frontiers | Frontiers in Sports and Active Living

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fspor.2023.1112739

OPEN ACCESS

EDITED BY
Sina David,
VU Amsterdam, Netherlands

REVIEWED BY
Silvio Rene Lorenzetti,
Swiss Federal Institute of Sport Magglingen, Switzerland,
Basilio Pueo,
University of Alicante, Spain

*CORRESPONDENCE
Valentina Camomilla,
✉ valentina.camomilla@uniroma4.it

SPECIALTY SECTION
This article was submitted to Sports Science, Technology and Engineering, a section of the journal Frontiers in Sports and Active Living

RECEIVED 30 November 2022
ACCEPTED 20 January 2023
PUBLISHED 09 February 2023

CITATION
Mascia G, De Lazzari B and Camomilla V (2023) Machine learning aided jump height estimate democratization through smartphone measures.
Front. Sports Act. Living 5:1112739.
doi: 10.3389/fspor.2023.1112739

Machine learning aided jump height estimate democratization through smartphone measures

Guido Mascia[1,2], Beatrice De Lazzari[1,2] and Valentina Camomilla[1,2]*

[1]Department of Movement, Human and Health Science, University of Rome "Foro Italico", Rome, Italy, [2]Interuniversity Centre of Bioengineering of the Human Neuromusculoskeletal System, University of Rome "Foro Italico", Rome, Italy

Introduction: The peak height reached in a countermovement jump is a well established performance parameter. Its estimate is often entrusted to force platforms or body-worn inertial sensors. To date, smartphones may possibly be used as an alternative for estimating jump height, since they natively embed inertial sensors.
Methods: For this purpose, 43 participants performed 4 countermovement jumps (172 in total) on two force platforms (gold standard). While jumping, participants held a smartphone in their hands, whose inertial sensor measures were recorded. After peak height was computed for both instrumentations, twenty-nine features were extracted, related to jump biomechanics and to signal time-frequency characteristics, as potential descriptors of soft tissues or involuntary arm swing artifacts. A training set (129 jumps – 75%) was created by randomly selecting elements from the initial dataset, the remaining ones being assigned to the test set (43 jumps – 25%). On the training set only, a Lasso regularization was applied to reduce the number of features, avoiding possible multicollinearity. A multi-layer perceptron with one hidden layer was trained for estimating the jump height from the reduced feature set. Hyperparameters optimization was performed on the multi-layer perceptron using a grid search approach with 5-fold cross validation. The best model was chosen according to the minimum negative mean absolute error.
Results: The multi-layer perceptron greatly improved the accuracy (4 cm) and precision (4 cm) of the estimates on the test set with respect to the raw smartphone measures estimates (18 and 16 cm, respectively). Permutation feature importance was performed on the trained model in order to establish the influence that each feature had on the outcome. The peak acceleration and the braking phase duration resulted the most influential features in the final model. Despite not being accurate enough, the height computed through raw smartphone measures was still among the most influential features.
Discussion: The study, implementing a smartphone-based method for jump height estimates, paves the way to method release to a broader audience, pursuing a democratization attempt.

KEYWORDS
neural network, accelerometer, gyroscope, strength and conditioning, modal analysis

Introduction

The human lower-body neuromuscular ability of generating power, its neuromuscular capacity, readiness, and fatigue can be identified through a vertical jump test (1,2). One of the most commonly used tests for such an analysis is the countermovement jump (CMJ) (1,3). The choice of the CMJ test is supported by its easy familiarization. Furthermore, the CMJ allows the extraction of a plethora of information about the ability of an individual to execute a stretch-shortening cycle (1,3,4). Such a mechanism, occurring at the muscle-tendon level, consists of an eccentric contraction followed by a concentric one of the same muscle group.

The elastic enhancement produced throughout the eccentric phase allows to augment the force produced during the concentric one (4). Among the many possible performance parameters biomechanically linked with the jump (5), the most widely analyzed for the CMJ is the reached peak height (1,6).

As a common practice for such an evaluation, force platforms (FPs) and motion capture systems represent the gold standard equipment to estimate the jumper center of mass (CoM) kinematics. The former exploits the ground reaction force (7), whereas the latter directly tracks the CoM displacement (8). Nonetheless, they are generally costly and cumbersome, with limited portability in an outdoor environment, eventually constraining measures to laboratory settings only. In the last few years, inertial measurement units (IMUs) proved to be a valuable alternative to FPs for jump height estimate (9–13). An IMU is composed of two triaxial sensors, namely an accelerometer and a gyroscope, measuring the net force acting on the sensors and their rate of change in angular velocity, respectively (14). The CoM behavior during a jump can be reliably tracked (15) using IMU-measured vertical acceleration by means of an appropriate sensor attachment and position (16), typically a belt worn at pelvis level, and mathematical manipulation (11). The rapid rise and the consequent cost reduction of smartphone devices (SP), which natively embed IMUs, suggests that such an analysis may be accomplished using them. Indeed, SPs are devices within everyone's reach, which can be thought to be an even more low-cost alternative if compared to commercially available IMUs systems (17). Smartphones would allow democratizing the access to vertical jump tests to every user owning one, especially if performed with no additional equipment (e.g., a belt or pocket holding, etc). Nevertheless, such embedded sensors were not developed for biomechanical analysis, and they do not necessarily present some required specifications, such as high sampling frequency or appropriate full scale range (17).

Smartphone IMUs have already been used to characterize jump activity: to detect it among other activities (18), to find possible correlations between jump mat variables and kinematic features in both CMJ and SJ (19), and to analyze drop jumps (20,21). Estimating jump height using SP-embedded IMU direct measures has not been attempted, to the best of our knowledge, aside from preliminary investigations forerunner of the current one (22,23). SP-focused studies were all based on video camera-based approaches (24,25). Among the alternatives available on the market, the most commonly used SP application is MyJump, whose reliability has been tested in several studies (26–32). Nonetheless, it uses a method relying on specific assumptions about the symmetry of the task execution, and it computes jump height using flight time, entailing the identification of take-off and landing instants of time. Developing jump height estimates through an SP-embedded IMUs, besides coping with low performance embedded sensors, requires framing the role of three factors and possibly compensate for their negative consequences: signal-processing procedures, sensor position, and presence of signal artifacts due to soft tissues (17,33).

Vertical jump height can be computed using three main *signal-processing procedures*: take-off velocity method (TOV), flight time method (FT), and direct tracking of CoM trajectory using motion capture systems (34). When using IMU data, only TOV and FT can be used. Evidence shows that TOV outperforms FT in terms of both biomechanical accuracy (35–38) and reliability (1,7,36,37). Indeed, FT involves the identification of two instants of time (take-off and landing), whether TOV requires only one (take-off). Moreover, FT is dependent on the hypothesis that the peak height is reached at exactly half the flight-time: this is not the case if the jumper naturally folds their legs, thus extending the flight-time, leading to a height overestimation (35–38). In the context of a self-evaluation with no external supervision, this issue could be critical. As regards experimental errors, estimating velocity from IMUs through numerical integration is usually affected by sensor errors. However, since CMJ entails a short integration interval ($\simeq$1.5 s from the movement onset to the take-off instant - (5)), this error can be considered negligible (12,39,40). Hypothesizing errors due to velocity estimation comparable or lower than those due to time events identification, preference is given to the TOV method, rather than the TF, since it: (i) involves the identification of a single time instant; (ii) is not affected by the jumper postural configuration.

The impact on jump height estimation accuracy of *sensor position* depends on several factors: first, measuring the acceleration at a single point close to the ideal CoM location (L3-L5 vertebrae) neglects the CoM displacement associated with the relative movements of body masses (41) (**Figure 1**). Second, if the sensing element is far from this point, e.g. attached to the sternum, the acceleration shape recorded during a jump is further different (**Figure 1**) (15). Nevertheless, the cumbersome nature of SPs makes their placement in proximity to the ideal human CoM position unsuitable. Holding the SP within the hands constitutes an ecological alternative, but may further increase this discrepancy. Achieving comparable results to the ones obtained through belt-worn IMUs (mean difference

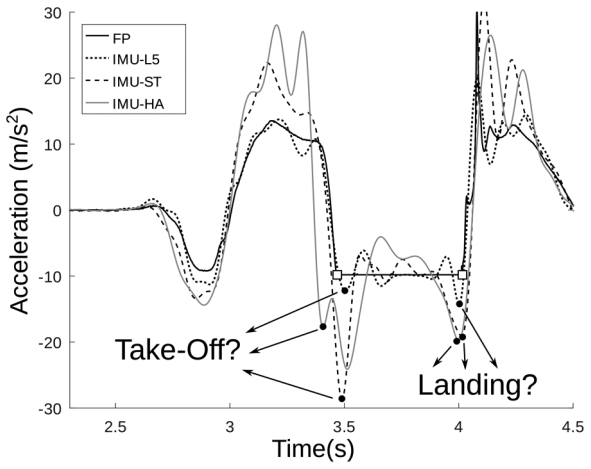

FIGURE 1
Example of vertical acceleration during CMJ, measured at different positions and instrumentations. Legend: Force Platform gold standard (continuous line, FP); IMU at L3-L5 vertebrae (dotted line, IMU-L5); IMU attached close to the sternum (dashed line, IMU-ST); IMU placed within the hands (gray continuous line, IMU-HA) resembling the configuration used in the study; the square markers identify the actual take-off and landing instants on the FP signal. IMU signals were low-pass filtered (Butterworth, 6th order, $f_c = 10$ Hz). It is possible to appreciate the wide variability of the key time instants (take-off and landing) depending on the sensor position. The depicted data were acquired through Vicon Nexus software using Vicon Blue-Trident IMUs (Vicon, Oxford, United Kingdom; sampling frequency: 1,000 sample/s; full scale range: accelerometer = $\pm$16 g, gyroscope = $\pm$2,000 deg/s).

≃5 cm - (10)) calls for developing ad hoc methods that compensate for the above-mentioned discrepancies.

Soft tissue artifacts, considered the most detrimental factor in human movement analysis (33), are of no second importance also for IMU measures (42,43). This is particularly true for high intensity movements, which presumably enhance wobbling due to the inertial whiplash of the soft tissues and of the attached sensor, depending on the sensor fixing (14,16). Although wobbling has been claimed as a secondary artifact cause, being accounted for a relatively small part of the total mean power (on average, 13% for 3 m/s running (44)), it can still be assumed to have an important role in disrupting the consistent estimate of both take-off and landing time instants (**Figure 1**), further reinforcing the choice of using TOV as preferential approach. Moreover, the disruptive effects brought to the signal waveform are potentially overlapped to the motor task in the frequency domain (44). As a promising alternative to characterize such effects, the use of modal analysis in the form of variational mode decomposition (VMD) could be beneficial. VMD is a numerical method able at discriminating a selected number of intrinsic mode functions, each centered around a specific central frequency (45). Such a technique could help in highlighting the oscillatory time-frequency traits of the analyzed signals, discriminating contributions attributable to different wobbling elements.

This study aims to estimate jump height using SP-embedded IMU measures. To cope with given uncertainties related to low performance sensors, sensor position, and soft tissues artifacts, jump height is predicted on the basis of selected information extracted from these measures. Features must be chosen as appropriate for the context they are describing: those biomechanically linked with the jump (5) would be reasonable candidates for obtaining a valid estimate; but it is also essential to embed features that can be predictive, and thus possibly compensate for, the wobbling oscillations mentioned above. This is in line with recent trends having data science emerging as a discipline capable of supporting findings related to sports-related issues through the use of automated methods (46–48).

A feature set was selected to estimate jump height through supervised learning. Gold standard height obtained via FP data and TOV method was used as reference outcome. Possible multicollinearity between the selected features was eventually tackled by using Lasso regularization. Considering only the reduced feature set, a machine learning approach was alternatively investigated training a multilayer perceptron neural network (MLP) through hyperparameters tuning performed via grid search. The influence that each feature exerted over the outcome was also evaluated using the permutation feature importance technique (PFI) (49).

Materials and methods

Experimental setup

Forty-three healthy sports science students were recruited (27M, 16F; mean ± SD: age = 25.9 ± 3.7 years; stature = 171 ± 10 cm; mass = 67.5 ± 10.9 kg). All participants signed an informed consent prior to the experimental session. Only physically active individuals were included, whereas individuals which underwent either lower limb surgery or injury in the six months prior to the experimental session were excluded from the study. The study was approved by the Internal Review Board of the University of Rome "Foro Italico" (No. CAR 94/2021). Participants held an SP in their hands (**Figure 2**) (Xiaomi Redmi 9T, Beijing Xiaomi Technology Company Limited, Beijing, China; sampling frequency = 128 samples/s; full scale range: accelerometer = ±8 g; gyroscope = ±360 deg/s). All SP-IMU data were collected using the app Phyphox (50), which was controlled remotely through the laboratory PC. Calibration tests were performed before each experimental session. These operations are essential to re-align the vertical acceleration signal with the world reference frame through sensor fusion algorithm, detailed in the "Data processing" section. Afterwards, each participant was instructed on how to properly perform one round of the experimental task according to the recommendations in (1). Then, they performed 4 CMJs onto two FPs (AMTI, Watertown, Massachusetts, USA; sampling frequency = 1,000 samples/s; size = 40 × 40 cm), one under each foot. Jumps were executed with the elbows at waist height, to limit arm swing inertial effects and to comfortably hold the SP with the hands (**Figure 2**). Each CMJ was visually inspected to verify if compliant with the prescribed recommendations and, if not, it was excluded. The task was executed over two FPs since: (i) a single one was too small (40 × 40 cm) for consistently recording a complete CMJ (static, loading, and landing); (ii) with two FPs the jumper attention was focused on task execution, rather than on landing on the FP. Each jump consisted of: (i) a static phase of a few seconds with the participant being in orthostatic position; (ii) a vocal command signaling the jump initiation; (iii) a second static phase as in (i).

Data processing

The SP-IMU offset and cross-axis sensitivity were computed and corrected according to Bergamini et al. (51) before each experimental

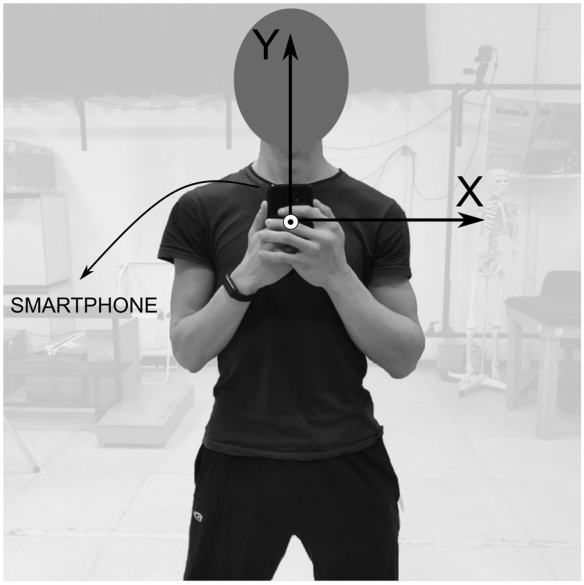

FIGURE 2
The experimental setup utilized for acquiring data from the SP device using the app Phyphox (remotely controlled via PC).

10.3389/fspor.2023.1112739

session. Vertical acceleration measures were aligned to the world reference to allow for a consistent gravity removal (11). Only the vertical component of the acceleration, a^{SP}, was considered for further computations.

In particular, gyroscope static bias was extrapolated from a 60 s static trial with the SP still on a flat surface, hence removed from each successive jump measure. For what concerns the accelerometer, three ad hoc 60 seconds acquisitions were performed; each consisted in aligning one of the three accelerometer axes with the gravity vector direction (51).

A comparable vertical acceleration measure, a^{FP}, was obtained from FP data. First, participant mass (m) was computed averaging the first second of the first static phase and dividing it by g ($g = 9.81\ m/s^2$). Consequently, the whole ground reaction force was divided by m, hence g was subtracted to obtain a^{FP}.

Peak jump height was computed through the TOV method for both instruments (h^{SP} and h^{FP}, respectively). For both data sets, vertical velocity, v, was computed through numerical integration of the corresponding acceleration from the CMJ onset (t_0) to the take-off (t_{TO}). More specifically: t_0 was computed as the time sample occurring 30 ms prior the first one deviating by 8 times the static phase standard deviation, following a similar approach as in Owen et al. (52); t_{TO} was chosen to be the first time sample such that $a \leq -g$. All data were processed and analyzed using GNU Octave (53).

Feature selection

A total of $M = 26$ features were extracted, as detailed in section "Model creation and evaluation". All the listed features are depicted in **Figures 3**, **4** and detailed in **Table 1**. All of them were extracted from a^{SP}, including the raw estimate of the height, h^{SP}. Nineteen jump-related variables (features from A to s) were inspired by Dowling and Vamos (5); three of them (u, W, and z), enlarging the description of power related variables were presented in Mascia and Camomilla (22); finally, the last three features are the central frequencies obtained processing a^{SP} via VMD (45), subdividing the signal into three intrinsic mode functions, each having a frequency spectrum centered around each of them. In particular, the high- and mid-central frequencies (f_1 and f_2) were assumed to be associated with the inertial effects due to wobbling masses, whereas the low-central frequency (f_3) was thought to be associated with the jump itself (**Figure 4**).

Model creation and evaluation

The final dataset was composed by 172 jumps, each associated with the corresponding 26 features computed from a^{SP}, and the peak height computed from a^{FP}, h^{FP}, considered as the dependent variable. Once data was arranged for all the jumps, the dataset was separated in two. More specifically, 75% of the jumps (129 examples) was used for creating the training set, whereas the remaining 25% (43 examples) was used as test set (54,55). The elements belonging to each of the two subsets were randomly selected.

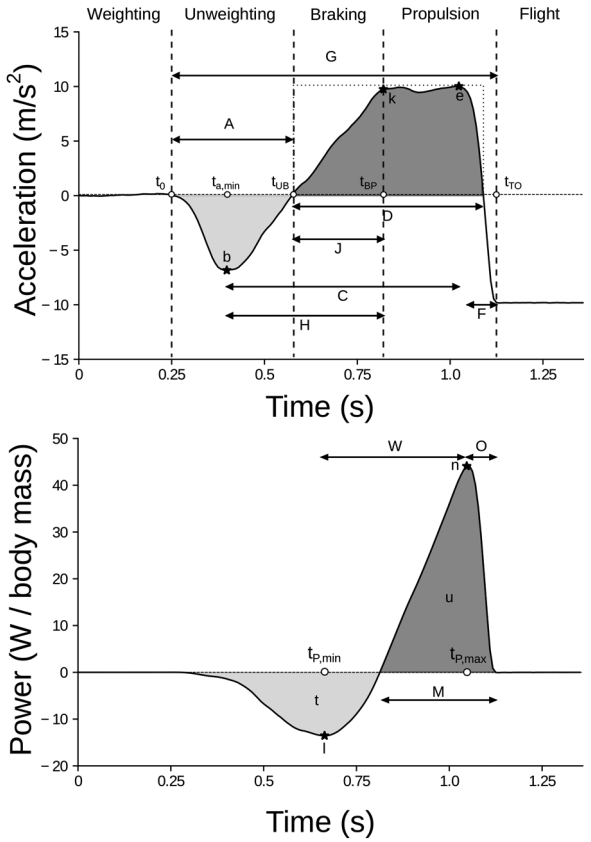

FIGURE 3
Visual depiction of the selected features, for a representative jump. Acceleration-related features are shown in the top panel, while power-related ones in the bottom panel. The vertical dotted lines in the top panel represent jump-phases transitions (1): the weighting phase lasts from the beginning to t_0 (jump onset); the unweighting phase lasts from t_0 to t_{UB}; the braking phase lasts from t_{UB} to t_{BP}; the propulsion phase lasts from t_{BP} to t_{TO}. Notice that t_{UB} and $t_{v,min}$ coincide. For the sake of clarity, only the former was depicted. Feature i cannot be represented as its numerical value was derived from further computations. The meaning of each feature is detailed in Table 1.

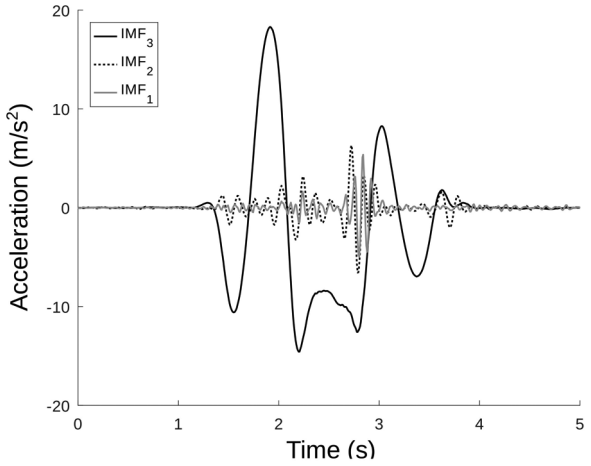

FIGURE 4
Intrinsic mode functions (IMFs) resulting from the application of the VMD algorithm on a representative CMJ. The black, dashed line represents the high frequency IMF_1; the gray continuous line represents the middle frequency IMF_2; the black continuous line represents the low frequency IMF_3. In this specific case: $f_1 = 6.06$ Hz; $f_2 = 3.49$ Hz; $f_3 = 0.55$ Hz.

TABLE 1 Detailed explanation of each of the analyzed features.

ID	Feature	Measure unit	Description
h^{SP}	Jump height	m	Height computed via TOV from a^{SP}
A	Unweighting phase duration	s	$[t_0, t_{UB}]$
b	Minimum acceleration	m/s^2	$a(t_{a,min})$
C	Time from minimum to maximum acceleration	s	$[t_{a,min}, t_{a,max}]$
D	Main positive impulse time	s	Time duration of positive acceleration from t_{UB} to the last positive sample prior t_{TO}
e	Maximum acceleration	m/s^2	$a(t_{a,max})$
F	Time from acceleration positive peak to take-off	s	$[t_{a,max}, t_{TO}]$
G	Ground contact duration	s	$[t_0, t_{TO}]$
H	Time from minimum acceleration to the end of breaking phase	s	$[t_{a,min}, t_{BP}]$
i	Maximum positive slope of acceleration	m/s^3	$\max(\mathrm{d}a(t)/\mathrm{d}t)$, $t \in [t_{a,min}, t_{a,max}]$
k	Acceleration at the end of the braking phase	m/s^2	$a(t_{BP})$
J	Time from negative peak velocity to the end of breaking phase	s	$[t_{v,min}, t_{BP}]$
l	Negative peak power	W/kg	$P(t_{P,min})$
M	Positive power duration	s	Self-explanatory
n	Positive peak power	W/kg	$P(t_{P,max})$
O	Time distance between positive peak power and take-off	s	$[t_{P,max}, t_{TO}]$
p	Mean slope between acceleration peaks	au	$p = (e - b)/C$
q	Shape factor	au	Ratio between the area under the curve from t_{UB} to the last positive sample prior t_{TO} (lasting D) and the one of a rectangle of sides D and e
r	Impulse ratio	au	$r = b/e$
s	Minimum negative velocity	m/s	$v(t_{v,min})$
u	Mean concentric power	W/kg	Average value of $P(t)$, $t \in [t_{BP}, t_{TO}]$
W	Power peaks delta time	s	$[t_{P,min}, t_{P,max}]$
z	Mean eccentric power	W/kg	Average value of $P(t)$, $t \in [t_0, t_{BP}]$
f_1	High central frequency	Hz	Highest VMD central frequency, associated with wobbling tissues and noise
f_2	Middle central frequency	Hz	Middle VMD central frequency, associated with wobbling tissues
f_3	Low central frequency	Hz	Lowest VMD central frequency, associated with the jump proper

Capital letters are for timings, small letters for the other parameters. Legend: au = arbitrary units; t_0,= jump onset time; t_{UB}, unbraking-braking phase transition time; t_{BP}, braking-propulsion phase transition time; t_{TO}, take-off time; $t_{a,min}$, minimum acceleration time; $t_{a,max}$, maximum acceleration time; $t_{v,min}$, minimum velocity time; $t_{P,min}$,= minimum power time; $t_{P,max}$, maximum power time.

Before training, each feature of the training set was normalized via z-score (55). The mean values and the standard deviations of each feature distribution were stored, so that they could be successively used for normalizing the test set features, accordingly.

In order to avoid possible multicollinearity between features, a feature reduction approach was performed on the training set using Lasso regularization (56). The regularization strength was set by selecting a value of $\alpha = 0.1$. The features that were excluded from such a shrinkage were not used for developing the machine learning algorithm.

A multilayer perceptron neural network (MLP) with one hidden layer was then used as machine learning architecture. A grid search with 5-fold cross validation was used for tuning the model hyperparameters (55,57). In particular, three hyperparameters were tuned: the activation function, the solver algorithm, and the number of neurons composing the hidden layer. The complete list of all the candidate hyperparameters for the grid search are presented in **Table 2**. The best and final model was chosen as the one with the combination of hyperparameters ensuring the best negative mean absolute error (58).

The height computed through the MLP was denoted as h^{MLP}, and compared also to the height directly estimated using the SP, h^{SP}.

To obtain a quantitative description of the influence that each feature had on the outcome, a permutation feature importance

TABLE 2 List of possible choices for each hyperparameter during the grid search.

Hyperparameter	Choices
Activation function	Identity, logistic, tanh, ReLU
Solver algorithm	lbfgs, sgd, adam
No. neurons in hidden layer	1–16

Legend: tanh, hyperbolic tangent; ReLU, rectified linear unit; lbfgs, limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm; sgd, stochastic gradient descent.

analysis (PFI) was performed on the trained model (49). Lets consider the model $\hat{f}$, the feature matrix X with j as feature index, the target h^{FP}, and the error $L[h^{FP}, \hat{f}(X)]$. In this study, the mean squared error (MSE) was selected as model error, since it enhances the prominence of feature importances being quadratic. Each permutation was performed on the training set only, meaning that each feature importance FI_j here presented belong to that set. Moreover, PFI was accomplished by splitting such a set in half, according to the recommendations proposed in Fisher et al. (59). The PFI algorithm follows:

1. Estimate the original model error: $e_0 = L[h^{FP}, \hat{f}(X)]$.
2. $\forall$ feature $j \in \{1, \ldots, M\}$:

   2.1 Generate the feature matrix with the permuted feature X^j_{perm}.
   2.2 Estimate the error for the feature matrix with the permuted jth feature: $e_j = L[h^{FP}, \hat{f}(X^j_{perm})]$.
   2.3 Compute the feature importance as: $FI_j = e_j/e_0$.

Statistical analysis

The performances of the criteria used for estimating h^{SP} and h^{MLP} were evaluated exploiting three metrics: (i) accuracy, considered as the root mean squared distance (RMSD) between the estimates and the true values; (ii) bias, computed as the average difference between the estimates and the true values; (iii) precision, representing the standard deviation of the differences, i.e., the dispersion of the error. Metrics (ii) and (iii) were directly computed from Bland-Altman plot analysis (60). Furthermore, the mean absolute error (MAE) was computed as an overall performance metrics for both the models.

Kendall's Tau test (61) was used for exploring possible heteroscedasticity of h^{SP} and h^{MLP} estimates (62,63). More specifically, the test was performed comparing the distribution of the averages versus the absolute differences of the gold standard and predicted values, as in Brehm et al. (63). If $\tau < 0.1$, data were considered homoscedastic; conversely, if $\tau \geq 0.1$, data were considered heteroscedastic.

Paired sample t-test was performed on the test set comparing the h^{FP} distribution with the h^{SP} and h^{MLP} ones, respectively. Finally, a linear regression analysis between h^{FP} with h^{SP} and h^{MLP}, respectively, was performed, so that a calibration between the systems could also be provided.

All the procedures regarding model creation and its performance evaluation have been accomplished through the *scikit-learn* environment (58).

Results

A total of 172 jumps were analyzed. Jumps height measured through the gold standard FP ranged from 10 cm up to 41 cm (25.6 $\pm$ 7.4 cm).

From the initial set of 26 features, only 17 were included in the final model after Lasso regularization. In particular, the features excluded were the following: A, C, D, k, m, n, o, p, and f_2, which meaning is explained in Table 1.

The grid search procedure analyzed a total number of 192 (4 activation functions $\times$ 3 solver algorithms $\times$ 16 neurons in the hidden layer) possible hyperparameter combinations. The best one (i.e., the one presenting the best negative mean absolute error among the 5 folds) was found to be the one having a logistic activation function with an hidden layer composed of 11 neurons, and its solver being the stochastic gradient descent algorithm. Among the 5 folds, the selected combination of hyperparameters presented an average negative mean absolute error ($\pm$SD) $= -0.8 \pm 0.0$ cm, with an average $R^2 = 0.978$. The results obtained for each of the 5 folds are detailed in Table 3.

Accuracy, computed as RMSD between h^{FP} and the estimates, improved of 4.5 times using the MLP model instead of SP-derived estimates. Similarly, using the MLP rather than the SP improved of 4 times the precision (SD of the differences between measure and estimate) and reduced of 4 times the MAE. The performance analysis for the analyzed models is detailed in Table 4. Moreover, Kendall's tau analysis performed on the average versus the absolute difference showed that h^{SP} distributions presented heteroscedasticity ($\tau = 0.38$). On the contrary, h^{MLP} distribution did show homoscedasticity ($\tau = -0.02$).

Paired sample t-test analysis showed significant differences between h^{FP} and h^{SP}, whereas no significant difference was found between h^{FP} and h^{MLP}. The results of such an analysis are detailed in Table 5. Finally, the results of the linear regression analysis for calibrating the two systems are provided in Table 6.

Bland-Altman plots of both SP- and MLP-heights estimated for the test set are presented in Figure 5, whereas the corresponding numerical values are reported in Table 7.

Permutation feature importance analysis highlighted the influence of each variable on the outcome. The five most influential features were e (maximum acceleration, FI = 9.64), J (time from negative peak velocity to the end of the braking phase, FI = 9.02), q (shape factor, FI = 7.63), s (minimum

TABLE 3 Scores of the grid search with cross validation for the best model. Notice that such results are related to the training set only.

	Negative MAE (cm)	R^2
Fold 1	−0.8	0.975
Fold 2	−0.7	0.981
Fold 3	−0.7	0.982
Fold 4	−0.8	0.975
Fold 5	−0.8	0.977
Mean	−0.8	0.978

TABLE 4 Results of the performance analysis for the analyzed models.

	Accuracy (cm)	Bias (cm)	Precision (cm)	MAE ± SD (cm)	τ
SP	18.0	−7.0	16.0	12.0 ± 13.0	0.38
MLP	4.0	−0.0	4.0	3.0 ± 3.0	−0.02

The accuracy is computed as the RMSD of h^{FP} and the estimates of each model; the bias is computed as the average difference of h^{FP} and the estimates of each model; the precision is computed as the standard deviation of the differences between h^{FP} and the estimates of each model; MAE is the mean absolute error (± standard deviation); τ is the Kendall's tau coefficient of the averages versus the absolute differences for each model.

TABLE 5 Results of the paired samples t-test analysis performed comparing h^{FP} with h^{SP} and h^{MLP}, respectively.

	t value	p value	Degrees of freedom	95% CI (cm)
SP	−2.770	0.008*	42	[− 12.2, − 1.8]
MLP	0.711	0.481	42	[− 1.0, 2.0]

Statistically significant differences are indicated with the superscript (*).

negative velocity, FI = 7.56), and h^{SP} (height computed from SP via TOV, FI = 6.39). A bar chart showing all outcomes of the PFI analysis for each variable is presented in Figure 6.

Discussion

In this study, a step towards jump height democratization using a smartphone through machine learning was proposed. While the SP alone entails unacceptable errors, the neural network model proposed here in the form of an MLP, reduced the error to produce outcomes that could be of use in an in-field setting.

The obtained results are promising considering that the low-cost of the IMU embedded into the smartphone allowed a feature extraction capable of predicting height estimates under the proposed assumptions. Indeed, the errors obtained are comparable to those typically obtained comparing commercially available IMU systems with FPs (mean difference ≃5 cm), as revised in Clemente et al. (10). For what concerns the accuracy, the MLP improved it bringing the discrepancy from 18 cm down to 4 cm. This improvement, as expressed in percentage of the true value, equivals obtaining 15.6% errors. The improvement brought by the MLP similarly improved precision from an initial value of 14 cm down to 4 cm, in line with what obtained in literature (10), acceptable depending on the subsequent application. Moreover, it is possible to appreciate that the MLP provided virtually no systematic bias. On the contrary, the average versus difference plot of h^{SP} (Figure 5) indicates that, on average, the use of TOV method leds to underestimated jump height values. Finally, an heteroscedastic

TABLE 6 Linear regression analysis performed between h^{FP} with h^{SP} and h^{MLP}, respectively.

	Intercept [95% CI] (cm) t value	h^{SP} [95% CI] (cm) t value	h^{MLP} [95% CI] (cm) t value	SEE (cm)	r
SP	17.668*** [13.944 21.391] 9.582	0.226*** [0.127 0.326] 4.594	—	6.257	0.583
MLP	5.681* [0.873 10.488] 2.386	—	0.789*** [0.600 0.977] 8.460	4.647	0.797

Legend: intercept, h^{SP}, and h^{MLP} are the coefficient of the regression analysis with 95% confidence intervals and t value; SEE is the standard error of the estimates, computed as the standard deviation of the regression residuals; r is the correlation coefficient. Significance level: (*) $p < 0.05$; (***) $p = 0$.

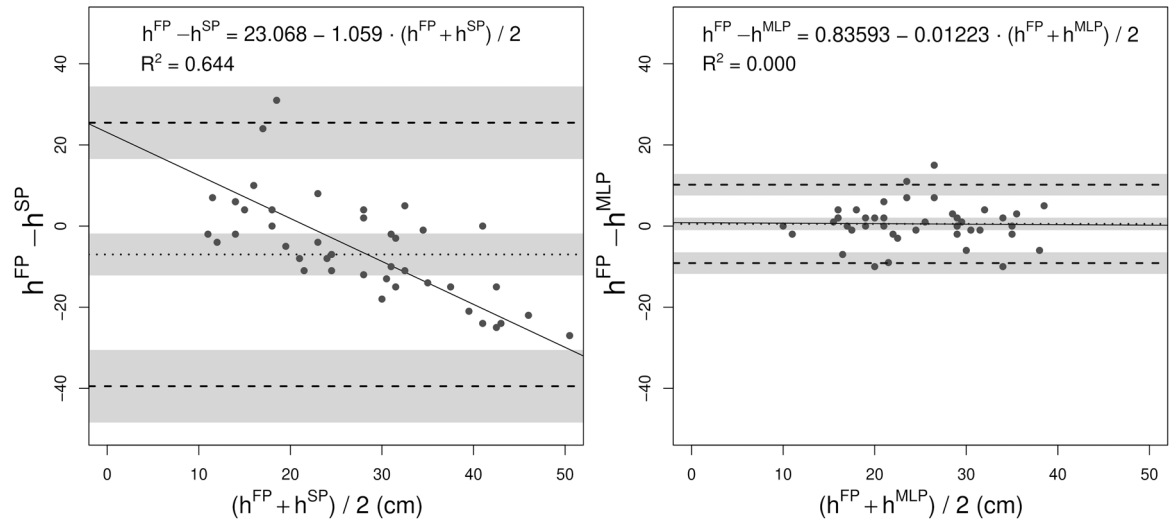

FIGURE 5
Bland-Altman analysis for the height estimates obtained directly from the SP with the TOV method (left) and through the MLP (right). The continuous line represents the bias (b) computed as the average difference between the gold standard and the estimate; the dot-dashed lines represent the lower and upper confidence intervals (UB and LB, respectively), computed as 1.96 standard deviations of the difference distribution away from the bias, in both directions. The gray shaded areas indicate the confidence interval for b, UB, and LB, respectively. The continuous line is the regression line of the average versus the differences, which equation is indicated in the above portion of the graph. The R^2 value refers to the latter equation.

TABLE 7 Bland-Altman analysis results for the two compared methods.

	Standardized bias	Bias [95% CI] (cm)	LB [95% CI] (cm)	UB [95% CI] (cm)
SP	−0.3	−7.0 [− 12.2, − 1.8]	−39.5 [− 30.5, − 48.4]	25.5 [34.4, 16.5]
MLP	0.0	0.5 [− 1.0, 2.0]	−9.1 [− 11.8, − 6.5]	10.2 [7.5, 12.8]

All the values and the related confidence intervals were computed according to what proposed in Giavarina (64). The corresponding graphical depiction of the values is presented in Figure 5.

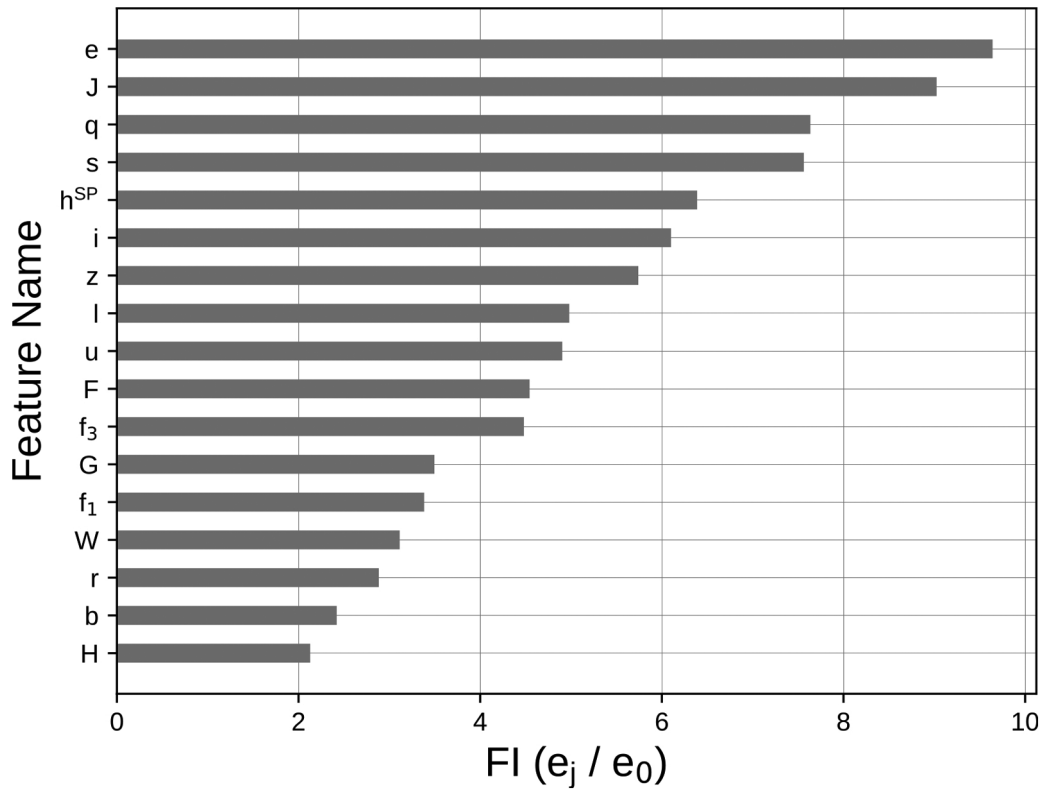

FIGURE 6
Barchart representing the outcome of the PFI analysis performed on all the variables selected by the Lasso regularization. Variables meaning is detailed in Table 1.

trend is clearly visible in the error of SP-computed jump heights (**Figure 5**), as confirmed by Kendall's Tau analysis ($\tau = 0.38$). On the contrary, there is no trend for MLP-estimates (**Figure 5**, $\tau = -0.02$). Estimates homoscedasticity corroborate model generalizability outside the training dataset.

Feature reduction, performed using Lasso regularization to avoid possible multicollinearity between the initial set of features, selected 17 features as model input. Permutation feature importance analysis proved this reduced feature set as influential of the outcome and ranked it in terms of influence on the model output (**Figure 6**). All the selected features, once the dataset was permuted (59), increased the MSE from a minimum of 2- to a maximum of about 9-times if compared with e_0 (see section "Model creation and evaluation"). Indeed, all the features exhibited a FI > 1, possibly indicating that Lasso regularization included only those features that could be predictive of jump height.

Among the five most important features, four of them (J, e, q, and s) belonged to two phases of the jump (braking and propulsion phases), with the fifth most important being the height computed through the TOV method from smartphone measures (h^{SP}).

Maximum acceleration (e, typically occuring in the propulsion phase of the CMJ), and the time from negative peak velocity to the end of the braking phase (i.e. the duration of the braking phase – J) were much more influential than all the others, increasing the MSE of about 9 times, when permuted. This could be seen as a mutual relationship, even though not collinear, between two distinct features of the CMJ. Indeed, the correlation between e and J in the training set was $r(e, J) = -0.52$, meaning that the shorter the braking phase, the bigger the magnitude of the recorded maximum acceleration. Moreover, the minimum negative velocity value (s) importance resulted prominent as well, having an impact of about 7-times on the MSE of the trained model. The impact of these three features together could be possibly linked to the biomechanical mechanism required for achieving the best possible jump height, that is, the stretch-shortening cycle (1,3,4).

The shape factor (q) after PFI influenced the MSE increasing it of about 7 times. This is in line with the claimed potential of this feature

to relate to an optimal force generation for achieving the maximum performance (5).

Although the height computed through smartphone alone (feature h^{SP}) is not sufficient for obtaining accurate estimates, it is still one of the most influential features as a first guess value to stay close to (FI = 6.39). Even though less influential than the above-mentioned features (FI = 4.48), the high central frequency (f_3) testifies the role of wobbling into disrupting jump height estimates. This is in line with similar results obtained using vertical acceleration measured from a belt-worn IMU at pelvic level: for this signal, the high central frequency was included in a model predictive of jump power, possibly compensating for the effects of soft tissue wobbling (22). This is also promising in the perspective of using VMD to identify the contribution of arm swing into vertical jump height, when voluntarily included in the jump action.

Such results must be analyzed in a wider framework. First, the sampling frequency, considered as a key player in jump height estimates quality, was lower (128 samples/s) than the 1,000 samples/s suggested as essential for a proper jump analysis (52). Nonetheless, testing the worst-case scenario was motivated by the democratization mission of this study, as results would then apply to any SP device available on the market, irrespective of its sensor sampling frequency. Second, holding the sensor in the hands was thought to be the most practical while repeatable experimental setup, with little perturbation to each jumper technique; however, it entails an undesired upper limb oscillatory movement. Despite the central frequency associated with this effect partially compensates for it, it can be hypothesized that placing the SP closer to the trunk and/or within an ad hoc harness could further enhance the prediction. Third, trial height ranged from a minimum of 10 cm up to a maximum of 41 cm, since none of the participants was an elìte athlete. Theoretically, the model output can be considered valid only within this range.

Among the strengths of this study, the model was devised using as gold standard reference jump heights that: (i) were computed from laboratory FP signals; (ii) were estimated through the TOV method, considered as the most reliable (36). Moreover, the approach presented here involves the use of features which are biomechanically related to the nature of the jump itself. Most importantly, they can be easily retrieved directly from the acceleration trace during the jump with simple mathematical manipulation.

The method could be embedded in the creation of a SP application, exploiting the current results to give runtime insights about jump height. The possibility to access the code and the methodology used (github.com/Maskul93/height-democratization) can also leverage actions to enlarge the current database and improve the estimates.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Internal Review Board of the University of Rome "Foro Italico" University of Rome "Foro Italico," piazza Lauro de Bosis 6, 00135, Rome, Italy. The patients/participants provided their written informed consent to participate in this study.

Author contributions

GM: Writing – review and editing, Writing – original draft, Visualization, Software, Methodology, Data Collection, Formal analysis, Data curation, Conceptualization. BDL: Writing – review and editing, Writing – original draft, Methodology, Data Collection, Formal analysis. VC: Writing – review and editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Data Collection, Formal analysis, Conceptualization. All authors contributed to the article and approved the submitted version.

Funding

Project Call: POR FESR Lazio 2014–2020 (Azione 1.2.1) Project Code: 20028AP000000095.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. McMahon JJ, Lake JP, Suchomel TJ. Vertical jump testing. In: Comfort P, Jones PA, McMahon JJ, editors. *Performance assessment in strength, conditioning*. Routledge (chap. 7).

2. Watkins CM, Barillas SR, Wong MA, Archer DC, Dobbs IJ, Lockie RG, et al. Determination of vertical jump as a measure of neuromuscular readiness, fatigue. *J Strength Cond Res*. (2017) 31:3305–10. doi: 10.1519/JSC.0000000000002231

3. Cormie P, McBride JM, McCaulley GO. Power-time, force-time,, velocity-time curve analysis of the countermovement jump: impact of training. *J Strength Cond Res*. (2009) 23:177–86. doi: 10.1519/JSC.0b013e3181889324

4. Komi PV. Stretch-shortening cycle: a powerful model to study normal and fatigued muscle. *J Biomech*. (2000) 33:1197–206. doi: 10.1016/S0021-9290(00)00064-6

5. Dowling JJ, Vamos L. Identification of kinetic and temporal factors related to vertical jump performance. *J Appl Biomech*. (1993) 9:95–110. doi: 10.1123/jab.9.2.95

6. Claudino JG, Cronin J, Mezêncio B, McMaster DT, McGuigan M, Tricoli V, et al. The countermovement jump to monitor neuromuscular status: a meta-analysis. *J Sci Med Sport*. (2017) 20:397–402. doi: 10.1016/j.jsams.2016.08.011

7. Linthorne NP. Analysis of standing vertical jumps using a force platform. *Am J Phys*. (2001) 69:1198–204. doi: 10.1119/1.1397460

8. Vanezis A, Lees A. A biomechanical analysis of good and poor performers of the vertical jump. *Ergonomics*. (2005) 48:1594–603. doi: 10.1080/00140130500101262

9. Casartelli N, Müller R, Maffiuletti NA. Validity and reliability of the myotest accelerometric system for the assessment of vertical jump height. *J Strength Cond Res*. (2010) 24:3186–93. doi: 10.1519/JSC.0b013e3181d8595c

10. Clemente F, Badicu G, Hasan UC, Akyildiz Z, Pino-Ortega J, Silva R, et al. Validity and reliability of inertial measurement units for jump height estimations: a systematic review. *Hum Mov*. (2022) 23:1–20. doi: 10.5114/hm.2023.111548

11. Rantalainen T, Finni T, Walker S. Jump height from inertial recordings: a tutorial for a sports scientist. *Scand J Med Sci Sports*. (2019) 30:38–45. doi: 10.1111/sms.13546

12. Picerno P, Camomilla V, Capranica L. Countermovement jump performance assessment using a wearable 3d inertial measurement unit. *J Sports Sci*. (2011) 29:139–46. doi: 10.1080/02640414.2010.523089

13. Spangler R, Rantalainen T, Gastin P, Wundersitz D. Inertial sensors are a valid tool to detect and consistently quantify jumping. *Int J Sports Med*. (2018) 39:802–808. doi: 10.1055/s-0044-100793

14. Camomilla V, Bergamini E, Fantozzi S, Vannozzi G. Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: a systematic review. *Sensors (Basel)* (2018) 18:873. doi: 10.3390/s18030873

15. Fathian R, Khandan A, Chiu LZF, Rouhani H. Assessment of countermovement jump with, without arm swing using a single inertial measurement unit. *Sports Biomech*. (2022):1–18. doi: 10.1080/14763141.2022.2032296

16. Forner-Cordero A, Mateu-Arce M, Forner-Cordero I, Alcántara E, Moreno JC, Pons JL. Study of the motion artefacts of skin-mounted inertial sensors under different attachment conditions. *Physiol Meas*. (2008) 29:N21–N31. doi: 10.1088/0967-3334/29/4/n01

17. Hughes GT, Camomilla V, Vanwanseele B, Harrison AJ, Fong DT, Bradshaw EJ. Novel technology in sports biomechanics: some words of caution. *Sports Biomech*. (2021):1–9. doi: 10.1080/14763141.2020.1869453

18. Li Y, Peng X, Zhou G, Zhao H. SmartJump: a continuous jump detection framework on smartphones. *IEEE Internet Comput*. (2020) 24:18–26. doi: 10.1109/mic.2020.2969610

19. Mateos-Angulo A, Galán-Mercant A, Cuesta-Vargas AI. Kinematic analysys by gender in different jump tests based on inertial sensors. *Rev Bras Med Esporte*. (2018) 24:263–7. doi: 10.1590/1517-869220182404186682

20. Schmidt M, Jaitner T, Nolte K, Rheinländer C, Wille S, Wehn N. A wearable inertial sensor unit for jump diagnosis in multiple athletes. In *Proceedings of the 2nd International Congress on Sports Sciences Research, Technology Support*. SCITEPRESS - Science, and Technology Publications. doi:10.5220/0005145902160220

21. Jaitner T, Schmidt M, Nolte K, Rheinländer C, Wille S, Wehn N. Vertical jump diagnosis for multiple athletes using a wearable inertial sensor unit. *Sports Technol*. (2015) 8:51–7. doi: 10.1080/19346182.2015.1117476

22. Mascia G, Camomilla V. An automated method for the estimate of vertical jump power through inertial measurement units. In *ISBS Proceedings Archive*. Vol. 39(1). Canberra, Australia: International Society of Biomechanics in Sports.

23. Mascia G, Valenti A, Camomilla V. Smartphone-based democratization of vertical jump height estimate. In *ISB 2021*, Stockholm.

24. Pueo B, Penichet-Tomas A, Jimenez-Olmedo JM. Validity, reliability, usefulness of smartphone, kinovea motion analysis software for direct measurement of vertical jump height. *Physiol Behav*. (2020) 227:113144. doi: 10.1016/j.physbeh.2020.113144

25. Webering F, Seeger L, Rother N, Blume H. Measuring vertical jump height using a smartphone camera with simultaneous gravity-based calibration. In *2021 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE. doi:10.1109/icce50685.2021.9427685

26. Bogataj Š, Pajek M, Andrašić S, Trajković N. Concurrent validity and reliability of my jump 2 app for measuring vertical jump height in recreationally active adults. *Appl Sci*. (2020) 10:3805. doi: 10.3390/app10113805

27. Carlos-Vivas J, Martin-Martinez JP, Hernandez-Mocholi MA, Perez-Gomez J. Validation of the iPhone app using the force platform to estimate vertical jump height. *J Sports Med Phys Fitness*. (2018) 58:227–232. doi: 10.23736/s0022-4707.16.06664-0

28. Driller M, Tavares F, McMaster D, O'Donnell S. Assessing a smartphone application to measure counter-movement jumps in recreational athletes. *Int J Sports Sci Coach*. (2017) 12:661–4. doi: 10.1177/1747954117727846

29. Gallardo-Fuentes F, Gallardo-Fuentes J, Ramırez-Campillo R, Balsalobre-Fernández C, Martınez C, Caniuqueo A, et al., Intersession and intrasession reliability and validity of the my jump app for measuring different jump actions in trained male and female athletes. *J Strength Cond Res*. (2016) 30:2049–56. doi: 10.1519/jsc.0000000000001304

30. Pueo B, Jimenez-Olmedo JM, Penichet-Tomas A, Bernal-Soriano MC. Inter-rater reliability of trained and untrained raters for measuring jump height with the myjump app. *J Phys Educ Sport*. (2018) 2018:821–824. doi: 10.7752/jpes.2018.02121

31. Rago V, Brito J, Figueiredo P, Carvalho T, Fernandes T, Fonseca P, et al. Countermovement jump analysis using different portable devices: implications for field testing. *Sports* (2018) 6:91. doi: 10.3390/sports6030091

32. Stanton R, Wintour S-A, Kean CO. Validity and intra-rater reliability of MyJump app on iPhone 6s in jump performance. *J Sci Med Sport*. (2017) 20:518–23. doi: 10.1016/j.jsams.2016.09.016

33. Camomilla V, Dumas R, Cappozzo A. Human movement analysis: the soft tissue artefact issue. *J Biomech*. (2017) 62:1–4. doi: 10.1016/j.jbiomech.2017.09.001

34. Nielsen ET, Jørgensen PB, Mechlenburg I, Sørensen H. Validation of an inertial measurement unit to determine countermovement jump height. *Asia-Pac J Sports Med Arthrosc Rehabil Technol*. (2019) 16:8–13. doi: 10.1016/j.asmart.2018.09.002

35. Aragón LF. Evaluation of four vertical jump tests: methodology, reliability, validity, and accuracy. *Meas Phys Educ Exerc Sci*. (2000) 4:215–28. doi: 10.1207/s15327841mpee0404_2

36. Moir GL. Three different methods of calculating vertical jump height from force platform data in men and women. *Meas Phys Educ Exerc Sci*. (2008) 12:207–18. doi: 10.1080/10913670802349766

37. Monnet T, Decatoire A, Lacouture P. Comparison of algorithms to determine jump height and flight time from body mounted accelerometers. *Sports Eng*. (2014) 17:249–59. doi: 10.1007/s12283-014-0155-1

38. Yamashita D, Murata M, Inaba Y. Effect of landing posture on jump height calculated from flight time. *Appl Sci*. (2020) 10:776. doi: 10.3390/app10030776

39. Veltink P, Slycke P, Hemssems J, Buschman R, Bultstra G, Hermens H. Three dimensional inertial sensing of foot movements for automatic tuning of a two-channel implantable drop-foot stimulator. *Med Eng Phys*. (2003) 25:21–8. doi: 10.1016/s1350-4533(02)00041-3

40. Woodman OJ. *An introduction to inertial navigation*. Tech. rep. doi:10.48456/TR-696

41. Lenzi D, Cappello A, Chiari L. Influence of body segment parameters and modeling assumptions on the estimate of center of mass trajectory. *J Biomech*. (2003) 36:1335–41. doi: 10.1016/s0021-9290(03)00151-9

42. Scalera GM, Ferrarin M, Marzegan A, Rabuffetti M. Assessment of stability of MIMU probes to skin-marker-based anatomical reference frames during locomotion tasks: effect of different locations on the lower limb. *Front Bioeng Biotechnol*. (2021) 9. doi: 10.3389/fbioe.2021.721900

43. Sheerin KR, Reid D, Besier TF. The measurement of tibial acceleration in runners—a review of the factors that can affect tibial acceleration during running and evidence-based guidelines for its use. *Gait Posture*. (2019) 67:12–24. doi: 10.1016/j.gaitpost.2018.09.017

44. Bonci T, Camomilla V, Dumas R, Chèze L, Cappozzo A. A soft tissue artefact model driven by proximal and distal joint kinematics. *J Biomech*. (2014) 47:2354–61. doi: 10.1016/j.jbiomech.2014.04.029

45. Dragomiretskiy K, Zosso D. Variational mode decomposition. *IEEE Trans Signal Process*. (2014) 62:531–44. doi: 10.1109/tsp.2013.2288675

46. Claudino JG, de Oliveira Capanema D, de Souza TV, Serrão JC, Pereira ACM, Nassis GP. Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Med Open*. (2019) 5. doi: 10.1186/s40798-019-0202-3

47. Cust EE, Sweeting AJ, Ball K, Robertson S. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *J Sports Sci*. (2018) 37:568–600. doi: 10.1080/02640414.2018.1521769

48. Zago M, Kleiner AFR, Federolf PA. Editorial: machine learning approaches to human movement analysis. *Front Bioeng Biotechnol*. (2021) 8. doi: 10.3389/fbioe.2020.638793

49. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/a:1010933404324

50. Staacks S, Hütz S, Heinke H, Stampfer C. Advanced tools for smartphone-based experiments: phyphox. *Phys Educ*. (2018) 53:045009. doi: 10.1088/1361-6552/aac05e

51. Bergamini E, Ligorio G, Summa A, Vannozzi G, Cappozzo A, Sabatini A. Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: accuracy assessment in manual and locomotion tasks. *Sensors* (2014) 14:18625–49. doi: 10.3390/s141018625

52. Owen NJ, Watkins J, Kilduff LP, Bevan HR, Bennett MA. Development of a criterion method to determine peak mechanical power output in a countermovement jump. *J Strength Cond Res*. (2014) 28:1552–8. doi: 10.1519/jsc.0000000000000311

53. Eaton JW, Bateman D, Hauberg S, Wehbring R. *GNU Octave version 6.1.0 manual: a high-level interactive language for numerical computations*. Available from: https://octave.org/doc/v6.1.0/

54. Kamruzzaman J, Begg R. Support vector machines, other pattern recognition approaches to the diagnosis of cerebral palsy gait. *IEEE Trans Biomed Eng*. (2006) 53:2479–90. doi: 10.1109/tbme.2006.883697

55. Halilaj E, Rajagopal A, Fiterau M, Hicks JL, Hastie TJ, Delp SL. Machine learning in human movement biomechanics: best practices, common pitfalls, new opportunities. *J Biomech*. (2018) 81:1–11. doi: 10.1016/j.jbiomech.2018.09.009

56. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B (Methodol)*. (1996) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x

57. Stone M. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Ser B (Methodol)*. (1974) 36:111–47. doi: 10.1111/j.2517-6161.1974.tb00994.x

58. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. (2011) 12:2825–30. doi: 10.48550/arXiv.1201.0490

59. Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J Mach Learn Res*. (2018) 20:1–81. doi: 10.48550/ARXIV.1801.01489

60. Bland J, Altman D. Measuring agreement in method comparison studies. *Stat Methods Med Res*. (1999) 8:135–60. doi: 10.1191/096228099673819272

61. Kendall MG. A new measure of rank correlation. *Biometrika*. (1938) 30:81–93. doi: 10.1093/biomet/30.1-2.81

62. Bland JM, Altman DG. Statistics notes: measurement error proportional to the mean. *BMJ* (1996) 313:106. doi: 10.1136/bmj.313.7049.106

63. Brehm M-A, Scholtes VA, Dallmeijer AJ, Twisk JW, Harlaar J. The importance of addressing heteroscedasticity in the reliability analysis of ratio-scaled variables: an example based on walking energy-cost measurements. *Dev Med Child Neurol*. (2011) 54:267–73. doi: 10.1111/j.1469-8749.2011.04164.x

64. Giavarina D. Understanding Bland Altman analysis. *Biochem Med*. (2015) 25:141–51. doi: 10.11613/bm.2015.015