL			LLPD			BLC 2010			BLC 2012		
	Win-draw	Win-defeat	Draw-defeat	Win-draw	Win-defeat	Draw-defeat	Win-draw	Win-defeat	Draw-defeat	Win-draw	Win-defeat	Draw-defeat
POL	99.2 ± 2.1	96.2 ± 3.9	97.7 ± 4.1	98.0 ± 3.8	96.7 ± 4.0	97.7 ± 4.3	99.6 ± 1.5	100.0 ± 0.0	100.0 ± 0.0	99.5 ± 1.7	100.0 ± 0.0	100.0 ± 0.0
NB	69.7 ± 17.2	89.2 ± 9.5	52.2 ± 19.5	75.9 ± 14.1	88.5 ± 8.8	60.9 ± 18.0	61.9 ± 13.4	81.6 ± 15.0	57.6 ± 17.2	69.5 ± 11.4	81.6 ± 15.0	48.6 ± 13.7
DT	82.5 ± 15.2	93.8 ± 6.4	80.6 ± 16.3	83.5 ± 17.4	92.8 ± 7.5	82.2 ± 14.2	73.1 ± 12.8	92.3 ± 11.9	61.7 ± 14.2	78.2 ± 11.3	92.3 ± 11.9	59.2 ± 15.7
MLP	85.2 ± 8.9	97.6 ± 4.3	80.6 ± 14.0	82.2 ± 10.8	95.9 ± 5.0	75.6 ± 11.2	72.7 ± 15.7	83.7 ± 10.8	65.6 ± 18.2	69.1 ± 9.4	83.7 ± 10.8	62.4 ± 15.3
RBF	76.6 ± 11.5	95.3 ± 4.7	62.4 ± 10.6	78.9 ± 13.1	93.3 ± 7.4	65.7 ± 14.8	65.6 ± 13.8	77.3 ± 8.8	61.8 ± 16.4	64.5 ± 10.6	77.3 ± 8.8	50.4 ± 11.8
SVM	74.8 ± 10.6	95.2 ± 4.4	67.8 ± 16.7	78.2 ± 12.2	91.1 ± 7.0	66.4 ± 16.4	70.4 ± 11.8	84.4 ± 10.1	65.8 ± 17.2	67.0 ± 9.8	84.4 ± 10.1	64.3 ± 13.0

Table 6

Average accuracy for classification with cross-validation technique for different datasets.

Datasets	Classifiers					
	POL	NB	DT	MLP	RBF	SVM
EPL	99.7	80.7	99.4	100.0	97.4	98.6
LLPD	99.5	83.1	99.4	100.0	97.3	93.4
BLC 2010	99.6	79.0	97.1	94.9	82.7	97.6
BLC 2012	98.9	82.0	97.0	95.2	84.3	97.2
Friedman rank	1.5	6.0	3.0	2.5	4.75	3.25
<i>p</i> -value			0.0103			

Table 7

Results for the metrics of the classifiers with the sliding window technique.

Datasets	Classifiers					
	POL	NB	DT	MLP	RBF	SVM
EPL	97.7	70.3	85.6	87.8	78.1	79.3
LLPD	97.5	75.1	86.1	84.5	79.3	78.6
BLC 2010	99.8	67.0	75.7	74.0	68.2	73.6
BLC 2012	99.8	66.5	76.6	71.8	64.1	71.9
Friedman rank	1.0	5.75	2.25	3.0	5.0	4.0
<i>p</i> -value			0.0031			

the dimensionality. The method is based on orthogonal transformation to convert data of possibly correlated variables into a set of values of linearly uncorrelated variables, which are named as principal components. The goal is to define the dominant patterns by calculating the maximum amount of variance with the fewest number of principal components. Therefore, the number of principal components should be less than or equal to the number of original variables. Principal component analysis has many applications and extensions and more details are presented by Jolliffe (2014) and Wold, Esbensen, and Geladi (1987).

Relief is a feature selection algorithm based on idea of estimating features according to how well their values discriminate between instances that are near each other. Considering a given instance, the algorithm searches for its two nearest neighbours, such as one from the same class, which can be named as “nearest hit”, and an the other from different class, which can be called as “nearest miss”. Although this algorithm does not classify redundant features and low numbers of training instances can provide wrong results, the method is noise-tolerant, can be applied on binary or continuous data and is not dependent on heuristics. More details are described by Kononenko (1994), Kira and Rendell (1992) and Hall (2000).

2.5. Validation method

The sliding window and cross-validation techniques were performed to compute the performance of the algorithms used to predict football matches results of championships. The investigated groups were win-draw, win-defeat and draw-defeat.

Cross-validation: We trained and tested the proposed approach using the 10-fold cross-validation method (Bruno et al., 2016). The matches were distributed into 10 subsets. Afterwards, we take one of the 10 subsets as test and the remaining 9 partitions as the training set. The accuracy of the model is then computed for this 1×9 separation. This procedure is repeated for each one of the 10 subsets, and so we end up with 10 values of accuracy. The final accuracy is computed as an average of that 10 values.

Sliding window: In this study, we also adopted the sliding window technique due to the fact that the time factor is considered in the resolution of result prediction problems (Vafaeipour, Rahbari, Rosen, Fazelpour, & Ansarirad, 2014). In this experiment, we di-

vided the data into sets with 4 rounds for training and 2 rounds for the testing of the classifiers. In each round 10 matches are played. This procedure was adopted for data from the 38 rounds of the championship, which resulted in 17 sets of data for evaluation. The final accuracy is computed as an average of the 17 sets.

2.6. Evaluation of proposed approach

Statistical analysis based on the Students *t*-test was used to measure the significance of features for discriminating between the championship data for win, draw and defeat. Features having a *p*-value of 0.05 or less are considered as statistically significant (Bhatia, Lam, Pang, Wang, & Ahuja, 2016).

In order to evaluate the proposed approach, the accuracy (AC) metric is also considered in this study (Dua, Singh, & Thompson, 2009; Fawcett, 2006). The AC measure is defined as the proportion of correct predictions related to the total number of evaluated samples:

$$AC = \frac{TP + TN}{P + N}, \quad (11)$$

where *TP* is the rate of true positive cases, *TN* is the rate of true negative cases, both calculated in relation to all positive (*P*) and all negatives (*N*) cases.

According to the proposed strategy by Zhu and Wu (2004) we evaluate the relationship between feature noise and classification accuracy and the impact of noise at different features for the proposed approach. For a dataset, we split it into a training set named as *X* and a test set named as *Y*. For evaluate the classifier, we used the data of *X* for training and *Y* for testing and denote the classification accuracy by clean training set vs clean test set (CvsC). Then, the features were manually corrupt based proposed method by Zhu and Wu (2004) and a new noisy training set *X'* was constructed from *X*. We train the classifier with data of *X'* and use the instances of *Y* for testing. We denote this classification accuracy by dirty training set vs clean test set (DvsC). After evaluated the dirty training set, we inserted the corresponding levels of feature noise into test set *Y* to produce a dirty test set *Y'* and use classifiers. We denote the classification accuracies by (CvsD) and (DvsD), respectively, clean training set vs dirty test set (CvsD) and dirty training set vs dirty test set (DvsD). The experiments were executed by using a set of evaluations, as shown in Fig. 3. In these experiments, we let the noise level be 0%, 20% and 40%, which were also adopted by Xiao, He, Jiang, and Liu (2010) and Zhu and Wu (2004).

3. Experimental results

The present section shows the evaluation of the proposed approach to predict match results in football championships. The Section 3.1 shows the results of the statistical analyses of the features and the prediction of matches results with the POL algorithm and traditional algorithms for championships. In Section 3.2 the impact of feature noise was evaluated for the algorithm POL, where various levels of noise were also added in the datasets. The use of selected features by the POL, PCA and Relief algorithms in the training phase were also investigated in Section 3.3 for the machine learning algorithms. Then, in Section 3.4 there is a comparison of accuracy values from different approaches for football matches results.

3.1. Analysis of the football championships with classifiers

A Students *t*-test was used for each feature of the football matches and the *p*-values obtained from each one can be used as a measure of how effective it is at separating groups. Fig. 4 presents the empirical cumulative distribution function (CDF) of the *p*-values for the investigated datasets.

Table 8Adjusted p -values for N vs N comparisons for all 15 hypotheses for each algorithm.

i	hypothesis	Unadjusted p	P_{Neme}	P_{Holm}	P_{Shaf}	P_{Berg}
1	POL vs .NB	.697294490218126E-4	0.010045941735327189	0.010045941735327189	0.010045941735327189	0.010045941735327189
2	NB vs .MLP	0.008150971593502683	0.12226457390254024	0.11411360230903755	0.08150971593502683	0.08150971593502683
3	POL vs .RBF	0.014019277113959921	0.21028915670939882	0.18225060248147898	0.14019277113959921	0.14019277113959921
4	NB vs .DT	0.023342202012890823	0.35013303019336234	0.2801064241546899	0.23342202012890823	0.16339541409023575
5	NB vs .SVM	0.03763531378731435	0.5645297068097153	0.41398845166045783	0.3763531378731435	0.22581188272388608
6	MLP vs .RBF	0.0889730117018133	1.3345951755271996	0.8897301170181331	0.8897301170181331	0.5338380702108798
7	POL vs .SVM	0.185876732365876	2.7881509854881403	1.6728905912928842	1.301137126561132	1.301137126561132
8	DT vs .RBF	0.185876732365876	2.7881509854881403	1.6728905912928842	1.301137126561132	1.301137126561132
9	POL vs .DT	0.25683925795785667	3.85258886936785	1.7978748057049967	1.7978748057049967	1.301137126561132
10	RBF vs .SVM	0.25683925795785667	3.85258886936785	1.7978748057049967	1.7978748057049967	1.301137126561132
11	NB vs .RBF	0.34470422200695766	5.170563330104365	1.7978748057049967	1.7978748057049967	1.3788168880278306
12	POL vs .MLP	0.449691797968891	6.745376969533366	1.798767191875564	1.798767191875564	1.3788168880278306
13	MLP vs .SVM	0.5707503880581739	8.561255820872608	1.798767191875564	1.798767191875564	1.7122511641745217
14	DT vs .MLP	0.7054569861112734	10.581854791669102	1.798767191875564	1.798767191875564	1.7122511641745217
15	DT vs .SVM	0.8501067391385257	12.751601087077887	1.798767191875564	1.798767191875564	1.7122511641745217

Table 9Adjusted p -values for N vs N comparisons of the classification algorithms over championships datasets with the sliding window technique.

i	Hypothesis	Unadjusted p	P_{Neme}	P_{Holm}	P_{Shaf}	P_{Berg}
1	POL vs .NB	3.298385207779531E-4	0.004947577811669297	0.004947577811669297	0.004947577811669297	0.004947577811669297
2	POL vs .RBF	0.0024969089151415957	0.03745363372712394	0.03495672481198234	0.024969089151415956	0.024969089151415956
3	NB vs .DT	0.008150971593502683	0.12226457390254024	0.10596263071553487	0.08150971593502683	0.08150971593502683
4	POL vs .SVM	0.023342202012890823	0.35013303019336234	0.2801064241546899	0.23342202012890823	0.16339541409023575
5	NB vs .MLP	0.03763531378731435	0.5645297068097153	0.41398845166045783	0.3763531378731435	0.26344719651120047
6	DT vs .RBF	0.03763531378731435	0.5645297068097153	0.41398845166045783	0.3763531378731435	0.26344719651120047
7	POL vs .MLP	0.13057001811573624	1.9585502717360435	1.1751301630416262	0.9139901268101537	0.7834201086944175
8	MLP vs .RBF	0.13057001811573624	1.9585502717360435	1.1751301630416262	0.9139901268101537	0.7834201086944175
9	NB vs .SVM	0.185876732365876	2.7881509854881403	1.301137126561132	1.301137126561132	0.7834201086944175
10	DT vs .SVM	0.185876732365876	2.7881509854881403	1.301137126561132	1.301137126561132	0.7834201086944175
11	POL vs .DT	0.34470422200695766	5.170563330104365	1.7235211100347883	1.3788168880278306	1.034112666020873
12	MLP vs .SVM	0.449691797968891	6.745376969533366	1.798767191875564	1.798767191875564	1.034112666020873
13	RBF vs .SVM	0.449691797968891	6.745376969533366	1.798767191875564	1.798767191875564	1.034112666020873
14	NB vs .RBF	0.5707503880581739	8.561255820872608	1.798767191875564	1.798767191875564	1.1415007761163478
15	DT vs .MLP	0.5707503880581739	8.561255820872608	1.798767191875564	1.798767191875564	1.1415007761163478

Table 10

Performance (%) of the learning machine algorithms with the selected features for the data from the championships with the cross-validation technique.

Dataset	Technique	NB	DT	MLP	RBF	SVM	Average
EPL	Baseline	80.7	99.4	100.0	97.4	98.6	95.2
	Relief	80.7	99.4	100.0	97.4	98.7	95.2
	PCA	80.7	99.4	100.0	97.8	98.7	95.3
	POL	81.1	99.6	100.0	99.7	100.0	96.0
LLPD	Baseline	83.1	99.4	100.0	97.3	93.4	94.6
	Relief	83.1	99.4	100.0	97.6	93.4	94.7
	PCA	82.5	99.4	100.0	98.1	92.7	94.5
	POL	87.0	99.4	100.0	99.4	98.3	96.8
BLC 2010	Baseline	79.0	97.1	94.9	82.7	97.6	90.2
	Relief	79.0	97.1	94.9	83.2	97.6	90.3
	PCA	68.8	70.5	71.4	70.9	74.8	71.2
	POL	82.6	98.2	98.8	98.6	98.7	95.3
BLC 2012	Baseline	82.1	97.0	95.2	84.3	97.3	91.1
	Relief	82.0	97.0	95.2	85.4	97.2	91.3
	PCA	69.9	68.9	68.6	71.7	74.6	70.7
	POL	83.0	97.2	98.5	97.6	98.3	94.9

Table 11

Classification with the learning machine algorithms applying selected features with the sliding window technique.

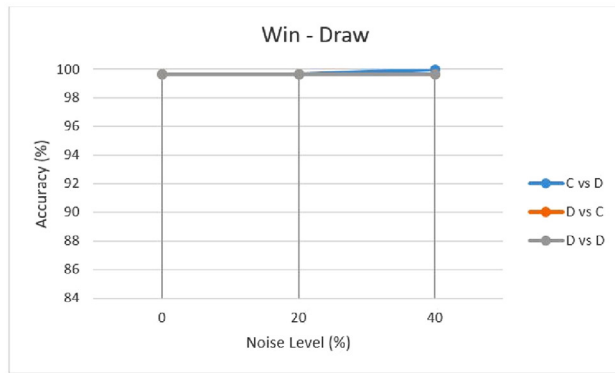
Dataset	Technique	NB	DT	MLP	RBF	SVM	Average
EPL	Baseline	70.3	85.6	87.8	78.1	79.3	80.2
	Relief	70.3	85.6	87.8	77.9	79.7	80.2
	PCA	71.1	87.7	93.0	81.0	81.3	82.8
	POL	71.0	90.1	99.1	93.9	84.7	87.7
LLPD	Baseline	75.1	86.1	84.6	79.3	78.6	80.7
	Relief	75.1	86.3	84.7	80.6	78.8	81.1
	PCA	75.1	86.2	89.4	81.9	79.2	82.3
	POL	74.9	90.9	96.0	93.5	83.0	87.6
BLC 2010	Baseline	67.0	75.7	74.0	68.2	73.6	71.7
	Relief	67.0	75.5	74.9	67.8	73.6	71.7
	PCA	59.6	59.7	57.6	62.7	63.9	60.7
	POL	71.7	91.2	97.8	95.4	89.2	88.9
BLC 2012	Baseline	66.5	76.6	71.8	61.1	71.9	70.1
	Relief	66.5	76.1	71.8	65.5	72.0	70.3
	PCA	61.1	56.4	58.4	61.6	61.9	59.9
	POL	72.1	92.7	97.2	95.2	88.3	89.1

In Fig. 4(a) we can see that more of 20% of the features have p -values close to zero and over 25% of the features have p -values smaller than 0.05 of the EPL dataset. For the LLPD dataset, 38% of the features have p -values smaller than 0.05 and can be used for discrimination of information. With the BLC 2010 dataset, there are around 25% of the features have p -values smaller than 0.05 for different groups. In the case of the BLC 2012 dataset, the win-draw and win-defeat groups present more of 20% of the features with p -values smaller than 0.05. Only the draw-defeat group of the BLC

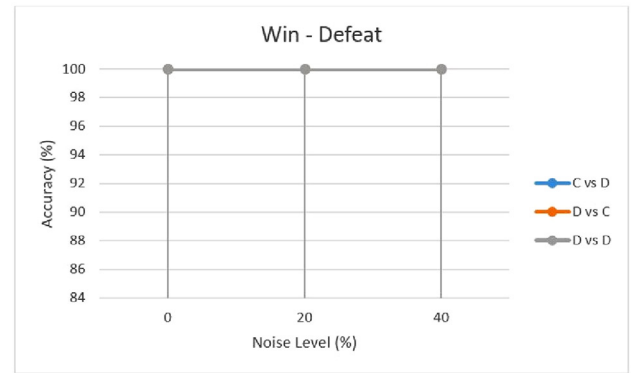
2012 dataset have 10% of the features that allows discrimination of data.

The average accuracy for the championships are shown in Tables 4 and 5 for the cross-validation and sliding window techniques, respectively.

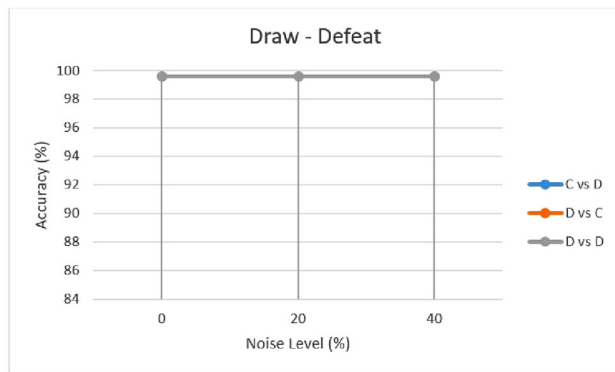
In Table 4 is presents the rates considering different metrics associated with the cross-validation technique. The POL algorithm provided relevant results for classification of these investigated championships. For EPL dataset, the accuracy value was approximately 99%. When we analysed the BLC 2010 dataset, we ob-



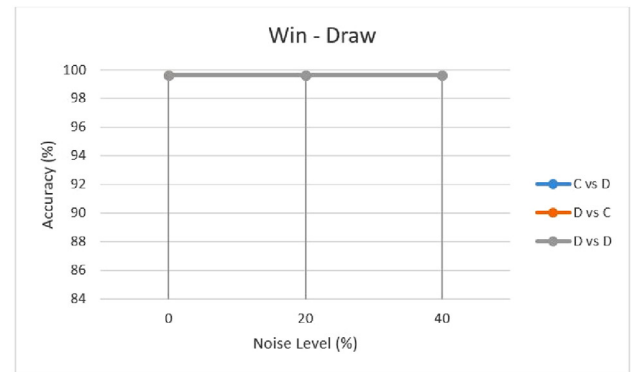
(a)



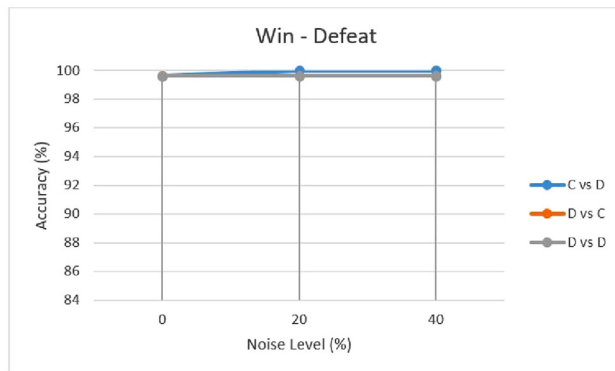
(b)



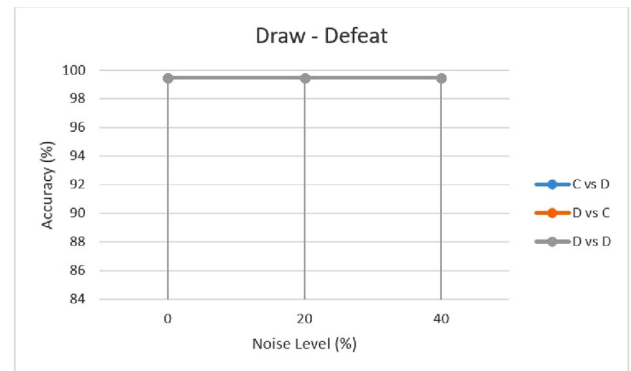
(c)



(d)



(e)

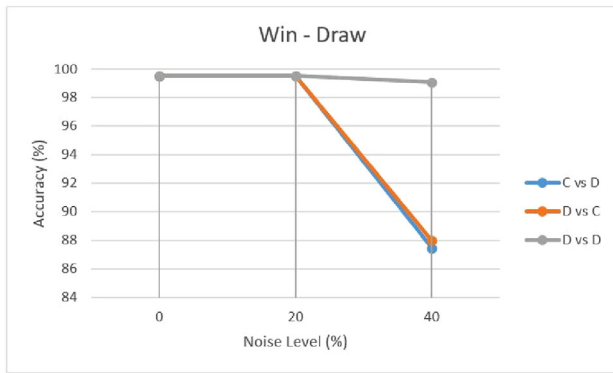


(f)

Fig. 5. Results (accuracy values) after addition of different noise levels, considering the datasets with the cross-validation technique: (a) win-draw of the EPL, (b) win-defeat of the EPL, (c) draw-defeat of the EPL, (d) win-draw of the LLPD, (e) win-defeat of the LLPD and (f) draw-defeat of the LLPD.

served that accuracy values are more discrepant between the POL, MLP, SVM and DT algorithms. Furthermore, only the POL algorithm presented ideal accuracy value for the win-defeat group for this dataset, with rates superior to 98% for all other groups and datasets. In Table 5 notes that the BN, DT, MLP, RBF and SVM algorithms provide results less promising than those obtained with the POL algorithm: the accuracy values in greater part of the datasets were over to 96%. All these rates are higher than those provided by other algorithms.

In Tables 6 and 7 are shown the accuracy values (average) calculated over the data (win-draw, win-defeat and draw-defeat) of each dataset. Following the significance test procedure of Garcia and Herrera (2008), we apply the Friedman test on the results provided by the classifiers associated with the cross-validation and sliding window techniques. The ranking are shown in tables with rows indicating the Friedman rank. The results show that the POL algorithm is ranked best that the NB, DT, MLP, RBF and SVM methods. The p -values of the tests considering the cross-validation and sliding window techniques are 0.0103 and 0.0031, respectively.



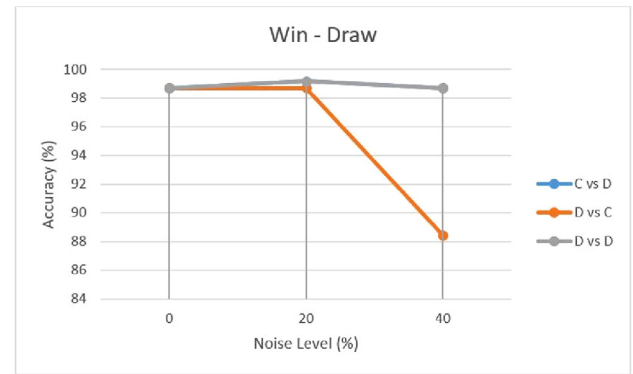
(a)



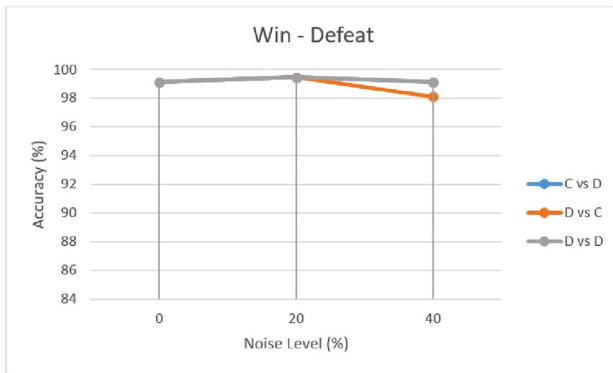
(b)



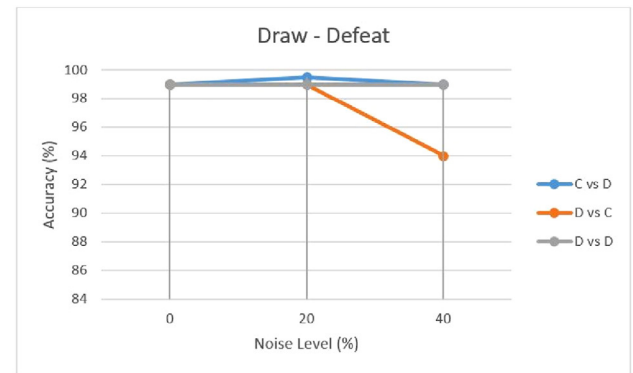
(c)



(d)



(e)



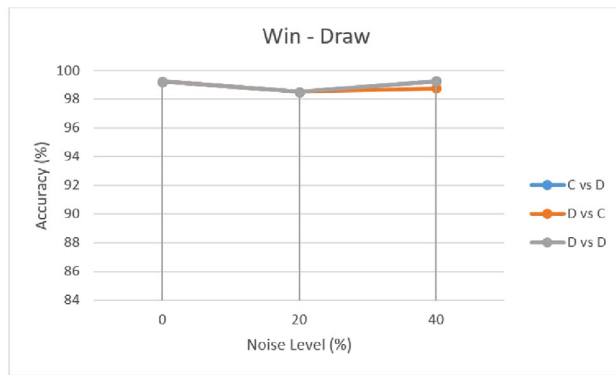
(f)

Fig. 6. Accuracy values after addition of different noise levels, considering the datasets with the cross-validation technique: (a) win-draw of the BLC 2010, (b) win-defeat of the BLC 2010, (c) draw-defeat of the BLC 2010, (d) win-draw of the BLC 2012, (e) win-defeat of the BLC 2012 and (f) draw-defeat of the BLC 2012.

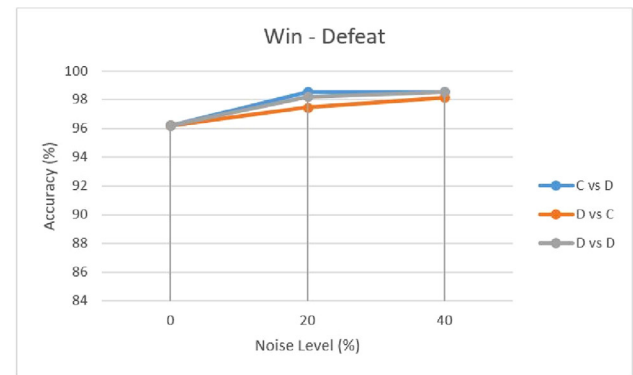
These p -values can indicate that we indeed have statistical significant differences between the methods.

The unadjusted values, adjusted p -values for Nemenyis, Holms, Shaffers, and Bergmann-Hommel's post-hoc procedures for N vs N comparisons for all possible pairs of algorithms are shown in Tables 8 and 9 for the cross-validation and sliding window techniques, and all datasets, respectively. In these experiments, if the adjusted p -value for an individual null hypothesis is smaller than 0.05, then this hypothesis is rejected (Trawinski, Smetek, Lasota, & Trawinski, 2014). The p -values without statistically significant dif-

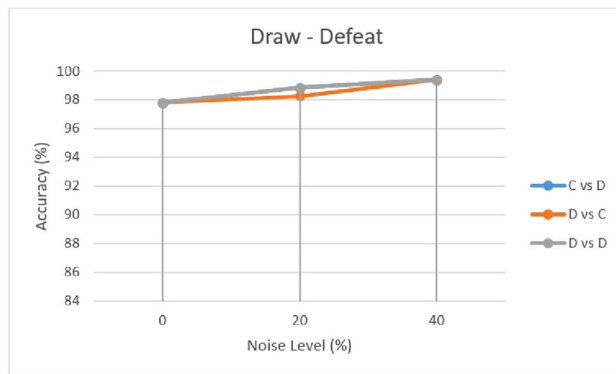
ferences between given pairs of models are marked with italics. Considering the cross-validation technique and the unadjusted values (Table 8), only 5 pairs (POL vs NB, NB vs MLP, POL vs RBF, NB vs DT and NB vs SVM) out of 15 have been rejected. When the sliding window technique is considered (Table 9), another combination was rejected: POL vs SVM, which results in 6 pairs. With the most powerful Shaffers, and Bergmann-Hommel's post-hoc procedures, only one pair (POL vs NB) models was rejected applying the cross-validation technique (Table 8), and two pairs (POL vs NB and POL vs RBF) were rejected with sliding window approach (Table 9).



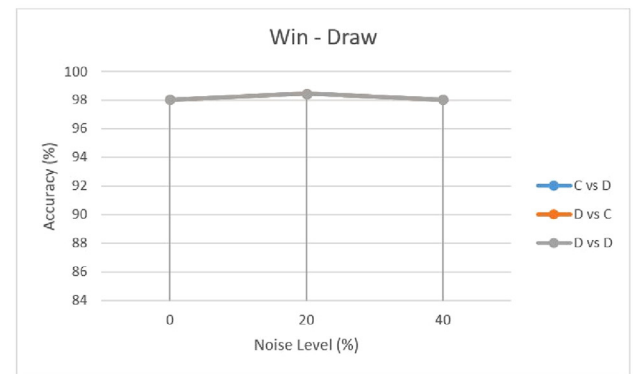
(a)



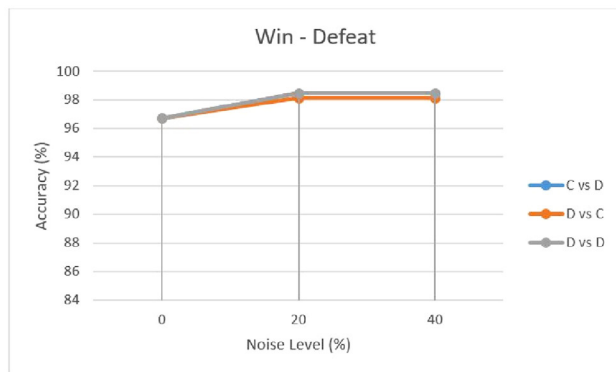
(b)



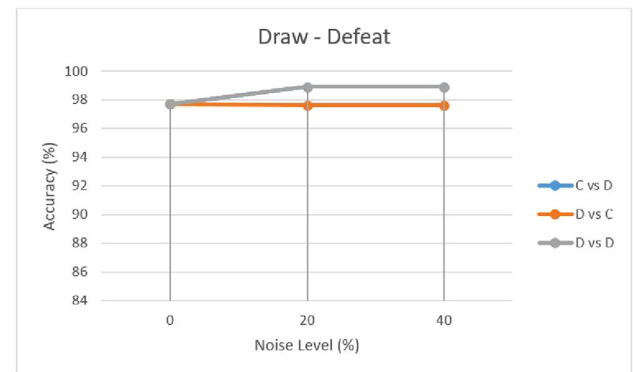
(c)



(d)



(e)



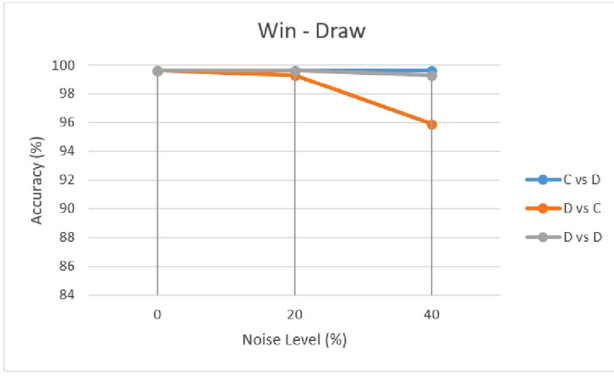
(f)

Fig. 7. Results after addition of different noise levels, considering the datasets with the sliding window technique: (a) win-draw of the EPL, (b) win-defeat of the EPL, (c) draw-defeat of the EPL, (d) win-draw of the LLPD, (e) win-defeat of the LLPD and (f) draw-defeat of the LLPD.

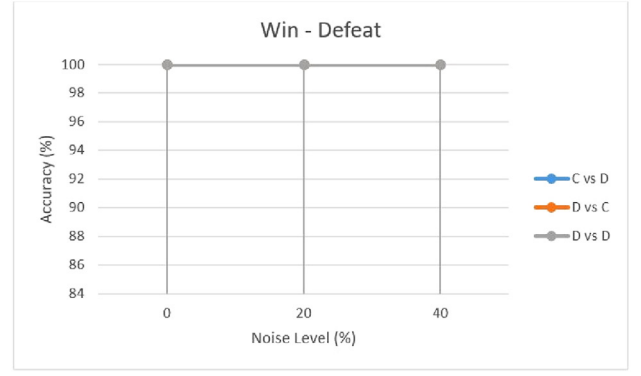
3.2. Analysis of the impact of noise at different features

In this subsection the results are presented with addition of different noise levels in the features in order to investigate the impact of this procedure on the robustness of the proposed approach. The performance of the model was defined with noise levels of 0%, 20% and 40%, as discussed by Xiao et al. (2010) and Zhu and Wu (2004). For each dataset, we considered the average accuracy for cross validation and sliding window techniques. Figs. 5 and 6 show the av-

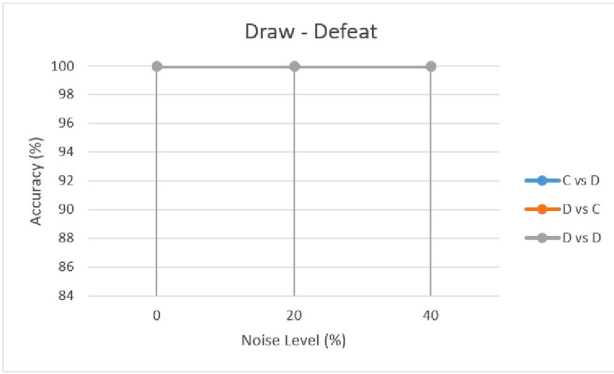
erage accuracy values of the POL algorithm with the datasets for cross-validation. In Figs. 7 and 8 we show the results of average accuracy values of the POL algorithm with the datasets for sliding window technique. Considering all these results, it is important to verify that our method can provide highly relevant accuracy rates (from 96% to 100%) even with up to 20% of added noise. For instance, when the datasets were tested with the cross-validation technique (Figs. 5 and 6), the results were closer to the idea values than those provided by the sliding window technique (Figs. 7 and



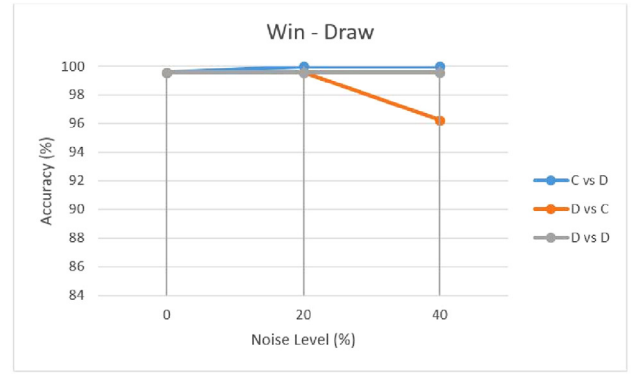
(a)



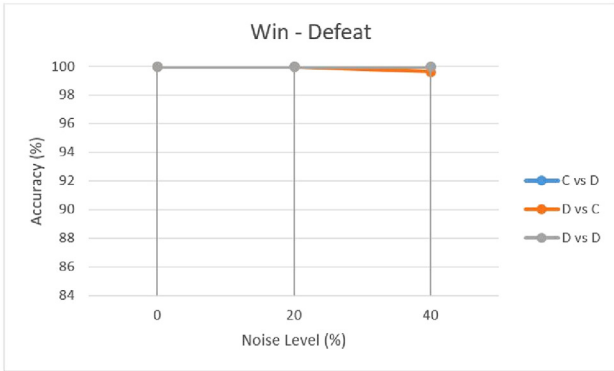
(b)



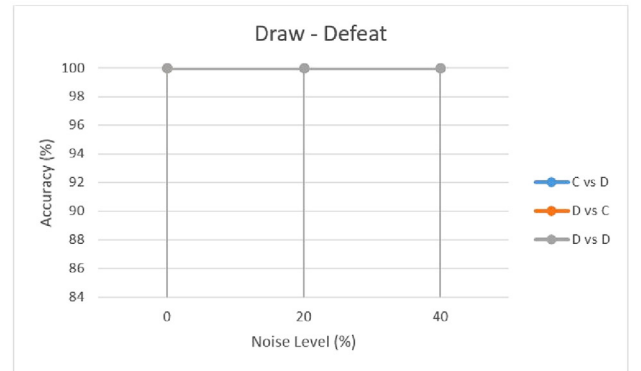
(c)



(d)



(e)



(f)

Fig. 8. Accuracy values after addition of different noise levels, considering the datasets with the sliding window technique: (a) win-draw of the BLC 2010, (b) win-defeat of the BLC 2010, (c) draw-defeat of the BLC 2010, (d) win-draw of the BLC 2012, (e) win-defeat of the BLC 2012 and (f) draw-defeat of the BLC 2012.

8). This can be verified for all cases of C vs D, D vs C and D vs D. However, we did not observe pattern for accuracy rates considering noise level of 40%. The lowest rate was with the BLC 2012 dataset (Fig. 6(d)), the win-draw group (cases of C vs D, D vs C and D vs D) and applying the cross-validation technique. Also, even with this noise level (40%), ideal values were verified with the sliding window technique, for instance: BLC 2010-dataset for win-defeat with cases of C vs D, D vs C and D vs D, Fig. 8(b); BLC 2012 dataset for draw-defeat with cases of C vs D, D vs C and D vs D, Fig. 8(f).

3.3. Feature selection and machine learning

Table 10 presents the average accuracy values for the win-draw, win-defeat and draw-defeat groups. The selected features by the algorithms were analysed applying the sliding window technique: the results are shown in Table 11.

As shown in Table 10, one notes that the performance of the classifiers were higher than the data analysed with the obtained features (baseline) from each championship. The strategy of using the selected features by the POL algorithm allowed to provide aver-

Table 12

Summary of studies for results prediction systems in football championships.

Reference	Competition	Dataset	Feature	Method	Accuracy
Ulmer and Fernandez (2013)	Accrued point system	English League championship 2002/2003 and 2011/2012	Fundamental	Baseline, Gaussian Naive Bayes, Hidden Markov Model, Multimodal Naïves Bayes, RBF SVM, RF, Linear SVM and One vs ALL SGD, NB, Bayesian Net, Logoost, K-NN, RF and ANN	0.52
Constantinou et al. (2013)	Eliminatory system	UEFA Champions League	Fundamental	NETICA software	0.68
Parinaz and Sadat (2013)	Accrued point system	Spanish Liga Championship	Fundamental and physiological	SVM	0.92
Igiri (2015)	Accrued point system	English League Championship	Fundamental	HIRP, LogitBoost, DTNB, FURIA, HyperPipes, J48, Naive Bayes, Multilayer Perceptron and RF	0.53
Tax and Joulstra (2015)	Accrued point system	Dutch League Championship	Fundamental and history	C5.0, JRip, RF, KNN, SVM and NB	0.56
Duarte et al. (2015)	Accrued point system	Portuguese League Championship	Life-history, fundamental and psychological	POL algorithm	0.58
Proposed approach	Accrued point system	EPL	Fundamental		0.99
		LLPD			0.99
		BLC 2010			0.99
		BLC 2012			0.98

age accuracy values of 96.0, 96.8, 95.3 and 94.9 for the EPL, LLPD, BLC 2010 and BLC 2012 datasets, respectively. This represents an increase of 0.8, 2.1, 5.1 and 3.7 in relation to the results considering the baseline features.

When the selected features were analysed with sliding window technique (Table 11), the average accuracy values were more interesting. For instance, the average accuracy values were 87.7, 87.6, 88.9 and 89.1 for EPL, LLPD, BLC 2010 and BLC 2012, respectively. These values are better than those obtained with the baseline features. Moreover, the POL algorithm was also compared to two other approaches for feature selection, such as Relief and PCA. In these comparative tests, our method also achieved the best performances, with variations of average accuracy among 24.2 (POL vs PCA considering data from BLC 2012 dataset and cross-validation technique) and 29.2 (POL vs PCA with data from BLC 2012 dataset and sliding window technique), as shown in Tables 10 and 11.

3.4. Related research

In the literature there is an expressive group of methods with goals of investigating the performance of teams as well as predict the football matches results. All these authors present a strategic methodology that may be adequate, although its consistency will only be evidenced by the final results. Complementary, in Table 12 is presented various methods for predicting results in football matches and their respective AC values, including the results presented in this study. In general, the POL algorithm has reached relevant results in the prediction of football matches. In these comparative tests, each author choose a strategic methodology that looks promising. However, we could not compare them in order to define the best one, for this would be a difficult task, not to say senseless. In fact, the different methodologies are rather complementary than ratable. Therefore, we believe that our method provided the desired robustness for similar applications. Furthermore, we contribute to future works, in which one could profit from valuable information contained in each test and references.

4. Conclusion

The study put forward herein presented a prediction method for football matches results based on information obtained by scouts during matches played in a championship. The data used for the verification of the method is based on the scout information,

which captures the fundamentals of a match. The data were sampled using the two techniques, cross-validation and sliding window, and evaluated by the classification algorithms. The empirical analysis show that the POL algorithm presented results that are relevant to our classification process for data. This resulted in accuracy values above 96% for the metrics of the classifiers with the cross-validation and sliding window techniques.

The POL algorithm is a supervised classifier and presents pertinent results with nonlinear data. This algorithm expands the input data in a superior space dimension, in a manner that allows for the adequate separation between the analysed classes. In this study different combinations were concatenated from the features of 3-dimensions in order to obtain the result. The robustness of the POL algorithm was also verified with addition of different noise levels on the features. The results were highly relevant (accuracy rates from 95% to 100%) even with up to 20% of added noise.

In the POL classifier, the main features were selected during the training and test stage for the definition of a single solution, which separates the classes in order to obtain the final result for the investigated groups. These most relevant features were employed in the analysis of classic learning algorithms, these allowed for an improvement in the obtained accuracy levels for the data from the championships (Section 3.3).

The prediction method shows itself to be extremely relevant reaching accuracy levels considered important for determining results from football matches, either directly with the POL classifier or indirectly with the feature selection for the machine learning algorithms. It is important to highlight that the accuracy indexes obtained were relevant to the data present in the literature for championships with an accrued point system.

The limitations of the POL classifier are related to the processing time of the algorithm as it is of an exponential order. In other words, the higher the number of features involved, higher will be the processing time. In this manner, it is not feasible at the present moment to use this in real time, in order to try and understand or change the direction of the match at the moment in which it is occurring. In this sense, future work will involve the investigation of these features in the context of exploring timeline information from a football match.

References

- Ajmera, P. K., & Holambe, R. S. H. (2010). Speaker recognition using auditory features polynomial classifier. *International Journal of Computer Applications*, 1(14), 86–91.

- Bhatia, K. S. S., Lam, A. C. L., Pang, S. W. A., Wang, D., & Ahuja, A. T. (2016). Feasibility study of texture analysis using ultrasound shear wave elastography to predict malignancy in thyroid nodules. *Ultrasound in Medicine Biology*, 42(7), 1671–1680.
- Brooks, J., Kerr, M., & Gutttag, J. (2016). Using machine learning to draw inferences from pass location data in soccer. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(5), 338–349.
- Bruno, D. O. T., do Nascimento, M. Z., Ramos, R. P., Batista, V. R., Neves, L. A., & Martins, A. S. (2016). Lbp operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues. *Expert Systems with Applications*, 55, 329–340.
- Campbell, W. M., Assaleh, K. T., & Broun, C. C. (2002). Speaker recognition with polynomial classifiers. *IEEE Transactions on Speech and Audio Processing*, 10(4), 205–212.
- Cardinal, B. J., Boonchaaythanasi, K., Cheung, S. Y., Lee, H., & Si, Q. (2016). The modern Olympic games 1896, 2016 and beyond: Can elite sport spectacles incite movement among the masses, or do they merely foster spectatoritis? *Journal of Physical Education, Recreation & Dance*, 87(7), 5–8. doi:10.1080/07303084.2016.1203670.
- Constantinou, A. C., Fenton, N. E., & Neil, M. (2013). Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks. *Knowledge-Based Systems*, 50, 60–86.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303–314. doi:10.1007/BF02551274.
- Broomhead, D. L. (1998). Radial basis functions, multi-variable functional interpolation and adaptive networks, 3(3), 321–355.
- Do Nascimento, M. Z., Martins, A. S., Neves, L. A., Ramos, R. P., Flores, E. L., & Carrijo, G. A. (2013). Classification of masses in mammographic image using wavelet domain features and polynomial classifier. *Expert Systems with Applications*, 40(15), 6213–6221.
- Dua, S., Singh, H., & Thompson, H. (2009). Associative classification of mammograms using weighted rules. *Expert Systems with Applications*, 36(5), 9250–9259.
- Duarte, L., Soares, C., & Teixeira, J. (2015). Previsão de resultados de jogos de futebol (Master's thesis). Faculdade da Engenharia da Universidade do Porto.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification* (2nd ed.). New York: Wiley-Interscience.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. ROC Analysis in Pattern Recognition.
- FIFA (2014). Official documents FIFA world cup. Accessed 03.05.16.
- Garcia, S., & Herrera, F. (2008). An extension on “statistical comparisons of classifiers over multiple data sets” for all pairwise comparisons. *Journal of Machine Learning Research*, 9(Dec), 2677–2694.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The Weka data mining software: An update. *SIGKDD Explorations Newsletter*, 11(1), 10–18.
- Hall, M. A. (2000). Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings of the seventeenth international conference on machine learning*, ICML '00 (pp. 359–366). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Hucaljuk, J., & Rakipović, A. (2011). Predicting football scores using machine learning techniques. In *MIPRO, 2011 proceedings of the 34th international convention* (pp. 1623–1627). IEEE.
- Igiri, C. P. (2015). Support vector machinebased prediction system for a football match result. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 17(3), 21–26.
- Jolliffe, I. (2014). Principal component analysis. *Wiley StatsRef: Statistics reference online*. John Wiley & Sons, Ltd. doi:10.1002/9781118445112.stat06472.
- Karabatak, M. (2015). A new classifier for breast cancer detection based on naïve Bayesian. *Measurement*, 72, 32–36.
- Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Proceedings of the ninth international workshop on machine learning* (pp. 249–256).
- Kononenko, I. (1994). Estimating attributes: analysis and extensions of relief. In *European conference on machine learning* (pp. 171–182). Springer.
- Moore, K. (2016). A second Maracanazo? The 2014 FIFA world cup in historical perspective. In *Sport in society* (pp. 1–17). doi:10.1080/17430437.2016.1158476.
- Olympic (2012). All facts London 2012. Accessed 03.05.16.
- Parinaz, O. F. E., & Sadat, M. F. (2013). Football result prediction with Bayesian network in spanish league-Barcelona team. *International Journal of Computer Theory and Engineering*, 5(5), 812–815.
- Park, B. J., Oh, S. K., & Kim, H. K. (2008). Design of polynomial neural network classifier for pattern classification with two classes. *Journal of Electrical Engineering & Technology*, 3(1), 108–114.
- Pendharkar, P., Khosrowpour, M., & Rodger, J. (2000). Application of Bayesian network classifiers and data envelopment analysis for mining breast cancer patterns. *Journal of Computer Information Systems*, 40(4), 127–132.
- Perin, C., Vuillemot, R., & Fekete, J. D. (2013). Soccerstories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2506–2515.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
- Ramirez-Villegas, J. F., & Ramirez-Moreno, D. F. (2012). Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions. *Neurocomputing*, 77(1), 82–100. doi:10.1016/j.neucom.2011.08.015.
- Sfeir, M. N. (2011). Laws of the game: (adapted from FIFA 2010–11). *World Literature Today*, 85.3 (May-June 2011), 38–40.
- Tax, N., & Joutstra, Y. (2015). Predicting the Dutch football competition using public data: A machine learning approach. *Transactions on Knowledge and Data Engineering*, 10(10), 1–13.
- Trawinski, B., Smetek, M., Lasota, T., & Trawinski, G. (2014). Evaluation of fuzzy system ensemble approach to predict from a data stream. In N. T. Nguyen, B. Attachoo, B. Trawinski, & K. Somboonviwat (Eds.), *Intelligent information and database systems: 6th Asian conference, ACIIDS 2014, Bangkok, Thailand, April 7–9, 2014, Proceedings, Part II* (pp. 137–146). Cham: Springer International Publishing. doi:10.1007/978-3-319-05458-2_15.
- Tüfekci, P. (2016). Prediction of football match results in Turkish super league games. In A. Abraham, K. Wegrzyn-Wolska, A. E. Hassanien, V. Snasel, & A. M. Alimi (Eds.), *Proceedings of the second international Afro-European conference for industrial advancement AECIA 2015* (pp. 515–526). Cham: Springer International Publishing. doi:10.1007/978-3-319-29504-6_48.
- Ulmer, B., & Fernandez, M. (2013). *Predicting soccer match results in the English Premier League*. Stanford (Ph.D. thesis).
- Vafaeipour, M., Rahbari, O., Rosen, M. A., Fazelpour, F., & Ansarirad, P. (2014). Application of sliding window technique for prediction of wind velocity time series. *International Journal of Energy and Environmental Engineering*, 5(2), 1–7.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1–3), 37–52.
- Xiao, J., He, C., Jiang, X., & Liu, D. (2010). A dynamic classifier ensemble selection approach for noise data. *Information Sciences*, 180(18), 3402–3421.
- Zhu, X., & Wu, X. (2004). Class noise vs. attribute noise: A quantitative study. *Artificial Intelligence Review*, 22(3), 177–210.