Analysis of key parameters for choosing a kind of sport based on human morphofunctional indicators using statistical and machine learning methods

Olga Pavlova[1,†], Vitalii Alekseiko[1,*,†], Valeriia Shvaiko[1,†], Nikita Vusatyi[1,†] , Houda El Bouhissi[2,†] and Roman Gakh[3,†]

[1] *Khmelnytskyi National University, Institutska str., 11, Khmelnytskyi, 29016, Ukraine*
[2] *LIMED Laboratory, Faculty of Exact Sciences,University of Bejaia, 06000, Bejaia, Algeria*
[3] *West Ukrainian National University, 11 Lvivska st., Ternopil, 46001, Ukraine*

Abstract

The work analyzes the influence of human morphofunctional indicators on the predisposition to a certain kind of sport. The existing decision support systems in the sports sphere are considered. The role of artificial intelligence for parameter analytics is determined. Statistical analysis of the data was performed. The statistical significance of each parameter was established using the t-test. The correlation value between sports, as well as between morphofunctional indicators using the Pearson and Spearman methodologies, is calculated for a more comprehensive understanding of the relationships. The importance of morphofunctional indicators for choosing a kind of sport is investigated using machine learning technologies, namely the random forest method is chosen.

Keywords

Kind of sport, correlation, data analysis, artificial intelligence, morphofunctional indicators

1. Introduction

In today's virtualized world, more and more people are striving to find the optimal sport for themselves, which will not only contribute to physical development, but also correspond to their individual characteristics and capabilities. However, the human factor often causes false decisions: people choose activities that do not meet their physiological characteristics or do not take into account potential risks, which can lead to injuries or loss of motivation. As a result, they quickly stop training without achieving the desired results.

There are many approaches to choosing a sport [1, 2] according to a person's preferences, but most studies are focused on the general choice of group or individual sports, as well as the availability of training in a particular region. However, in addition to these undoubtedly

The Second International Conference of Young Scientists on Artificial Intelligence for Sustainable Development (YAISD), May 8-9, 2025, Ternopil, Ukraine

* Corresponding author.
† These authors contributed equally.
pavlovao@khmnu.edu.ua (O. Pavlova); vitalii.alekseiko@gmail.com (V. Alekseiko); lera.schvajko@gmail.com (V. Shvaiko); mixggua@gmail.com (N. Vusatyi); houda.elbouhissi@gmail.com (H. El Bouhissi); r.hakh@wunu.edu.ua (R. Gakh)
0000-0001-7019-0354 (O. Pavlova); 0000-0003-1562-9154 (V. Alekseiko); 0009-0001-4267-3479 (V. Shvaiko); 0009-0005-1518-7087 (N. Vusatyi); 0000-0003-3239-8255 (H. El Bouhissi); 0000-0002-3937-2210 (R. Gakh)

important parameters, it is also advisable to take into account the morphofunctional characteristics of a person.

The aim of the work is to identify the correlation between morphofunctional indicators and sports, as well as to establish the importance of each indicator.

At this stage, artificial intelligence (AI) can play a significant role, which is a powerful tool for selecting a sport based on human morphofunctional indicators. Using large amounts of data, the system can analyze the physical characteristics of professional athletes in various sports and forms the necessary parameters and standards on the basis of this. This allows to make an objective and scientifically based choice of optimal activity for a particular person.

In addition, AI is able to analyze individual user data such as height, weight, body composition, endurance level, etc., and offer the most suitable sport. The system based on AI can also assess potential risks, anticipate possible health problems, and make recommendations about exertion, training intensity, and safety techniques. This approach not only increases the effectiveness of sports, but also contributes to the long-term involvement of people in an active lifestyle, helping them to avoid overloads and injuries.

Thus, the introduction of AI in the field of sports opens up new opportunities for a personalized approach to the choice of physical activities. This allows everyone to find the best option for maintaining health, developing physical abilities and achieving sports goals.

2. Literature Review

In our previous works [3, 4] the indicators-based decision support system and method for choosing kind of sport were proposed. Also, the indicators and values of their impact on different kinds of sport were determined. Moreover, the analysis of the relevant studies describing using AI for decision support in choosing kind of sport was conducted.

The article [5] demonstrates the potential of AI to improve sports performance analysis and decision-making during training and competition. The study [6] provides a practical perspective on the application of AI in real-world health tracking systems that can be integrated into sports applications. The research [7] describes the integration of AI to facilitate more accurate analysis and prediction of sports training results using mobile sensors and deep neural networks. The article [8] discusses current trends and practical solutions in sports technologies based on the use of artificial intelligence. The research [9] analyzes which technological innovations influence the development of sports applications and how AI contributes to their improvement. The article [10] demonstrates the ability to detect and classify different types of sports activities from live video streams using convolutional neural networks. The study [11] describes the development of a mobile application for personalized coaching of runners that uses AI to analyze data from sensors and biometric indicators to optimize the training process. The article [12] identifies AI's impact on China's sports industry. The study [13] is devoted to the legal aspects of implementing AI systems in the sports sector under the European Union legislation.

The research [14] analyzes morphofunctional indicators of an elite Chilean mountain runner. It was provided laboratory experiments to improve training strategy. The article [15] reveals medical point of view. This study focuses on analyzing the effects of isometric and isotonic exercise training on the morpho-functional parameters of the right ventricle in Olympic athletes. Also it was analyzed machine learning approaches [16, 17, 18] and its usage for health monitoring [19]. However, the considered studies do not solve the problem of morphofunctional indicators' analysis and choosing a kind of sport on basis of them.

On the basis of provided study it was formulated hypothesis accordingly to the topic of the research: There is a significant difference between the indicators' means of two kind of sports.

3. Methodology

The study used expert data from specialists in the field of physical education and sports. A detailed description of the data used is given in previous works [3, 4].

The t-test is used to test a statistical hypothesis, which is used to compare the means of two groups and determine whether the difference is statistically significant. The formula 1 describes if the difference between the means of two groups is significant [20].

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \tag{1}$$

Where:

$\overline{X}_1, \overline{X}_2$ – sample means of groups 1 and 2;

s_1^2, s_2^2 – sample variances of groups 1 and 2;

n_1, n_2 – sample sizes of groups 1 and 2.

The larger the values, the greater the difference between the two groups.

Pearson and Spearman methodologies were used to determine the correlation between parameters. Comparing the two approaches allows for a more comprehensive assessment due to the possibility of analyzing both linear and monotonic dependencies.

Pearson's correlation coefficient determines the linear relationship between two variables:

$$r = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{(X_i - \overline{X})^2} \cdot \sqrt{(Y_i - \overline{Y})^2}}, \tag{2}$$

Where:

X_i, Y_i – individual data points;

$\overline{X}, \overline{Y}$ – the mean values of X and Y;

In formula (2) the numerator represents the covariance between X and Y, and the denominator normalizes the values by their standard deviations. The r value 1 is represent perfect positive correlation, 0 – no correlation and –1 – perfect negative correlation.

The Spearman correlation measures the monotonic relationship between two variables by ranking the data before calculation:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}, \tag{3}$$

Where:

d_i – the difference between the ranks of corresponding values X_i and Y_i;

n – the number of data points.

Similar to Pearson correlation, the ρ value 1 is represent perfect increasing monotonic relationship, 0 – no monotonic relationship and –1 – perfect decreasing monotonic relationship.

4. Results and Discussions

All morphofunctional indicators used in the research are presented in Table 1 with description. For better visualization on plots it is used numbering. Also Table 1 shows T-test results. All indicators have extremely small p-values (< 0.05), meaning they are statistically significant.

Table 1
T-test results

Indicator number	Indicator	t statistics	p value
1	Height, cm	11.21	$5.03 \cdot 10^{-11}$
2	Weight-height index (body mass index), weight (g), height (cm)	6.59	$8.19 \cdot 10^{-7}$
3	Muscle development index	14.06	$4.39 \cdot 10^{-13}$
4	Ratio of arm span to body length while standing, cm	4.89	$5.52 \cdot 10^{-5}$
5	Running 30 m, s	10.79	$1.1 \cdot 10^{-10}$
6	Standing long jump, cm	11.31	$4.23 \cdot 10^{-11}$
7	Throwing a stuffed ball over a distance (1 kg), m	21.65	$2.93 \cdot 10^{-17}$
8	Sitting raises in 60 seconds, number	12.59	$4.62 \cdot 10^{-12}$
9	Flexion and extension of arms in a prone position, number	13.48	$1.09 \cdot 10^{-12}$
10	Standing torso tilt (torso forward tilt from a sitting position), cm	19.19	$4.58 \cdot 10^{-16}$
11	Shuttle race (4x9 m), s	15.95	$2.84 \cdot 10^{-14}$
12	Reaction speed (fishing stick with centimeter markings), cm	26.05	$4.12 \cdot 10^{-19}$
13	Jump rope in 60 seconds, number	16.88	$8.06 \cdot 10^{-15}$
14	Ruler twist (difference from shoulder width), cm	80.00	$1.19 \cdot 10^{-30}$

The smallest p-value is observed for indicator 14. This indicates an extremely strong difference between the groups. Indicators 7 and 12 also have very small p-values, indicating clear group differences. Indicator 14 has the highest t-statistic (80.00). Indicators 7, 10, and 12 also have high t-statistics. Indicators 4 and 2 have the lowest t-statistics, but still show significant differences. Thus, all indicators show statistically significant differences between the two groups.

Relationships between kind of sports on the basis of important for them morphofunctional human indicators are shown in Figure 1 as a adjency matrix heatmap. To check for the presence of relationships, a binary classification was used, where 0 means the absence of such indicators, and 1 indicates their presence.

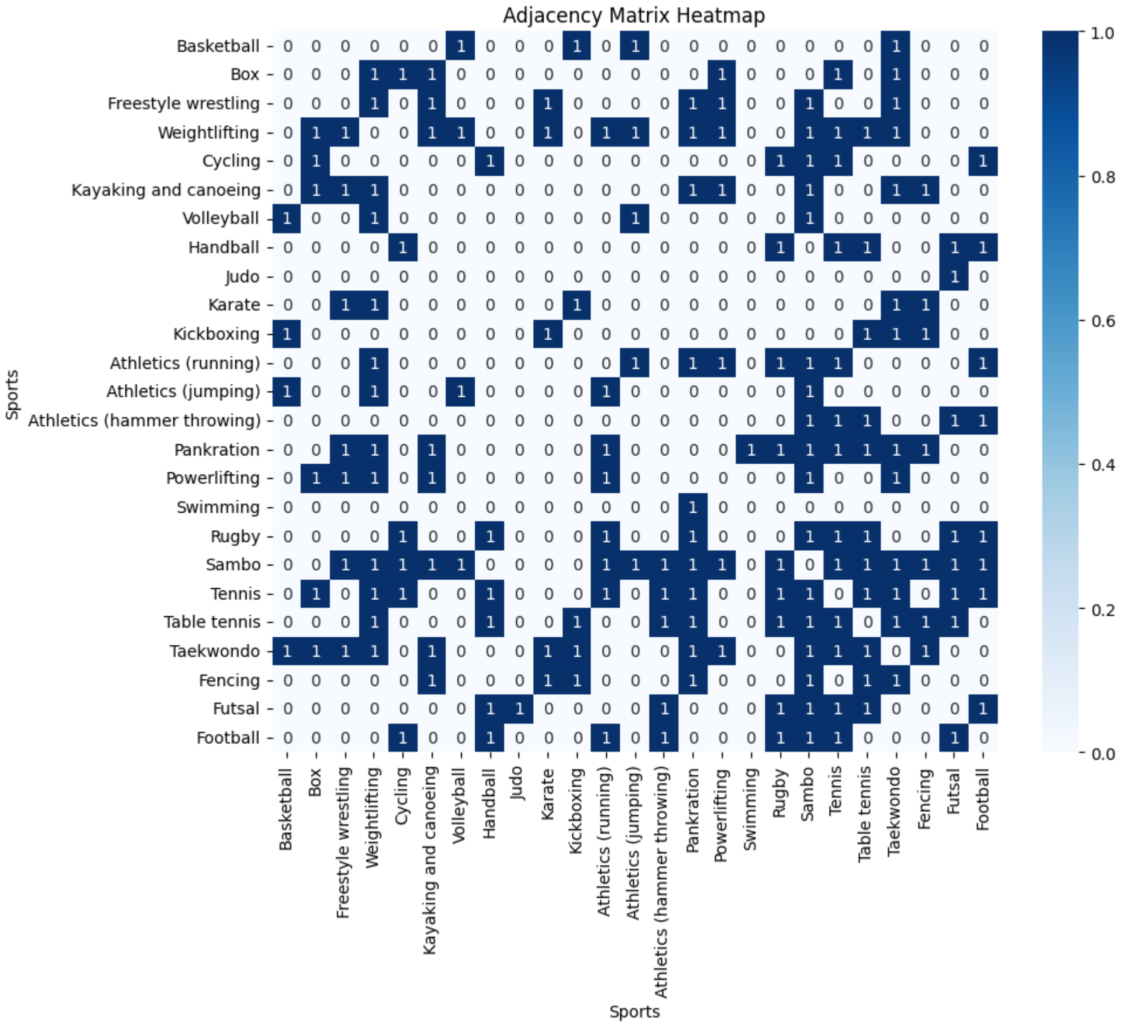

Figure 1: Adjacency matrix heatmap for kinds of sport.

Figure 2 shows boxplot of morphofunctional indicators. Each square represents the interquartile range (IQR). The median indicates the central value. The whiskers represent the range of most of the data points.

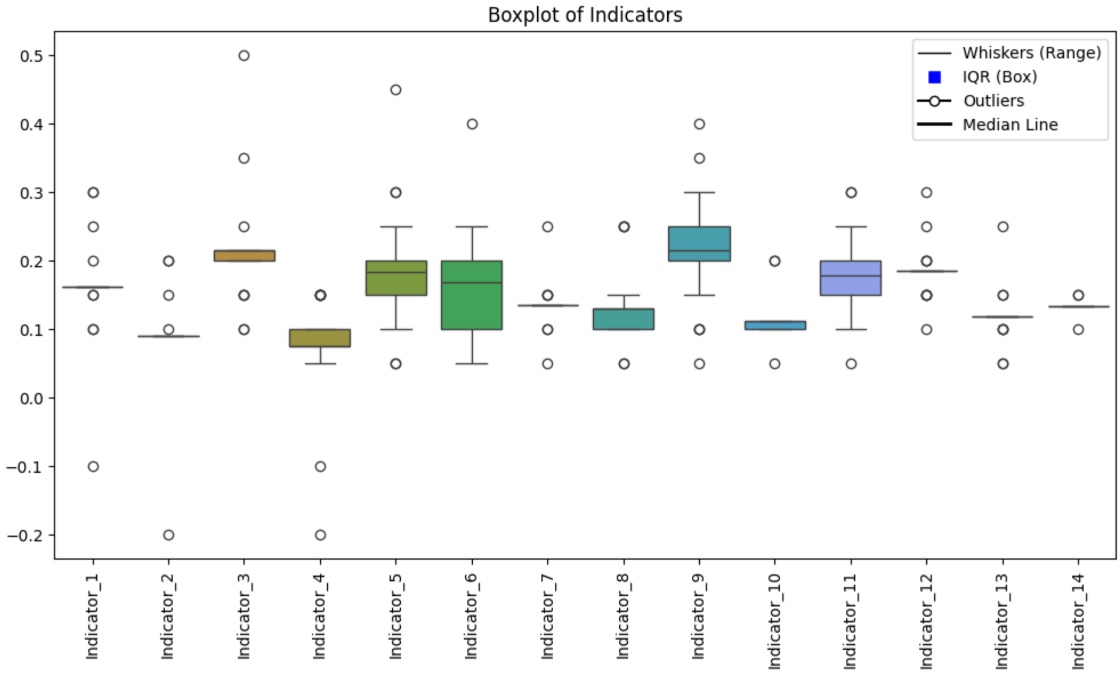

Figure 2: Boxplot of morphofunctional indicators.

Figure 3 shows a histogram of indicators. X-axis represents the range of values for a particular indicator and Y-axis represents the number of observations.

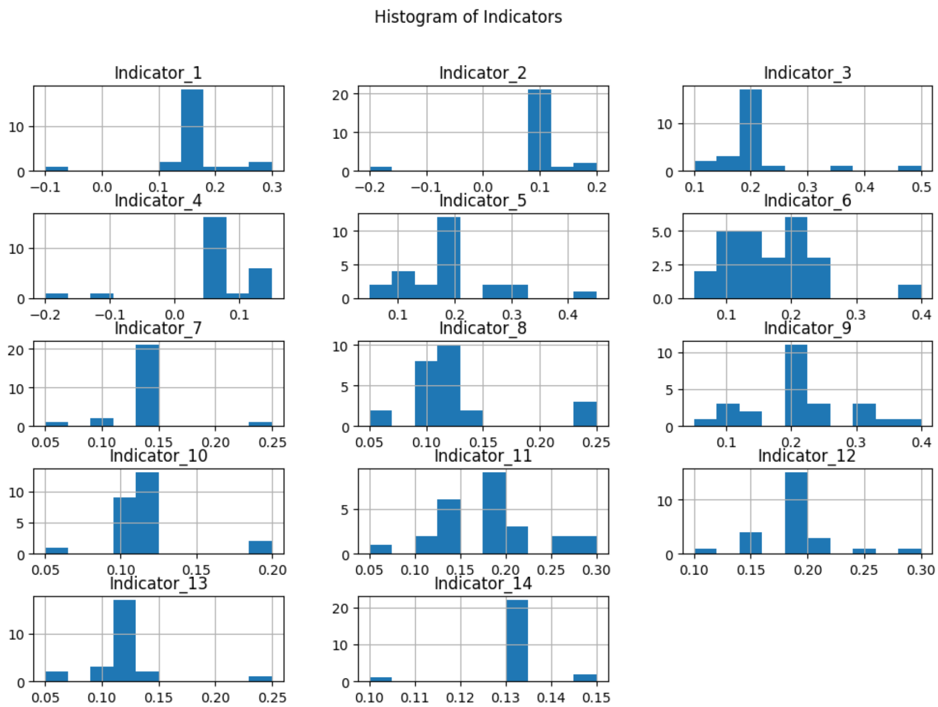

Figure 3: Histogram of indicators.

The boxplot also shows outliers. It means values that are significantly different from the majority of the data. It can be seen from the chart, that Indicators 6 and 9 have larger outliers compared to the other indicators. Indicators with taller squares and longer whiskers (e.g. indicators 5, 6, and 9) have higher variability. Indicators with shorter rectangles (e.g. indicator 10) have lower variability. It means the data is more consistent. If the median is closer to the bottom of the rectangle, the data is positively skewed. If the median is closer to the top, the data is negatively skewed. Thus, for some indicators, there is a skew. It means that the data is asymmetrically distributed.

The correlation matrix is presented in Figure 4. It displays the results of the Pearson correlation. According to this approach, linear relationships between indicators are measured in the range from −1 to 1, where 1 reflects a perfect positive correlation, 0 indicates no correlation, and −1 indicates a perfect negative correlation.

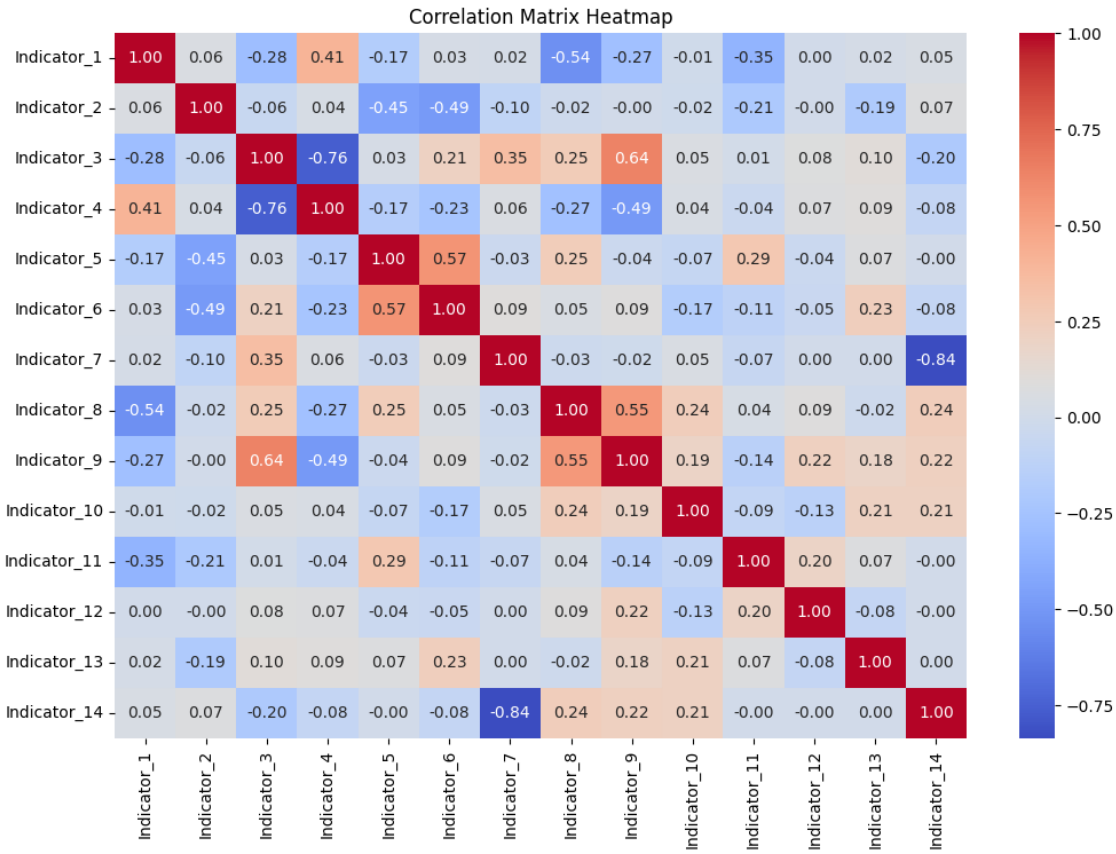

Figure 4: Pearson correlation matrix heatmap.

The Spearman correlation matrix is presented in Figure 5. It measures monotonic relationships, which, however, do not necessarily have to be linear.

The results of the study showed that some indicators have strong positive and negative relationships. Thus, indicators 7 and 14 demonstrate a strong negative correlation (−0.84), Indicator 3 and 4 demonstrate a strong positive correlation (0.76). Most correlations are weak or moderate.

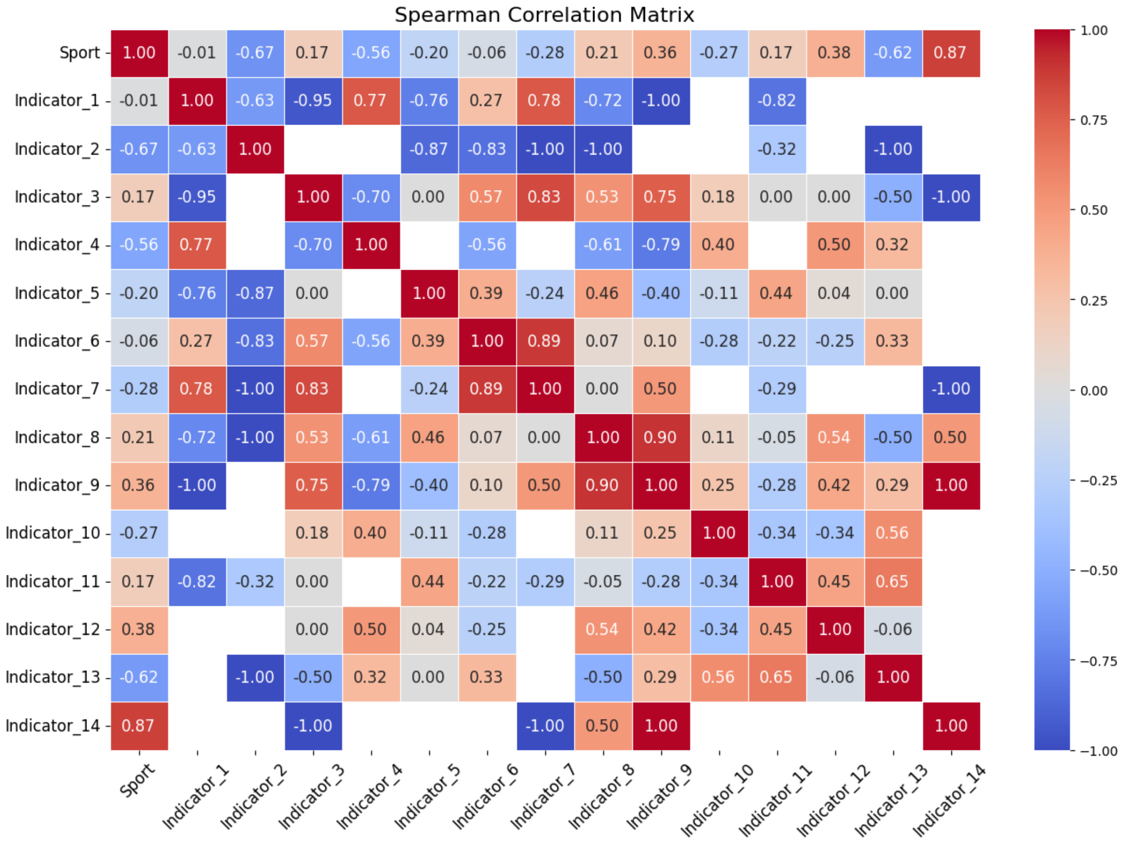

Figure 5: Spearman correlation matrix heatmap.

Spearman correlation provides a slightly different view of the relationships. Stronger correlations are observed between some indicators. The «Sport» column suggests which indicators contribute more to the classification. Indicator 14 has a high Spearman correlation with sport (0.87), which means that it can strongly influence the classification.

Pearson correlation gives more accurate results under linear relationships. If the data has strong outliers or nonlinear trends, Spearman correlation is more reliable. Spearman correlation also helps to rank indicators that influence the choice of sports.

It was provided Random Forest Classification to identify the most important indicators for determining the most suitable kind of sport on the basis of morphofunctional indicators. Results are presented in Figure 6. Thus indicators 6, 5, 9 and 11 have more significant impact then other.

5. Conclusions

As a result of the conducted research, it was possible to establish dependencies between sports, based on morphofunctional indicators, which play a key role for them. Correlations between morphofunctional indicators were also established, which contributes to a comprehensive understanding of the criteria on the basis of which the choice of sport is made.

The conducted data analysis allowed us to confirm the hypothesis that there is a difference between the parameters for different sports.

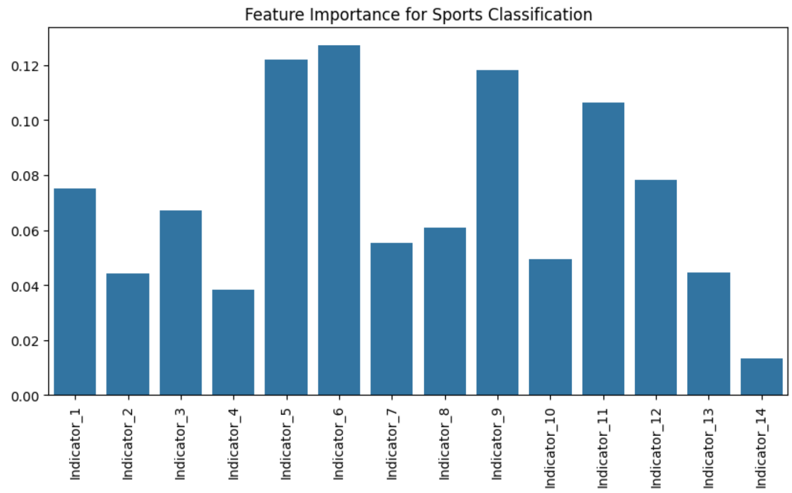

Figure 6: Feature importance for determining kind of sport on the basis of morphofunctional indicators due to Random Forest Classifier.

The analysis of the importance of indicators for making a decision about the predisposition to a sport, performed using AI technologies, allows us to capture hidden patterns and more accurately select sports.

The further efforts of the authors will be focused on approaches that provides intellectual analysis of human morphofunctional indicators and choosing a kind of sport based on them using AI technologies.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: grammar and spelling check; DeepL Translate in order to: some phrases translation into English. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Choosing the right sport for you (for teens). (January 2021). https://kidshealth.org/en/teens/find-sports.html

[2] Venture, J. (2022, May 10). What kind of sport are you into? 5 great tips on how to choose the right one. Troy Media. Re-inventing How Canadians Stay in Touch. https://troymedia.com/joint-venture/what-kind-of-sport-are-you-into-5-great-tips-on-how-to-choose-the-right-one/

[3] O. Pavlova, O. Soltyk, V. Shvaiko, J. Ilchyshyna, H. El Bouhissi. Human Morphofunctional Indicators-Based Decision Support System for Choosing Kind of Sport. MoMLeT+DS 2023. 322-333.

[4] T. Hovorushchenko, O. Pavlova, V. Shvaiko, A. Kuzmin, V. Kulbachnyi. Method and system for choosing kinds of sport based on human morphofunctional indicators. ITTAP 2023. 661-671.

[5] Akan, S., & Varlı, S. (2022). Use of deep learning in soccer videos analysis: survey. Multimedia Systems, 29(3), 897–915. https://doi.org/10.1007/s00530-022-01027-0.

[6] psico-smart.com. (28 August, 2024). Integration of artificial intelligence in health and wellness tracking systems. https://psico-smart.com/en/blogs/blog-integration-of-artificial-intelligence-in-health-and-wellness-tracking-systems-163046.

[7] Lu, L., Xi, Y., & Zhang, X. (2023). Application of mobile sensors based on deep neural networks in sports psychological detection and prevention. Preventive Medicine, 173, 107613. https://doi.org/10.1016/j.ypmed.2023.107613.

[8] Harper, G. (2025, February 7). AI in Sports: A Complete Guide. TopDevelopers.co. https://www.topdevelopers.co/blog/ai-in-sports/.

[9] Seçkin, A. Ç., Ateş, B., & Seçkin, M. (2023). Review on Wearable technology in Sports: Concepts, challenges and opportunities. Applied Sciences, 13(18), 10399. https://doi.org/10.3390/app131810399.

[10] Peppas, K., Tsolakis, A. C., Krinidis, S., & Tzovaras, D. (2020). Real-Time physical activity recognition on smart mobile devices using convolutional neural networks. Applied Sciences, 10(23), 8482. https://doi.org/10.3390/app10238482.

[11] Pashaie, S., Mohammadi, S., & Golmohammadi, H. (2024). Unlocking athlete potential: The evolution of coaching strategies through artificial intelligence. Proceedings of the Institution of Mechanical Engineers Part P Journal of Sports Engineering and Technology. https://doi.org/10.1177/17543371241300889.

[12] Lu, L., Yang, S., & Li, Q. (2024). The interaction of digital economy, artificial intelligence and sports industry development --based on China PVAR analysis of provincial panel data. Heliyon, 10(4), e25688. https://doi.org/10.1016/j.heliyon.2024.e25688.

[13] A. Orlando, "AI for Sport in the EU Legal Framework," 2022 IEEE International Workshop on Sport, Technology and Research (STAR), Trento – Cavalese, Italy, 2022, pp. 100-105, doi: 10.1109/STAR53492.2022.9860029.

[14] C. Nieto-Jiménez, E. Baez-San Martin, D. Rojas-Valverde, R. Yáñez-Sepúlveda, & M. Tuesta (2025). Morpho-functional characterization of an elite Chilean mountain runner: Insights from a high-performance case study. Journal of Human Sport and Exercise , 20(2), 409-418. https://doi.org/10.55860/7xs51s84

[15] G. Di Gioia, D. Polito, S. P. Crispino, V. Maestrini, A. Nenna, A. Segreti, M. R. Squeo, E. Lemme, & A. Pelliccia (2023). Influence of isometric versus isotonic exercise training on right ventricular morpho-functional parameters in Olympic athletes. Echocardiography, 41(1). https://doi.org/10.1111/echo.15731

[16] M. Sirola, & J. E. Hulsund (2021). Machine-Learning Methods in Prognosis of Ageing Phenomena in Nuclear Power Plant Components. International Journal of Computing, 20(1), 11-21. https://doi.org/10.47839/ijc.20.1.2086

[17] Roman Lynnyk, Victoria Vysotska, Yurii Matseliukh, Yevhen Burov, Lyubomyr Demkiv, Andrij Zaverbnyj, Anatoliy Sachenko, Inna Shylinska, Iryna Yevseyeva, Oksana Bihun. DDOS Attacks Analysis Based on Machine Learning in Challenges of Global Changes. CEUR Workshop Proceedings (CEUR-WS.org) MoMLeT+DS 2020 Modern Machine Learning Technologies and Data Science Workshop 2020, pp. 159-171. ISSN 1613-0073.

[18] H. Lipyanina, V. Maksymovych, A. Sachenko, T. Lendyuk, A. Fomenko, I. Kit, (2020). Assessing the Investment Risk of Virtual IT Company Based on Machine Learning. In: Babichev, S., Peleshko, D., Vynokurova, O. (eds) Data Stream Mining & Processing. DSMP

2020. Communications in Computer and Information Science, vol 1158. Springer, Cham. https://doi.org/10.1007/978-3-030-61656-4_11
[19] S. Shafi Bhat, V. Selvam, & G. A. Ansari (2023). Predicting Life Style of Early Diabetes Mellitus using Machine Learning Technique. International Journal of Computing, 22(3), 345-351. https://doi.org/10.47839/ijc.22.3.3230
[20] Hayes, A. (2024, October 4). T-Test: What it is with multiple formulas and when to use them. Investopedia. https://www.investopedia.com/terms/t/t-test.asp.