SPORTS BIOMECHANICS
https://doi.org/10.1080/14763141.2022.2159870

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Comparison of different machine learning models to enhance sacral acceleration-based estimations of running stride temporal variables and peak vertical ground reaction force

Aurélien Patoz[a,b], Thibault Lussiana[b,c,d], Bastiaan Breine[b,e], Cyrille Gindre[b,c] and Davide Malatesta[a]

[a]Institute of Sport Sciences, University of Lausanne, Lausanne, Switzerland; [b]Research and Development Department, Volodalen Swiss Sport Lab, Aigle, Switzerland; [c]Research and Development Department, Chavéria, France; [d]Research Unit EA3920 Prognostic Markers and Regulatory Factors of Cardiovascular Diseases and Exercise Performance, Health, Innovation platform, University of Franche-Comté, Besançon, France; [e]Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

ABSTRACT

Machine learning (ML) was used to predict contact (t_c) and flight (t_f) time, duty factor (DF) and peak vertical force ($F_{v,max}$) from IMU-based estimations. One hundred runners ran on an instrumented treadmill (9–13 km/h) while wearing a sacral-mounted IMU. Linear regression (LR), support vector regression and two-layer neural-network were trained (80 participants) using IMU-based estimations, running speed, stride frequency and body mass. Predictions (remaining 20 participants) were compared to gold standard (kinetic data collected using the force plate) by calculating the mean absolute percentage error (MAPE). MAPEs of $F_{v,max}$ did not significantly differ among its estimation and predictions ($P = 0.37$), while prediction MAPEs for t_c, t_f and DF were significantly smaller than corresponding estimation MAPEs ($P \leq 0.003$). There were no significant differences among prediction MAPEs obtained from the three ML models ($P \geq 0.80$). Errors of the ML models were equal to or smaller than (≤32%) the smallest real difference for the four variables, while errors of the estimations were not (15–45%), indicating that ML models were sufficiently accurate to detect a clinically important difference. The simplest ML model (LR) should be used to improve the accuracy of the IMU-based estimations. These improvements may be beneficial when monitoring running-related injury risk factors in real-world settings.

ARTICLE HISTORY
Received 16 March 2022
Accepted 13 December 2022

KEYWORDS
Biomechanics; inertial measurement unit; duty factor; contact time; running injuries

Introduction

While providing many health benefits, running is also associated with lower limb overuse injuries (Fredette et al., 2021; Hreljac et al., 2000; Hreljac, 2004; Nielsen et al., 2012). These injuries often occur when a repetitive stress is applied to the system beyond its maximum tolerance (Hreljac, 2004). The peak vertical ground reaction force ($F_{v,max}$), contact time (t_c) and duty factor (DF), i.e., the product of t_c and stride frequency (SF) (Folland et al., 2017; Minetti, 1998), were shown to play a role in running-related injury

CONTACT Aurélien Patoz aurelien.patoz@unil.ch

development (Edwards, 2018; Kiernan et al., 2018; Lenhart et al., 2014; Malisoux et al., 2022; Matijevich et al., 2019; Sasimontonkul et al., 2007; Scott & Winter, 1990). Flight time (t_f) might also play a role as it takes both the vertical ground reaction force and its time of production into account (Appendix).

These variables have often been estimated using inertial measurement units (IMUs) (Chew et al., 2018; Day et al., 2021; Falbriard et al., 2018; Lee et al., 2010; Norris et al., 2014; Patoz et al., 2022), which are effective devices to longitudinally monitor these variables outside of a laboratory (Camomilla et al., 2018). However, obtaining accurate estimations based on IMU data depends on several factors such as the number of sensors, sensor position, or signal filtering (Alcantara et al., 2021). For instance, error on t_c was ~10 ms when using foot-worn inertial sensors (Chew et al., 2018; Falbriard et al., 2018). Using a single sacral-mounted IMU to estimate t_c, t_f and $F_{v,\max}$ led to root mean square errors (RMSEs) of 20 ms and 0.15BW compared to gold standard values (force plate) (Patoz et al., 2022). Similarly, Day et al. (2021) reported Pearson correlation coefficients (r) of ~0.65 between IMU estimations and gold standard values for t_c and $F_{v,\max}$. A sacral-mounted IMU is a natural choice because it approximates the location of the centre of mass (Napier et al., 2020) but led to error two times larger for t_c. However, applying advanced analysis methods such as machine learning (ML) on top of these estimations may provide more accurate predictions.

ML was used to explain the differences of gait patterns between high- and low-mileage runners (Xu et al., 2022) as well as to estimate biomechanical variables based on IMU data (Alcantara et al., 2021; Derie et al., 2020; Matijevich et al., 2020; Wouda et al., 2018). ML has the advantage to provide an analytical model which is trained and tested using different subsets of the dataset (Halilaj et al., 2018) and built from physics-based variables, i.e., variables that demonstrated to provide changes in running biomechanics (Alcantara et al., 2021). The modelling of the relationships between clinical outcomes and biomechanical measures was attempted using ML models like linear regressions (LRs), support vector machines and artificial neural networks (NNs) (Backes et al., 2020; Halilaj et al., 2018). Though limited to linear relationships, LRs are widely used because the regression coefficients are useful for model interpretability (Chambers, 1992). On the other hand, support vector machines and NNs are used to model non-linear relationships. Although they usually provide better accuracies than LRs, their coefficients are difficult to interpret because of their large numbers (Halilaj et al., 2018). Therefore, using both basic and complex ML models might illustrate the tradeoff between interpretability and accuracy and give the option to prioritise between the former and the latter.

Hence, the purpose of this study was to apply ML to predict t_c, t_f, DF and $F_{v,\max}$ from their respective IMU-based estimations. It was hypothesised that further applying ML to these IMU-based estimations should provide predictions with higher accuracies than those previously reported for the estimations (Patoz et al., 2022). Errors of the ML models were also compared to the smallest real difference (SRD) for the four variables, i.e., it was investigated if the ML models were sufficiently accurate to detect a clinically important difference. The comparison among the predictions of several ML models would allow defining which model has the best tradeoff between interpretability and accuracy.

Materials and methods

Participant characteristics

An existing database of 100 recreational runners (Patoz et al., 2022) (females: 27, age: 29 ± 7 years, height: 169 ± 5 cm, body mass: 61 ± 6 kg and weekly running distance: 22 ± 16 km; males: 73, age: 30 ± 8 years, height: 180 ± 6 cm, body mass: 71 ± 7 kg and weekly running distance: 38 ± 24 km) was used in the present study. Participants were required to run at least once a week and to not have current or recent lower-extremity injury (≤1 month) to be involved in this study. The local Ethics Committee of the XXX approved the study protocol prior to data collection (XXX) and adhered to the latest version of the Declaration of Helsinki of the World Medical Association. Written informed consent was obtained for all subjects.

Experimental procedure, data collection and estimations from inertial measurement unit data

The experimental procedure, data collection and IMU-based estimations have already been described elsewhere (Patoz et al., 2022) and are briefly summarised herein.

An IMU of 9.4 g (Movesense sensor, Suunto, Vantaa, Finland) was attached to the sacrum of participants. Then, after a warm-up run of 7-min (9–13 km/h) on an instrumented treadmill (Arsalis T150–FMT-MED, Louvain-la-Neuve, Belgium), three 1-min running trials (9, 11 and 13 km/h) were recorded in a randomised order. These speeds were chosen because they represent the most commonly adopted speeds of recreational runners (Selinger et al., 2022). Data analysis was performed on the IMU and kinetic data corresponding to the first 10 strides following the 30-s mark. IMU and kinetic data were not exactly synchronised (technical limitation), but the same 10 strides were used for each running trial of each participant because the synchronisation delay between IMU and kinetic data was small (≤50 ms).

A home-made iOS application running on an iPhone SE (Apple, Cupertino, CA, USA) was used to collect IMU data (saturation range: ±8*g*) at 208 Hz (manufacturing specification). IMU data were then transferred to a personal computer for post processing.

Kinetic data were collected at 200 Hz using the force plate embedded into the treadmill (Arsalis, Louvain-la-Neuve, Belgium) together with the Vicon Nexus software (v2.9.3, Vicon, Oxford, UK). The Visual3D Professional software (v6.01.12, C-Motion Inc., Germantown, MD, USA) was used to process the 3D ground reaction forces (analog signal), which were first exported in .c3d format. The forces were low-pass filtered at 20 Hz using a fourth-order Butterworth filter.

Gold standard t_c and t_f were given by the time during which the vertical ground reaction force was above and below 20 N, respectively (Smith et al., 2015). Gold standard DF was given by the product of t_c and SF. Gold standard $F_{v,\max}$ was given by the maximum of the vertical ground reaction force during t_c and was expressed in body weight units. The gold standard variables were computed within Visual3D and given as the average over 10 analysed strides.

A custom c++ code (ISO/IEC, 2020) was used to process IMU data and has already been described elsewhere (Patoz et al., 2022). Briefly, the vertical ground reaction force

was approximated by the vertical acceleration (previously reoriented and filtered using a truncated Fourier series to 5 Hz) multiplied by body mass. Then, t_c, t_f, DF and $F_{v,\max}$ were estimated as in the gold standard case but using the approximated vertical ground reaction force. In other words, t_c and t_f were given by the time during which the approximated vertical ground reaction force was above and below 20 N, respectively, DF was given by the product of t_c and SF, and $F_{v,\max}$ was given by the maximum of the approximated vertical ground reaction force during t_c. The custom c++ code provided the estimated variables as the average over the 10 analysed strides.

Predicted variables obtained using machine learning models

Three ML models: LR, support vector regression (SVR) – the regression analog of support vector machine – with the radial basis function kernel, and two-layer NN (NN2), were constructed to predict t_c, t_f, DF and $F_{v,\max}$ using a train/test method (80–20% split; 80 and 20 runners in the training and testing set, respectively). All the running trials from one subject were included in only one set to ensure that the models generalise well to new data and a similar distribution of male (72.5%) and female (27.5%) was maintained in both subsets to avoid introducing bias in the model during training (Halilaj et al., 2018). For each variable predicted by the three models, four features were used as predictors: running speed, runner's body mass, SF and corresponding IMU-based estimation. This choice follows from their relationship with changes in running biomechanics (Alcantara et al., 2021; Nagahara et al., 2018; Nilsson & Thorstensson, 1989) and to keep the models relatively simple. The SF included in the features was the IMU-based estimation and was almost identical to the gold standard (Figure 1). The features were standardised by removing the mean and by scaling to unit variance. The different models were trained using a 5-fold cross-validation approach for hyperparameter optimisation. Hyperparameters are given in Table 1. The trained models were used to make predictions on the testing set, which was previously standardised based on the mean and standard deviation (SD) of the training set, leading to a total of 60 predictions (three running speeds × 20 individuals). The accuracy between gold standard and predicted values was quantified using r, RMSE and mean absolute percentage error (MAPE). Besides, RMSE was compared to the SRD to evaluate if the precision of a model is sufficient to detect a clinically important difference. Indeed, SRD can be defined as the smallest change that indicates a clinically important difference and was calculated as $\text{SRD} = 1.96\sigma$, where σ is the within-subject standard deviation of the gold standard values. The analysis was performed using Python (v3.7.4, available at http://www.python.org).

Statistical analysis

All data are presented as mean ± SD. To examine the presence of systematic bias between gold standard t_c, t_f, DF and $F_{v,\max}$ values and corresponding predicted or estimated values, Bland-Altman plots were constructed (Atkinson & Nevill, 1998; Bland & Altman, 1995). In case of a systematic bias, a positive value indicates the estimated or predicted variable is overestimated. In addition, lower and upper limit of agreements and 95%

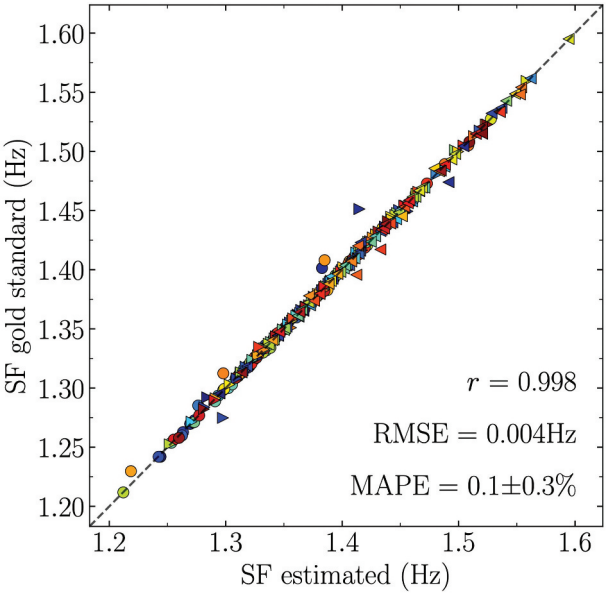

Figure 1. Gold standard (obtained using force plate data) stride frequency (SF) as function of estimated SF (obtained using inertial measurement unit data, no machine learning) for the entire set of data and corresponding Pearson correlation coefficient (*r*), root mean square error (RMSE) and mean absolute percentage error (MAPE). Each point represents the value for a given participant-running speed combination (300 points: three running speeds x 100 runners). Colors represent different participants while the three symbols represent different running speeds (o: 9 km/h, ▷: 11 km/h, ◁: 13 km/h).

Table 1. Hyperparameters optimised during the 5-fold cross validation for the three machine learning models employed.

Machine learning model	Hyperparameter	Values
Linear regression	Intercept in the model	True and False
Support vector regression	C (inversely proportional to the strength of the regularisation)	20 points (logarithmic scale between 0.001 and 10,000)
	Epsilon (specifies the epsilon-tube within which no penalty is associated in the training loss function with points predicted within a distance epsilon from the actual value)	20 points (logarithmic scale between 0.001 and 100)
Two-layer neural network	Activation function of the first layer	relu, tanh, sigmoid and softmax
	Dimensionality of the inner layer	8, 16, 32 and 64
	Batch size	2, 4, 8 and 16
	Loss function	mean absolute error and mean squared error

confidence intervals were calculated. Moreover, residual plots were inspected and no obvious deviations from homoscedasticity and normality were observed; therefore, one-way [model (no model vs LR vs SVR vs NN2)] repeated measures ANOVA with

Mauchly's correction for sphericity and employing Holm corrections for pairwise post hoc comparisons were used to compare MAPE between models. This comparison was possible because an MAPE was calculated for each estimation/prediction made in the testing set. Statistical analysis was performed using Jamovi (v1.6.23, available at https://www.jamovi.org) with a level of significance set at $P \leq 0.05$.

Results

Participant characteristics and biomechanical variables within training and testing sets

Participant characteristics were not significantly different between training and testing sets ($P \geq 0.24$; Table 2). Gold standard values in the training set and gold standard, estimated (using IMU data, no ML) and predicted values in the testing set are reported in Tables 3 and 4, respectively.

Accuracy of the machine learning models (predictions) and estimations

The ML models predicted t_c with an r of 0.89 ± 0.01, RMSE of 12.2 ± 0.2 ms and MAPE of 3.6 ± 0.1% (mean ± SD for the three models). As for t_f, the r, RMSE and MAPE were 0.86 ± 0.01, 11.7 ± 0.4 ms and 9.3 ± 0.4%. DF was predicted with an r of 0.84 ± 0.03, RMSE of 1.7 ± 0.1%, and MAPE of 3.6 ± 0.2%. As for $F_{v,\max}$, the r, RMSE and MAPE were 0.77 ± 0.01, 0.13 ± 0.01BW and 3.8 ± 0.1% (Figure 2). For completeness, Figure 2 also depicts the gold standard as function of estimated values for the testing set together with their corresponding r, RMSE and MAPE.

A significant model effect was reported for the MAPE of t_c, t_f and DF ($P \leq 0.001$) but not of $F_{v,\max}$ ($P = 0.37$). Post hoc tests revealed that the MAPEs obtained using the three ML models were significantly smaller than the MAPE obtained without ML for t_c, t_f and

Table 2. Participant characteristics for the training (80 runners) and testing (20 runners) sets.

Characteristics	Training set	Testing set	*p*-value
Sex	M = 58; F = 22	M = 15; F = 5	NA
Age (yr)	30 ± 7	30 ± 8	0.96
Height (cm)	177 ± 8	177 ± 7	0.89
Body mass (kg)	68 ± 8	70 ± 6	0.29
Running distance (km/week)	32 ± 24	39 ± 20	0.24

The values are presented as mean ± standard deviation. M: male, F: female and NA: not applicable.

Table 3. Gold standard (obtained using force plate data) contact time (t_c), flight time (t_f), duty factor (DF) and peak vertical ground reaction force ($F_{v,\max}$) for the training set (80 runners) at three running speeds.

Running speed (km/h)	t_c (ms)	t_f (ms)	DF (%)	$F_{v,\max}$ (BW)
9	277 ± 23	95 ± 23	37.3 ± 2.9	2.4 ± 0.2
11	249 ± 20	113 ± 19	34.4 ± 2.4	2.5 ± 0.2
13	227 ± 17	124 ± 17	32.3 ± 2.2	2.6 ± 0.2

The values are presented as mean ± standard deviation.

Table 4. Gold standard (obtained using force plate data) contact time (t_c), flight time (t_f), duty factor (DF) and peak vertical ground reaction force ($F_{v,max}$) as well as corresponding estimated (obtained using inertial measurement unit data, no machine learning) and predicted [obtained using three machine learning models: linear regression (LR), support vector regression with the radial basis function kernel (SVR) and two-layer neural network (NN2)] values for the testing set (20 runners) at three running speeds.

Variable	Running speed (km/h)	Gold standard	Estimated	Predicted LR	Predicted SVR	Predicted NN2
t_c (ms)	9	282 ± 18	268 ± 14	279 ± 13	278 ± 13	278 ± 13
	11	253 ± 17	257 ± 13	253 ± 12	251 ± 15	252 ± 11
	13	229 ± 14	246 ± 11	229 ± 9	227 ± 10	228 ± 9
t_f (ms)	9	86 ± 19	100 ± 9	89 ± 15	91 ± 14	89 ± 17
	11	105 ± 21	101 ± 9	105 ± 16	107 ± 15	108 ± 16
	13	117 ± 18	100 ± 8	118 ± 13	119 ± 13	119 ± 14
DF (%)	9	38.4 ± 2.3	36.4 ± 1.0	38.0 ± 1.7	37.5 ± 1.3	37.8 ± 1.3
	11	35.4 ± 2.4	35.9 ± 0.9	35.3 ± 1.6	34.9 ± 1.5	35.4 ± 1.5
	13	33.2 ± 2.1	35.6 ± 0.7	33.1 ± 1.4	32.8 ± 1.4	33.1 ± 1.4
$F_{v,max}$ (BW)	9	2.35 ± 0.15	2.39 ± 0.10	2.34 ± 0.11	2.34 ± 0.11	2.33 ± 0.12
	11	2.47 ± 0.20	2.45 ± 0.11	2.48 ± 0.13	2.48 ± 0.13	2.50 ± 0.13
	13	2.59 ± 0.18	2.47 ± 0.09	2.60 ± 0.10	2.59 ± 0.11	2.58 ± 0.09

The values are presented as mean ± standard deviation.

DF ($P \leq 0.003$; Figure 2). However, there was no significant difference among the MAPEs obtained using the three ML models for these three variables ($P \geq 0.80$).

Bland-Altman plots between gold standard and predicted or estimated values are given in Figure 3, and systematic bias as well as lower and upper limit agreements are reported in Table 5. The smallest bias was reported for LR.

Accuracy improvement between the predictions and estimations

Using ML allowed increasing r by 28 ± 1%, 59 ± 2%, 65 ± 5% and 15 ± 1%, for t_c, t_f, DF and $F_{v,max}$, respectively, compared to those obtained from IMU-based estimations. As for the RMSEs, they decreased by 37 ± 1%, 39 ± 2%, 37 ± 4% and 16 ± 4% for t_c, t_f, DF and $F_{v,max}$, respectively, while the MAPEs decreased by 40 ± 1%, 40 ± 3%, 41 ± 3% and 9 ± 1% (Table 6).

Ability to detect a clinically important difference

SRD was equal to 13.2 ms, 15.4 ms, 1.8% and 0.13BW for t_c, t_f, DF and $F_{v,max}$, respectively. RMSE of the ML models were equal to or smaller than (≤32%) the SRDs of the four variables. However, RMSE of the estimated values were larger than the SRDs of the four variables (15–45%).

Optimal coefficients of the linear regression models

The optimal coefficients obtained for the predictors used in the LR models are given in Table 7. Among all predictors, SF did not contribute significantly to the predictions ($P \geq 0.69$; Table 7). Hence, new LR models which did not include SF as a predictor were optimised, and optimal coefficients are reported in Table 8. These new LR models predicted t_c with an r of 0.88, RMSE of 12.9 ms and MAPE of 3.9 ± 3.3%. As for t_f, r,

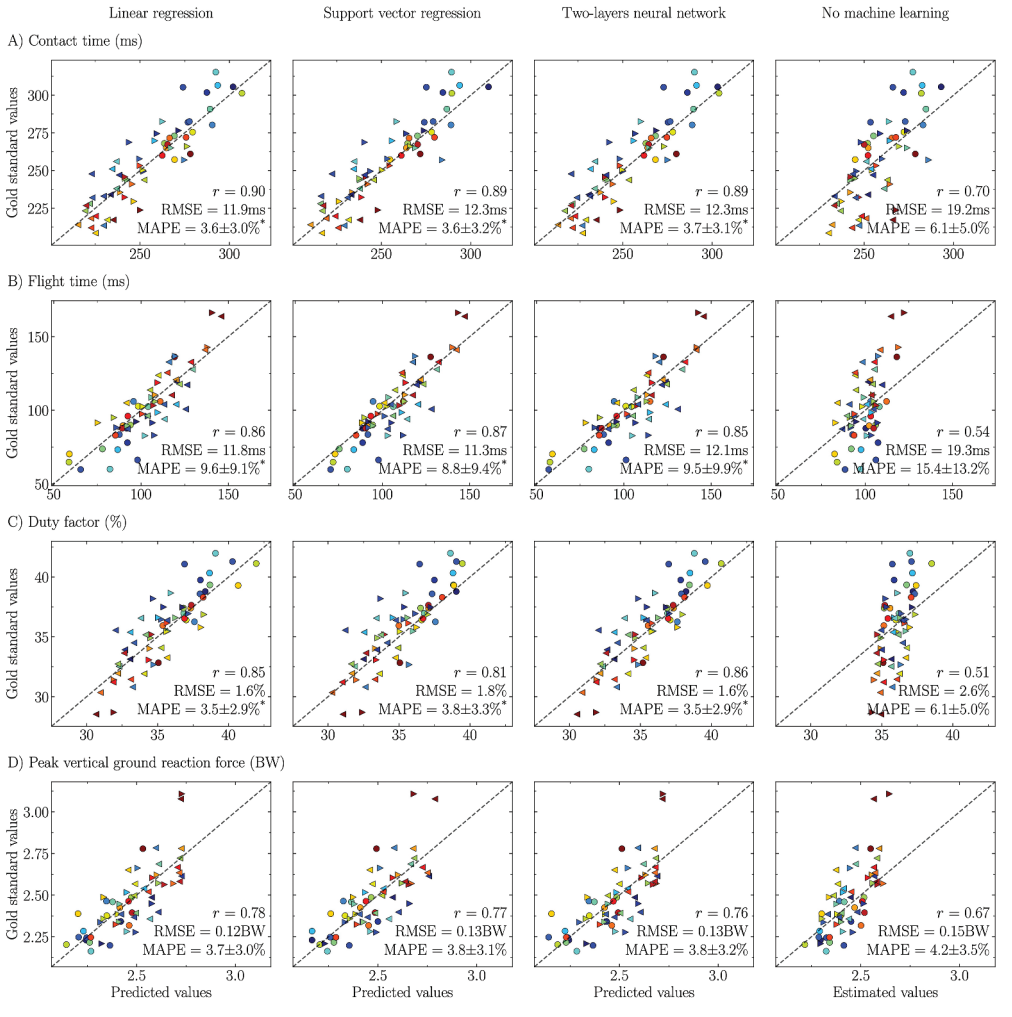

Figure 2. Gold standard (obtained using force plate data) as function of predicted (obtained using three different machine learning models) and estimated (obtained using inertial measurement unit data, no machine learning) (A) contact time, (B) flight time, (C) duty factor and (D) peak vertical ground reaction force for the testing set and corresponding Pearson correlation coefficient (*r*), root mean square error (RMSE) and mean absolute percentage error (MAPE). The one-way repeated measures ANOVA revealed a significant model effect (no model vs linear regression vs support vector regression with the radial basis function kernel vs two-layer neural network) for contact time, flight time and duty factor when comparing the MAPE among the models. *significant difference ($P \leq 0.003$) between the MAPE of the predictions obtained using a given machine learning model and the MAPE of the estimations obtained using inertial measurement unit data, as determined by Holm post hoc tests. Each point represents the value for a given participant-running speed combination (60 points: three running speeds x 20 runners). Colors represent different participants while the three symbols represent different running speeds (o: 9 km/h, ▷: 11 km/h, ◁: 13 km/h).

RMSE and MAPE were 0.85, 12.2 ms and 9.9 ± 8.9%. DF was predicted with an *r* of 0.82, RMSE of 1.8% and MAPE of 4.1 ± 3.1%, while $F_{v,\max}$ with an *r*, RMSE and MAPE of 0.77, 0.13BW and 3.8 ± 3.0%.

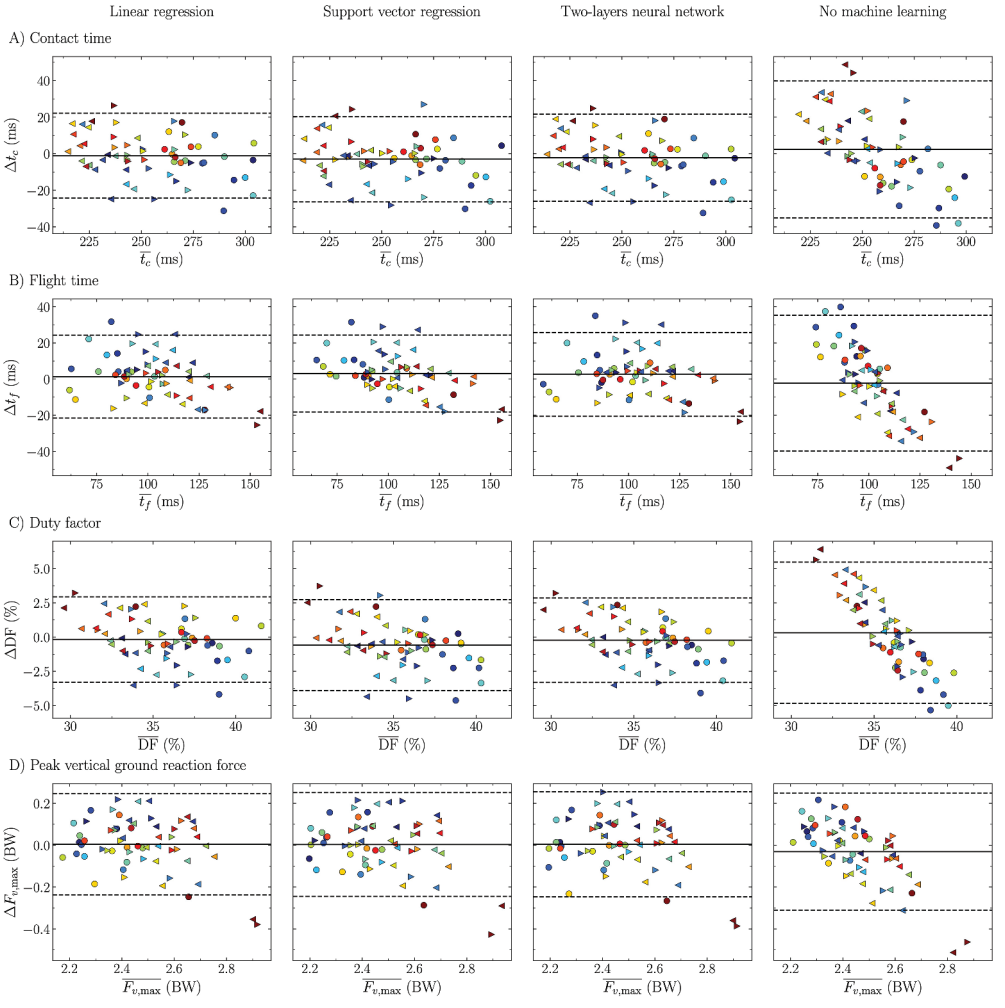

Figure 3. Comparison between gold standard (obtained using force plate data) and predicted (obtained using three different machine learning models) as well as estimated (obtained using inertial measurement unit data, no machine learning) (A) contact time, (B) flight time, (C) duty factor and (D) peak vertical ground reaction force for the testing set [differences (Δ) as a function of mean values together with systematic bias (solid line) as well as lower and upper limit of agreements (dashed lines), i.e., a Bland-Altman plot]. Each point represents the value for a given participant-running speed combination (60 points: three running speeds x 20 runners). Colors represent different participants while the three symbols represent different running speeds (o: 9 km/h, ▷: 11 km/h, ◁: 13 km/h). For systematic bias, positive values indicate the estimated or predicted variable is overestimated.

Discussion and implications

The purpose of the present study was to apply ML to predict t_c, t_f, DF and $F_{v,\max}$ from their respective IMU-based estimations. According to the hypothesis, further applying ML to IMU-based estimations of t_c, t_f, DF and $F_{v,\max}$ increased the accuracy of their predictions. However, the enhancement was not significant for $F_{v,\max}$. The simplest ML model (LR) was characterised by a similar prediction accuracy than more complicated

Table 5. Systematic bias, lower limit of agreement (Lloa) and upper limit of agreement (Uloa) between contact time (t_c), flight time (t_f), duty factor (DF) and peak vertical ground reaction force ($F_{v,\max}$) obtained using a force plate (gold standard method) and machine learning models (predictions; linear regression, support vector regression with the radial basis function kernel and two-layer neural network) as well as an inertial measurement unit (estimations; no machine learning) for the testing set (20 runners).

Variable	Method	Systematic bias	Lloa	Uloa
t_c (ms)	Linear regression	−1.0 [−4.0, 2.0]	−24.3 [−29.4, −19.1]	22.2 [17.1, 27.3]
	Support vector regression	−3.0 [−6.0, 0.0]	−26.3 [−31.4, −21.1]	20.3 [15.2, 25.5]
	Two-layer neural network	−2.2 [−5.3, 0.9]	−26.0 [−31.3, −20.7]	21.7 [16.4, 26.9]
	No machine learning	2.3 [−2.5, 7.2]	−35.1 [−43.4, −26.9]	39.8 [31.5, 48.0]
t_f (ms)	Linear regression	1.3 [−1.6, 4.3]	−21.6 [−26.6, −16.5]	24.2 [19.2, 29.3]
	Support vector regression	3.1 [0.3, 5.8]	−18.3 [−23.0, −13.6]	24.4 [19.7, 29.1]
	Two-layer neural network	2.6 [−0.4, 5.6]	−20.5 [−25.6, −15.4]	25.8 [20.7, 30.9]
	No machine learning	−2.2 [−7.1, 2.6]	−39.8 [−48.0, −31.5]	35.3 [27.0, 43.6]
DF (%)	Linear regression	−0.2 [−0.6, 0.2]	−3.3 [−4.0, −2.6]	2.9 [2.3, 3.6]
	Support vector regression	−0.6 [−1.0, −0.2]	−3.9 [−4.6, −3.2]	2.7 [2.0, 3.5]
	Two-layer neural network	−0.2 [−0.6, 0.2]	−3.3 [−4.0, −2.6]	2.9 [2.2, 3.5]
	No machine learning	0.3 [−0.3, 1.0]	−4.8 [−6.0, −3.7]	5.5 [4.3, 6.6]
$F_{v,\max}$ (BW)	Linear regression	0.00 [−0.03, 0.04]	−0.24 [−0.29, −0.18]	0.25 [0.19, 0.30]
	Support vector regression	0.00 [−0.03, 0.04]	−0.24 [−0.30, −0.19]	0.25 [0.20, 0.31]
	Two-layer neural network	0.00 [−0.03, 0.04]	−0.25 [−0.30, −0.19]	0.26 [0.20, 0.31]
	No machine learning	−0.03 [−0.07, 0.01]	−0.31 [−0.37, −0.25]	0.25 [0.19, 0.31]

Confidence intervals of 95% are given in square brackets [lower, upper]. For systematic bias, positive values indicate the estimated or predicted variable is overestimated.

Table 6. Percentage difference of the Pearson correlation coefficient (r), root mean square error (RMSE) and mean absolute percentage error (MAPE) between those obtained using estimations based on inertial-measurement unit data and those obtained using a machine learning model among linear regression, support vector regression with the radial basis function kernel and two-layer neural network, for four predicted variables, i.e., contact time, flight time, duty factor and peak vertical ground reaction force.

Variable	Metrics	Linear regression (%)	Support vector regression (%)	Two-layer neural network (%)
Contact time	r	29	27	27
	RMSE	−38	−36	−36
	MAPE	−40	−42	−40
Flight time	r	59	61	57
	RMSE	−39	−41	−37
	MAPE	−38	−43	−38
Duty factor	r	67	59	69
	RMSE	−40	−32	−40
	MAPE	−42	−37	−43
Peak vertical ground reaction force	r	16	15	13
	RMSE	−20	−13	−13
	MAPE	−11	−9	−8

models (SVR and NN2). Therefore, the simplest ML model (LR) should be used to improve the accuracy of the estimations of t_c, t_f, DF and $F_{v,\max}$ obtained using a sacral-mounted IMU across a range of running speeds. These improvements may be beneficial when monitoring running-related injury risk factors in real-world settings.

ML was able to improve the prediction accuracy, as reported by the higher r and lower RMSE and MAPE compared to those of the IMU-based estimations (Figures 2 and 3 and Table 6). Nonetheless, the enhancement reported for $F_{v,\max}$ was not significant. Using more complicated ML models (SVR and NN2) did not further improve the prediction accuracy

Table 7. Optimal coefficients, standard error and *p*-values (*P*) obtained for the predictors used in the linear regression models, i.e., intercept, runner's body mass, stride frequency, running speed and estimated variable obtained using inertial measurement unit data, constructed to predict contact time (t_c), flight time (t_f), duty factor (DF) and peak vertical ground reaction force ($F_{v,max}$).

Variable	Metrics	Coefficient	Standard error	*P*
t_c (ms)	intercept	251.08	56.09	**<0.001**
	runner's body mass	5.40	0.11	**<0.001**
	stride frequency	9.40	23.22	0.69
	running speed	−13.71	0.60	**<0.001**
	estimated t_c	22.20	0.10	**<0.001**
t_f (ms)	intercept	110.64	28.24	**<0.001**
	runner's body mass	−5.33	0.11	**<0.001**
	stride frequency	−3.35	15.80	0.83
	running speed	13.80	0.61	**<0.001**
	estimated t_f	12.50	0.10	**<0.001**
DF (%)	intercept	34.68	4.67	**<0.001**
	runner's body mass	0.74	0.02	**<0.001**
	stride frequency	0.67	1.92	0.73
	running speed	−1.92	0.08	**<0.001**
	estimated DF	1.56	0.10	**<0.001**
$F_{v,max}$ (BW)	intercept	2.51	0.27	**<0.001**
	runner's body mass	0.01	0.00	**<0.001**
	stride frequency	−0.01	0.14	0.93
	running speed	0.07	0.01	**<0.001**
	estimated $F_{v,max}$	0.15	0.07	**0.02**

Significant coefficients ($P \leq 0.05$) are reported in bold font.

Table 8. Optimal coefficients, standard error and *p*-values (*P*) obtained for the linear regression models which excluded stride frequency as a predictor to predict contact time (t_c), flight time (t_f), duty factor (DF) and peak vertical ground reaction force ($F_{v,max}$).

Variable	Metrics	Coefficient	Standard error	*P*
t_c (ms)	intercept	251.08	20.84	**<0.001**
	runner's body mass	5.29	0.11	**<0.001**
	running speed	−13.27	0.64	**<0.001**
	estimated t_c	14.42	0.06	**<0.001**
t_f (ms)	intercept	110.64	13.43	**<0.001**
	runner's body mass	−4.79	0.11	**<0.001**
	running speed	12.27	0.53	**<0.001**
	estimated t_f	13.88	0.09	**<0.001**
DF (%)	intercept	34.68	4.17	**<0.001**
	runner's body mass	0.63	0.02	**<0.001**
	running speed	−1.61	0.08	**<0.001**
	estimated DF	1.56	0.11	**<0.001**
$F_{v,max}$ (BW)	intercept	2.51	0.19	**<0.001**
	runner's body mass	0.01	0.00	**<0.001**
	running speed	0.07	0.01	**<0.001**
	estimated $F_{v,max}$	0.15	0.07	**0.02**

Significant coefficients ($P \leq 0.05$) are reported in bold font.

compared to the simple LR (Figure 2 and Table 6). These results corroborate previous findings which observed similar errors for LR and quantile regression forest when predicting t_c, $F_{v,max}$ and vertical impulse with an accelerometer (Alcantara et al., 2021). Moreover, the present RMSE and MAPE of t_c and $F_{v,max}$ were similar to those previously obtained (t_c: ~10 ms and ~4% and $F_{v,max}$: ~0.14BW and ~4%) using a different algorithm to estimate t_c and $F_{v,max}$ from IMU data (Alcantara et al., 2021). Nonetheless, these previous results might suffer from generalisation due to the small sample size ($N = 37$). Using three inertial

sensors placed on the lower limb (two on lower leg and one on pelvis), Wouda et al. (2018) achieved a 3% error with a NN (10–14 km/h), which is similar to the present accuracy (MAPE ~4%, Figure 2). Despite their low prediction error, their results were harder to interpret because of the experimental setup (three IMUs instead of one) and more complicated ML model than the model employed herein. Practically, the improvements reported herein may be beneficial for practitioners seeking to monitor running-related injury risk factors in real-world settings, though keeping in mind that there exists only limited evidence for most running-related injury-specific risk factors (Willwacher et al., 2022). Besides, as asymmetry level might be an important factor to consider for injured runner (Russell Esposito et al., 2015), an ML model should be used to predict the biomechanical variables of the right and left lower limbs separately. Moreover, as the biomechanical variables of an injured lower limb might give different values than the ones used in the current training set (healthy individuals), the ML model should further be trained using injured runners and by separating the values of the biomechanical variables of the injured and non-injured lower limb in the training process.

ML was able to decrease the confidence limits (95% confidence intervals and lower and upper limit of agreements) compared to those of the IMU-based estimations (Table 5). In addition, the systematic bias reported for the simple linear regression was smaller than the bias obtained without ML (Table 5). Moreover, Figure 3 suggests that the IMU-based estimations have a proportional bias (i.e., the error depends on the value of the estimated parameter). This proportional bias drastically decreased when using ML. Hence, these results strengthen the use of ML to obtain more accurate predictions.

SF was not reported as a predictor of the four variables ($P \geq 0.69$; Table 7). Hence, LRs which consider only the body mass, running speed and IMU-based estimation should be used to improve the prediction accuracy (see Table 8 for the coefficients). The optimal coefficients of an ML model might be specific to the IMU-based estimations used in the training and testing sets because each algorithm used to obtain these IMU-based estimations might have its own bias. Hence, these LRs can be used to predict t_c, t_f and DF, and $F_{v,\max}$ to a lower extent, as long as the IMU-based estimations were obtained using the present algorithm, which is described elsewhere (Patoz et al., 2022). Nonetheless, further studies should try to create an ML model based on IMU-based estimations obtained from different algorithms, so that its usage could largely be generalised.

Previously, ML was also used to predict the vertical impulse from its IMU-based estimation as well as body mass, running speed and step frequency (Alcantara et al., 2021). The authors reported an almost perfect correlation between gold standard and predicted vertical impulse values ($r = 0.995$) and obtained that the intercept and step frequency of the LR were the only significant predictors of the vertical impulse. However, this was not necessarily needed. Indeed, as the integral of the vertical external forces during a running step is null (Equation 1):

$$\int_0^{t_c} F_z(t)dt - mg\left(t_c + t_f\right) = 0, \tag{1}$$

we get

$$t_{\text{step}} = \frac{\int_0^{t_c} F_z(t)dt}{mg} = \frac{I_z}{mg}, \tag{2}$$

where $t_{step} = t_c + t_f$ and I_z represent the step time and vertical impulse, respectively. Therefore, according to Equation (2), the step frequency, i.e., the inverse of t_{step}, is given by the inverse of the vertical impulse expressed in body weight units. Hence, the model created by Alcantara et al. (2021) to predict the vertical impulse was redundant and not necessarily needed. First, the vertical impulse is directly given by t_{step} and thus by the inverse of the step frequency (Equation 2). Second, they assumed that the step frequency estimated using IMU data is a valid surrogate to its gold standard counterpart (they used the step frequency estimated using IMU data as a predictor for the vertical impulse, t_c, and $F_{z,max}$). Thus, they already indirectly assumed that the estimated vertical impulse, i.e., t_{step} (the inverse of the step frequency), is equivalent to its gold standard counterpart. In the present study, gold standard and estimated SF were shown to be equivalent ($r = 0.998$; Figure 1), which corroborates what has just been explained. Indeed, t_{step} could be approximated by half of the stride time because small symmetry indices ≤4% were previously reported for t_{step} of competitive, recreational and novice runners at running speeds ranging from 8 to 12 km/h (Mo et al., 2020).

As expected, as gold standard and estimated SF were equivalent ($r = 0.998$; Figure 1), similar MAPEs were reported between t_c and DF (~4%; Figure 2). Thus, the DF prediction is almost only dependent on the t_c prediction. Finally, it is worth mentioning that using predicted t_c values and IMU-based estimations of SF to predict DF instead of constructing a specific LR led to a slightly larger prediction accuracy. Indeed, using predicted t_c values from the LR reported in Table 8, DF was predicted with an r of 0.82, RMSE of 1.8% and MAPE of 3.9 ± 3.3%.

The strength of the present results is due to the large dataset employed ($N = 100$). This dataset allows better generalisation of the results than those previously obtained with the smaller cohorts of 37 runners (Alcantara et al., 2021), though the generalisation might not apply to populations not represented in the training set. Hence, further studies should include a broader population (increase N) by including elite athletes and less experienced runners. Moreover, injured runners should also be included in the training set, and the values of the biomechanical variables of the left and right lower limb should be separated in the training process, especially in the case of an asymmetry-based injury. In this case, the dataset would contain as much different running gaits as possible, which would make the trained ML models as much generalisable as possible. Further studies could also apply other ML models and even more complex models such as deep learning models, though their complexity makes them very difficult to interpret (Halilaj et al., 2018). Furthermore, running trials were performed only at level, endurance speeds and on a treadmill. However, predictions obtained using ML might also perform well overground because spatiotemporal parameters between treadmill and overground running are largely comparable (Van Hooren et al., 2020). Nonetheless, running speed must be known to use ML models. In real-life situation, the ML model could use the instantaneous running speed provided by the gps of the smartwatch or smartphone to predict the biomechanical variables in real-time. Finally, further studies should focus on improving the predictions by using additional conditions (i.e., faster speeds, positive and negative slopes and different types of ground) when training the ML models.

Conclusion

Further applying ML to IMU-based estimations of t_c, t_f, DF and $F_{v,\max}$ increased the accuracy of their predictions, though the enhancement was not significant for $F_{v,\max}$. The simplest ML model (LR) was characterised by a similar prediction accuracy than more complicated models (SVR and NN2). Moreover, errors of the ML models were equal to or smaller than the SRD for the four variables, while errors of the estimations were not, indicating that ML models were sufficiently accurate to detect a clinically important difference. Therefore, the simplest ML model (LR) should be used to improve the accuracy of the estimations of t_c, t_f, DF and $F_{v,\max}$ obtained using a sacral-mounted IMU across a range of running speeds. These improvements may be beneficial for practitioners seeking to monitor running-related injury risk factors in real-world settings.

Acknowledgments

The authors warmly thank the participants for their time and cooperation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the Innosuisse grant no. 35793.1 IP-LS.

ORCID

Aurélien Patoz http://orcid.org/0000-0002-6949-7989
Thibault Lussiana http://orcid.org/0000-0002-1782-401X
Bastiaan Breine http://orcid.org/0000-0002-7959-7721
Davide Malatesta http://orcid.org/0000-0003-3905-5642

Author contributions

Conceptualization, A.P., T.L., C.G. and D.M.; methodology, A.P., T.L., C.G. and D.M.; investigation, A.P., T.L. and B.B.; formal analysis, A.P. and B.B.; writing—original draft preparation, A.P.; writing—review and editing, A.P., T.L., B.B., C.G. and D.M.; supervision, A.P., T.L., C.G. and D.M.

Data availability statement

The datasets and codes supporting this article are available online at https://github.com/aurelienPatoz/predictions-of-tc-tf-DF-Fvmax-using-machine-learning.

References

Alcantara, R. S., Day, E. M., Hahn, M. E., & Grabowski, A. M. (2021). Sacral acceleration can predict whole-body kinetics and stride kinematics across running speeds. *PeerJ*, *9*, e11199. https://doi.org/10.7717/peerj.11199

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, *26*(4), 217–238. https://doi.org/10.2165/00007256-199826040-00002

Backes, A., Skejø, S. D., Gette, P., Nielsen, R. Ø., Sørensen, H., Morio, C., & Malisoux, L. (2020). Predicting cumulative load during running using field-based measures. *Scandinavian Journal of Medicine & Science in Sports*, *30*(12), 2399–2407. https://doi.org/10.1111/sms.13796

Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement: Why plotting difference against standard method is misleading. *The Lancet*, *346*(8982), 1085–1087. https://doi.org/10.1016/s0140-6736(95)91748-9

Camomilla, V., Bergamini, E., Fantozzi, S., & Vannozzi, G. (2018). Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review. *Sensors*, *18*(3), 873. https://doi.org/10.3390/s18030873

Chambers, J. (1992). Chapter 4: linear models. In J. Chambers & T. Hastie (Eds.), *Statistical models in S*. Wadsworth & Brooks/Cole.

Chew, D. -K., Ngoh, K.J. -H., Gouwanda, D., & Gopalai, A. A. (2018). Estimating running spatial and temporal parameters using an inertial sensor. *Sports Engineering*, *21*(2), 115–122. https://doi.org/10.1007/s12283-017-0255-9

Day, E. M., Alcantara, R. S., McGeehan, M. A., Grabowski, A. M., & Hahn, M. E. (2021). Low-pass filter cutoff frequency affects sacral-mounted inertial measurement unit estimations of peak vertical ground reaction force and contact time during treadmill running. *Journal of Biomechanics*, *119*, 110323. https://doi.org/10.1016/j.jbiomech.2021.110323

Derie, R., Robberechts, P., Van den Berghe, P., Gerlo, J., De Clercq, D., Segers, V., & Davis, J. (2020). Tibial acceleration-based prediction of maximal vertical loading rate during overground running: A machine learning approach. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.00033

Edwards, W. B. (2018). Modeling overuse injuries in sport as a mechanical fatigue phenomenon. *Exercise and Sport Sciences Reviews*, *46*(4), 224–231. https://doi.org/10.1249/JES.0000000000000163

Falbriard, M., Meyer, F., Mariani, B., Millet, G. P., & Aminian, K. (2018). Accurate estimation of running temporal parameters using foot-worn inertial sensors. *Frontiers in Physiology*, *9*(610). https://doi.org/10.3389/fphys.2018.00610

Folland, J. P., Allen, S. J., Black, M. I., Handsaker, J. C., & Forrester, S. E. (2017). Running technique is an important component of running economy and performance. *Medicine & Science in Sports & Exercise*, *49*(7), 1412–1423. https://doi.org/10.1249/MSS.0000000000001245

Fredette, A., Roy, J. S., Perreault, K., Dupuis, F., Napier, C., & Esculier, J. F. (2021). The association between running injuries and training parameters: A systematic review. *Journal of Athletic Training*, *57*(7), 650–671. https://doi.org/10.4085/1062-6050-0195.21

Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J. L., Hastie, T. J., & Delp, S. L. (2018). Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *Journal of Biomechanics*, *81*, 1–11. https://doi.org/10.1016/j.jbiomech.2018.09.009

Heise, G. D., & Martin, P. E. (2001). Are variations in running economy in humans associated with ground reaction force characteristics? *European Journal of Applied Physiology*, *84*(5), 438–442. https://doi.org/10.1007/s004210100394

Hreljac, A. (2004). Impact and overuse injuries in runners. *Medicine and Science in Sports and Exercise*, *36*(5), 845–849. https://doi.org/10.1249/01.mss.0000126803.66636.dd

Hreljac, A., Marshall, R. N., & Hume, P. A. (2000). Evaluation of lower extremity overuse injury potential in runners. *Medicine and Science in Sports and Exercise*, *32*(9), 1635–1641. https://doi.org/10.1097/00005768-200009000-00018

International standard ISO/IEC. (2020). Programming languages—C++. Geneva, Switzerland: International Organization for Standardization

Kiernan, D., Hawkins, D. A., Manoukian, M. A. C., McKallip, M., Oelsner, L., Caskey, C. F., & Coolbaugh, C. L. (2018). Accelerometer-based prediction of running injury in national collegiate athletic association track athletes. *Journal of Biomechanics*, *73*, 201–209. https://doi.org/10.1016/j.jbiomech.2018.04.001

Lee, J. B., Mellifont, R. B., & Burkett, B. J. (2010). The use of a single inertial sensor to identify stride, step, and stance durations of running gait. *Journal of Science and Medicine in Sport*, *13*(2), 270–273. https://doi.org/10.1016/j.jsams.2009.01.005

Lenhart, R. L., Thelen, D. G., Wille, C. M., Chumanov, E. S., & Heiderscheit, B. C. (2014). Increasing running step rate reduces patellofemoral joint forces. *Medicine and Science in Sports and Exercise*, *46*(3), 557–564. https://doi.org/10.1249/MSS.0b013e3182a78c3a

Malisoux, L., Gette, P., Delattre, N., Urhausen, A., & Theisen, D. (2022). Spatiotemporal and ground-reaction force characteristics as risk factors for running-related injury: A secondary analysis of a randomized trial including 800+ recreational runners. *The American Journal of Sports Medicine*, *50*(2), 537–544. https://doi.org/10.1177/03635465211063909

Matijevich, E. S., Branscombe, L. M., Scott, L. R., & Zelik, K. E. (2019). Ground reaction force metrics are not strongly correlated with tibial bone load when running across speeds and slopes: Implications for science, sport and wearable tech. *PLoS One*, *14*(1), e0210000. https://doi.org/10.1371/journal.pone.0210000

Matijevich, E. S., Scott, L. R., Volgyesi, P., Derry, K. H., & Zelik, K. E. (2020). Combining wearable sensor signals, machine learning and biomechanics to estimate tibial bone force and damage during running. *Human Movement Science*, *74*, 102690. https://doi.org/10.1016/j.humov.2020.102690

Minetti, A. E. (1998). A model equation for the prediction of mechanical internal work of terrestrial locomotion. *Journal of Biomechanics*, *31*(5), 463–468. https://doi.org/10.1016/S0021-9290(98)00038-4

Mo, S., Lau, F. O. Y., Lok, A. K. Y., Chan, Z. Y. S., Zhang, J. H., Shum, G., & Cheung, R. T. H. (2020). Bilateral asymmetry of running gait in competitive, recreational and novice runners at different speeds. *Human Movement Science*, *71*, 102600. https://doi.org/10.1016/j.humov.2020.102600

Nagahara, R., Takai, Y., Kanehisa, H., & Fukunaga, T. (2018). Vertical impulse as a determinant of combination of step length and frequency during sprinting. *International Journal of Sports Medicine*, *39*(04), 282–290. https://doi.org/10.1055/s-0043-122739

Napier, C., Jiang, X., MacLean, C. L., Menon, C., & Hunt, M. A. (2020). The use of a single sacral marker method to approximate the centre of mass trajectory during treadmill running. *Journal of Biomechanics*, *108*, 109886. https://doi.org/10.1016/j.jbiomech.2020.109886

Nielsen, R. O., Buist, I., Sørensen, H., Lind, M., & Rasmussen, S. (2012). Training errors and running related injuries: A systematic review. *International Journal of Sports Physical Therapy*, *7*(1), 58–75.

Nilsson, J., & Thorstensson, A. (1989). Ground reaction forces at different speeds of human walking and running. *Acta Physiologica Scandinavica*, *136*(2), 217–227. https://doi.org/10.1111/j.1748-1716.1989.tb08655.x

Norris, M., Anderson, R., & Kenny, I. C. (2014). Method analysis of accelerometers and gyroscopes in running gait: A systematic review. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, *228*(1), 3–15. https://doi.org/10.1177/1754337113502472

Patoz, A., Lussiana, T., Breine, B., Gindre, C., & Malatesta, D. (2022). A single sacral-mounted inertial measurement unit to estimate peak vertical ground reaction force, contact time, and flight time in running. *Sensors*, *22*(3), 784. https://doi.org/10.3390/s22030784

Russell Esposito, E., Choi, H. S., Owens, J. G., Blanck, R. V., & Wilken, J. M. (2015). Biomechanical response to ankle-foot orthosis stiffness during running. *Clinical Biomechanics*, *30*(10), 1125–1132. https://doi.org/10.1016/j.clinbiomech.2015.08.014

Sasimontonkul, S., Bay, B. K., & Pavol, M. J. (2007). Bone contact forces on the distal tibia during the stance phase of running. *Journal of Biomechanics*, *40*(15), 3503–3509. https://doi.org/10.1016/j.jbiomech.2007.05.024

Scott, S. H., & Winter, D. A. (1990). Internal forces of chronic running injury sites. *Medicine and Science in Sports and Exercise*, *22*(3), 357–369. https://doi.org/10.1249/00005768-199006000-00013

Selinger, J. C., Hicks, J. L., Jackson, R. W., Wall-Scheffler, C. M., Chang, D., & Delp, S. L. (2022). Running in the wild: Energetics explain ecological running speeds. *Current Biology*, *32*(10), 2309–2315. https://doi.org/10.1016/j.cub.2022.03.076

Smith, L., Preece, S., Mason, D., & Bramah, C. (2015). A comparison of kinematic algorithms to estimate gait events during overground running. *Gait & Posture*, *41*(1), 39–43. https://doi.org/10.1016/j.gaitpost.2014.08.009

Van Hooren, B., Fuller, J. T., Buckley, J. D., Miller, J. R., Sewell, K., Rao, G., Willy, R. W. … Willy, R. W. (2020). Is motorized treadmill running biomechanically comparable to overground running? A systematic review and meta-analysis of cross-over studies. *Sports Medicine*, *50*(4), 785–813. https://doi.org/10.1007/s40279-019-01237-z

Willwacher, S., Kurz, M., Robbin, J., Thelen, M., Hamill, J., Kelly, L., & Mai, P. (2022). Running-related biomechanical risk factors for overuse injuries in distance runners: a systematic review considering injury specificity and the potentials for future research. *Sports Medicine*, *52*(8), 1863–1877. https://doi.org/10.1007/s40279-022-01666-3

Wouda, F. J., Giuberti, M., Bellusci, G., Maartens, E., Reenalda, J., van Beijnum, B.-J.F., & Veltink, P. H. (2018). Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors. *Frontiers in Physiology*, *9*(218). https://doi.org/10.3389/fphys.2018.00218

Xu, D., Quan, W., Zhou, H., Sun, D., Baker, J. S., & Gu, Y. (2022). Explaining the differences of gait patterns between high and low-mileage runners with machine learning. *Scientific reports*, *12*(1), 2981. https://doi.org/10.1038/s41598-022-07054-1

Appendix. The Relation between Flight Time and Net Vertical Impulse

The integral of the vertical external forces during a running step is null. Hence, t_f relates to the net vertical impulse ($I_{z,\text{net}}$), i.e., the integral of the vertical ground reaction force (F_z), which is above body weight during t_c (Equations A1 and A2) (Heise & Martin, 2001)

$$\int_0^{t_c}(F_z(t) - mg)dt - mg\,t_f = 0, \tag{A1}$$

$$t_f = \frac{\int_0^{t_c}(F_z(t) - mg)dt}{mg} = \frac{I_{z,\text{net}}}{mg}. \tag{A2}$$

Therefore, t_f takes both the vertical ground reaction force and its time of production into account. Hence, t_f might play a role in running-related injury development.