Journal of Physical Education and Sport ® (JPES), Vol. 23 (issue 4), Art 123, pp. 983 - 989, April 2023
online ISSN: 2247 - 806X; p-ISSN: 2247 – 8051; ISSN - L = 2247 - 8051 © JPES

Original Article

Evaluation of neural network feature and function settings on the model performance and accuracy

MICHAL BOZDĚCH[1], RENÁTA VYCHODILOVÁ[2]
[1,2,]Department of Physical Education and Social Sciences, Masaryk University, CZECH REPUBLIC

Published online: April 30, 2023
(Accepted for publication April 15, 2023)
DOI:10.7752/jpes.2023.04123

Abstract:
Both in general and in sport, the neural network is one of the most frequently used type of the artificial intelligence algorithms. Due to their high price, wealthy sports clubs could only afford to use it on a daily basis in the past. The situation has changed with the development of affordable softwares and their manuals. Although sports club managers can employ them more nowadays, there is still a problem how to prepare the data ideally and to set up the most efficient model algorithm. In the executed studies and literature sources, the function setting has been examined and emphasised more, whereas the set-up of the feature setting has been rather neglected. It has only been recommended to improve the model, but there have not yet been sufficient observation. The current study aims to determine if the features or the function settings have a greater effect on the model accuracy. The initial feature dataset (n = 18882) was obtained from publicly available sources. Each of the six different feature settings consisted of 96 models. A total of 384 models were created, in which their testing accuracy and the percentage difference between the training and the testing phases were further analyzed. No statistically significant differences were found in the accuracy of the function settings, but statistically significant differences were found in the feature settings. Based on the results, the current study concludes that the feature settings are a more important factor to increase the model accuracy than the function settings, especially the reduction of the number of the outputs. Furthermore, the study found that the variables Weight, Height, and Age had the highest frequency of the occurrence of the normalized importance. Therefore, they can be identified as among the most important features to predict the final rank. The results of this study suggest that more emphasis should be put on the feature setting and not just on the function setting when preparing the first model of artificial intelligence.
Key words*:* **ANN; DNN; inputs, outputs selection and extraction; hidden and output layer activation function, optimization algorithm**

Introduction

The term *Artificial Intelligence* (AI) was first used by McCarthy et al. (1955). However, its historical development can be traced back to 1920, when the word *Robot* first appeared in the play named "R.U.R." (Rossum's Universal Robots), written by Karel Čapek. The current AI development is referred to as the fourth wave, and it has not reached its peak yet. If AI architects have a basic knowledge of the investigated phenomenon (e.g. performance, muscle strength, risk of injury), and awareness of the application and limitations of relevant AI models, further rapid development can be expected in this field (not only in sports) (Claudino et al., 2019; Ramkumar et al., 2022). Due to a technological development and growing interest in sports, a huge amount of structured and unstructured data has emerged that can be used to predict future events (Horvat & Job, 2020). The basic concept of the AI models is to use the data to automate selected tasks (Ramkumar et al., 2022). However, it is not recommended to make important decisions on the basis of the AI results only. They should still be taken as an additional decision-making tool. The use of AI in sports is still considered a relatively new, but rapidly developing technology capable of simulating human intelligence with its applications**,** limitations, strengths and weaknesses (Claudino et al., 2019; Glikson & Woolley, 2020). However, it still needs to be explored thoroughly. Machine learning (ML), which is an integral part of AI, refers to the automated detection of the patterns in data sets (Woschank et al., 2020). ML was first understood to allow computers to learn without being explicitly programmed (Samuel, 1959). The present definition describes ML as a computer programming process optimising the performance criteria that use the data (Alpaydim, 2010). ML is typically classified as Supervised Learning, Unsupervised Learning and Reinforcement Learning. In literature, additional divisions can be found, such as Semi-Supervised Learning, Transduction, and Learning to Learn Algorithms and some other (Alpaydim, 2010; Kotsiantis, 2007; Nasteski, 2017). The current study is a typical Supervised Learning type of ML because the model learns from a labeled dataset. To be more precise, it is a type of Classification model, because the outputs are discrete (Gianey & Choudhary, 2018; Nasteski, 2017). The correct use of model selection, evaluation, and algorithm selection techniques of machine learning is vital eveywhere , both in sport and the commercial sector and in academic research (Raschka, 2018). In sports, ML algorithms can help athletes,

Corresponding Author: MICHAL BOZDĚCH, E-mail: michal.bozdech@fsps.muni.cz

coaches and managers to do things, including to predict team and individual performances (in real-time, but also as a long-term prognosis), to make sport-betting decisions, to predict career trajectory, to identify talented athletes and tactical patterns, to analyse and improve motion patterns, to prepare individual training plans and to prevent possible injuries (Barron et al., 2018; Bonilla et al., 2022; Horvat & Job, 2020; Imas et al., 2018; Muazu Musa et al., 2019; Priymak et al., 2020).

Koseler and Stephan (2017) found out that the Support Vector Machine and k-Nearest Neighbor were the most frequently used ML algorithms in baseball analysis. They also concluded that the artificial neural networks (ANN) are the most frequently used model in baseball. This was also confirmed by Horvat and Job (2020), Claudino et al. (2019) (not only in baseball). Bunker and Susnjak (2022) also concluded that the ANN does not necessarily perform significantly better than the other ML models. The results of the model accuracy show that similar models are interchangeable. However, it is necessary to take into consideration that each predictive model has its advantages and disadvantages. Thus, it is not appropriate to compare the results of different studies, as they were obtained from different datasets (Haghighat et al., 2014; Horvat & Job, 2020).

An important preliminary step that significantly contributes to the model accuracy is feature selection and extraction. This can be described as the process of identifying, adding and removing irrelevant and redundant features and thus making the model faster, more efficient and more accurate (Han et al., 2012; Yu & Liu, 2004). As already mentioned, adding features to the model (e.g. a combination of two other features, goals per number of played games) is another technique for maximizing the relevance of the model (Jović et al., 2015). However, it is important to know that various ML algorithms with a higher number of features can suffer from overfitting and low accuracy (Lin et al., 2014). Although there are various methods for reducing feature dimensions, it is difficult to conclude which method provides the best results (Horvat & Job, 2020). Similarly, Raschka (2018) states that machine learning involves a lot of experimentation to gradually fine-tune the model's accuracy. In the best situation, it does not experiment with one algorithm only (the best, or the most logical for the given situation), but with several which are compared with each other. This is possible because the data is from the same dataset.

In the past, only wealthier sports clubs could afford to pay specialists (employees or services of another company) for the data analysis to subsequently apply the results successfully in practice. The situation is changing with the availability of literature sources and the development of software with a simple and intuitive interface (such as SPSS, MATLAB, Statistica, etc.). Many sports managers (coaches, stakeholders, data analysts, or sports enthusiasts) can take the advantage of AI, without the need to know a suitable programming language or to pay for data analyst services. Although it is possible to find out some exceptions, most studies do not present a comprehensive description of how to predict (via accuracy) the results for the comparison using the initial feature set and feature selection as this is the know-how of the ML architect. The studies mostly are missing a detailed description of the features, such as adding and removing irrelevant and redundant features or compared accuracy results to different model settings, according to the number of hidden layers, types of activation functions, output layers, training types and optimisation algorithms. For this reason, the current study aims to demonstrate the application of various methods to increase the accuracy of the predicted results of the dataset. The goal of the study was to quantify differences between two approaches (feature and function setting) and enable the findings to be used by students, athletes, sports managers and coaches as well as the general public.

Material & methods

Data collection

The dataset contained 18882 data points??items" from a total of 1929 players. The data was obtained from the publicly available and validated (from the sports association) website (www.fortuna.cz). The set inclusion criteria were as follows: seasons 2014/2015–2018/2019, the overall statistics of soccer players (of 16 teams) for the top Czech league; only forwards, defenders, and midfielder playing positions. The excluded variable was the nationality and all the variables related to the goalkeepers (as their stats differed significantly from other player positions, e. g. height, number of goals, yellow and red cards received). Missing data was omitted. The above seasons were only analysed due to the change in the rules and the objectives of this study.

Dependent variables (for the 1st and 2nd phases of this study) were nominal (player positions), ordinal (number of games played, numbers of goals, yellow and red cards received) and of scale (age, height, weight). They were obtained and calculated up to the last day of the season. Five consecutive seasons have been analysed separately (not cumulatively) as it provides a greater accuracy (Horvat & Job, 2020).

Procedure

The only output variable was the final team/players ranking at the end of the season (ordinal scale; 1st to 16th rank for 1st to 4th phases; by quarters for 5th phase; 1st rank and the others rank was used for 6th phase).

The default settings of the Artificial Neural Network (ANN) if there was only 1 hidden layer, or the Deep Neural Network (DNN) if there were 2 hidden layers, were always set as the Multilayer Perceptron (suitable for solving non-linear problems); Standardised Rescaling of Covariates; 70% of partitions placed into training and 30% to testing the model; the maximum number of the epochs was 10,000. Where the output was the players/teams' rank at the end of the season (as a ordinal variable or discrete) and the input was a continuous (the decimal age calculated on the last day of the season; the player's height and the weight), and discrete (the number of played games; the number of goals; the number of yellow and red cards) variables. Then (in the 3rd phase of research), the ratios Goal/games, Goal/yellow cards, Goal/red cards, and Age/games were calculated.

The research contained 6 individual computed and analyzed phases of the machine learning modelling using Artificial Neural Network (ANN) and Deep Neural Network (DNN), in which the AI models' output and input features were gradually, systematically and purposefully manipulated. More precisely, 1st phase contained only the raw data. In the second phase, the NN models without outliers were calculated. The newly calculated variables were added to the model in the third phase. The variables with the least effect on the model accuracy were removed from the model in the fourth phase. The fifth phase consisted of the reduction (from 16 to 4) outputs. The last (the sixth) phase continued with the reduction OF the number of outcomes from 4 to 2 (1st rank and the other ranks). Although modelling the accurate machine learning algorithms do not have a one-size-fits-all approach, this is a recommended procedure to refine the model accuracy (Horvat & Job, 2020; MathWorks, 2020).

Each of the six individual phases contained the same and predetermined order set of the individual NN model functions. More precisely, the models contain two hidden layers, two activation functions (Sigmoid, Hyberbolic tangen,), four output layers (Softmax, Identity, Sigmoid, Hyberbolic tangent), three types of training (Batch, Minibatch, Online), and two optimisation algorithms (Gradient descent, Scaled conjugate gradient). The combination of the algorithms that cannot be combined were not involved. There was a total of 64 possible models in each research phase. The training and testing accuracy were obtained for all calculated models (n = 384). Further, the results of the percentage difference (%diff) between the training (T_1) and testing (T_2) model accuracy was calculated with the formula: $(|T_1 - T_2|) / ((T_1 + T_2) / 2)$.

According to their percentage of the importance of individual variables, the most accurate models from testing accuracy were thoroughly analysed using the Receiver Operator Characteristic (ROC) and the Area Under the ROC Curve (AUC).

Statistical analysis

The Kruskal-Wallis test was used to compare the results of the model accuracy between different functions (up to 96 individual models) and feature settings (1st to 6th phases of the research), because the assumption for the parametric analysis of the variance test was violated according to the results of Levene's test. The level of the statistical significance was set at $p < 0.05$ and also adjusted for multiple comparisons (using Bonferroni correction). For the statistical analyses and to calculate the models, IBM SPSS software version 28.0.0 for Windows (IBM SPSS Inc. Chicago, IL, USA) was used. No coding language was used in this study, only a basic use of the syntax.

Results & Discussion

There was 96 different ways to modify a given NN model. The testing accuracy of the values of the individual models are presented in Fig. 1. Table 1 only contains the models with the lowest and the highest calculated accuracy, gradually according to the recommended procedure for working with the data when creating a NN model. Raw data only; after removing outliers from the continuous variables; adding possible relevant features; removing redundant and irrelevant features (based on the results of the importance); modification of the number of the outputs (based on the quarter of the final rank at the end of the season). The results of the last phase are not listed in Fig. 1 and Table 1, but are more thoroughly analysed separately (in Table 2).

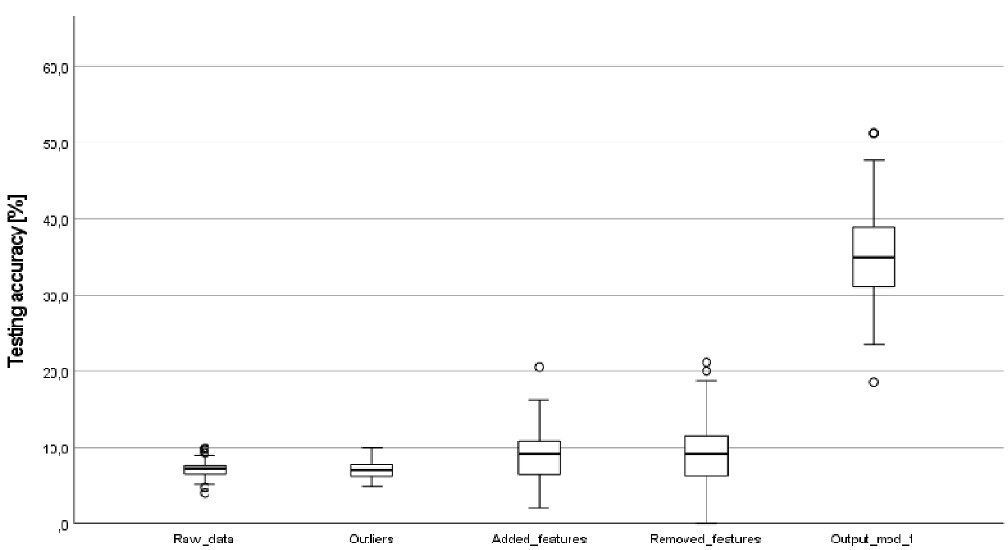

Fig. 1. An overview of accuracy results from model testing in individual phases.

The box plot in Fig. 1 shows the different dispersion of the testing accuracy according to the various phases of the feature settings. This could be described as the first phase containing the lowest prediction accuracy, which tended to increase with each subsequent phase. However, with the increasing model accuracy, the heterogeneity of the results also increased.

Table 1. an overview of the most accurate models in the four individual phases of strengthening.

Model	Architecture			Training		Classification [%]	
	Hidden Layer	Activation Function	Output Layer	Type of Training	Optimization algorithm	Training	Testing
(1) Raw_LA	2	Tanh	Tanh	Batch	GD	8.2	4.0
(1) Raw_MA	1	Tanh	Identity	Minibatch	GD	12.1	9.9*
(2) Outliers_LA	2	Tanh	Identity	Online	GD	8.4	4.8
(2) Outliers_MA	1	Tanh	Identity	Online	GD	10.9	10.0
(3) Add features_LA	1	Tanh	Tanh	Online	GD	13.1	2.1
(3) Add features_MA	2	Tanh	Identity	Minibatch	GD	8.8	20.5
(4) Remove features_LA	1	Sigmoid	Sigmoid	Batch	GD	11.6	0.0*
(4) Remove features_MA	2	Sigmoid	Identity	Batch	SCG	7.8	21.1
(5) Output mod_1_LA	2	Tanh	Softmax	Batch	SCG	39.8	18.6
(5) Output mod_1_MA	1	Sigmoid	Sigmoid	Batch	SCG	42.9	51.3

Note: LA … The least accurate model; MA … the most accurate model; Tanh … Hyperbolic tangent; SCG … Scaled conjugate gradient; GD … Gradient descent; * … in case of duplicate results, larger or smaller values from the model training decided.

The increasing accuracy of the model with the increasing number of features contradicts the conclusions stated by Horvat and Job (2020) who found out that with the increasing number of the features (and seasons), the outcome prediction decreased. This opposite conclusion may be because the model from this study did not reach the greatest possible accuracy. In Fig. 1, it can be further seen that after removing the least important features from the model, neither the accuracy nor the difference changed significantly. This was also evident when analysing the percentage of differences between the training and testing phases.

According to the results of the descriptive analysis, it can be argued that the accuracy of the testing phase is similar between the raw data ($\bar{x} = 7.2 \pm 1.2$; skewness = 0.1; kurtosis = 1.0) and the data without outliers ($\bar{x} = 7.1 \pm 1.1$; skewness = 0.2; kurtosis = -0.2). It increases slightly when the features are added ($\bar{x} = 9.0 \pm 3.6$; skewness = 0.5; kurtosis = 0.7), but does not change much after removing the excess features ($\bar{x} = 9.2 \pm 4.5$; skewness = 0.4; kurtosis = 0.2). The prominent increase in the model accuracy came after the first reduction of the number of outputs (from 16 to 4 outputs; Output_mod_1, $\bar{x} = 35.3 \pm 6.5$; skewness = 0.1; kurtosis = 0.3). It is not necessary to characterize the descriptive statistics of the last phases of this study (Output_mod_2) as all calculated results of the model testing reached 100% model accuracy.

From the above mentioned, it is possible to confirm the conclusions by Horvat and Job (2020) that adding a feature (e.g. combining initial features) contributes to a greater accuracy of the NN model. But a much more significant increase in the model accuracy is achieved by reducing the number of the outputs. This was also confirmed by the Bunker and Susnjak study (2022) where the authors state that sports with a larger number of possible outcomes provide a lower model accuracy, which was most evident in the sixth phase of the current study where the dependent variable was adjusted from 4 to 2 outputs. This caused 41 models to reach 100% accuracy (in model testing phase). Weight (36.59%, n = 15) was the variable with the highest number of the occurrences of 100% Normalised importance. Then, Height (17.07%, n = 7), Age (14.63%, n = 6), Goal/red (9.76%, n = 4), Games played (7.32%, n = 3), Age/games (7.32%, n = 3), Goal/yellow (4.88%, n = 2), number of goals (2.44%, n = 1), Goal/games (0.00%, n = 0), respectively. If the model accuracy was further refined, the last 3 variables would be excluded from the NN modelling. The number of goals was only found once as a feature with the highest normalised importance because compared to other sports, fewer goals are scored in soccer, or because all player positions were added to the calculated NN models. It may be suggested for future research studies that a separate models should be calculated for each player's position and the occurrence of their normalised importance should be compared.

After fine-tuning the model accuracy, it is typical that another type of performance-metric optimisation is performed, such as the Area Under the Curve curve and the Receiver Operating Characteristic curve (Raschka, 2018). Therefore, these 41 most accurate models were further evaluated using the Receiver Operating

Characteristic Curve and the Area Under the Curve, which ranged from 0.125 to 1.000. The following Table 2 provides an overview of the model architecture and training settings of the NN models (n = 14) that achieved 100% accuracy and the Area Under the Curve was equal to 1.00.

Table 2. Overview of the model settings with 100% accuracy and 1.00 AUC.

Model	**Architecture**			**Training**	
	No. Hidden Layer	**Activation Function**	**Output Layer**	**Type of Training**	**Optimization algorithm**
[1]	1	Hyberbolic tangent	Softmax	Batch	Scaled conjugate gradient
[2]	1	Hyberbolic tangent	Softmax	Batch	Gradient descent
[3]	1	Hyberbolic tangent	Sigmoid	Batch	Scaled conjugate gradient
[4]	1	Sigmoid	Softmax	Batch	Gradient descent
[5]	1	Sigmoid	Sigmoid	Online	Gradient descent
[6]	1	Sigmoid	Sigmoid	Minibatch	Gradient descent
[7]	2	Hyberbolic tangent	Softmax	Batch	Scaled conjugate gradient
[8]	2	Hyberbolic tangent	Softmax	Online	Gradient descent
[9]	2	Hyberbolic tangent	Softmax	Minibatch	Gradient descent
[10]	2	Hyberbolic tangent	Sigmoid	Batch	Scaled conjugate gradient
[11]	2	Hyberbolic tangent	Sigmoid	Batch	Gradient descent
[12]	2	Hyberbolic tangent	Sigmoid	Minibatch	Gradient descent
[13]	2	Sigmoid	Softmax	Online	Gradient descent
[14]	2	Sigmoid	Softmax	Minibatch	Gradient descent

It seems that the best models are composed of two hidden layers and Hyberbolic tangent as the activation function. For the output layer, it does not matter whether you use the Sigmoid or Softmax activation function. The most appropriate combination of individual elements was not found for the type of training and optimisation algorithm. Therefore, it may more depend on the data set and the situation for which the NN model is modelled. It is worth mentioning, that Table 2 does not contain an Identity and Hyperbolic tangent output layer. Activation functions transform the neuron's activation level into the output signal. Several activation functions can significantly affect the model accuracy (Karlik & Olgac, 2011). Studies focusing on the model accuracy (according to the activation function) indicate that although the Hyperbolic Tangent Function is similar to the Sigmoid function, the Sigmoid functions are used most commonly and different settings of the activation function improve the Artificial neural network (ANN) accuracy (Buhmann, 2003; Karlik & Olgac, 2011; Widrow & Lehr, 1990). For the Deep neural network (DNN), the Hyperbolic tangents as activation functions for both layers of neurons provide a better accuracy (Karlik & Olgac, 2011).

The fact is that there is not one but 14 models with 100% accuracy, and the Area Under the Curve with 1.00 was very unexpected. For comparison, Hucaljuk and Rakipović (2011) used several machine learning algorithms to predict outcomes in the Champions' League and achieved a maximum accuracy of 68.8%. Similar accuracy was described by Horvat and Job (2020) in their extensive review, where the average maximum accuracy of predicting soccer match outcomes was 72.43% (with 2 outliers) with a 54.55–93.00% range of the average maximum accuracy. In general, it is easier to predict the results of individuals than teams, where the outcomes are affected by multiple variables that are not measurable or difficult to measure (Horvat and Job, 2020). Another reason why the model accuracy was higher than in other studies could be based on the fact that a classification model type was used to predict the outcome (Delen et al., 2012; Elfrink, 2018; Soto Valero, 2016). It is important to note that it is not appropriate to compare the model accuracy with other studies as either they come from different datasets or the models were focused on predicting different outcomes (Bunker & Susnjak, 2022; Horvat & Job, 2020).

Fig. 1 indicated significant differences between the individual research phases, which was confirmed by the Kruskal-Wallis test ($p < .001$). After adjusting for multiple comparisons testing, the post hoc test also found out statistically significant differences in 8/15 cases (53.3%). More precisely, these were the differences between 5th phase and every other research phase (except 6th phase), and 6th phase and every other research phase (except 5th phase). No statistically significant differences were found between the testing model accuracy and the type of the NN model setting in any of the research phases; $p > 0.05$. These results suggest that since there were no statistical differences in the accuracy between the settings of the individual model algorithms in each research phase, the feature setting (especially the output setting) is more important for achieving a greater accuracy.

As mentioned above, the neural network (NN) is the most widely used machine learning model. In practice, the neural network may not be the most appropriate method for future studies with a similar focus, because the dataset only consists of the end-of-season data (when the results are already known). Therefore, it would be appropriate to try other methods and procedures for predicting the final ranking of the team. Thus, the methods such as Forecasting and Bayes Theorem could provide more relevant predictions.

The inability to compare the model accuracy of two different algorithms - since they do not come from the same dataset - significantly complicates the work of researchers and data scientists. Therefore, it would be suitable to invent (or derive) methods (approaches, strategies) for this type of analysis (e.g based on the effect size testing).

Conclusions

Due to a huge popularity of the artificial intelligence and machine learning, the amount of statistical softwares where it is possible to create a model is growing. For these reasons, this study focused on quantifying the effect of the feature and function settings, and on the accuracy of the classification of the neural network model. The results suggest that since there were no differences in the model accuracy between the settings of the individual algorithms (in every research phase), in contrast to the feature settings where statistically significant differences were found, it can be stated that the feature settings are more important for better model accuracy. In particular, the lower number of the output variables the model has, the better effect on model accuracy can be observed. Therefore, it would be appropriate to prepare the model features and consult with an expert on a given (sports) issue what the model should predict and whether the results of the model are applicable in practice. From 41 different algorithms (which reached 100% model accuracy), The weight, height, and age were the most often features with the normalised importance occurrence. This suggests that not the number of the goals (not even the other derivatives arising from this variable), but these variables are better for predicting whether a team will win in their soccer league.

The advantage but also disadvantage of the AI models is that they can find even small trends. However, this does not mean that the creators of the NN algorithms should not spend time on data preparation, or the correct selection and combination of variables. On the contrary, the conclusions from this study show that this effort can contribute not only to the general validity, but mainly to the accuracy of the model itself. Without the application of these steps, there is a high probability that the model will show a low accuracy, which was not caused by the absence of the trend, but does not prevent the presence of disturbing variables or the omission of important variables. Therefore, the AI model must not be set up only by experienced AI architects (e.g. from mathematical sciences) who do not have the knowledge and experience from the sports environment, as they can simply miss important factors or find a trend using irrelevant variables that is not applicable in practice, even if it will have a high model accuracy. Therefore, instead of learning and deepening the knowledge about the application and setting the various AI functions, it is advisable to focus on the practical preparation, which should consist of understanding the task, and how it is intended to be achieved.

Conflicts of interest - The authors have no conflicts of interest to declare.

References:

Alpaydim, E. (2010). *Introduction to Machine Learning*. The MIT Press.

Barron, D., Ball, G., Robins, M., & Sunderland, C. (2018). Artificial neural networks and player recruitment in professional soccer. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0205818

Bonilla, D. A., Peralta-Alzate, J. O., Bonilla-Henao, J. A., Urrutia-Mosquera, W., Cannataro, R., Kočí, J., & Petro, J. L. (2022). Unsupervised machine learning analysis of the anthropometric characteristics and maturity status of young colombian athletes. *Journal of Physical Education and Sport, 22*(1), 256–265. https://doi.org/10.7752/jpes.2022.01033

Buhmann, M. D. (2003). *Radial basis functions: theory and implementations* (Vol. 12). Cambridge university press.

Bunker, R., & Susnjak, T. (2022). The application of machine learning techniques for predicting results in team sport: a review. *Journal of Artificial Intelligence Research* 73, 1285-1322. https://doi.org/10.1613/jair.1.13509

Claudino, J. G., Capanema, D. D. O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports medicine-open, 5*, 1-12.

Čapek, K. (1920). *RUR (Rossum's Universal Robots)*. Aventinum.

Delen, D., Cogdell, D., & Kasap, N. (2012). A comparative analysis of data mining methods in predicting NCAA bowl outcomes. *International Journal of Forecasting*, *28*(2), 543–552. https://doi.org/10.1016/j.ijforecast.2011.05.002

Elfrink, T. (2018). *Predicting the outcomes of MLB games with a machine learning approach*. Vrije universiteit.

Gianey, H. K., & Choudhary, R. (2018). Comprehensive Review on supervised machine learning Algorithms. *Proceedings - 2017 International Conference on Machine Learning and Data Science, MLDS 2017*, *2018-January*, 38–43. https://doi.org/10.1109/MLDS.2017.11

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Haghighat, M., Rastegari, H., & Nourafza, N. (2014). A Review of Data Mining Techniques for Result Prediction in Sports. *Advances in Computer Science: An International Journal*, *2*(6), 7–12. https://www.researchgate.net/publication/262560138

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Concepts and Techniques* (3rd ed.). Elsevier.

Horvat, T., & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(5). https://doi.org/10.1002/widm.1380

Horvat, T., Job, J., & Medved, V. (2018). Prediction of euroleague games based on supervised classification algorithm k-nearest neighbours. *Proceedings of the 6th International Congress on Sport Sciences Research and Technology Support (IcSPORTS 2018)*, 203–207. https://doi.org/10.5220/0006893502030207

Hucaljuk, J., & Rakipović, A. (2011). Predicting football scores using machine learning techniques. *Proceedings of the 34th International Convention MIPRO*. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5967321&isnumber=5967009

Imas, Y., Khmelnitska, I., Khurtyk, D., Korobeynikov, G., Spivak, M., & Kovtun, V. (2018). Neural network modeling of diagonal stride technique of highly qualified skiers with hearing impairments. *Journal of Physical Education and Sport, 18*(Suppl. 2), 1217–1222. https://doi.org/10.7752/jpes.2018.s2181

Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. *Proceedings of 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. https://doi.org/doi: 10.1109/MIPRO.2015.7160458.

Karlik, B., & Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, *111*(122). https://www.researchgate.net/publication/228813985

Koseler, K., & Stephan, M. (2017). Machine learning applications in baseball: a systematic literature review. *Applied Artificial Intelligence*, *31*(9–10), 745–763. https://doi.org/10.1080/08839514.2018.1442991

Kotsiantis, S. B. (2007). Supervised machine learning: a review of classification techniques. *Informatica*, *31*, 249–268.

Lin, J., Short, L., & Sundaresan, V. (2014). *Predicting National Basketball Association Winners*.

MathWorks. (2020, March 4). *Introduction to Machine Learning, Part 4: Getting Started with Machine Learning*. https://www.mathworks.com/videos/introduction-to-machine-learning-part-4-getting-started-with-machine-learning-1542879650900.html

McCarthy, J., Minsky M. L., Rochester, N., Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine, 27*(4), 12–14.

Muazu Musa, R., Abdul Majeed, A. P. P., Taha, Z., Abdullah, M. R., Husin Musawi Maliki, A. B., & Azura Kosni, N. (2019). The application of Artificial Neural Network and k-Nearest Neighbour classification models in the scouting of high-performance archers from a selected fitness and motor skill performance parameters. *Science and Sports*, *34*(4), e241–e249. https://doi.org/10.1016/j.scispo.2019.02.006

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*, *4*, 51–62. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05

Priymak, S., Kolomiets, N., & Goletc, V. (2020). Forecasting the game role of volleyball players in accordance with the methodology of artificial intelligence. *Journal of Physical Education and Sport*, *20*(1), 179–185. https://doi.org/10.7752/jpes.2020.01024

Ramkumar, P. N., Luu, B. C., Haeberle, H. S., Karnuta, J. M., Nwachukwu, B. U., & Williams, R. J. (2022). Sports medicine and artificial intelligence: a primer. *The American Journal of Sports Medicine, 50*(4), 1166-1174.

Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. http://arxiv.org/abs/1811.12808

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal*, *3*(3), 535–554.

Soto Valero, C. (2016). Predicting win-loss outcomes in MLB regular season games-a comparative study using data mining methods. *International Journal of Computer Science in Sport*, *15*(2), 91–112. https://doi.org/10.1515/ijcss-2016-0007

Widrow, B., & Lehr, M. A. (1990). 30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation. *Proceedings of the IEEE*, *78*(9), 1415–1442. https://doi.org/10.1109/5.58323

Woschank, M., Rauch, E., & Zsifkovits, H. (2020). A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics. *Sustainability (Switzerland)*, *12*(9). https://doi.org/10.3390/su12093760

Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *The Journal of Machine Learning Research*, *5*, 1205–1224.