ORIGINAL ARTICLE

Modeling the gameplay actions of elite volleyball players and teams based on statistical match reports

Sergii Iermakov[1ABCDE], Tetiana Yermakova[1ABCDE], Krzysztof Prusik[2CDE]

[1] *Kharkiv State Academy of Design and Arts, Ukraine*
[2] *Gdansk University of Physical Education and Sport, Poland*

Authors' Contribution: A – Study design; B – Data collection; C – Statistical analysis; D – Manuscript Preparation; E – Funds Collection

Abstract

Background and Study Aim In modern sports analysis statistical modeling of gameplay actions based on match data is becoming a key tool for optimizing training processes and tactical preparation. The aim of the research is to create models of volleyball players' actions based on statistical reports of the 2022 World Championship matches.

Material and Methods The study used statistical data on the World Volleyball Championship matches among men. The data was extracted from open internet sources and converted into tables in CSV format. These tables were processed in the PyCharm programming environment using Python code. The pandas library was used for data analysis and statistical operations, and 'scikit-learn' for machine learning.

Results Models are presented that best predict the results for teams and volleyball players. Important features for teams have been identified, indicating the successful execution of game elements for the team. The regression equations for the team represent a linear combination of various gameplay metrics that affect the total number of points the team scores in a match. They also emphasize the importance of action elements. Linear regression equations predict the total number of points a volleyball player scores based on various statistical indicators.

Conclusions It is recommended to use statistical modeling to optimize training and tactical strategies based on key gameplay metrics. Linear regression equations can assist in evaluating the effectiveness of a player and team. Regular data updates will ensure the relevance of models for better match preparation. Consideration should be given to the possibilities of implementing analytical tools based on the developed models into training programs to optimize the team's preparation for future matches.

Keywords: volleyball, model, modeling, statistical report, regression

Introduction

In the modern sports world, statistical analysis and modeling of gameplay actions are becoming increasingly valuable tools for coaches, analysts, and athletes. In volleyball, as in many other team sports, these tools allow for a deeper understanding of game dynamics, optimizing strategies, and improving player performance. Based on statistical match reports, researchers can identify key factors that influence a team's success and also predict future outcomes.

In the context of the aforementioned observations, many scholars and researchers in the field of sports actively employ statistical methods to analyze various aspects of the game. Several studies [1, 2] focus on the impact of factors such as player's position, seasonal variations, and altitude on soccer players' performance. Others [3, 4] apply statistical and neural models to predict success in soccer. Research [5, 6] delves into the influence of ball possession on physical activity in different sports like hockey and futsal. Some studies [7, 8] provide an in-depth analysis of statistics in beach volleyball and badminton. Meanwhile, a few studies [9, 10, 11] explore the key success factors in rugby and basketball. These studies collectively underscore the profound and diverse influence of statistics on athletic achievements across various disciplines.

Just as statistics play a pivotal role in analyzing individual performance and game dynamics, it also becomes an integral tool in analyzing and predicting the performance of teams as a whole. This is especially relevant in team sports, where every player's action can influence the final match outcome.

The use of statistical reports in sports is becoming an increasingly valuable tool for analyzing and predicting team performance. Several studies [12, 13] delve into statistical data from LaLiga, pinpointing key determinants of a team's championship success. Complementary research [14, 15] explores the influence of athletic performance and other statistical metrics on Serie A team outcomes. A focus on the reliability and practicality of statistical systems, including the Champdas Master Match Analysis System, and the prediction of football

doi:10.15561/health.2023.0202

match results using data is presented in works [16, 17]. Further broadening the perspective, studies [18, 19] assess player movement and statistical markers in leagues like the NRLW Premiership and UEFA Champions League. Another study [20], although discussing statistics, primarily centers on DNA Motif, making it somewhat peripheral to the main theme. Collectively, these studies emphasize the profound role of statistical analysis in contemporary sports.

While statistical analysis in soccer and other popular team sports continues to evolve, volleyball is not left behind. This sport, although having its unique characteristics, also actively employs statistical methods for optimizing training and tactical strategies.

In modern volleyball, statistical analysis is pivotal for grasping the game's dynamics and strategies. Several studies [21] focus on the techniques and effectiveness of blocks in women's volleyball across various skill levels. Research [22] sets performance benchmarks for elite men's beach volleyball. Studies [23, 24] delve into the sequence and diversity of game combinations in women's volleyball, showcasing distinct performance patterns. A series of works [25, 26] probe the interplay between ball reception and attack efficiency in men's volleyball, including the impact of rotation and reception zones on offensive actions. Deep insights into how ball reception and the receiver's position shape attack strategies, offering an ecological perspective on in-game interactions, are presented in studies [27, 28]. Concluding the discussion, a comprehensive review [29] underscores the significance of match analysis in ball team sports, highlighting the indispensable role of statistical scrutiny in decoding game dynamics.

After a deep analysis of volleyball, it's worth noting that similar methods of statistical analysis and modeling are applied in other sports as well. In various disciplines, sports analysts and researchers aim to optimize the performance of teams and players using advanced methods and technologies.

Modeling in sports has emerged as a pivotal instrument for performance analysis and optimization. A cluster of studies [30, 31, 32, 33] harness mathematical and Bayesian frameworks to discern optimal team structures and dissect factors influencing performance in team-centric sports. Technological methodologies to evaluate athletes' agility and forecast match results are the focus of research [34, 35]. Psychological dimensions, encompassing the interplay between emotions and performance as well as spectator motivation and allegiance, are explored in works [36, 37]. Emphasizing the nuances of context and the primacy of data preprocessing for match analysis across diverse sports are studies [38, 39]. Collectively, these studies underscore the intricate and multifarious nature of modeling in contemporary sports.

Modeling in volleyball, grounded in match data, is on the rise as a means to refine training and tactical approaches. A set of studies [40, 41] employ stochastic and Bayesian frameworks, including Markov chains and set difference prediction models, to dissect game dynamics. The integration of cutting-edge technologies, such as artificial intelligence and wireless communication networks, in analyzing and fine-tuning volleyball techniques and strategies is the focus of research [42, 43]. Endriani et al. [44] put forth a lower pass model anchored in the Umbrella Learning paradigm. Meanwhile, a series of works [45, 46, 47] leverage random matrix models to gauge the efficacy of collegiate volleyball training and offer a quantitative portrayal of offensive player maneuvers. Rounding off the discussion, a study [48] unveils an algorithm tailored for green planning in the context of the volleyball premier league. Collectively, these studies spotlight the diverse and avant-garde methodologies in volleyball modeling today.

Transitioning from the general context of sports modeling to more specialized disciplines, volleyball deserves special attention. In this sport, analysts and researchers actively employ advanced methods and technologies for in-depth game analysis and optimization of training processes. Regression models in volleyball play a pivotal role in understanding various aspects of player performance and game dynamics. One study [49] examines the influence of biological maturation on anthropometric and physical indicators in male adolescents playing volleyball. Fuchs et al. [50, 51] analyzes movement characteristics during spike approaches in women and the relationship between general jump types and the performance of spike jumps in elite volleyball players of both genders, emphasizing the importance of physical preparation and movement mechanics. Another study [52] explores the relationship between sleep, subjective well-being, and injury risk in college female volleyball players, while research [25] investigates the correlation between ball reception and performance in high-level men's volleyball. Yet another study [53] underscores the influence of visual behavior on decision-making during volleyball blocking. Overall, these studies highlight the multifaceted and complex nature of regression analysis in volleyball, encompassing physiological, mechanical, and psychological aspects of the game.

Modern sport is at the crossroads of innovation and technological progress. Statistical analysis and modeling have become an integral part of sports analysis, providing coaches, analysts, and athletes with valuable insights and strategic advantages. In various sports, whether it's football, volleyball, or any other team discipline, deep data analysis allows teams to identify their strengths and weaknesses, optimize training processes, and predict match outcomes. Such research not only helps teams achieve better results but also contributes to the development of the discipline itself, introducing new methods and approaches. In light of this relevance and the need for deep analysis, the aim of our research is to create models of volleyball players' actions based on statistical reports from the 2022 World Championship matches.

Materials and Methods

Data Sources

Data on the matches of the Men's Volleyball World Championship in 2022 were extracted from open sources [54]. Group C of the preliminary games was selected, involving 4 teams: Poland, USA, Mexico, and Bulgaria. In total, data on 56 volleyball players from 6 matches were extracted.

Procedure

The results of statistical reports on matches were converted into tables in CSV format. The data tables were scaled to a common scale, making them comparable. For this, the normalization method [38] was used. As an example, Table 1 shows the normalized indicators of the match between Mexico and Bulgaria. A special HTML code with an embedded script was also used to convert report indicators into normalized data (normalized data can also be obtained in Excel or other programs). In the PyCharm programming environment, the data was processed using Python code.

Methodology for Developing a Regression Model to Forecast Team Indicators.

The primary objective was to create a model capable of predicting the total points (Total_points) of a team based on its statistical indicators in various aspects of the game.

• Step 1: Data Loading and Preprocessing

A new data table was created, containing the statistical indicators of each team based on the results of each match (Table 2). The data was saved in the CSV file.

• Step 2: Building the regression model

For data processing in the PyCharm programming environment, Python code was used. The following libraries were applied in the code: pandas, sklearn.ensemble, RandomForestRegressor, and sklearn.model_selection. The «Random Forest» model [55] was used to analyze and predict the effectiveness of team actions based on their statistical indicators.

Analysis objective and main stages. The goal of the analysis is to determine the importance of various features, such as «Serve aces», «Attack points», «Block points», and others, in the context of their influence on the team's total points.

Main stages of the analysis:

1. Feature and Target Variable Definition: Various statistical indicators of the team, such as «Serve aces», «Attack points», «Block points», and others, were chosen as features. The target variable is the team's total points (Total points).
2. Model Creation: The 'Random Forest Regressor' model from the 'scikit-learn' library is applied. This method allows assessing the importance of each feature in the context of its influence on the target variable.
3. Model Training: The model is trained on all available data.
4. Model Evaluation: The mean squared error (MSE) of the model is evaluated using 3-fold cross-validation.
5. Feature Importance Extraction: After training the model, the importance of each feature is extracted. This allows determining which features most strongly influence the team's total points.

Model Evaluation and Results. The results of the models for different teams are presented in Table 3. Based on the comparison of results, conclusions were drawn about which features most significantly influence the team's performance.

• Step 3. Predicting Match Success Based on Metrics

Option 1.

Within this research stage, in the PyCharm programming environment, Python code was developed to analyze statistical data of volleyball matches. The primary goal was to predict the overall success of the team based on key game metrics.

Main Analysis Stages:

1. Data Loading: Initial data on volleyball matches are loaded from a CSV file.
2. Data Preparation: Features and the target variable are defined. Data is split into training and test sets.
3. Modeling: A linear regression model is applied for training based on the training data set.
4. Performance Evaluation: The mean squared error (MSE) is calculated to assess the accuracy of the model's predictions on the test set.
5. Results Interpretation: Based on the obtained data, a regression equation is formed, reflecting the relationship between the features and the target variable.

As a result of executing the code, a CSV file with analysis results is created, including MSE and the regression equation (Table 4). This approach provides the opportunity for a deep analysis of the influence of various game metrics on the overall success of a volleyball team.

Option 2.

Analysis and prediction of volleyball game success based on key game metrics. Within this research, code was developed for a detailed analysis and evaluation of the importance of various volleyball game metrics using the linear regression method. The initial data, containing statistical information about matches of various teams, was extracted from CSV files.

Table 1. Normalized indicators of the match between Mexico and Bulgaria

ID	Label	Role	Country	Total points	Serve total	Serve aces	Receiving total	Attack total	Attack points	Attack %total	Block points
29	MEX1lib	lib	Mexico	0.03725	0.00000	0.01500	0.02105	0.00000	0.00060	0.00000	0.00060
30	MEX3lib	lib	Mexico	0.12146	0.00000	0.01500	0.10526	0.00000	0.00060	0.00000	0.00060
31	MEX4doi	doi	Mexico	0.01620	0.00000	0.01500	0.00000	0.00000	0.00060	0.00000	0.00060
32	MEX6doi	doi	Mexico	0.57726	0.10000	0.04800	0.14211	0.14211	0.14400	0.00044	0.00060
33	MEX7dia	dia	Mexico	0.69371	0.12105	0.04800	0.00000	0.20000	0.30000	0.00066	0.02400
34	MEX8sva	sva	Mexico	0.02146	0.00526	0.01500	0.00000	0.00000	0.00060	0.00000	0.00060
35	MEX9cen	cen	Mexico	0.29406	0.08947	0.01200	0.01579	0.06842	0.06000	0.00038	0.04800
36	MEX12doi	doi	Mexico	0.38120	0.07368	0.01200	0.11579	0.09474	0.08400	0.00039	0.00060
37	MEX14cen	cen	Mexico	0.18894	0.10000	0.01500	0.00526	0.03158	0.03600	0.00050	0.00060
38	MEX15sva	sva	Mexico	0.10705	0.06316	0.01500	0.00526	0.01053	0.01200	0.00050	0.00060
39	MEX16cen	cen	Mexico	0.01620	0.00000	0.01500	0.00000	0.00000	0.00060	0.00000	0.00060
40	MEX18doi	doi	Mexico	0.10852	0.01579	0.01500	0.03158	0.02105	0.02400	0.00050	0.00060
41	MEX20dia	dia	Mexico	0.01620	0.00000	0.01500	0.00000	0.00000	0.00060	0.00000	0.00060
42	MEX21cen	cen	Mexico	0.02146	0.00000	0.01500	0.00000	0.00526	0.00060	0.00000	0.00060
		Mean		0.17373	0.02971	0.01745	0.02457	0.04514	0.05418	0.00015	0.00252
		Std Deviation		0.20501	0.03017	0.00751	0.04764	0.07084	0.10788	0.00024	0.00544
43	BUL1doi	doi	Bulgaria	0.01650	0.00000	0.01500	0.00000	0.00000	0.00075	0.00000	0.00075
44	BUL2cen	cen	Bulgaria	0.01650	0.00000	0.01500	0.00000	0.00000	0.00075	0.00000	0.00075
45	BUL3cen	cen	Bulgaria	0.09528	0.02857	0.01500	0.00000	0.00571	0.01500	0.00100	0.03000
46	BUL4doi	doi	Bulgaria	0.02221	0.00000	0.01500	0.00571	0.00000	0.00075	0.00000	0.00075
47	BUL5cen	cen	Bulgaria	0.31496	0.13143	0.01500	0.01143	0.05143	0.09000	0.00067	0.01500
48	BUL7sva	sva	Bulgaria	0.03364	0.01714	0.01500	0.00000	0.00000	0.00075	0.00000	0.00075
49	BUL8doi	doi	Bulgaria	0.01650	0.00000	0.01500	0.00000	0.00000	0.00075	0.00000	0.00075
50	BUL9sva	sva	Bulgaria	0.14214	0.08000	0.01500	0.00571	0.01143	0.00000	0.00000	0.03000
51	BUL11cen	cen	Bulgaria	0.21972	0.04571	0.01500	0.01714	0.05143	0.06000	0.00044	0.03000
52	BUL14doi	doi	Bulgaria	0.59986	0.09143	0.01500	0.16000	0.13714	0.19500	0.00054	0.00075
53	BUL16lib	lib	Bulgaria	0.15364	0.00000	0.01500	0.13714	0.00000	0.00075	0.00000	0.00075
54	BUL18doi	doi	Bulgaria	0.08507	0.01143	0.01500	0.02857	0.02857	0.00075	0.00000	0.00075
55	BUL19dia	dia	Bulgaria	0.68843	0.12000	0.03000	0.02286	0.20000	0.30000	0.00057	0.01500
56	BUL23doi	doi	Bulgaria	0.74411	0.09143	0.03000	0.14286	0.19429	0.27000	0.00053	0.01500
		Mean		0.2249	0.04408	0.01714	0.03796	0.04857	0.0668	0.00027	0.01007
		Std Deviation		0.26156	0.04877	0.00545	0.05979	0.07327	0.10755	0.00034	0.01228

Note: The data in the table were obtained and normalized using specialized HTML code with an embedded script. ID - the serial number of the volleyball player in the general list of 56 names; Label - a conditional designation of the volleyball player: the first 3 letters - the short name of the country: MEX - Mexico, BUL - Bulgaria; numbers - the number of the volleyball player; the last 3 letters - the position of the volleyball player by court zones: Zone 1: Setter (sva); Zone 2: Right-side hitter - HitterR (doi); Zone 3: Middle blocker (cen); Zone 4: Opposite (dia); Zone 5: Outside hitter - HitterO (doi); Zone 6: Central blocker - Middle blocker/Libero (cen or lib). Note: Conditional abbreviations are provided in parentheses. Total points: Total number of points. Serve total: Total number of serves. Serve aces: Number of ace serves. Receiving total: Total number of receptions. Attack total: Total number of attacks. Attack points: Number of points scored during an attack. Attack %total: Percentage of the total number of attacks. Block points: Number of points scored during blocking.

Main Analysis Stages:

1. Data Loading: Data for the selected teams is loaded from CSV files.
2. Feature Definition: Various game metrics, such as the total number of serves, successful attacks, blocks, etc., are chosen as features. The target variable is the total number of team points.
3. Modeling: A linear regression model is applied for training on the data of each team.
4. Results Interpretation: Feature weights are extracted from the model, allowing the determination of the importance of each for predicting the total number of points.

The analysis results are saved in a CSV file (Table 5), where rows represent different features, and columns represent feature weights for each of the

Table 2. Normalized statistical indicators of team actions

Game N	Teams	Total points	Serve total	Serve aces	Receiving total	Attack total	Attack points	Attack %total	Block points	Result sets
1	Poland	0.64447	0.14186	0.02459	0.06977	0.13721	0.24590	0.00055	0.02459	3
1	USA	0.61748	0.11860	0.03443	0.09767	0.13023	0.20656	0.00048	0.02951	1
2	Poland	0.39945	0.09302	0.01475	0.01163	0.05814	0.16721	0.00060	0.05410	3
2	Mexico	0.32180	0.04419	0.00984	0.05116	0.08837	0.12295	0.00037	0.00492	0
3	Poland	0.47627	0.09535	0.05902	0.02093	0.07907	0.18689	0.00058	0.03443	3
3	Bulgaria	0.37256	0.04651	0.00984	0.07442	0.07907	0.13770	0.00043	0.02459	0
4	USA	0.35454	0.07674	0.01475	0.00000	0.05581	0.18689	0.00068	0.01967	3
4	Mexico	0.30147	0.02791	0.00000	0.05581	0.06977	0.13279	0.00044	0.01475	0
5	USA	0.50480	0.09767	0.02951	0.04419	0.11163	0.18689	0.00048	0.03443	3
5	Bulgaria	0.50084	0.08140	0.00984	0.07442	0.12326	0.18197	0.00044	0.02951	0
6	Mexico	0.80468	0.17442	0.04426	0.11860	0.17674	0.26557	0.00050	0.02459	3
6	Bulgaria	0.87840	0.17442	0.02459	0.13953	0.20000	0.30000	0.00052	0.03934	2

Note: As in Table 1.

Table 3. Model indicators: mean squared error (MSE) and feature importance

Team	Mean MSE	Importance						
		Serve aces	Attack points	Block points	Serve total	Receiving total	Attack total	Attack %total
Poland	0.01855	0.10809	0.11765	0.19558	0.06214	0.08235	0.15294	0.28125
USA	0.02658	0.12941	0.05882	0.08998	0.07059	0.08235	0.12941	0.43943
Mexico	0.09318	0.16469	0.16471	0.15291	0.09416	0.09412	0.16471	0.16471
Bulgaria	0.07127	0.13908	0.21176	0.15294	0.09762	0.06919	0.16471	0.16471

Note: As in Table 1.

Table 4. Data for predicting team success (using Poland as an example) based on key game metrics (MSE, regression equation)

Parameter	Value
MSE	0.02174
Coefficients	[1.09968, 0.13459, -0.61978, 0.06223, 0.22418, 0.52879, 0.00099]
Intercept	0.35511
Equation	$y = 0.35511 + 1.09968 \times x_1 + 0.13459 \times x_2 - 0.61978 \times x_3 + 0.06223 \times x_4 + 0.22418 \times x_5 + 0.52879 \times x_6 + 0.00099 \times x_7$

Note. MSE - mean squared error; Intercept - (or y-axis intersection) - the constant term of the regression equation (this is the value that the predicted variable takes when all independent variables are zero); y - Total points; x_1 - Serve aces; x_2 - Attack points; x_3 - Block points; x_4 - Serve total; x_5 - Receiving total; x_6 - Attack total; x_7 - Attack %total.

Table 5. Normalized statistical indicators of team actions: Poland and USA; Mexico and Bulgaria

Sign/Math	POL-USA		MEX-BUL	
	POL	USA	MEX	BUL
Attack_%total	-0.0006912	-0.0006971	0.0012557	0.0010874
Attack_points	1.0030827	0.8938832	1.3507186	1.3336897
Attack_total	1.016222	1.5087634	0.8270077	0.7258307
Block_points	-0.4644407	0.1914848	0.1896939	0.0997666
Receiving_total	0.7034869	0.6885175	0.6387063	0.9596867
Serve_aces	0.4257201	0.1660849	0.3210757	0.2174071
Serve_total	0.5515817	1.4494683	1.2232252	1.0510816

considered teams. This provides the opportunity to compare the importance of various metrics between teams and draw conclusions about the key aspects of the game that influence success in a match.

This approach can be a valuable tool for coaches and data analysis specialists in volleyball, helping to better understand which game elements require special attention during the training process and preparation for upcoming matches.

Preliminary processing and organization of volleyball players' data.

For a detailed analysis of the statistical indicators of volleyball players, preliminary data processing and organization were carried out. The initial data was extracted from open sources containing statistical reports on 6 matches. Based on this data, a general table was created, including information about each of the 56 volleyball players who participated in 3 matches, totaling 168 rows of data.

As an illustration, Table 1 is an excerpt from this general table, containing data only for one match between the teams of Mexico and Bulgaria. After compiling the general table, the data was structured and transformed in such a way that it could be easily used for further analysis.

Within this study, four separate codes were developed in the Python programming language. These codes are intended for processing, structuring, and analyzing data about volleyball players and their statistical indicators in games. The 'pandas' library, which is one of the main tools for working with data in Python, was actively used to accomplish these tasks.

Within the research, five stages of data processing about volleyball players were developed:

- Stage 1: This stage performs the initial data loading from a CSV file, organizes them by the unique code of the volleyball player, and saves the result in a new CSV file.
- Stage 2: At this stage, a new DataFrame is created, in which data for each volleyball player is combined from different rows into one, based on the unique player code. Processed data is saved in a separate CSV file.
- Stage 3: Here, certain columns are split into several new columns, using a comma as a delimiter. After that, the original columns are deleted, and the data is saved in a new file.
- Stage 4: At this stage, data is loaded and processed in such a way that only the first value is selected from rows with multiple values. The result is saved in the CSV file.
- Stage 5: This stage analyzes the data, highlighting the top 3 players in each role based on their total points for three matches. Data is grouped by roles, summed up, and sorted. The analysis results are saved in the CSV file (Table 6).

Table 6 presents the best volleyball players of various positions. However, it's worth noting that for players in the 'lib' and 'sva' positions, many metrics are missing in the statistical reports. This is due to the specific roles they play in the game. Specifically, 'lib' players cannot participate in attacks, and 'sva' players do not participate in receiving (the 'Receiving total' element). As a result, most of their metrics in Table 6 are equal to 0.

These stages ensure effective preprocessing and structuring of data about volleyball players, which is an integral part of preparing for detailed analysis and modeling. This approach allows for the most efficient use of available data to obtain valuable information about the performance of volleyball players.

In the next stage, an analysis is conducted aimed at determining how various volleyball players' performance indicators (such as the number of successful serves, receptions, etc.) influence their total points in matches. The linear regression method is used for this purpose.

Main steps of the stage:

1. Data Loading: Original data about volleyball players and their statistical indicators are loaded from a CSV file.
2. Feature and Target Variable Definition: Various game performance indicators of volleyball players are chosen as features, and their total points in matches serve as the target variable.
3. Linear Regression Application: Based on the selected features and the target variable, a linear regression model is constructed for each player's position (role).
4. Regression Equation Creation: As a result of the analysis, regression equations are formed, showing how game indicators influence the overall performance of volleyball players in different positions.

The results of this analysis are saved in the CSV file (Table 7), providing volleyball coaches and analysts with valuable information about which aspects of the game most strongly influence the overall performance of volleyball players in different positions.

Table 7 showcases data fragments for volleyball players of different roles, with data presented only for one player of each role. For other volleyball players of these positions, the data is similar. This table reflects the results of linear regression for randomly selected players.

Description of the table columns:

- Athlete Code: A unique code of the player, which also indicates his role in the team (for example, «doi» in the code «POL14doi»).
- Regression Equation: An equation predicting the total points of a player based on statistical indicators (X1, X2, etc.). The coefficients of the equation indicate the influence of each indicator

Table 6. Top 3 volleyball players in each position (role) based on the total points over three matches.

Match N	Athletes List N	Athlete Code	Role	Country	Total points_1	Total points_2	Total points_3	Serve total_1	Serve total_2	Serve total_3	Serve aces_1	Serve aces_2	Serve aces_3	Receiving total_1	Receiving total_2	Receiving total_3	Attack total_1	Attack total_2	Attack total_3	Attack points_1	Attack points_2	Attack points_3	Attack %total_1	Attack %total_2	Attack %total_3	Block points_1	Block points_2	Block points_3	Sets Won_1	Sets Won_2	Sets Won_3	Sets Lost_1	Sets Lost_2	Sets Lost_3	Total Points Sum
3, 1, 2	12	POL20cen	cen	Poland	10	11	6	12	17	5	2	1	1	1	0	0	8	14	5	5	8	3	63	57	60	3	2	2	3	3	3	0	1	0	27
6, 3, 5	51	BUL11cen	cen	Bulgaria	6	5	8	8	4	7	0	0	0	3	0	0	9	5	8	4	3	4	44	60	50	2	2	4	2	0	0	3	3	3	19
6, 5, 3	47	BUL5cen	cen	Bulgaria	8	5	5	23	12	13	1	0	1	2	0	0	9	9	5	6	5	2	67	56	40	1	0	2	2	0	0	3	3	3	18
6, 4, 2	33	MEX7dia	dia	Mexico	31	8	14	23	9	9	4	0	0	0	0	0	38	21	25	25	8	14	66	38	56	2	0	0	3	0	0	2	3	3	53
5, 6, 3	55	BUL19dia	dia	Bulgaria	15	23	9	15	21	10	1	2	1	4	4	4	29	35	25	12	20	8	41	57	32	2	1	0	0	2	0	3	3	3	47
3, 1, 2	3	POL6dia	dia	Poland	13	20	3	12	23	3	4	2	0	1	2	0	16	30	4	7	17	3	44	57	75	2	1	0	3	3	3	0	1	0	36
1, 5, 4	19	USA8doi	doi	USA	16	13	11	14	13	8	1	2	0	26	11	9	29	22	15	14	10	9	48	45	60	1	1	2	1	3	3	3	0	0	40
4, 1, 5	16	USA2doi	doi	USA	15	14	11	13	12	10	1	1	1	14	32	17	18	21	22	12	12	8	67	57	36	2	1	2	3	1	3	0	3	0	40
5, 3, 6	56	BUL23doi	doi	Bulgaria	7	8	21	11	5	16	1	0	2	27	33	25	12	16	34	6	8	18	50	50	53	0	0	1	0	0	2	3	3	3	36
5, 6, 3	53	BUL16lib	lib	Bulgaria	0	0	0	0	0	0	0	0	0	17	24	12	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	3	3	3	0
6, 2, 4	29	MEX1lib	lib	Mexico	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	2	3	3	0
6, 2, 4	30	MEX3lib	lib	Mexico	0	0	0	0	0	0	0	0	0	20	11	21	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	2	3	3	0
5, 1, 4	20	USA11sva	sva	USA	6	0	3	6	0	13	1	0	1	0	0	0	1	0	1	1	0	1	100	0	100	4	0	1	3	1	3	0	3	0	9
5, 3, 6	50	BUL9sva	sva	Bulgaria	3	0	3	11	8	14	1	0	1	0	0	1	3	0	2	2	0	0	67	0	0	0	0	2	0	0	2	3	3	3	6
4, 2, 6	38	MEX15sva	sva	Mexico	2	3	1	6	10	12	0	2	0	0	0	1	3	0	2	2	0	1	67	0	50	0	1	0	0	0	3	3	3	2	6

Note: As in Table 1; Match N: match number; Athletes List N: athlete's list number; Athlete Code: athlete's code; Role: athlete's position; Country: country; numbers at the end of column headers indicate the match number in which the athlete participated (a total of 3 matches); Sets Won: sets won; Sets Lost: sets lost; Total Points Sum: total points sum.

Table 7. Regression equations for volleyball players of specific roles (positions)

Athlete Code	Regression Equation	Predicted Total Points
POL14doi	$y = -0.03345 + -6.68246 \times x_1 + 10.45318 \times x_2 + 1.36997 \times x_3 + 2.95688 \times x_4 + 2.61856 \times x_5 + -2.72975 \times x_6 + 8.76889 \times x_7 + 0.33013 \times x_8 + 5.33036 \times x_9$	1.32887
USA5dia	$y = 0.60845 + -13.95346 \times x_1 + -18.55422 \times x_2 + -3.96732 \times x_3 + 2.11181 \times x_4 + 4.04327 \times x_5 + -2.13618 \times x_6 + 7.49829 \times x_7 + 14.12933 \times x_8 + 2.30154 \times x_9$	0.25299
MEX9cen	$y = 0.09693 + -0.46850 \times x_1 + 1.05904 \times x_2 + -5.03671 \times x_3 + -7.10744 \times x_4 + -32.54464 \times x_5 + 40.98940 \times x_6 + 1.62731 \times x_7 + 2.88691 \times x_8 + 11.12398 \times x_9$	0.68562
BUL7sva	$y = -0.12667 + 10.81818 \times x_1 + 5.65259 \times x_2 + -4.58106 \times x_3 + 9.33697 \times x_4 + 0.00000 \times x_5 + -2.17450 \times x_6 + -0.02436 \times x_7 + 1.05836 \times x_8 + 2.35115 \times x_9$	0.09805
POL17lib	$y = 0.05250 + 0.00000^* \times x_1 + 0.00000 \times x_2 + 0.00000 \times x_3 + 1.00578 \times x_4 + 1.00183 \times x_5 + 1.00108 \times x_6 + 0.00000 \times x_7 + 0.00000 \times x_8 + 0.00000 \times x_9$	0.33333

NOTE: x_1 corresponds to 'Serve aces_1'; x_2 corresponds to 'Serve aces_2'; x_3 corresponds to 'Serve aces_3'; x_4 corresponds to 'Receiving total_1'; x_5 corresponds to 'Receiving total_2'; x_6 corresponds to 'Receiving total_3'; x_7 corresponds to 'Serve total_1'; x_8 corresponds to 'Serve total_2'; x_9 corresponds to 'Serve total_3'; Numbers in the form "_1", "_2", "_3" refer to the match, respectively 1, 2, 3.

on the total score.

- Predicted Total Points: The forecasted total points based on the regression equation.

At the end of the table, a note is provided, explaining that each of the indicators (x_1, x_2, etc.) corresponds to a specific game statistic.

Example of analysis: Consider the player with the code «POL14doi». His regression equation indicates the dependence of his total points on the number of aces on the serve, successful receptions, and the total number of serves in different matches. The coefficients of the equation emphasize that the most significant influence on his results is the number of successful receptions in the first and second matches.

For processing and analyzing data obtained from tables, the PyCharm programming environment and the Python programming language were used.

Main stages of the analysis:

1. Data loading and preliminary processing: Using the 'pandas' library, the data was loaded, cleaned, and prepared for further analysis.
2. Statistical analysis: Using the same library, basic statistical indicators, such as mean values and standard deviations, were calculated.
3. Levene's Test: To check the homogeneity of variances between groups, Levene's Test was applied. This was done using the 'stats 'module from the 'scipy' library.

The results of this analysis provide important information about the distribution of data, their central tendency and variability, as well as the homogeneity of variances between groups.

Results

Based on the available statistical reports of the 2022 World Championship matches provided on [championat.com], an analysis of the teams' actions was conducted. Normalized statistical indicators of the teams' actions in 6 matches are presented in Table 2.

Analyzing the indicators of the match between the teams of Bulgaria and Mexico, the following conclusions can be drawn:

- Total points: On average, Bulgaria scored more points than Mexico. However, Bulgaria also had a higher standard deviation, indicating a greater spread in the data.
- Serves: Bulgaria had better serve indicators, but the difference was not significant.
- Aces: Although Mexico had more aces on average, its data spread was also higher.
- Reception: Bulgaria, on average, received the ball better, but with a larger data spread.
- Attack: Bulgaria had better attack indicators, both in total and in points for the attack.
- % successful attack: The indicators of both teams were very close, but Bulgaria had a slight advantage.
- Blocking: Bulgaria also had better blocking indicators, but Mexico's data spread was higher.

Overall, Bulgaria performed better in most indicators, but it should be noted that they also had a greater data spread for some parameters. Similarly, the actions of other teams can be analyzed.

Based on the comparison of model results for different teams (Table 3), the following conclusions can be drawn:

1.Mean Squared Error (MSE):

- Team Poland has the best MSE value, indicating high accuracy of the model in predicting this team's results.
- Team USA follows, then Bulgaria.
- Team Mexico has the highest error value, indicating the lowest accuracy of the model in predicting this team's results.

2.Importance of features:

- Serve aces: This feature is most important for Team Mexico and least important for Team Poland.
- Attack points: This feature is most important for Team Bulgaria and least important for Team USA.
- Block points: This feature is most important for Team Poland and least important for Team USA.
- Serve total: This feature is most important for Team Poland and least important for both Team USA and Mexico.
- Receiving total: This feature has roughly equal importance for all teams, but Team Poland stands out slightly.
- Attack total: This feature is most important for both Team Poland and Mexico, and least important for Team USA.
- Attack %total: This feature is most important for Team USA, and its importance is roughly equal for other teams.

3.General Conclusions:

- The model predicts the results for Team Poland best and for Team Mexico worst.
- Different features have varying degrees of importance for different teams, which may reflect their unique strategies and playing styles.
- For instance, for Team USA, the most important feature is Attack %total, which might indicate the key role of attack efficiency in their strategy. Whereas for Team Bulgaria, the key feature is the number of successful attacks.
- This analysis can help coaches and analysts better understand the strategies and playing styles of different teams, as well as identify the key factors of their success.

The obtained results are consistent with the data from the statistical report of matches based on the '% of points' indicator [54]:
Poland - 100%, USA - 66.7%, Mexico - 22.2%, Bulgaria - 11.1%.

Based on the analysis conducted and comparison with real data from the statistical report of matches by the '% points' indicator from the [54] website, the following conclusions can be drawn:

1. *Poland:* Model showed the best results for the Poland team, which is consistent with real data where Poland has an indicator of 100%.
2. *USA*: The model also demonstrates good results for the USA team, which matches real data where the USA has an indicator of 66.7%.
3. *Mexico*: Despite the model showing the lowest accuracy for the Mexico team, real data confirms the team's low indicator of 22.2%.
4. *Bulgaria*: The model showed average results for the Bulgaria team, which also aligns with real data where Bulgaria has an indicator of 11.1%.

Thus, the results obtained using the model align well with real data from the statistical report. This confirms the correctness and effectiveness of the applied approach and analysis methodology.

The regression equation (Table 4) for the Poland team represents a linear combination of various game metrics that are presumed to influence the team's total points in a match.

From the equation, the following conclusions can be drawn:

Serve aces (x_1) have the highest positive coefficient, indicating that successful serves make a significant contribution to the team's total points.

Block points (x_3) have a negative coefficient, which might be counterintuitive. This could suggest that there are certain peculiarities in the data or that blocks are not always a key success factor for the Poland team.

Attack %total (x_7) has a very small coefficient, suggesting that the percentage of successful attacks doesn't greatly influence the overall result, at least in the context of the other metrics considered.

The other variables also contribute to the forecast, but their influence is less compared to 'Serve aces' in serving.

The obtained regression equations for each team (using the example of the Poland team – table 4) are as follows:

Team Poland:
$y = 0.35511 + 1.09968 \times x_1 + 0.13459 \times x_2 - 0.61978 \times x_3 + 0.06223 \times x_4 + 0.22418 \times x_5 + 0.52879 \times x_6 + 0.00099 \times x_7$

Team USA:
$y = 0.22231 + 0.37322 \times x_1 + 0.0 \times x_2 + 0.37322 \times x_3 + 0.52923 \times x_4 + 1.11738 \times x_5 + 1.41146 \times x_6 - 0.00506 \times x_7$

Team Mexico:
$y = 0.00923 + 0.39394 \times x_1 + 1.18182 \times x_2 + 0.08758 \times x_3 + 1.30403 \times x_4 + 0.55887 \times x_5 + 0.9521 \times x_6 + 0.00053 \times x_7$

Team Bulgaria:
$y = 0.00024 + 0.16862 \times x_1 + 1.34933 \times x_2 + 0.11238 \times x_3 + 1.06341 \times x_4 + 0.74434 \times x_5 + 0.8773 \times x_6 + 0.00091 \times x_7$

where: y - Total points; x_1 - Serve aces; x_2 - Attack points; x_3 - Block points; x_4 - Serve total; x_5 - Receiving total; x_6 - Attack total; x_7 - Attack %total

Analysis of the regression equations for various teams reveals the following:

Team Poland:

The regression equation for the Poland team emphasizes the importance of serve aces (x_1) with a coefficient of 1.09968, indicating a significant influence of successful serves on the total number of points. However, it's interesting to note that blocks (x_3) have a negative coefficient, which might suggest specific game strategies or statistical anomalies.

Team USA:

For the USA team, the most significant contributions to the total number of points come from attacks (x_6) with a coefficient of 1.41146, as well as receptions (x_5) with a coefficient of 1.11738.

This might suggest that the USA team focuses on offensive actions and quality serve reception.

Team Mexico:

For Team Mexico, the most significant contributions to the total points are made by attacking actions (x_2) and serves (x_4) with coefficients of 1.18182 and 1.30403 respectively. This may reflect an aggressive playing style of the team, emphasizing strong attacking hits and serves.

Team Bulgaria:

For Team Bulgaria, the most significant contribution to the total points is made by attacking actions (x_2) with a coefficient of 1.34933. This indicates that Team Bulgaria focuses on the attack, possibly having strong attacking players in the lineup.

Overall, the analysis of regression equations for different teams allows us to identify key points and features of each team's strategy. However, it should be remembered that regression coefficients depend on the available data and can change depending on specific matches and seasons.

The indicators of features and their impact on the results of an individual match can be evaluated based on the match data, for example, Poland-USA or Mexico and Bulgaria (Table 5).

Based on the feature weights for Poland and the USA, presented in Table 5, the following conclusions can be drawn:

• *Team Poland:*

– The features Attack_total and Attack_points have the most significant positive influence on the total number of points (Total_points). This emphasizes the importance of successful attacks and points scored during attacks for the Polish team.
– The features Receiving_total and Serve_total also influence Total_points, but their weights are smaller compared to Attack_total and Attack_points.
– The negative weight of the feature Attack_%total may indicate an inverse relationship between this indicator and Total_points.

At first glance, it may seem counterintuitive that an increase in the percentage of successful attacks might be associated with a decrease in the total number of points. This could be explained by various tactical and strategic decisions of the Polish team during the match.

Game Strategy: The team can choose a more conservative strategy, making fewer risky attacks, which leads to a higher percentage of successful attacks, but a smaller total number of attacks and, consequently, fewer points.

Quality of the Opponent: If the team plays against a strong opponent with good defense, it can have a higher percentage of successful attacks (since it chooses only the most confident attacks), but fewer total attacks and, therefore, fewer points.

Unsuccessful Attacks: Not all attacks are successful. Some of them can be repelled or blocked by the opponent, which does not add points to the team.

• *Team USA:*

– The most significant influence on Total_points is exerted by the features Attack_total and Serve_total. At the same time, Attack_total has the highest positive weight, confirming the importance of successful attacks and serves for this team.
– The feature Attack_points also has a positive weight, but it is smaller compared to Attack_total and Serve_total.
– The features Receiving_total and Block_points have a positive impact, but their weights are smaller compared to the previous ones.
– The indicator Serve_aces has the least influence on Total_points for the USA.
– Based on the feature weights for Mexico and Bulgaria (Table 5), the following conclusions can be drawn:

Team Mexico:

The features Attack_points and Serve_total have the most significant influence on the total number of points (Total_points), both having positive weights. This may indicate the importance of successful attacks and service for the Mexican team.

Features Attack_total and Receiving_total also positively influence Total_points, but their weights are less compared to Attack_points and Serve_total.

Serve_aces have the least influence on Total_points for Mexico.

Team Bulgaria:

The feature Attack_points has the most significant influence on Total_points, having a positive weight. This may indicate the importance of successful attacks for the Bulgarian team.

Features Serve_total and Receiving_total also positively influence, but their weights are less compared to Attack_points.

Attack_total also influences Total_points, but its weight is less compared to the previous ones.

Other features, such as Serve_aces, Block_points, and Attack_%total, have the least influence on Total_points for Bulgaria.

In general, for both teams, the most valuable indicators are related to the attack, emphasizing the importance of aggressive and effective gameplay in the attack to achieve high results. However, it should be noted that for Mexico, service is also essential, while for Bulgaria, this indicator is less critical.

Discussion

In modern sports, data analysis plays a pivotal role in forecasting and understanding game strategies. In volleyball, as in other team sports, certain metrics can serve as indicators of a team's success. The aim of our research is to create models of volleyball players' actions based on statistical reports from the 2022 World Championship matches. In this study, we employ machine learning methods to analyze and predict success in volleyball matches based on key game metrics. In this regard, the importance of statistical analysis in sports games is confirmed in studies by Akyildiz et al. [1], Guan et al. [4], Cunniffe et al. [5], Nunes et al. [7], and Yi et al. [11]. Moreover, statistical reports in sports are often used to analyze and forecast team performance [12, 14, 18, 20, 56, 57]. In relation to volleyball, statistical analysis is a crucial element in understanding the dynamics and strategy of the game.

Statistical analysis in volleyball

Several researchers have analyzed various aspects of the game: Echeverria et al. et al. [21] examined the effectiveness of blocks in women's volleyball, while Giatsis [22] studied performance criteria in men's beach volleyball. Special attention is given to the relationship between ball reception and attack effectiveness, as described in the works of Lopez et al. [25, 26], and the sequence of game combinations researched by Hileno et al. [23] and Laporta et al. [24]. Rocha et al. [27] and Rodrigues Rocha et al. [28] analyzed the influence of a player's position on building an attack. Sarmento et al. [29] emphasized the importance of statistical analysis for a deep understanding of game processes.

In our research, we used statistical reports on matches of elite teams in the World Championship. The descriptive statistics obtained from data processing provided an insight into the activities of teams and volleyball players. This allowed for the assessment of each player's contribution to the team's success.

While previous studies have primarily focused on specific aspects of volleyball, such as block effectiveness, performance criteria, or the relationship between ball reception and attack effectiveness, our research offers a comprehensive view of elite team dynamics in the World Championship. By utilizing statistical reports from these matches, we have been able to provide a more holistic understanding of team activities and individual player contributions. This approach not only gives a broader perspective but also allows for a more detailed assessment of each player's impact on the team's success, filling a gap that existed in prior research.

Modeling in volleyball

In modern volleyball, data-based modeling is critically important for enhancing training methodologies and game strategies. Aldous et al. [40] and Ntzoufras et al. [41] are developing stochastic and Bayesian methods for analyzing game dynamics, while Dai et al. [42] and Yuan et al. [43] are exploring the role of technological innovations in analyzing volleyball techniques. Endriani et al. [44] proposes a model based on Umbrella Learning, while Leng et al. [45], Li et al. [46], and Wang et al. [47] use random matrix models for analyzing learning and gameplay actions. Salimifard et al. [48], in turn, is developing an algorithm for scheduling in the volleyball premier league. All these works demonstrate innovative and diverse approaches to modeling in contemporary volleyball.

Our research presents an approach to extracting data from open internet sources and converting it into CSV format tables. The table data was processed in the PyCharm programming environment using Python code. In the PyCharm programming environment, using Python code, models of teams and players were also created. The 'pandas' library and 'scikit-learn' for machine learning were used to analyze data and perform statistical operations.

While the existing body of research in modern volleyball modeling has predominantly revolved around advanced stochastic methods, Bayesian analyses, technological innovations, and specific algorithms for game dynamics and scheduling, our study introduces a novel and practical approach to data extraction and processing. By harnessing open internet sources and converting this data into CSV format tables, we've streamlined the data acquisition process. Furthermore, our utilization of the PyCharm programming environment, combined with Python coding, offers a more integrated and efficient method for creating team and player models. The incorporation of the 'pandas' library and 'scikit-learn' for machine learning further distinguishes our research, enabling more robust data analysis and statistical operations, thereby bridging the gap between raw data acquisition and actionable insights in volleyball training and strategies.

Regression models in volleyball

In modern volleyball, regression models become a tool for a deep understanding of player performance and game dynamics. Albaladejo-Saura et al. [49] focuses on the influence of biological maturation on the physical preparation of adolescents, while Fuchs et al. [50, 51] analyzes the mechanics of movement and jump shots in elite volleyball players. Haraldsdottir et al. [52] emphasizes the relationship between sleep quality, well-being, and the risk of injury in female student volleyball players, and Lopez et al. [25] explores the relationship between ball reception and performance in men's volleyball. Vila-Maldonado et al. [53] accentuates the role of visual perception in blocking.

In our study, regression equations were created for teams and players. This allowed us to assess the contribution of each player to the team's success. Models are presented that best predict results for the team and volleyball players. Features of importance for teams are identified, which indicate the successful execution of game elements for the team. The regression equations for the team represent a linear combination of various game metrics that affect the team's total points in a match. They also emphasize the importance of action elements. Linear regression equations predict the total number of points for a volleyball player based on various statistical indicators.

While the contemporary research landscape in volleyball has seen a surge in specialized regression models focusing on aspects like biological maturation, movement mechanics, sleep quality, and visual perception, our study offers a more holistic and directly actionable approach to understanding team and player performance. By crafting regression equations tailored for both teams and individual players, we've been able to quantify the specific contributions of each player to their team's overall success. Our models stand out in their ability to predict results with precision, identifying key features that determine successful gameplay. Moreover, our linear regression equations provide a comprehensive view, linking various statistical indicators to the total points scored by a player. This approach not only offers insights into individual performance but also underscores the significance of specific action elements in the broader game dynamics, bridging the gap between individual metrics and team success.

In the era of digital transformation in sports, statistical modeling and analytics become integral tools for achieving high results. Our research emphasizes the importance of applying these methods in volleyball, where every action element, every player's movement can be decisive in a match. The recommendations presented in our study aim to help coaches and teams better understand and use statistical data to enhance their training and tactical strategies. It is especially crucial to regularly update data to ensure models remain relevant and reflect the real dynamics of the game. Implementing analytical tools based on the developed models into training programs can be a revolutionary step in preparing teams for future matches. This will allow teams to stay one step ahead of their opponents, adapting to changing game conditions and optimizing their strategies based on scientific data. In conclusion, statistical modeling and analytics in volleyball not only enhance the understanding of the game but also provide teams with specific tools to achieve success on the global stage.

Conclusions

In our study, a thorough analysis of statistical reports on matches of elite teams in the World Championship was conducted. This analysis allowed for a detailed understanding of the activities of both individual teams and individual volleyball players. Special attention was paid to assessing the contribution of each player to the overall success of the team.

The use of modern technological solutions, such as extracting data from open internet sources and subsequent processing in the PyCharm programming environment, ensured high accuracy and relevance of the obtained data. The use of the Python programming language, combined with the 'pandas' and ‘scikit-learn’ libraries, facilitated effective machine learning and statistical analysis.

The main emphasis of the study was on creating regression models for teams and players. These models not only allowed for an assessment of the contribution of each player to the team's success but also identified key game metrics that most strongly influence the total number of team points in a match. The linear regression equations developed in the study provide the ability to predict the total number of points a volleyball player will score based on various statistical indicators.

Our research underscores the importance of statistical modeling in modern volleyball and demonstrates the practical value of applying modern technological solutions to analyze game dynamics and the performance of teams and players. The results of our research can be applied by coaches and analytics specialists to optimize training processes, develop tactical strategies, and predict the results of future matches. Especially important is the identification of key game metrics that can serve as indicators of a team's success.

References

1. Akyildiz Z, Yuksel Y, Birgonul Y, Ceylan HI, Oliveira R, Cene E, et al. Comparison of external load measures per minute based on match seasonal periods and playing positions of a soccer Turkish super league team. *BMC Musculoskeletal Disorders*, 2023;24(1): 656. https://doi.org/10.1186/s12891-023-06787-y
2. Alanis A, Salas O, Salas K, Quintero I, Carranza Y, Salazar L. Playing at altitude. Performance of a Mexican professional football team at different level of altitude. *Apunts Sports Medicine*, 2022;57(215): 100391. https://doi.org/10.1016/j.apunsm.2022.100391
3. Alpaslan M, Akcan F. The effect of some statistical parameters on soccer success using the example of the Turkey Super League. *Pakistan Journal of Medical & Health Sciences*, 2021;15(2): 879–882.
4. Guan S, Wang X. Optimization analysis of football match prediction model based on neural network. *Neural Computing & Applications*, 2022;34(4): 2525–2541. https://doi.org/10.1007/s00521-021-05930-x
5. Cunniffe E, Connor M, Beato M, Grainger A, McConnell W, McCarthy Persson U, et al. The influence of possession status on the physical output of male international hockey players. *International Journal of Sports Science & Coaching*, 2022;17(2): 412–422. https://doi.org/10.1177/17479541211033958
6. Oliveira GP, Volpini de Barros JA, Menezes RP, Morato MP. The fifth player in the brazilian futsal elite: between a greater chance to finish and the setback danger. *Revista Brasileira De Futsal E Futebol*, 2021;13(56): 630–641.
7. Nunes RFH, Carvalho RR, Palermo L, Souza MP, Char M, Nakamura FY. Match analysis and heart rate of top-level female beach volleyball players during international and national competitions. *Journal of Sports Medicine and Physical Fitness*, 2020;60(2): 189–197. https://doi.org/10.23736/S0022-4707.19.10042-4
8. Torres-Luque G, Ivan Fernandez-Garcia A, Carlos Blanca-Torres J, Kondric M, Cabello-Manrique D. Statistical Differences in Set Analysis in Badminton at the RIO 2016 Olympic Games. *Frontiers in Psychology*, 2019;10: 731. https://doi.org/10.3389/fpsyg.2019.00731
9. Vaz L, Hendricks S, Kraak W. Statistical Review and Match Analysis of Rugby World Cups Finals. *Journal of Human Kinetics*, 2019;66(1): 247–256. https://doi.org/10.2478/hukin-2018-0061
10. Mikolajec K, Banys D, Zurowska-Cegielska J, Zawartka M, Gryko K. How to Win the Basketball Euroleague? Game Performance Determining Sports Results During 2003-2016 Matches. *Journal of Human Kinetics*, 2021;77(1): 287–296. https://doi.org/10.2478/hukin-2021-0050
11. Yi Q, Zhang S, Fang W, Gomez-Ruano MA. Modeling the Keys to Team's Success in the Women's Chinese Basketball Association. *Frontiers in Psychology*, 2021;12: 671860. https://doi.org/10.3389/fpsyg.2021.671860
12. Del Coso J, Brito de Souza D, Lopez-Del Campo R, Blanco-Pita H, Resta R. The football championship is won when playing away: difference in match statistics between the winner and the second-place team in LaLiga. *International Journal of Performance Analysis in Sport*, 2020;20(5): 879–891. https://doi.org/10.1080/24748668.2020.1801201
13. Souza DB, Lopez-Del Campob R, Blanco-Pita H, Resta R, Del Coso J. A new paradigm to understand success in professional football: analysis of match statistics in *LaLiga* for 8 complete seasons. *International Journal of Performance Analysis in Sport*, 2019;19(4): 543–555. https://doi.org/10.1080/24748668.2019.1632580
14. Longo UG, Sofi F, Candela V, Dinu M, Cimmino M, Massaroni C, et al. Performance Activities and Match Outcomes of Professional Soccer Teams during the 2016/2017 Serie A Season. *Medicina-Lithuania*, 2019;55(8): 469. https://doi.org/10.3390/medicina55080469
15. Longo UG, Sofi F, Candela V, Risi Ambrogioni L, Pagliai G, Massaroni C, et al. The influence of athletic performance on the highest positions of the final ranking during 2017/2018 Serie A season. *Bmc Sports Science Medicine and Rehabilitation*, 2021;13(1): 32. https://doi.org/10.1186/s13102-021-00259-3
16. Gong B, Cui Y, Gai Y, Yi Q, Gomez MA. The Validity and Reliability of Live Football Match Statistics From Champdas Master Match Analysis System. *Frontiers in Psychology*, 2019;10: 1339. https://doi.org/10.3389/fpsyg.2019.01339
17. Wheatcroft E. Forecasting football matches by predicting match statistics. *Journal of Sports Analytics*, 2021;7(2): 77–97. https://doi.org/10.3233/JSA-200462
18. Newans T, Bellinger P, Buxton S, Quinn K, Minahan C. Movement Patterns and Match Statistics in the National Rugby League Women's (NRLW) Premiership. *Frontiers in Sports and Active Living*, 2021;3: 618913. https://doi.org/10.3389/fspor.2021.618913
19. Yi Q, Gomez MA, Liu H, Sampaio J. Variation of match statistics and football teams' match performance in the group stage of the uefa champions league from 2010 to 2017. *Kinesiology*, 2019;51(2): 170–180. https://doi.org/10.26582/k.51.2.4
20. Kopp W, Vingron M. DNA Motif Match Statistics Without Poisson Approximation. *Journal of Computational Biology*, 2019;26(8): 846–865. https://doi.org/10.1089/cmb.2018.0144
21. Echeverria C, Ortega E, Palao JM. Normative Profile of the Efficacy and Way of Execution for the Block in Women's Volleyball from Under-14 to Elite Levels. *Montenegrin Journal of Sports Science and Medicine*, 2020;9(1): 41–47. https://doi.org/10.26773/mjssm.200306
22. Giatsis G. Beach volleyball performance benchmarks in men's high level. *Journal of Human Sport and Exercise*, 2023;18(2): 446–461. https://doi.org/10.14198/jhse.2023.182.15
23. Hileno R, Arasanz M, Garcia-de-Alcaraz A. The Sequencing of Game Complexes in Women's

Volleyball. *Frontiers in Psychology*, 2020;11: 739. https://doi.org/10.3389/fpsyg.2020.00739
24. Laporta L, Araripe Medeiros AI, Vargas N, Castro H de O, Bessa C, Joao PV, et al. Coexistence of Distinct Performance Models in High-Level Women's Volleyball. *Journal of Human Kinetics*, 2021;78(1): 161–173. https://doi.org/10.2478/hukin-2021-0048161
25. Lopez E, Diez-Vega I, Molina JJ. Reception and performance in high level male volleyball: A relational study. *Journal of Human Sport and Exercise*, 2022;17(2): 409–423. https://doi.org/10.14198/jhse.2022.172.16
26. Lopez E, Molina JJ, Diaz-Bento MS, Diez-Vega I. Spike performance in K1: influence of rotation and reception area on high level men's volleyball teams. *Retos-Nuevas Tendencias En Educacion Fisica Deporte Y Recreacion*, 2023;(48): 213–221.
27. Rocha ACR, Laporta L, Rodrigues GP, de Lira CAB, Castro H de O, Costa G de CT. Is it possible for the reception and the player-receiver to influence the offensive construction in volleyball? An ecological vision and Intragame procedural relationships in high-level teams. *International Journal of Performance Analysis in Sport*, 2023;23(4): 319–333. https://doi.org/10.1080/24748668.2023.2229204
28. Rocha ACR, Laporta L, Andre Barbosa De Lira C, Modenesi H, Figueiredo LS, Costa GDCT. Complex I in male elite volleyball: an interactional analysis according to reception location. *International Journal of Performance Analysis in Sport*, 2022;22(1): 77–89. https://doi.org/10.1080/24748668.2021.2003961
29. Sarmento H, Clemente FM, Afonso J, Araujo D, Fachada M, Nobre P, et al. Match Analysis in Team Ball Sports: An Umbrella Review of Systematic Reviews and Meta-Analyses. *Sports Medicine-Open*, 2022;8(1): 66. https://doi.org/10.1186/s40798-022-00454-7
30. Budak G, Kara I. A multiobjective mathematical model to form the best team at sports clubs: team harmony and player performance objectives. *Team Performance Management*, 2022;28(3/4): 145–159. https://doi.org/10.1108/TPM-06-2021-0047
31. Wang H, Dong C, Fu Y. Optimization analysis of sport pattern driven by machine learning and multi-agent. *Neural Computing & Applications*, 2021;33(4): 1067–1077. https://doi.org/10.1007/s00521-020-05022-2
32. Wang J. Mining and Prediction of Large Sport Tournament Data Based on Bayesian Network Models for Online Data. *Wireless Communications & Mobile Computing*, 2022;2022: 1211015. https://doi.org/10.1155/2022/1211015
33. Fioravanti F, Delbianco F, Tohme F. The relative importance of ability, luck and motivation in team sports: a Bayesian model of performance in the English Rugby Premiership. *Statistical Methods and Applications*, 2023;32(3): 715–731. https://doi.org/10.1007/s10260-022-00677-8
34. Duthie GM, Robertson S, Thornton HR. A GNSS-based method to define athlete manoeuvrability in field-based team sports. *Plos One*, 2021;16(11): e0260363. https://doi.org/10.1371/journal.pone.0260363
35. Hsu YC. Using Convolutional Neural Network and Candlestick Representation to Predict Sports Match Outcomes. *Applied Sciences-Basel*, 2021;11(14): 6594. https://doi.org/10.3390/app11146594
36. Campo M, Champely S, Louvet B, Rosnet E, Ferrand C, Pauketat JVT, et al. Group-Based Emotions: Evidence for Emotion-Performance Relationships in Team Sports. *Research Quarterly for Exercise and Sport*, 2019;90(1): 54–63. https://doi.org/10.1080/02701367.2018.1563274
37. Kim S, Morgan A, Assaker G. Examining the relationship between sport spectator motivation, involvement, and loyalty: a structural model in the context of Australian Rules football. *Sport in Society*, 2021;24(6): 1006–1032. https://doi.org/10.1080/17430437.2020.1720658
38. Phatak AA, Mehta S, Wieland FG, Jamil M, Connor M, Bassek M, et al. Context is key: normalization as a novel approach to sport specific preprocessing of KPI's for match analysis in soccer. *Scientific Reports*, 2022;12(1): 1117. https://doi.org/10.1038/s41598-022-05089-y
39. Singh A, Scarf P, Baker R. A unified theory for bivariate scores in possessive ball-sports: The case of handball. *European Journal of Operational Research*, 2023;304(3): 1099–1112. https://doi.org/10.1016/j.ejor.2022.05.010
40. Aldous DJ, Cruz M. A real-world Markov chain arising in recreational volleyball. *Involve, a Journal of Mathematics*, 2021;14(5): 829–852. https://doi.org/10.2140/involve.2021.14.829
41. Ntzoufras I, Palaskas V, Drikos S. Bayesian models for prediction of the set-difference in volleyball. *Ima Journal of Management Mathematics*, 2021;32(4): 491–518. https://doi.org/10.1093/imaman/dpab007
42. Dai X, Li S. Application Analysis of Wearable Technology and Equipment Based on Artificial Intelligence in Volleyball. *Mathematical Problems in Engineering*, 2021;2021: 5572389. https://doi.org/10.1155/2021/5572389
43. Yuan Z, Zhang Y, Li B, Jin X. A Diagnostic Model of Volleyball Techniques and Tactics Based on Wireless Communication Network. *Computational and Mathematical Methods in Medicine*, 2022;2022: 2185908. https://doi.org/10.1155/2022/2185908
44. Endriani D, Sitompul H, Mursid R, Dewi R. Development of a Lower Passing Model for Volleyball Based Umbrella Learning Approach. *International Journal of Education in Mathematics Science and Technology*, 2022;10(3): 681–694. https://doi.org/10.46328/ijemst.2508
45. Leng S, Shao M. A Study on the Effect of the Club Model on the Effectiveness of College Volleyball Teaching Based on a Random Matrix Model. *Mathematical Problems in Engineering*, 2022;2022: 5681412. https://doi.org/10.1155/2022/5681412
46. Li F, Jia N, Wang H, Zheng H. Nonlinear Random Matrix Model and Research for Quantitative Representation of Volleyball Attacker's Action Links.

Mathematical Problems in Engineering, 2022;2022: 2279813. https://doi.org/10.1155/2022/2279813
47. Wang T, Zheng H, Li F, Jia N, Cai Z. Evaluation Algorithm of Volleyball Players' Competitive Ability Based on the Random Matrix Model. *Mathematical Problems in Engineering*, 2022;2022: 6967379. https://doi.org/10.1155/2022/6967379
48. Salimifard K, Li J, Mohammadi D, Moghdani R. A multi objective volleyball premier league algorithm for green scheduling identical parallel machines with splitting jobs. *Applied Intelligence*, 2021;51(7): 4143–4161. https://doi.org/10.1007/s10489-020-02027-1
49. Albaladejo-Saura M, Vaquero-Cristobal R, Garcia-Roca JA, Esparza-Ros F. Influence of biological maturation status on selected anthropometric and physical fitness variables in adolescent male volleyball players. *Peerj*, 2022;10: e13216. https://doi.org/10.7717/peerj.13216
50. Fuchs PX, Fusco A, Bell JW, von Duvillard SP, Cortis C, Wagner H. Movement characteristics of volleyball spike jump performance in females. *Journal of Science and Medicine in Sport*, 2019;22(7): 833–837. https://doi.org/10.1016/j.jsams.2019.01.002
51. Fuchs PX, Mitteregger J, Hoelbling D, Menzel HJK, Bell JW, von Duvillard SP, et al. Relationship between General Jump Types and Spike Jump Performance in Elite Female and Male Volleyball Players. *Applied Sciences-Basel*, 2021;11(3): 1105. https://doi.org/10.3390/app11031105
52. Haraldsdottir K, Sanfilippo J, McKay L, Watson AM. Decreased Sleep and Subjective Well-Being as Independent Predictors of Injury in Female Collegiate Volleyball Players. *Orthopaedic Journal of Sports Medicine*, 2021;9(9): 23259671211029285. https://doi.org/10.1177/23259671211029285
53. Vila-Maldonado S, Saez-Gallego NM, Garcia-Lopez LM, Contreras OR. Influence of visual behaviour on decision making in volleyball blocking. *Revista Internacional De Medicina Y Ciencias De La Actividad Fisica Y Del Deporte*, 2019;19(75): 489–504. https://doi.org/10.15366/rimcafd2019.75.007
54. Volleyball. [Internet]; 2023 Aug 02 [cited 2023 Aug 25]. Available from: https://www.championat.com/volleyball/_oth.html
55. Breiman L. Random Forests. *Machine Learning*, 2001;45(1): 5–32. https://doi.org/10.1023/A:1010933404324
56. Kozinc Z, Plesa J, Sarabon N. Questionable Utility of the Eccentric Utilization Ratio in Relation to the Performance of Volleyball Players. *International Journal of Environmental Research and Public Health*, 2021;18(22): 11754. https://doi.org/10.3390/ijerph182211754
57. Nimmagadda SL, Mullins A, Reiners T, Mani N. Design Science Guided Sports Information System Framework Development for Sports Data Analytics. *Amcis 2020 Proceedings*, 2020.

Information about the authors:

Sergii Iermakov; (Corresponding Author); https://orcid.org/0000-0002-5039-4517; sportart@gmail.com; Department of Methodologies of Cross-Cultural Practices, Kharkiv State Academy of Design and Arts; Kharkiv, Ukraine.

Tetiana Yermakova; https://orcid.org/0000-0002-3081-0229; yermakova2015@gmail.com; Department of Pedagogy, Ukrainian and Foreign Philology; Kharkiv State Academy of Design and Arts; Kharkiv, Ukraine.

Krzysztof Prusik; https://orcid.org/0000-0002-9273-3126; prusik6471@gmail.com; Department of Sport, Gdansk University of Physical Education and Sport; Poland.

Cite this article as:
Iermakov S, Yermakova T, Prusik K. Modeling the gameplay actions of elite volleyball players and teams based on statistical match reports. *Pedagogy of Health*, 2023;2(2):50–64.
https://doi.org/10.15561/health.2023.0202

Received: 10.10.2023
Accepted: 14.11.2023; Published: 30.12.2023