

RESEARCH

Open Access



A toolbox for volleyball data analytics: a case study on the Italian women's league

Andrea Accornero¹, Pasquale Cascarano^{2*}, Maurizio Napolitano³, Davide Mazzanti⁴ and Gustavo Marfia²

*Correspondence:
pasquale.cascarano2@unibo.it

¹ Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

² Department of the Arts, University of Bologna, Bologna, Italy

³ Digital Commons Lab, Fondazione Bruno Kessler, Trento, Italy

⁴ AIAPAV - Associazione Italiana Allenatori Pallavolo, Rome, Italy

Abstract

Data analytics plays a central role in volleyball, offering new ways to enhance team performance and inform match strategies. This paper presents an open-source Python-based toolbox that extends the PyDataVolley library, enabling advanced processing, visualization, and analysis of scouting data. The toolbox includes Machine Learning Clustering algorithms, Multi-Criteria Decision Analysis approaches, and Markov Chain models, and is validated using datasets from the 2023–2024 and 2024–2025 seasons of the Italian Women's Serie A2 Championship. A case study on the 2025 Italian Cup final between Consolini Volley and Trentino Volley highlights the practical impact of the toolbox. Throughout the 2024–2025 season, Consolini Volley consistently showed superior performance metrics compared to Trentino Volley. However, in the final match, Trentino Volley implemented a targeted tactical strategy informed by the insights from the toolbox, which effectively challenged Consolini's gameplay. As a result, Consolini Volley exhibited a general drop across all main performance indicators during the final.

Keywords: Sports analytics, Volleyball, Scouting data, Data volleyball, Machine learning, Clustering, Multi-criteria decision analysis, Markov chain models, Performance metrics

Introduction

In the past few years, sports data analytics [1, 2] has increased exponentially, delivering insightful solutions to maximize team performance, support strategic decision-making, and prevent players' injuries [3–6].

The current literature illustrates that academic researchers have primarily focused on sports such as football (soccer) and basketball [7–10]. These are invasion games [11] in which two opposing sides seek to invade each other's territory and score in a goal or over a line. They are oppositional games, with possession able to change at any moment, demanding constant tactical adaptation and coordinated team movement. Rather than being divided into discrete units, these games unfold through continuous, fluid sequences that evolve dynamically in space and time. The use of spatiotemporal tracking technologies [12, 13] has enabled the development of complex data-driven models to analyze players' positioning, passing interactions, and scoring opportunities in real-time, driving informed decision-making in coaching and performance optimization [1, 14].

In field games like baseball [11], the players score by striking a ball and running to designated areas, while the opposing team fields the ball and attempts to prevent scoring or record outs. These sports are typically played on large outdoor fields with clearly defined positions and strategic placement of players. In net games like volleyball [11], the opponents are separated by a net or wall and aim to send an object, such as a ball or shuttle, into the opponent's area so that it cannot be returned successfully. Success depends on placement, timing, and control. Both baseball and volleyball share a discrete and structured gameplay [5, 15–17], offering rich opportunities for analytical modeling, although following a different paradigm [1]. These sports consist of segmented and well-defined actions (e.g., a pitch and a rally, for baseball and volleyball respectively), allowing data analysts to model transitions, evaluate player decision-making, and assess performances within repeated tactical patterns.

Volleyball is one of the most widely played team sports globally [18], with a rich history that spans more than a century. Serve, reception, set, attack, block, and defense are the fundamental game actions within a volleyball match. Each of these actions presents a range of technical variations, however, the success of their executions is dependent not only on the physical capabilities of the volleyball players [19, 20] but also on their tactical skills [21]. This underscores the importance of developing tools that support adaptive tactical systems in volleyball [22, 23].

In this context, data analysis can play a critical role by providing coaches with objective insights into player performance and match dynamics [5, 24, 25], enabling them to make informed decisions and guide players toward more effective tactical choices. However, different from sports such as soccer, basketball, and baseball, all of which have embraced sophisticated analytical models, many volleyball technical and tactical components have yet to be fully examined through advanced data-driven methodologies [5]. This gap depends on different factors: limited commercial exposure, restricted data accessibility, and an as-yet emerging software ecosystem for the processing, visualization, and analysis of volleyball data at scale with sophisticated tools [26]. Nevertheless, the availability of detailed scouting reports makes volleyball an excellent candidate for modern data-driven analytical techniques [25].

Data Volley,¹ is the standard software to help the technical staff to compile professional volleyball scouting files (in.dvw format), providing a structured record of every technical action performed during the game. In addition, the software offers several tools for data analysis and visualization. It is widely used by top teams and federations for performance evaluation and match preparation [25]. However, as a closed-source commercial tool, it lacks flexibility: users cannot customize features or integrate new analytical models, and the absence of an open API limits accessibility for researchers and smaller clubs. To overcome these challenges, the Open Volley project² was developed as an open-source toolbox allowing the parsing and processing of.dvw files generated by Data Volley for data visualization and analysis. However, Open Volley also presents constraints: its functionality is primarily built in R [27], which restricts its compatibility with advanced data-driven models that are usually implemented in

¹ <https://www.dataproject.com/Products/EU/it/Volleyball/DataVolley4>.

² <https://github.com/openvolley>.

other programming languages such as Python [28, 29]. These limitations underscore the need for a versatile framework to unlock the potential of volleyball scouting data.

In this work, we present a novel open-source toolbox for the visualization and analysis of volleyball data (see Fig. 1). More specifically, we provide a pipeline to encode multiple scouting files exported from Data Volley in.dvw format into a unified, structured database using the.csv format which can be easily imported into programming environments such as Python. Furthermore, a Python library is developed, enabling users to visualize data through an interface that resembles the one provided by Data Volley. In collaboration with the coaching staff of Trentino Volley, a team competing in the Italian Women's Serie A2, a set of novel performance metrics is defined to numerically represent the team performance across the various technical and tactical aspects of the game. These metrics are validated to ensure they accurately reflect the quality and outcomes of a team's play. Furthermore, they are analyzed over time to detect performance trends and temporal dynamics throughout the season. In addition, Multi-Criteria Decision Analysis [30] and Machine Learning-based Clustering [31] techniques are provided as data analysis tools, as well as Markov chain models [32] which can be used to represent the rally dynamics, capture recurring transition patterns and estimate outcome probabilities from different game states. All these tools extend the capabilities of the existing open-source project PyDataVolley [33] unlocking deeper data analysis. The toolbox is validated on a novel dataset of scouting files from the 2023–2024 and 2024–2025 seasons of the Italian Women's Serie A2 Championship. To validate the proposed open-source toolbox, a detailed case study

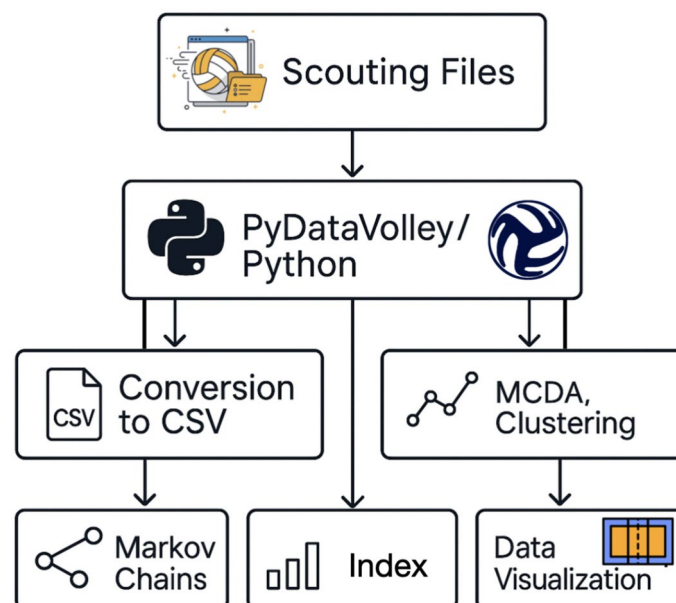


Fig. 1 Overview of the proposed toolbox. Scouting files exported from Data Volley are processed using a Python-based toolkit. The toolbox includes CSV conversion, computation of performance indicators, application of machine learning clustering and multi-criteria decision analysis techniques, Markov chain modeling, and generation of interactive data visualizations.

is conducted on the 2025 Coppa Italia Serie A2 final, played in Bologna in February, between Trentino Volley³ and Consolini Volley.⁴

The paper is organized as follows. In Sect. 2 we provide an overview of the relevant literature concerning volleyball data analysis. In Sect. 3, we describe the proposed toolbox, detailing the data processing pipeline, visualization modules, performance metrics, clustering techniques, decision analysis models, and Markov chain modeling. In Sect. 4, we present the analysis of two seasons of the Italian Women's Serie A2 Championship and report the findings of a case study on the 2025 Serie A2 Italian Cup final. In Sect. 5 limitations and future works are discussed. Finally, Sect. 6 draws the conclusions.

Related work

Match analysis plays a crucial role in identifying strengths and weaknesses in volleyball performance and in translating this information into targeted technical and tactical training content. As highlighted in [34], match analysis is valuable for guiding training programs, enabling coaches to tailor practice sessions based on performance data.

Volleyball data analysis has evolved significantly in recent years but remains relatively underrepresented compared to other more popular sports such as football, basketball, and baseball [7, 9, 10, 15, 35–37]. For example, [35] employed big data analytics and classification and regression trees to model scoring probabilities in basketball under high-pressure conditions. Similarly, using a data-driven approach, [36, 37] investigated team and player scoring probabilities across different areas of the basketball court. In football, [38, 39] proposed statistical models for predicting match outcomes and tournament progression. All these frameworks share a data-driven methodology; however, unlike our work, they are specifically designed for sports other than volleyball, such as football and basketball.

The seminal work presented in [40] established the role of statistical methods in volleyball performance analysis and inspired early quantitative approaches to match analysis. Probabilistic modeling has attracted researchers working on volleyball data. In [5, 17, 41–44] Markov Chains and Bayesian models are employed to model rally structures and to quantify individual player contributions in elite volleyball. More specifically, [43] developed a hierarchical framework incorporating team-specific offensive and defensive effects, derived from in-game efficiency statistics, while [17] quantified how changes in technical actions affect win probabilities and recommended skill prioritization to improve team performance. Moreover, in [17] the authors proposed a Monte Carlo simulation-based decision-support system to evaluate and test the effectiveness of different strategic options in match contexts. A different paradigm was developed in [45] where the authors applied an inductive logic programming method to discover spatiotemporal attack strategies and to identify interpretable patterns linking player positions and skill sequences to successful rallies.

A significant body of recent literature has explored the usage of machine learning in volleyball data analysis. For instance, the authors in [24, 46] developed models based machine learning models to predict match outcomes using detailed match datasets

³ <https://www.trentinovolley.it/it/femminile/roster>.

⁴ <https://www.volleysangiovanni.it/>

based on player features and team aggregates suggesting that the latter are better for the accuracy, while [47] applied edge computing for real-time tactical patterns forecasting. Furthermore, [48] demonstrated that data quality is the most critical factor influencing the accuracy of the prediction, while the choice of the machine learning model has less impact. More advanced intelligent systems have also been proposed for tracking and tactical recognition from match videos [49–51]. For instance, [50] presented a transformer-based framework for fully automated multi-player tracking and skill recognition and, moreover, in [51] the authors introduced a wearable sensor system that uses machine learning approaches to perform real-time action recognition. Finally, these paradigms have been used for identifying risk factors for overuse injuries [52] through daily monitoring of jump load and wellness. Additionally, they have been used for personalized training [53], whereby incorporating biomechanical and electromyographic data into learning pipelines, training recommendations are provided based on muscle activation patterns.

The Data Volley software has been largely used for tactical analysis [21, 25]. Notably, in [21], the software identified attacking as the most crucial action for set success, especially when supported by versatile settings and reduced error rates, with minor differences between elite junior and senior women's volleyball. As part of the Open Volley project, an open-source R implementation of Data Volley was developed.⁵ This version offers tools for importing, parsing, and analysing volleyball scouting files in `dvw` (and `vsm`) format using the official Data Volley notation. It supports simple filtering plays by action type, computing performance indicators (e.g. success rates, efficiency, serve win rate, side-out/break point percentages), and generating visualizations such as distributions and heatmaps.

Despite these advances, the availability of open-source and customizable tools that support both data visualization and analysis remains limited. To the best of our knowledge, no study in the literature has proposed a unified toolbox capable of incorporating both machine learning-based methods and advanced statistical approaches. Furthermore, as noted by [5], the diversity of data sources and analytical tools used across studies prevent from meaningful comparisons between literature findings. One of the primary goals of the toolbox proposed in this work is precisely to standardize volleyball data analysis practices, enabling more consistent, scalable, and reproducible workflows. Finally, no previous study has performed an analysis on very recent scouting data from the Italian Serie A, one of the most competitive and technically rich volleyball leagues in the world.

Methods

In this section, we first describe the dataset used and the scouting file structure. Then, we describe the main modules of the toolbox, including data conversion, visualization, custom performance metrics, machine learning clustering techniques, multi-criteria decision analysis, and Markov chain modeling.

⁵ <https://github.com/openvolley/datavolley>.

Scouting files and dataset

Volleyball is a team sport played between two teams of six players on a rectangular court divided by a net. The game is structured into rallies, each beginning with a serve and ending when the ball touches the ground, a fault is committed, or the referee stops play. Teams aim to win rallies by grounding the ball in the opponent's court or forcing an error. The match is typically played in a best-of-five sets format, and the first team to win three sets wins the match. Players rotate clockwise through six positions, organized into a front and back row. Each team has players with different roles, including a setter, two outside and one opposite hitter, two middle blockers, and the libero who substitutes the middle blocker when he/she is in the back row.

In Volleyball, the data collection is typically carried out by the scoutmen who are responsible for recording every ball contact using standardized alphanumeric codes and assigning evaluations based on predefined technical criteria. Data can be collected either manually, using papers or specialized software such as Data Volley, VolleyStation,⁶ and VolleyMetrics,⁷ or through emerging automated systems based on computer vision [49–51, 54]. Although these technologies currently lack the precision and detail provided by human scouts.

The main output of data collection is the scouting file, where each ball contact is encoded in a standardized format. The most widely adopted coding standard for this is Data Volley. However, some evaluations may reflect the scoutman's interpretation or team-specific standards. In this study, all files were collected and reviewed by Trentino Volley's staff to ensure uniformity.

These scouting files are usually used in the pre-match to analyze opponents, during matches to adjust strategies in real-time, and in the post-match to assess performance and guide training. According to the official.dvw notation in DataVolley, a scouting file is composed of three different parts: main, advanced, and extended codes. The main code considers five components: the team identifier (* for home, *a* for away), the player's number, the game action performed (e.g., S for serve, R for reception, A for attack, B for block, D for defense, E for set), a game action-specific descriptor, and a performance evaluation (e.g., #, +, !, /, -, =) reflecting the quality of the action. It is important to note that team scouts register primarily their own team's data; therefore, when a game action receives dual evaluations (one for each team), priority is given to the evaluation relevant to their own team. For example, a winning spike with a touch on the opponent's block is recorded as an attack “#” for the attacking team and as a block “=” (point lost) for the defending team. This approach enables the calculation of specific performance indicators for each team and for every game action. In Table 1 we report a short explanation for the evaluation symbols for each game play.

Along with the main code, each action may have an extended code (max 4 elements) identifying tactical combinations, starting and landing zones, and/or sub-zones of the ball. The extended code (max 3 elements) contains details such as action type, players involved (e.g., receiver or blockers), and special contextual tags. Additionally, metadata provides essential match dates, teams, set scores, roles, and rotations.

⁶ <https://volleystation.com/>

⁷ <https://portal.volleymetrics.hudl.com.>

Table 1 Evaluation symbols for game actions (S = Serve, R = Reception, E = Set, A = Attack, B = Block, D = Defence) as used in Data Volley scouting system.

Symbol	General meaning	S	R	E	A	B	D
#	Winning action	Ace	Perfect reception (all attack combinations are available)	Perfect set (single or no bloc)	Attack point	Block point	Dig leading to direct point
+	Positive action	Serve causes severe trouble	Good reception (not all attack combinations are available)	Good set (the opponent has only one attack)	Positive attack (defence out of system)	Positive block touch	Good dig allowing full attack
!	Context-dependent	Serve with slight advantage	Acceptable reception	Acceptable set	Attack kept in play with some advantage	Block touch with no clear advantage	Dig keeping rally alive
/	Context-dependent	Overpass favouring serving team	Overpass to opponent	Poor set limiting options	Attack blocked (point to opponent)	Net fault	Poor dig giving advantage to opponent
-	Clear negative	Easy serve	Poor reception	Bad set	Weak attack easily defended	Failed block attempt	Poor dig
=	Error	Service error	Reception error	Setting error	Attack error	Block error	Dig error (missed ball)

The dataset for this study comprises detailed scouting data from the full 2023–2024 season and part of the 2024–2025 season of the Italian Women’s Serie A2 Championship. Provided by the technical staff of Trentino Volley, the data were originally recorded in the official Data Volley.dvw format. The competition features two groups (Group A and Group B) of 10 teams each. The dataset includes all 180 matches from the 2023–2024 season (8780 rallies, 231,697 ball touches, 285 players) and, for the 2024–2025 season, only the matches played by Trentino Volley (16 matches, 809 rallies, 23,050 touches, 116 players) and Consolini Volley (15 matches, 766 rallies, 18,759 touches, 120 players), as well as the Italian Serie A2 Cup final between Trentino Volley and Consolini Volley.

PyDataVolley

PyDataVolley [33] is a Python library developed within the OpenVolley [55] ecosystem. We extended it to support more advanced analytical workflows, improve compatibility, and address limitations in flexibility and metadata handling. Below, we provide more details about the implemented tools.

Conversion to CSV

The original implementation of PyDataVolley already allowed basic parsing of DataVolley (.dv/.dvw) files, however, we also implemented a converting system from.dvw files to.csv files by exploring well-established Python libraries like Pandas. Furthermore, dynamic parsing of setter calls and offensive combinations were implemented, allowing consistent interpretation of custom codes used by scoutmen across different leagues. Player roles were added to the dataset to enhance role-specific analyses, and

additional metadata fields such as matchday, group, and final scores were integrated to support large-scale competition studies. The concept of ball positioning during sets was introduced, distinguishing setting zones (front, near, over, back) to enable spatial tactical analysis. The zone mapping was refined by including end-subzones for more precise action localization. In addition, new statistical metrics were developed: Expected Side-out Rate, Expected Breakpoint Rate, and Set Assist Rate providing deeper insights into key phases of play. Specialized filtering functions were also implemented, enabling the extraction of attacks after perfect receptions and first transition attacks, supporting targeted analysis of offensive efficiency. Finally, many improvements were made to the data extraction process, including automatic encoding detection for file robustness.

The current version of the library is designed to translate raw scouting files into structured formats that can be easily integrated into modern Python-based data science pipelines.

Data Visualization

Using the available data, various game situations were graphically represented through Python, employing the Bokeh library [56] to create interactive visualizations. The goal was to transform numerical and categorical information into a more intuitive format, thereby facilitating the analysis of game dynamics.

Performance Metrics

The limited ability to customize performance indicators is a major drawback of current tools used by coaching staff [5]. In practice, coaches may wish to adjust the weight assigned to specific actions according to strategic priorities or training objectives. Our toolbox overcomes this limitation by enabling the definition of custom efficiency metrics, where positive and negative evaluations can be specified for each fundamental game action (serve, attack, reception, block, and defense). These metrics can be based either on established volleyball standards or on the specific requirements expressed by the coaching staff. The efficiency metrics are computed according to the following formula:

$$\text{Efficiency} = \frac{\text{Positive Evaluations} - \text{Negative Evaluations}}{\text{Total Attempts}}.$$

This calculation provides a normalized performance metric, balancing successful and unsuccessful actions relative to the total number of attempts for each action.

In our analysis, the attack actions were further segmented into three categories: Overall Attack, First Ball Side-Out (FBSO), namely the first attack after the reception, and Attack After Service Turn (AST), namely the first attack after the opponent's attack. This segmentation enables a clearer distinction between side-out and break-point phases. Furthermore, the selection of positive and negative evaluations for each game action was guided by two main criteria: (i) consistency with volleyball performance standards [43], where positive outcomes are those that directly lead to scoring opportunities or limit the opponent's options, and negative outcomes are those that result in a point for the opponent or a clear tactical disadvantage; and (ii) context-specific feedback from Trentino Volley's coaching staff, who adjusted these classifications based on their tactical priorities and experience (e.g., drawing on the evaluation codes used in match scouting, as shown in Table 1). For example, in attack-related metrics (Overall Attack, FBSO, AST),

only winning attacks (#) are counted as positive, while errors (=) and blocked attacks resulting in opponent points (/) are considered negative. For serve, both direct points (#) and highly disruptive serves (+) are positive, whereas service errors (=) are negative. In block, the inclusion of touches that keep the rally under control (+) reflects the Trentino Volley coaching staff's emphasis on slowing down the opponent's offense, even without an immediate point. For defensive actions, only those enabling an effective counterattack (#, +) are considered positive, while plays ending the rally in favor of the opponent (=) are negative. We remark that the implemented tool allows users to define, according to their own criteria, which events are considered positive and which are considered negative for each action. In Table 2, we summarize, for each game action, the evaluations classified as positive or negative in the following experimentation.

Moreover, we point out that, in our implementation, these efficiency indicators can be computed either globally at the team level or individually for each player, enabling both collective and personalized performance assessment. Finally, these metrics can be analyzed both statically, by aggregating metrics across the entire season, and progressively, by computing cumulative metrics match-by-match to monitor performance trends and tactical evolution over time.

Clustering Techniques

The current version of the toolbox provides machine learning-based clustering techniques including:

- K-Means (partition-based clustering), which divides teams into well-defined groups by minimizing within-cluster variance [57];
- Gaussian Mixture Model (model-based probabilistic clustering), which allows for more flexible cluster shapes by assigning each point a probability of belonging to each cluster [58];
- Agglomerative Clustering (hierarchical clustering), which uses a bottom-up hierarchical strategy to identify structures in the data without the need to specify initial centroids [59].

All clustering models were implemented using the `scikit-learn` library, leveraging the `KMeans`, `GaussianMixture`, and `AgglomerativeClustering` classes. Data normalization was performed before training with `StandardScaler` to ensure comparability across features before model fitting.

Table 2 Positive and negative evaluations used for the computation of the efficiency metrics for each game action

Skill	Positive evaluations	Negative evaluations
Overall Attack	#	=, /
Serve	#, +	=
Reception	#	=
Block	#, +	=
Defense	#, +	=
FBSO	#	=, /
AST	#	=, /

Multi-Criteria Decision Analysis Techniques

The proposed toolbox also integrates several MCDA techniques [30], namely the Analytic Hierarchy Process (AHP) [60], the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [61], the Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE), and the Multi-Objective Optimization on the Basis of Ratio Analysis (MOORA) [62]. These methods enable the construction of team rankings by aggregating multiple performance indicators into a single composite score. Specifically, after defining and computing the performance metrics for each team, the MCDA techniques were applied to combine several technical indicators (overall attack, serve, reception, block, defense, FBSO, and AST) into one overall rating. In this study, each indicator was assigned an equal weight, although the toolbox allows full customization based on coaching preferences.

In the implemented AHP (using the `ahpy` package), the decision problem was structured with the team ranking as the overall goal, the technical indicators as criteria, and the teams as alternatives. All criteria were given equal importance through uniform pairwise comparisons, and the resulting weights were aggregated with the team performance scores. Moreover, TOPSIS (implemented via `scikit-criteria`) ranked teams by calculating each team's relative distance from an ideal performance profile (best value in each indicator) and from a worst profile (worst value in each indicator), with higher rankings assigned to teams closer to the ideal and farther from the worst. In addition, PROMETHEE (implemented via `pyDecision`) used an outranking method, comparing each team pair across all indicators to calculate positive and negative preference flows, which were then combined into a net preference score to produce the final ranking. Finally, MOORA (implemented via `pyDecision`) normalized all performance indicators, combined them into a single composite score, and ranked teams directly from these aggregated normalized values, providing a simple yet robust ranking method.

Markov Chain Modeling

Markov Chain modeling is well-suited to the nature of the dataset [41], as each rally, which ends with the point scored by a team, results from a sequence of ball contacts. More precisely, each ball contact during a rally is treated as a state in a Markov chain, while the rally outcome (point won/lost) represents an absorbing state.

A Markov Chain [32, 41] is a stochastic process where the probability of transitioning to the next state depends solely on the current state. This characteristic can be expressed by the following formulation:

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i).$$

This formulation highlights that the system's future evolution X_{n+1} relies only on its present state X_n . The states are categorized as Transient States, namely intermediate stages of the system, such as a reception, set, or attack during a rally, and Absorbing States, final states where the process ends, such as the rally conclusion when a point is awarded.

The Markov chain sequences are constructed as follows. Each ball contact is encoded using abbreviations: serve is labeled as `SV`, while other skills like block (`B`), dig (`D`), set (`S`), and reception (`R`) are combined with the evaluation symbol (`#`, `+`, `!`, `/`, `-`, `=`). Attack actions are referred to as `A` plus the attack type code (e.g., `XX` stands for a fast ball for the outside hitter player), and rally conclusions are marked with `P`. States are

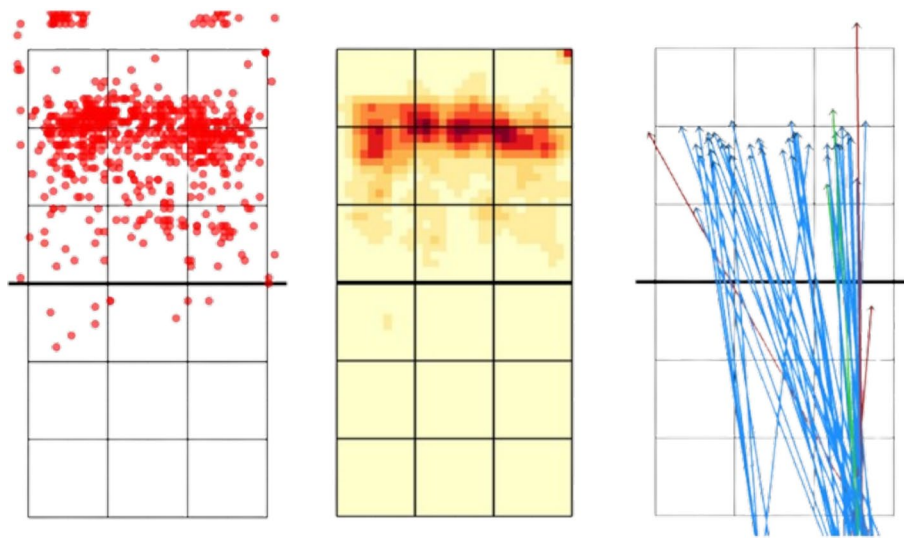


Fig. 2 Examples of three visual representations of volleyball serves. From left to right: scatter plot of serve landing points, heat map showing serve distribution density, and trajectory plot illustrating serve directions

progressively concatenated within the same possession, for example, a sequence of R#, S+, AXX becomes R#, R# . S+, R# . S+ . AXX. A prefix (e.g., H_ for home team, V_ for visitors) is added to each state to identify team possession. Finally, the rally sequences are grouped by match and point identifiers, generating an ordered progression of states for Markov modeling.

The goal is to determine, for each transient state, the probability that the rally will end in one of the two absorbing states (H_P or V_P, namely point for home team and visitors, respectively) [32, 41]. For example, if starting from a state s , the probability of absorption into H_P is 0.87, it means that the home team has an 87% chance of winning the rally from that situation. The absorption probabilities, $v_H(s)$ and $v_V(s)$, satisfy the system of linear equations:

$$v_H(s) = \sum_{t \in S} p_{st} v_H(t), \quad v_V(s) = \sum_{t \in S} p_{st} v_V(t),$$

where p_{st} are the transition probabilities [32, 41]. In our implementation, the system is solved using the iterative method described [63].

Results and discussion

In this section, we present the results obtained by applying the proposed toolbox to scouting data from the Italian Women's Serie A2 Championship. The validation was conducted on two datasets: (i) the complete 2023–2024 season, comprising all 180 league matches across Groups A and B, and (ii) a partial dataset from the 2024–2025 season, including all matches played by Trentino Volley and Consolini Volley. Using these datasets, we first analyze overall trends and team performance profiles across the league. Subsequently, we focus on a detailed case study of the 2025 Italian Cup final between Consolini Volley and Trentino Volley, in order to highlight the impact of data-driven tactical preparation.

Italian women's serie A2: season 2023–2024

The toolbox was initially validated using the complete 2023–2024 season data from the Italian Women's Serie A2 Championship. We here show the results referring to Group A.

Concerning the visualization module within the implemented toolbox, in Fig. 2 we provide three different visualizations of volleyball serves. The scatterplot (left) displays the landing points of all serves. The heatmap representation (middle) highlights areas of high serve density. Finally, the trajectory plot (right) illustrates the direction and path of each serve highlighting in red the wrong serves. The data correspond to three different sets played by Bartoccini-Fortinfissi Perugia, the team that finished first in Group A at the end of the 2023–2024 season. These visualizations are inspired by those commonly used in Data Volley and are designed to ensure the toolbox offers a level of usability comparable to that of a widely adopted professional tool.

We now show the results obtained by dividing the teams based on their efficiency metrics (Overall Attack, Serve, Reception, Bloc, Defence, FBSO and AST) through machine learning clustering techniques. We set the number of clusters (k) equal to three to reflect the natural segmentation of the league: promotion playoffs (from 1st to 4th), mid-rank (from 5th to 6th), and play-out zone (from 7th to 10th). However, we tested different values for the number of clusters ($k = 2, 3, 4$). In particular, for $k = 2$ the three models tended to merge top and mid-ranked teams, oversimplifying the league structure. By setting $k = 4$, the additional group fragmented the bottom zone without improving interpretability. The best balance in terms of interpretability and validation metrics was obtained for $k = 3$. In particular, the Adjusted Rand Index (K-Means vs Gaussian: 0.8226; K-Means vs Agglomerative: 0.6599; Gaussian vs Agglomerative: 0.6567) demonstrated a substantial level of agreement across clustering methods. The results reported in Table 3 confirm that the three algorithms generally produce consistent groupings, with only minor differences observed.

We focused on Perugia's dynamic progression across clusters: in the first three matchdays, all three clustering methods placed the team in mid-to-low levels.

Table 3 Comparison of cluster assignments ($k = 3$) using K-Means (KM), Gaussian Mixture Model (GMM), and Agglomerative Clustering (AC) for each team in Group A. The number in parentheses indicates the team's final ranking.

Team (Final Rank)	KM	GMM	AC
BARTOCCINI-FORTINFISSI PERUGIA (1)	2	2	2
FUTURA GIOVANI BUSTO ARSIZIO (2)	2	2	2
CDA VOLLEY TALMASSONS FVG (4)	2	2	2
CITTÀ DI MESSINA (3)	1	1	2
TECNOTEAM ALBESE VOLLEY COMO (5)	1	1	1
VTB FCRDIL BOLOGNA (7)	1	1	1
VALSABBINA MILLENIUM BRESCIA (6)	0	0	0
VOLLEY SOVERATO (8)	0	0	0
NUOVOLI ALTAFRATTE PADOVA (9)	0	1	0
SIRDECO VOLLEY PESCARA (10)	0	0	0

The number in parentheses indicates the team's final ranking

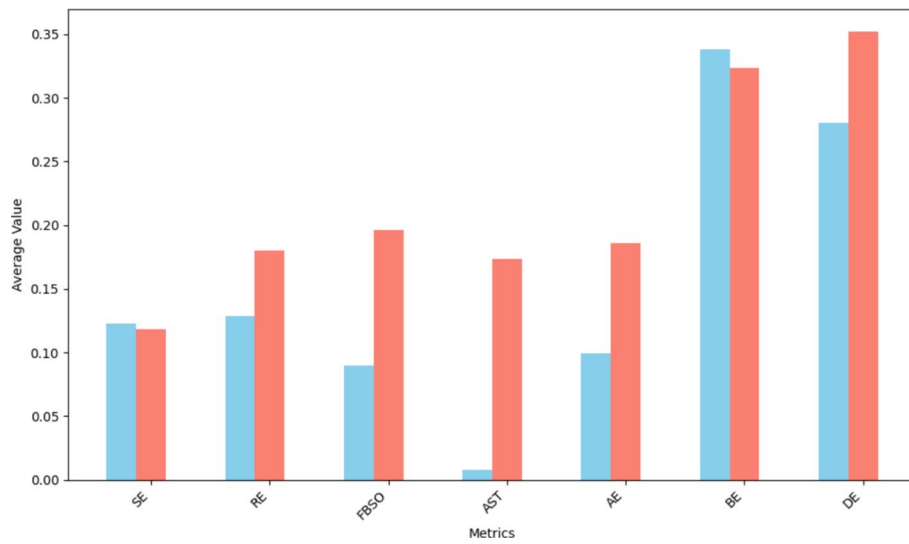


Fig. 3 Comparison of performance indicators, namely Serve Efficiency (SE), Reception Efficiency (RE), First Ball Side-Out (FBSO), Attack After Service Turn (AST), Overall Attack Efficiency (AE), Block Efficiency (BE), and Defence Efficiency (DE), for Bartoccini-Fortinfissi Perugia across two periods of the 2023–2024 season: the first three matchdays (blue) and the remainder of the regular season (red).

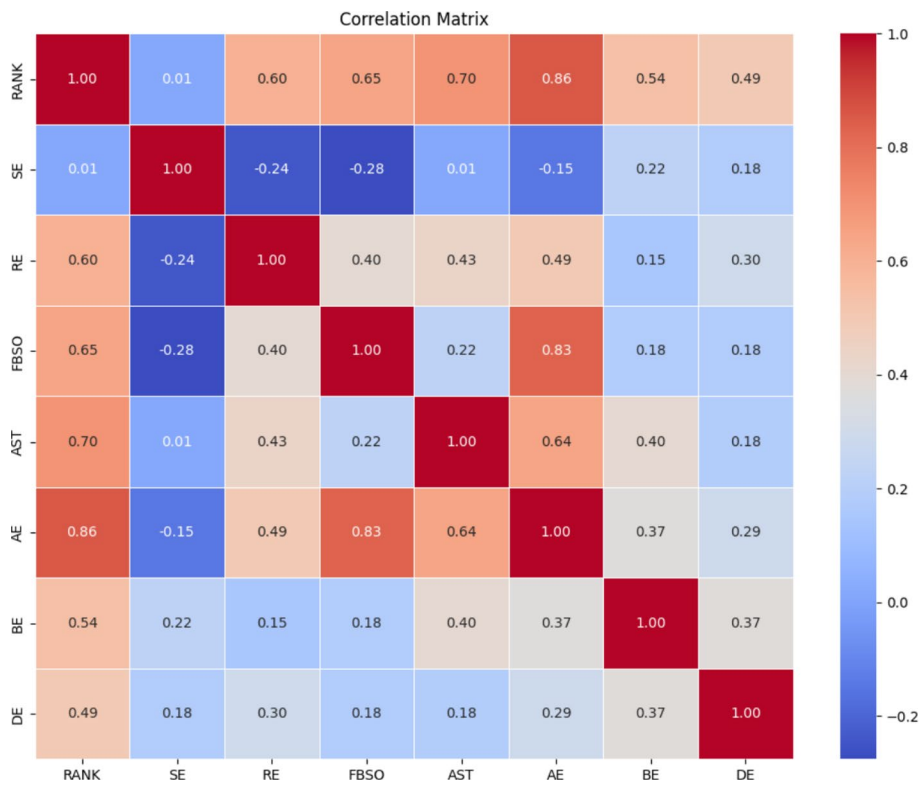


Fig. 4 Correlation matrix between the final aggregated ranking (RANK) and the performance efficiency indicators: serve efficiency (SE), reception efficiency (RE), first-ball side-out (FBSO), attack after service turn (AST), overall attack efficiency (AE), block efficiency (BE), and defense efficiency (DE).

However, from matchday 4 onwards, Perugia consistently ranks in the top tier, driven by a marked improvement in efficiency metrics. This improvement in key indicators is shown in Fig. 3. Notably, the most substantial changes occur in offensive metrics, while performance in blocking and serving remains stable or slightly declines in the latter part of the season. This is coherent with existing literature [21], revealing that attacking performances have a great impact on success.

To better corroborate these findings, the correlation matrix in Fig. 4 highlights the relationship between the final aggregated ranking and the performance efficiency indicators. Specifically, after computing individual rankings using AHP, TOPSIS, PROMETHEE, and MOORA, we derived a consolidated ranking by averaging, for each team, the positions obtained across the four methods. This procedure ensures that the final order reflects a robust consensus rather than the bias of a single technique. Among all variables, AE emerges as the strongest determinant of success ($r = 0.86$), confirming that offensive performance is the main driver of team ranking. AST ($r = 0.70$) and FBSO ($r = 0.65$) also show substantial correlations, emphasizing the importance of converting both serve-serve and post-serve situations into points. RE presents a moderate-to-high correlation ($r = 0.60$), consistent with its role as a prerequisite for effective offensive organization. Conversely, BE ($r = 0.54$) and DE ($r = 0.49$) display only moderate associations with the ranking, indicating that while they contribute to performance, they are less decisive in distinguishing top-ranked teams. Finally, SE exhibits an almost negligible correlation ($r = 0.01$), suggesting that serving effectiveness had limited influence on competitive outcomes in this dataset. Finally, when looking at the relationships between indicators, several noteworthy patterns emerge. AE and FBSO correlate strongly ($r = 0.83$), indicating that teams effective in side-out situations tend also to have high overall attacking efficiency. Similarly, AE and AST correlate positively ($r = 0.64$), reflecting the continuity between attack effectiveness in reception and in transition phases. RE shows moderate associations with AE ($r = 0.49$) and AST ($r = 0.43$), supporting its role as a facilitator for offensive success. In contrast, defensive skills are less interrelated with offensive ones: block efficiency correlates only moderately with AE ($r = 0.37$), while defense efficiency shows weak-to-moderate links with the rest of the indicators ($r \leq 0.30$). Finally, serve efficiency stands out for its lack of meaningful correlation with almost all other metrics, suggesting that serving performance is largely independent of the broader efficiency profile of the teams.

Italian women's serie A2: season 2024–2025

This section presents the outcomes of a case study based on the 2024–2025 season data, carried out in collaboration with the Trentino Volley technical staff, to support match preparation for the Serie A2 Italian Cup final between Trentino Volley and Consolini Volley. The goal was to assess both season-long trends and match-specific dynamics in order to support pre-match preparation and post-match evaluation.

The Trentino Volley technical staff was provided with a report about Consolini Volley's performances during the 2024–2025 season. All the results have been computed by using the implemented toolbox. We here present a small batch of the reported outcomes. The radar chart depicted in Fig. 5 illustrates the overall performance profile of the two Italian

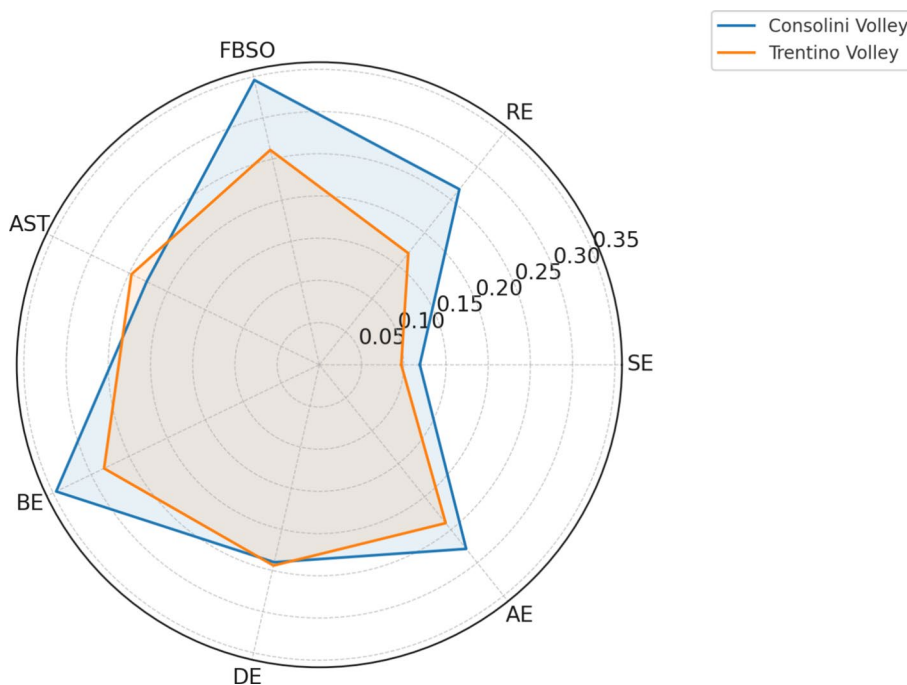


Fig. 5 Radar plot comparing performance efficiency indicators between Consolini Volley (blue) and Trentino Volley (orange). The axes represent serve efficiency (SE), reception efficiency (RE), first-ball side-out (FBSO), attack after service turn (AST), overall attack efficiency (AE), block efficiency (BE), and defense efficiency (DE).

Table 4 Absorption probabilities of selected rally states for Consolini Volley compared to the opponent. Each state corresponds to a specific sequence of reception (R), set (S), and attack (A) actions. The probabilities indicate the likelihood that the rally ends in favor of Consolini Volley versus the opposing team.

State	Consolini volley	Opponent
R+.S+.AXX	0.751689	0.248311
R#.S#.AVV	0.290849	0.709151
R#.S#.AC1	0.291344	0.708656

Each state corresponds to a specific sequence of reception (R), set (S), and attack (A) actions. The probabilities indicate the likelihood that the rally ends in favor of Consolini Volley versus the opposing team

Cup finalists throughout the 2024–2025 season. Consolini Volley demonstrates superior values in most fundamentals, especially in blocking and attacking. This radar chart provided a high-level tactical overview and served as a reference point for evaluating how these trends would manifest in the final match.

Furthermore, the Markov Chain analysis integrated into our toolbox revealed the following outcomes for Consolini Volley: 954 short rallies (fewer than 6 ball touches) won vs 796 lost, 230 medium rallies (6 to 12 ball touches) won vs 180 lost, and 92 long rallies (more than 12 ball touches) won vs 57 lost. In terms of percentages, Consolini Volley won 54.5% of short rallies, 56.1% of medium rallies, and 61.7% of long rallies, demonstrating a growing advantage as rallies extended in length and highlighting the team’s effectiveness in prolonged exchanges.

Analyzing the most frequent scoring patterns for Consolini Volley, it is evident that situations involving the attack code XX (fastball for the outside hitter) are particularly effective. In Table 4, we report some selected rally states along with their corresponding absorption probabilities, computed through the Markov Chain model. Each state represents a specific sequence of actions of Consolini Volley during the rally. Our focus on different attack types confirms that the XX attack leads to a high level of scoring for Consolini Volley. Conversely, attacks with code C1 (middle blocker attack) and VV (pipe attack, i.e. an attack from the middle part of the back line) appear to be less effective in generating points.

In Table 5 we report the performance metrics computed with the data of the Italian Cup final for both teams, along with the delta values from their respective season averages. We observe that Consolini Volley maintained relatively high absolute values, particularly in blocking and attacking, however it is notable that the team experienced a performance decline across all indicators, despite still outperforming Trentino Volley in several metrics. In contrast, Trentino Volley demonstrated a positive deviation in the service category (+0.1856) which, however, did not have a decisive impact on the outcome of the match. This aligns with previous findings [64], which showed that service efficiency does not have a significant impact in women's volleyball. More recent research, however, has demonstrated that it plays a decisive role in men's volleyball [65].

Overall, the results suggest that while Consolini Volley kept a solid performance profile, Trentino Volley's informed approach, guided by the analysis generated using our toolbox, effectively challenged Consolini Volley's game. These findings align with recent theoretical evidence [34] suggesting that informed, data-driven strategies lead to enhanced performance outcomes.

Limitations and future works

This section discusses the limitations of our study and highlights directions for future research. First, the analysis relies on scouting data, where the coding of game actions and the assignment of efficiency scores are inevitably subject to the interpretation of

Table 5 Comparison of match efficiency indicators between Consolini Volley and Trentino Volley in the Serie A2 Italian Cup. For each game action, the absolute efficiency value is reported along with the variation (superscript) relative to the respective 2024-2025 seasonal average. Positive variations are highlighted in *italic*, while negative variations are shown in bold.

Efficiency	Consolini volley	Trentino volley
SERVICE	0.1039 ^{-0.0131}	0.2826 ^{+0.1856}
RECEPTION	0.2222 ^{-0.0578}	0.1978 ^{+0.0288}
ATTACK	0.2478 ^{-0.0262}	0.1575 ^{-0.0825}
BLOCK	0.3567 ^{-0.0223}	0.1591 ^{-0.1239}
DIG	0.2031 ^{-0.0459}	0.2690 ^{+0.0250}
FBSO	0.2034 ^{-0.1296}	0.1767 ^{-0.0843}
AST	0.0470 ^{-0.1840}	0.0000 ^{-0.2470}

For each game action, the absolute efficiency value is reported along with the variation (superscript) relative to the respective 2024–2025 seasonal average. Positive variations are highlighted in *italic*, while negative variations are shown in bold

the scout. Although standardized protocols reduce variability, a degree of subjectivity remains and may introduce bias into the dataset. Future work will focus on reducing subjectivity through the use of automated tracking systems based on artificial intelligence paradigms. In addition, the toolbox has been developed and tested using scouting files obtained by Data Volley, which raises questions of compatibility with other scouting formats. Extending the framework will require adaptations and validation across diverse competitions and datasets.

The dataset analyzed provides valuable insights into the Italian Women's Serie A2 Championship, however its scope is geographically and competitively restricted. Expanding the analysis to include international leagues would allow for comparative assessments across different competitive contexts and playing styles, enhancing the generalizability of the findings.

Moreover, the present work focuses primarily on descriptive and exploratory analyses, without systematically assessing the statistical significance of observed differences. While the toolbox highlights relevant performance patterns, further studies should integrate inferential statistical methods or resampling techniques to rigorously test the robustness of the results.

Finally, in terms of future developments, the proposed toolbox could be extended by integrating advanced machine learning models for predictive purposes (e.g., match outcome prediction or player performance forecasting).

Conclusion

This work presents an open-source Python-based toolbox for advanced volleyball data analysis, offering a flexible and extensible alternative to existing tools like DataVolley. By combining traditional statistical indicators with techniques such as machine learning-based clustering, Multi-Criteria Decision Analysis, and Markov chain modeling, the toolbox enables data-driven interpretation of volleyball match dynamics.

The validation on data from the 2023–2024 and 2024–2025 seasons of the Italian Women's Serie A2 Championship confirmed the reliability and scalability of the proposed approach. The case study on the 2025 Italian Cup final particularly highlights the toolbox's practical value: despite Consolini Volley's superior season averages in several key areas, Trentino Volley leveraged strategic insights generated through the toolbox to challenge their opponent's performance. Indeed, Consolini Volley experienced a measurable drop across all performance indicators in the final match.

These results confirm the potential of accessible, customizable, and reproducible data analytics tools in supporting competitive volleyball coaching and match preparation.

Author Contributions

Conceptualization: A.C., P.C., M.N. Methodology: A.C., P.C., M.N. Software: A.C., P.C., M.N. Validation: A.C., P.C., M.N., D.M. Formal Analysis: A.C., P.C., M.N., D.M. Investigation: A.C., P.C., M.N. Data Curation: A.C., M.N. Writing—Original Draft: A.C., P.C. Writing—Review and Editing: A.C., P.C., M.N., D.M., G.M. Visualization: A.C., P.C. Supervision: P.C., M.N. Project Administration: G.M. Resources: D.M., G.M.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Received: 25 April 2025 Accepted: 7 September 2025

Published online: 24 October 2025

References

1. Morgulev E, Azar OH, Lidor R. Sports analytics and the big-data era. *Int J Data Sci Anal.* 2018;5(4):213–22.
2. Kaur A, Kaur R, Jagdev G. Analyzing and exploring the impact of big data analytics in sports sector. *SN Comput Sci.* 2021;2(3):184.
3. Tsai C-W, Lai C-F, Chao H-C, Vasilakos AV. Big data analytics: a survey. *J Big Data.* 2015;2:1–32.
4. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *J Big Data.* 2015;2:1–21.
5. Silva M, Marcelino R, Lacerda D, João PV. Match analysis in volleyball: a systematic review. *Monten J Sports Sci Med.* 2016;5(1):35.
6. Shah M, Shah A, Patel K, Kshirsagar A, Sanghvi S, Sojitra V. Predictive analytics, strategic game analysis, and injury prevention in sports: the role of big data and artificial intelligence. *Mach Learn Comput Sci Eng.* 2025;1(1):1–25.
7. Sarlis V, Tjortjis C. Sports analytics—evaluation of basketball players and team performance. *Inf Syst.* 2020;93:101562.
8. Sarlis V, Chatziliadis V, Tjortjis C, Mandalidis D. A data science approach analysing the impact of injuries on basketball player and team performance. *Inf Syst.* 2021;99:101750.
9. Rocchetti M, Berveglieri F, Cappiello G. Football data analysis: the predictive power of expected goals. In: *Proceedings of the 25 Th Annual European GAME'ON Conference on Simulation, AI and in Computer Games 2024.*
10. Goes F, Meerhoff L, Bueno M, Rodrigues D, Moura F, Brink M, et al. Unlocking the potential of big data to support tactical performance analysis in professional soccer: a systematic review. *Eur J Sport Sci.* 2021;21(4):481–96.
11. O'Connor J, Alfrey L, Penney D. Rethinking the classification of games and sports in physical education: a response to changes in sport and participation. *Phys Educ Sport Pedagog.* 2024;29(3):315–28.
12. Liu J, Huang G, Hyyppä J, Li J, Gong X, Jiang X. A survey on location and motion tracking technologies, methodologies and applications in precision sports. *Expert Syst Appl.* 2023;229:120492.
13. Gudmundsson J, Horton M. Spatio-temporal analysis of team sports. *ACM Comput Surv (CSUR).* 2017;50(2):1–34.
14. Jud M, Thalmann S. AI in digital sports coaching—a systematic review. *Manag Sport Leisure.* 2025;2025:1–17.
15. Mizels J, Erickson B, Chalmers P. Current state of data and analytics research in baseball. *Curr Rev Musculoskelet Med.* 2022;15(4):283–90.
16. Baumer B, Zimbalist A. *The Sabermetric Revolution: assessing the Growth of Analytics in Baseball.* Philadelphia: University of Pennsylvania Press; 2014.
17. Almujaheed S, Ongor N, Tigmo J, Sagoo N. Sports analytics: designing a volleyball game analysis decision-support tool using big data. In: *2013 IEEE Systems and Information Engineering Design Symposium, 2013.* p. 19–24. IEEE
18. Reeser JC, Bahr R. *Handbook of Sports Medicine and Science, Volleyball.* Hoboken: Wiley; 2017.
19. Quiroga ME, García-Manso JM, Rodríguez-Ruiz D, Sarmiento S, Saa Y, Moreno MP. Relation between in-game role and service characteristics in elite women's volleyball. *J Streng Cond Res.* 2010;24(9):2316–21.
20. Monteiro R, Mesquita I, Marcelino R. Relationship between the set outcome and the dig and attack efficacy in elite male volleyball game. *Int J Perform Anal Sport.* 2009;9(3):294–305.
21. Inkinen V, Häyrynen M, Linnamo V. Technical and tactical analysis of women's volleyball. *Biomed Hum Kinet.* 2013;5(1):43.
22. Rodríguez-Ruiz D, Quiroga ME, Miralles JA, Sarmiento S, Saá Y, García-Manso JM. Study of the technical and tactical variables determining set win or loss in top-level european men's volleyball. *J Q Anal Sports.* 2011;7(1):1.
23. Lobietti R, DI Michele R, Merni F. Relationships between performance parameters and final ranking in professional volleyball. In: *Dancs H, Hughes M, O'Donoghue P. editors. Proceedings of WCPAS 7, 2006.* p. 474–483.
24. Lin H-Y, Lien J-J, Chang Y-N. Machine learning in sports analytics: Volleyball match outcome prediction. In: *2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2024.* p. 1–6. IEEE
25. Harabagiu N. The importance of using the "data volley" software and of the "data video" system in the tactical training of the middle blocker for official games. *Gymn Sci J Edu Sports Health.* 2020;21(1):34–41.
26. Harabagiu N. The methodology of applying the "data volley" programme of statistical analysis within volleyball sports competitions. *Știința Culturii Fizice.* 2020;2(36):153–61.
27. Chambers JM. *Software for Data Analysis: programming with R.* Berlin: Springer. 2008. p. 2.
28. Raschka S, Patterson J, Nolet C. Machine learning in python: main developments and technology trends in data science, machine learning, and artificial intelligence. *Information.* 2020;11(4):193.
29. McKinney W. *Python for Data Analysis: Data Wrangling with Pandas, Numpy, and Jupyter.* " O'Reilly Media, Inc.," 2022.
30. Ishizaka A, Nemery P. *Multi-criteria Decision Analysis: methods and Software.* Hoboken: Wiley; 2013.
31. Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, et al. A review of clustering techniques and developments. *Neurocomputing.* 2017;267:664–81.
32. Ching W-K, Ng MK. Markov chains. *Models Algorithms Appl.* 2006;650:111–39.
33. OpenVolley Project: PyDataVolley—a Python parser for DataVolley files. <https://github.com/openvolley/pydatavolley>. Accessed: 2025-04-25 (2025).
34. Junior NKM. Match analysis for elaborate the volleyball training: a review. *J Sports Games.* 2021;3(1):1–7.

35. Zuccolotto P, Manisera M, Sandri M. Big data analytics for modeling scoring probability in basketball: the effect of shooting under high-pressure conditions. *Int J Sports Sci Coach*. 2018;13(4):569–89.
36. Zuccolotto P, Sandri M, Manisera M. Spatial performance indicators and graphs in basketball. *Soc Indic Res*. 2021;156(2):725–38.
37. Carlesso ML, Cappozzo A, Manisera M, Zuccolotto P. Scoring probability maps in the basketball court with indicator kriging estimation. *Comput Stat*. 2024;40:1–21.
38. Groll A, Ley C, Schauberger G, Eetvelde H. A hybrid random forest to predict soccer matches in international tournaments. *J Quant Anal Sports*. 2019;15(4):271–87.
39. Karlis D, Ntzoufras I. Bayesian modelling of football outcomes: using the skellam's distribution for the goal difference. *IMA J Manag Math*. 2009;20(2):133–45.
40. Eom HJ, Schutz RW. Statistical analyses of volleyball team performance. *Res Q Exerc Sport*. 1992;63(1):11–8.
41. Powers S, Stancil L, Consiglio N. Estimating individual contributions to team success in women's college volleyball. *J Quant Anal Sports*. 2025;21:117–35.
42. Fellingham GW. Evaluating the performance of elite level volleyball players. *J Quant Anal Sports*. 2022;18(1):15–34.
43. Gabrio A. Bayesian hierarchical models for the prediction of volleyball results. *J Appl Stat*. 2021;48(2):301–21.
44. Palao JM. Side-out success and ways that points are obtained in women's college volleyball. *J Sports Anal*. 2018;4(4):243–50.
45. Van Haaren J, Ben Shitrit H, Davis J, Fua P. Analyzing volleyball match data from the 2014 world championships using machine learning techniques. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. p. 627–634
46. Sanghvi D, Deshpande P, Shanbhogue S, Shah V. Analyzing and predicting ncaa volleyball match outcome using machine learning techniques. In: *ICAI Workshops*, 2021. p. 99–116.
47. Liu Q, Liu Q. Prediction of volleyball competition using machine learning and edge intelligence. *Mob Inf Syst*. 2021;2021(1):5595833.
48. Wenninger S, Link D, Lames M. Performance of machine learning models in application to beach volleyball data. *Int J Comput Sci Sport*. 2020;19(1):24–36.
49. Zhao K, Jiang W, Jin X, Xiao X. Artificial intelligence system based on the layout effect of both sides in volleyball matches. *J Intell Fuzzy Syst*. 2021;40(2):3075–84.
50. Jiang L, Yang Z, Gang L. Transformer-based multi-player tracking and skill recognition framework for volleyball analytics. *IEEE Access* 2025.
51. Salim FA, Postma DB, Haider F, Luz S, Beijnum B-JFv, Reidsma D. Enhancing volleyball training: empowering athletes and coaches through advanced sensing and analysis. *Front Sports Act Liv*. 2024;6: 1326807.
52. Leeuw A-W, Zwaard S, Baar R, Knobbe A. Personalized machine learning approach to injury monitoring in elite volleyball players. *Eur J Sport Sci*. 2022;22(4):511–20.
53. Dai X, Li S. Volleyball data analysis system and method based on machine learning. *Wirel Commun Mob Comput*. 2021;2021(1):9943067.
54. Chen H-T, Tsai W-J, Lee S-Y, Yu J-Y. Ball tracking and 3d trajectory approximation with applications to tactics analysis from single-camera volleyball sequences. *Multimed Tools Appl*. 2012;60:641–67.
55. Project O. OpenVolley: open Source Tools for Volleyball Data Analysis. <https://github.com/openvolley>. Accessed: April 2025 2025.
56. Jolly K. *Hands-on Data Visualization with Bokeh: interactive Web Plotting for Python Using Bokeh*. Packt Publishing Ltd, 2018.
57. Jain AK. Data clustering: 50 years beyond k-means. *Pattern Recogn Lett*. 2010;31(8):651–66.
58. Reynolds DA. Gaussian mixture models. *Encycl Biometr*. 2009;741(659–663):3.
59. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *J Classif*. 2014;31:274–95.
60. Darko A, Chan APC, Ameyaw EE, Owusu EK, Pärn E, Edwards DJ. Review of application of analytic hierarchy process (ahp) in construction. *Int J Constr Manag*. 2019;19(5):436–52.
61. Uzun B, Taiwo M, Syidanova A, Uzun Ozsahin D. The technique for order of preference by similarity to ideal solution (topsis). *Appl Multi Criteria Decis Anal Environ Civil Eng*. 2021;2021:25–30.
62. Thakkar JJ, Thakkar JJ. Multi-objective optimization on the basis of ratio analysis method (moora). *Multi Criteria Decis Mak*. 2021;2021:191–198.
63. Ermon S, Gomes C, Sabharwal A, Selman B. Designing fast absorbing markov chains. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2014. p. 28.
64. González-Silva J, Fernández-Echeverría C, Conejero M, Moreno MP. Characteristics of serve, reception and set that determine the setting efficacy in men's volleyball. *Front Psychol*. 2020;11:222.
65. Pawlik D, Bobula G, Mroczek D. The effectiveness and types of serves used in elite women's and men's volleyball in the 2021/2022 season. *Sci Rep*. 2024;14(1):17286.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.