Data in Brief 57 (2024) 111116

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A new dataset for spatial, temporal, and tactical analysis of female handball player movements

Raul Montoliu[a,*], Pere Urbón-Bayes[b], Gabriel Daza[c], Abraham Batalla-Gavalda[d]

[a] Institute of New Imaging Technologies. Universitat Jaume I. Av. Sos Baynat, 12560 Castellon, Spain
[b] Universitat Oberta de Catalunya, Spain
[c] National Institute of Physical Education of Catalonia (INEFC), University of Barcelona (UB), Avda. Estadi, Av. de l'Estadi 12-22, Anella Olímpica E-08038, Barcelona, Spain
[d] EUSES Escola Universitària de la Salut i l'Esport, Rovira i Virgili University, 43870 Tarragona, Spain. Department of Education and Specific Didactics. Universitat Jaume I. Av. Sos Baynat, 12560 Castellon, Spain

ARTICLE INFO

Article history:
Received 26 June 2024
Revised 22 October 2024
Accepted 30 October 2024
Available online 6 November 2024

Dataset link: A New Dataset for Spatial, Temporal and Tactical Analysis of Female Handball Player Movements (Original data)

Keywords:
Sport
Handball
Position
Game analysis
Team sports
Time motion

ABSTRACT

This paper presents a comprehensive dataset detailing the precise indoor positioning of players from a female amateur handball team across 10 real matches. Utilizing Ultra-Wideband (UWB) technology, the dataset captures each player's x and y coordinates every second throughout the games. Additionally, a preliminary game analysis is included, specifying the initiation and termination times of each team's possession. This analysis provides a variety of labels crucial for training machine learning algorithms, encompassing distinctions such as attack or defense, structured or unstructured play, goal outcomes, and the differentiation between counter-attacks and static phases. In total, the dataset comprises 84691 positioning measurements, offering a valuable resource for in-depth study and analysis of player dynamics and game strategies in female handball.

* Corresponding author
E-mail address: montoliu@uji.es (R. Montoliu).

https://doi.org/10.1016/j.dib.2024.111116
2352-3409/

Specifications Table

Subject	Sports Sciences, Therapy, and Medicine
Specific subject area	Indoor positioning data for female amateur handball players using UWB technology, analyzed for game dynamics and strategy insights.
Type of data	Analyzed, Filtered (Note: Raw data was analyzed and filtered to obtain player's indoor positions)
Data collection	Players wore sensors using Ultra-Wideband (UWB) technology to obtain real-time indoor positions. The dataset reports X, and Y coordinates per second for each player across 10 matches. Additionally, it includes possession start and end times, and detailed events during each possession.
Data source location	Data was collected in several Handball courts in Catalonia, Spain. Data is stored at the Institute of New Imaging Technologies, Castellon, Spain and fully accessible by Zenodo.
Data accessibility	Repository name: Zenodo Data identification number: https://doi.org/10.5281/zenodo.12065335 Direct URL to data: https://zenodo.org/records/12065335 Instructions for accessing these data: The files are in csv and excel formats. They can easily and anonymized accessed by Zenodo repository.

1. Value of the Data

- The dataset includes X and Y positioning, analyzed second by second, with temporal and action values on the court.
- Knowing the positioning second by second allows for understanding the behavior and movement patterns of different players in the game.
- This dataset can be used by other researchers to compare movement patterns by positions or even action-reaction situations between players.
- It can serve as a reference for researchers aiming to analyze matches through space occupation or those interested in tactical analysis.
- This dataset can provide methodological assistance for data processing to obtain spatial parameters in the offensive and defensive behavior of women's handball.

2. Background

Handball is a dynamic team sport played on an indoor court with goals at each end, known for its fast pace, physicality, and exciting gameplay. The objective is to throw a ball into the opponent's goal while preventing them from doing the same. Handball teams consist of seven players each, including a goalkeeper. The game involves quick passes, dribbling, and strategic positioning to create scoring opportunities.

Although handball is less popular compared to sports like European football or basketball, it has attracted increasing attention from researchers in recent years. Wagner et al. [1] conducted a bibliographic review analyzing the specific characteristics of handball teams, suggesting that changes in intensity, specific player skills, hand-to-hand confrontations, positive mental abilities, and social factors are key analyzable aspects. These aspects encompass the coordination, endurance, strength, and cognition of the athletes. Supporting this, Przednowek et al. [2] analyzed handball players compared to a normal population control group in psychomotor ability tests, demonstrating that handball players had better reaction times and movement patterns than the control group subjects.

The physiological demands of handball were further explored in a systematic review by García-Sánchez et al. [3], which examined the performance of 1175 elite handball players (1042 men and 133 women). The review detailed metrics such as the distance traveled and throwing actions, distinguishing between genders and positions. Focusing on women's handball, a significant study by Kniubaite et al. [4] analyzed 14 competitive matches of an elite women's team, comparing the internal and external load of national and international competitions. This study found differences in various load variables and established reference values, concluding that elite players face high physical and physiological demands during competition.

Machine learning and data analysis have also been applied to women's handball. Manchado et al. [6] examined the horizontal movement and sprint acceleration profiles of elite female players, along with heart rates during matches. They concluded that acceleration is influenced by aerobic performance and player position, suggesting that sprint and endurance training should be tailored to positional demands. Caporal et al. [7] identified key parameters that differentiate high-level from scholar-level 13–14-year-old Brazilian female players, proposing a mathematical model to identify handball talent. Oytin et al. [8] assessed various machine learning models to predict athletic performance, finding that a radial-basis function neural network was most effective for predicting outcomes in countermovement jumps, sprints, shuttle runs, and agility tests. Van Der Tillaar et al. [9] used an inertial measurement unit (IMU) worn on the wrist and machine learning techniques to classify different types of handball throws and predict ball velocity, achieving an accuracy of 80–87% when identifying approach and throw types.

Unlike more popular sports, such as soccer or basketball, there are very few comprehensive databases on handball. This dataset has been created to fill this gap. A noteworthy dataset on handball is [5], where the authors obtained similar data to ours, but from video recordings, unlike our dataset, which uses UWB sensors. Our dataset also includes detailed information about team possessions and events during those possessions. It primarily includes positional data that can be used for tactical analysis, providing insights into team strategies, formations, and patterns of play. Coaches and analysts can identify effective offensive and defensive tactics, such as player positioning during set pieces, transitions between attack and defense, and defensive structures. Additionally, the dataset can be used for player performance evaluation, allowing coaches to assess individual performance based on positioning, movement efficiency, game involvement, and adherence to tactical instructions. This data helps identify player strengths and weaknesses, informing coaching decisions such as substitutions or adjustments in player roles.

3. Data Description

The dataset consists of three different *csv* (comma-separated values) files. We have also included the same data using *Microsoft Excel* files (i.e. in format *xlsx*) since, for some users, this file type can be more convenient.

The first file is called **positions.csv** (and **positions.xlsx**) and comprises the position data of the players across all games. Each line of the file has the following information:

- **game**: the unique identifier of the game (from 1 to 10).
- **player**: the player's unique identifier in the field during the possession (from 0 to 5).
- **period:** 1 for the first period and 2 for the second one.
- **time**: an integer number showing the time (in seconds) when the position data was recorded. A 1 in this field represents the first second of the first period if the field period has been set to 1, or the first second of the second period if the field period has been set to 2.
- **x**: a float number showing the x coordinate in meters. The x-axis corresponds to the sideline of the handball court.
- **y**: a float number showing the y coordinate in meters. The y-axis corresponds to the baseline of the handball court.

The edge of the coordinates is located at the bottom left corner of the court. Table 1 shows 10 rows of this file as an example. In total, the positions file has 84691 rows.

The player identifier (player field) is unique only within a specific possession. Once possession changes, the same player may be assigned a different ID, meaning a player can have

Table 1
10 lines of the positions file.

game	player	period	time	x	y
1	player_0	1	481	26.061	7.3981
1	player_1	1	481	28.411	14.150
1	player_2	1	481	31.424	15.972
1	player_3	1	481	25.320	10.710
1	player_4	1	481	27.631	5.927
1	player_5	1	481	24.256	4.774
1	player_0	1	482	23.569	9.4016
1	player_1	1	482	25.494	14.845
1	player_2	1	482	28.993	16.965
1	player_3	1	482	22.291	10.041

different IDs across multiple possessions in the dataset. The focus of the analysis is on tactical movements during each possession rather than tracking individual performance across the entire dataset. Since only the six on-court players are considered (excluding goalkeepers), the Player ID ranges from 0 to 5.

The proposed dataset also included the game segmented in possessions and a set of labels explaining what happened in those possessions. Possession is defined as the period from when the team gets the ball until the moment when it loses it. The format of the file **possessions.csv** (and **possessions.xlsx**) is as follows:

- **game:** the unique identifier of the game (from 1 to 10).
- **possesion_id**: The unique identifier of the possession.
- **period**: 1 or 2.
- **time_start**: absolute in-game time the segment starts in seconds.
- **time_end**: absolute in-game time the segment ends in seconds.
- **game_phases:** AT for an attack, DEF for a defense.
- **score_team_a:** score of the team wearing the sensors (team A).
- **score_team_b:** score of the opposite team (team B).
- **in_game_players_team_a:** number of players on the court for team A.
- **in_game_players_team_b:** number of players on the court for team B.
- **passive_alert:** 1 if during the possession a passive alert has been raised by the referee, 0 otherwise.
- **number_of_passes:** number of passes executed after the passive alert was raised. According to Handball rules, when the referee raises the passive alert, the team has a maximum of 4 passes. If the team makes more than 4, it loses the ball.
- **passive_alert_result:** the result of the passive warning situation. 1 when the possession finishes in goal, 0 otherwise.
- **tactical_situation:** label indicating the current segment tactical situation, either *equal* (both teams have the same number of players on the court), *superiority* (the team studied has more players), or *inferiority* (the team studied has less players).
- **time_in_seconds:** duration in seconds for the segment.
- **fk_count:** number of free kicks during the segment. A free kick is a technical fault decided by the referee.
- **pen_count:** number of penalties (7-meter throw), either one or zero.
- **tm_count:** number of times outs during the segment, ideally one or zero, but could be two in some rare cases.
- **prev_tm:** zero if this is the first time out of the possession, one otherwise. For this dataset, this is always zero.
- **post_tm:** one if there are more timeouts in the possession, zero otherwise. For this dataset, this is always zero.
- **sequences:** number of sequences of the possession.

- **possesion_result:** possession success result, *True* if in case of an AT possession, a goal was achieved, *False* otherwise. *True* in DEF when a goal was prevented, *False* otherwise.
- **offense_type:** Type of AT in play, either *counter-attack* or *normal* for a static attack**.**
- **misses:** number of missed passes during an AT segment.
- **throws:** number of throws during an AT segment.
- **goal:** number of goals during an AT segment, zero or one.
- **throw_zone:** Throw zone from the last shot in the segment.
- **organized_game:** *True* if the play during the segment was considered to be organized, *False* otherwise.

Table 2 displays the data for a single row of this file. Specifically, it details the possession with ID 10A, which begins at second 481 of the first game and first period, ending at second 522. This was an attacking possession when the score was 7 to 5. The possession concluded with a goal and is classified as an organized attack.

Table 2
A row from the possessions.csv file corresponding to the possession with ID 10A.

game	1	time_in_seconds	34
period	1	fk_count	1
possession_id	10A	pen_count	0
time_start	481	tm_count	0
time_end	522	prev_tm	0
game_phases	AT	post_tm	0
score_team_a	7	sequences	2
score_team_b	5	possession_result	TRUE
in_game_players_team_a	7	offense_type	normal
in_game_players_team_b	7	misses	1
passive_alert	0	throws	1
number_of_passes		goal	1
passive_alert_result		throw_zone	left_in
tactical_situation	equal	organized_game	yes

In handball, possessions can be segmented into short periods called sequences. For instance, a team's attack may begin with a fast transition as they attempt to score quickly. However, this effort might be thwarted by effective defense. Following this, there may be a phase of slower play, during which the attacking team adjusts to position the most suitable player for an attack. This sequence often culminates in a shift to a fast attack as the players aim to score. In this case, the possession can be divided into three sequences. The file **sequences.csv** (or **sequences.xlsx**) details the internal segmentation of each possession, using the same format as the possession file but with an added field for the sequence ID.

Python source code is available to plot team trajectories in each possession. Fig. 1 shows four examples of the plots that can be generated using the provided source code. The first row of Fig. 1 illustrates two attacking possessions (IDs 11A and 15A), while the second row shows two defensive possessions (IDs 11D and 17D). All four possessions are from game 2, period 1.

Fig. 2 displays the file and folder structure of the dataset. The **csv** folder contains the CSV files, and the **excel** folder contains the equivalent files in XLSX format. The **src** folder stores the Python source code, where the **main.py** script is used to generate the plots. The dataset also includes two important files in the root directory: **readme.txt**, which provides instructions for running the code, and **handball.yml**, which is necessary for correctly installing the dependencies needed to run the software.

4. Experimental Design, Materials and Methods

The study sample consisted of 14 amateur female handball players belonging to the maximum regional category of Catalunya (*Liga Catalana Sénior Femenina*). The average age of the

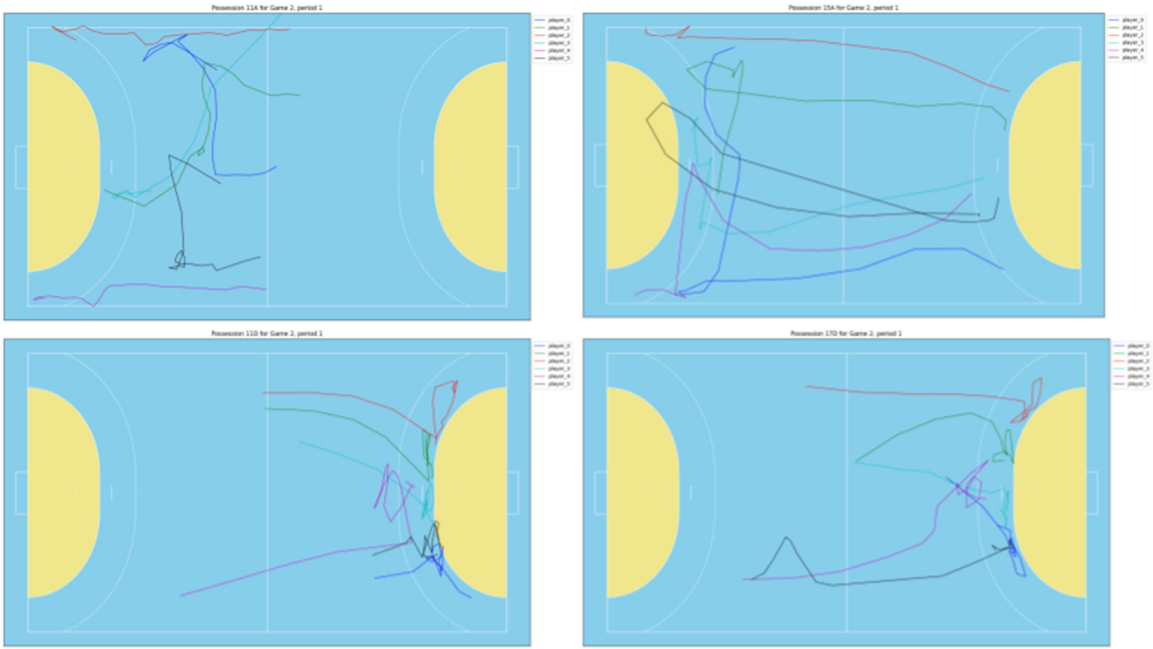

Fig. 1. Four examples of the plots that can be obtained from the included source code. The first row shows the players' trajectories of possessions with ID 11A and 15A. The second row shows the ones for the possessions with ID 11D and 17D.

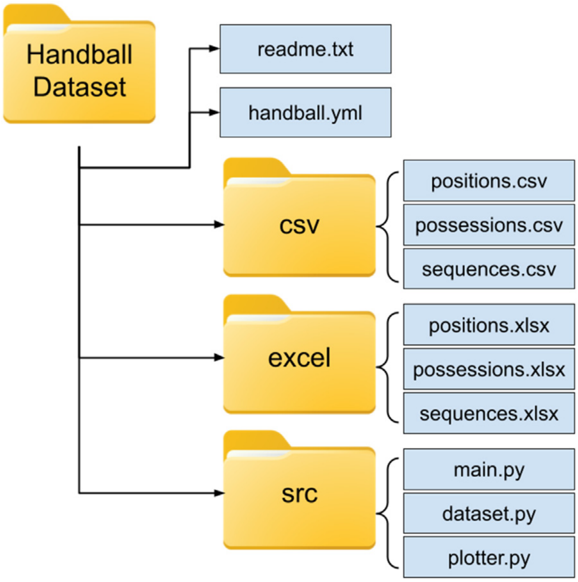

Fig. 2. A scheme of the files and folders included with the proposed dataset.

players is 21.99±3.41 years old. The data was collected in 10 consecutive official matches. The players had four 2-hour weekly training sessions and one match at the weekend. In addition, during the study, special care was taken to ensure that participants did not engage in moderate or high-intensity physical activity apart from the assessment or training sessions. The participants had no personal or family history of cardiac pathology, nor had they suffered any injury that could alter regular sports practice in the 6 months before the study.

During competitive matches, the kinematic demands of each player were monitored using a WIMU Pro™device (RealTrack Systems, Almería, Spain), with Ultra-Wideband (UWB) technology to estimate the indoor positioning of the players. The devices were placed on the players' tops just before the start of the warm-up. At the same time, the matches were recorded with two video cameras (Panasonic -HC-V380EG-K Camcorder 2.51 MP Full HD, Kadoma, Osaka, Japan), which were synchronized by acoustic and visual signals with the UWB devices. The video cameras were placed in an elevated position from the stands, which allowed at least half of the track to be recorded without having to be moved (Fig. 3).

Subsequently, the information recorded by the devices was downloaded using the SPRO™ v.990 software (RealTrack Systems, Almería, Spain). Once WIMU files were downloaded, a set of Python scripts was used to transform the original data into the file format presented in this paper.

Possession time starts and ends were obtained by the authors of the paper by watching the videos of the games. The rest of the fields were filled with the support of a professional handball coach.

As the primary goal of the article is to present the structure, content, and potential applications of the dataset, it does not include specific analyses or hypothesis testing where a power calculation would typically be relevant. Since the dataset includes information on ball possessions, it is feasible to train a machine learning algorithm to distinguish between structured and unstructured attacks. There are a total of 730 possessions, and if we were to conduct a specific analysis focused on differentiating between these types of attacks, a power calculation could indeed be considered in that context. For example, assuming a small effect size (odds ratio = 1.5), an alpha level of 0.05, and a desired power of 0.80, the required total sample size would be

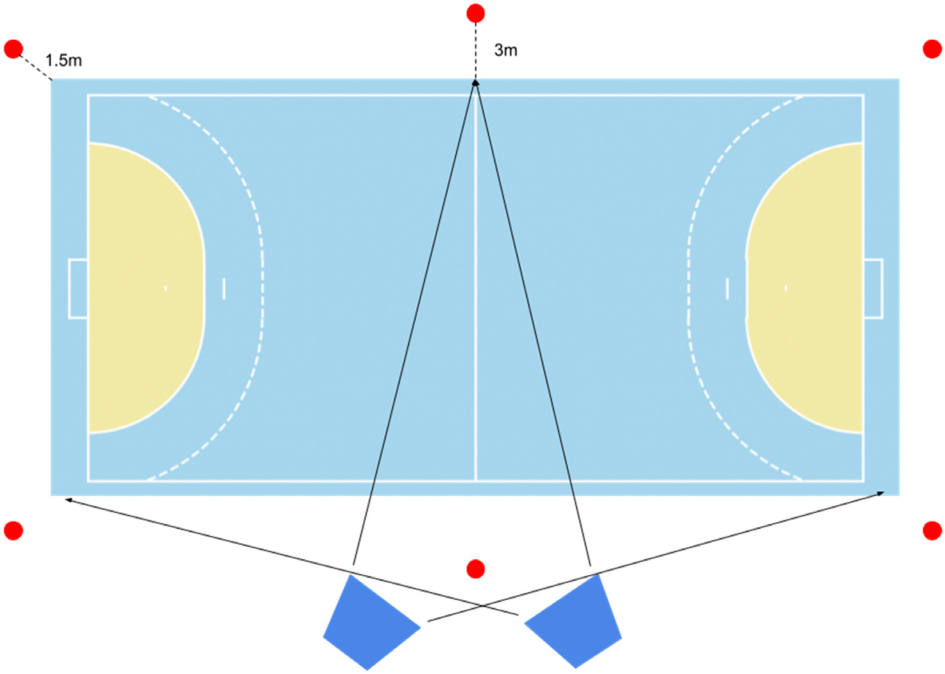

Fig. 3. A schematic representation of the sensors used to capture the data. Red circles indicate the approximate position of the 6 UWB antennas needed to trilaterate the position of the players wearing the UWB tag. Two video cameras were used to record the matches.[1]

308. If we increase the desired power to 0.95, the required sample size would increase to 503. However, with our current total sample size of 730 and an effect size of 1.5, we achieve a power of 0.979. These calculations were performed using the G*Power 2.1.9.7 software.

Limitations

Since the dataset was obtained from real situations, some aspects were out of our control. For instance, not always the 6 players on the court wore the positioning sensor due to sensor failures most of the time. Such possessions where not all the players wore the localization sensors have been discarded.

Goalkeepers were not included in this study.

The UWB antennas were consistently placed in approximately the same position for each game. Therefore, a localization coordinate (x, y) obtained during one match would be slightly shifted to (x+e, y+e) in subsequent matches, where "e" represents a small number (less than 30 cm). This means that while the antennas are situated nearly identically in each game, there exists a minor variation in the measured location, which is less than 30 cm.

Ethics Statement

None of the participants received any financial or in-kind reward for their collaboration in the study. They also signed an informed consent form, and a protocol was established for the delivery and explanation of the results. The study was approved by the ethics committee of the Consell Català de l'Esport (024-CEICGC-2022). This study was conducted taking into account the principles of the Declaration of Helsinki for research involving human subjects (Declaración de Helsinki de la AMM, 2013), as well as the criteria established in the Biomedical Research Act (Ley 14/2007, de 3 de julio, de Investigación biomédica., 2007) and under the Spanish Data Protection Act (Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales., 2018).

Credit author statement

Raúl Montoliu: Conceptualization, Software, Writing - Original Draft, Visualization, Methodology. **Pere Urbón:** Software, Formal analysis, Data Curation, Writing - Review & Editing, Visualization, Conceptualization. **Abraham Batalla:** Conceptualization, Resources, Writing - Review & Editing, Methodology. **Gabriel Daza:** Validation, Writing - Review & Editing, Conceptualization.

Data availability

A New Dataset for Spatial, Temporal and Tactical Analysis of Female Handball Player Movements (Original data) (Zenodo)

Acknowledgements

This work was supported in part by the Agencia Estatal de Investigación (AEI)/10.13039/501100011033/Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea (UE), under Grant PID2021-1226420B-C44. We also thank the players wearing the sensors for their help.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Wagner, T. Finkenzeller, S. Würth, S.P. von Duvillard, Individual and team performance in team-handball: a review, J. Sports Sci. Med. 13 (4) (2014) 808–816.
[2] K. Przednowek, M. Śliż, J. Lenik, B. Dziadek, S. Cieszkowski, P. Lenik, D. Kopeć, K. Wardak, K.H. Przednowek, Psychomotor abilities of professional handball players, Int. J. Environ. Res. Public Health 16 (11) (2019) 1909, doi:10.3390/ijerph16111909.
[3] C. García-Sánchez, R.M. Navarro, C. Karcher, A. de la Rubia, Physical demands during official competitions in elite handball: a systematic review, Int. J. Environ. Res. Public Health 20 (4) (2023) 3353, doi:10.3390/ijerph20043353.
[4] A. Kniubaite, A. Skarbalius, F.M. Clemente, D. Conte, Quantification of external and internal match loads in elite female team handball, Biol. Sport 36 (4) (2019) 311–316, doi:10.5114/biolsport.2019.88753.
[5] Omar A. Mures, Javier Taibo, Emilio J. Padrón, Jose A. Iglesias-Guitian, A comprehensive handball dynamics dataset for game situation classification, Data Brief 52 (2024), doi:10.1016/j.dib.2023.109848.
[6] C. Manchado, J. Pers, F. Navarro, A. Han, E. Sung, P. Platen, Time-motion analysis in women's team handball: Importance of aerobic performance, J. Human Sport Exercise 8 (2) (2013) 376–390, doi:10.4100/jhse.2012.82.06.
[7] G. Caporal, A.C.A. Gaya, A. Pedretti, J.B. Mello, School–Age talent identification in female handball by a mathematical model: equation proposed from anthropometry, maturity offset, fitness and technical field-tests, Res. Q. Exercise Sport (2024) 1–9, doi:10.1080/02701367.2024.2358962.
[8] M. Oytun, C. Tinazci, B. Sekeroglu, C. Acikada, H.U. Yavuz, Performance prediction and evaluation in female handball players using machine learning models, IEEE Access 8 (2020) 116321–116335, doi:10.1109/ACCESS.2020.3004182.
[9] R. van den Tillaar, S. Bhandurge, T. Stewart, Can machine learning with IMUs be used to detect different throws and estimate ball velocity in team handball? Sensors 21 (2021) 2288, doi:10.3390/s21072288.