REVIEW ARTICLE

DOI: 10.54392/ijpefs2643

International Journal of
PHYSICAL EDUCATION, FITNESS AND SPORTS

Physics-Informed Explainable Machine Learning for Estimating Lower-Limb Loading Across Running Speeds using Open Three-Dimensional Biomechanics Data

Samuel Asante [a], Archibald Danquah-Amoah [b,*], Alex Abrokwa Fosuhene [c], Dominic Akwasi Bonah [c]

[a] Department of Computer Science, Accra Technical University, Ghana.
[b] Department of Medical Laboratory Technology, Biomedical Equipment Engineering. Accra Technical University, Ghana.
[c] Department of Science Laboratory Technology, Accra Technical University, Ghana.
* Corresponding Author E-mail: adanquah-amoah@atu.edu.gh
DOI: https://doi.org/10.54392/ijpefs2643
Received: 17-04-2026; Revised: 24-10-2026; Accepted: 06-11-2026; Published: 20-11-2026

Abstract: Peak vertical ground reaction force is a common marker of running load, but it is normally measured on force plates that runners do not have outside the laboratory. This study asked whether peak force in runners excluded from model development can be estimated from marker-based kinematics alone, and whether a physical model of running adds to machine learning. Previously collected treadmill trials of 39 runners at 2.5, 3.5 and 4.5 m/s from a public biomechanics dataset gave 101 participant-speed trials. Outcome measures were derived exclusively from force signals, and predictors were derived exclusively from marker trajectories, with contact and flight times detected kinematically. Linear, kernel and tree-based learners were compared with a calibrated spring-mass estimate and with a hybrid in which machine learning corrected the spring-mass residual, using leave-one-runner-out validation with all tuning nested inside the training runners. Kinematic contact time agreed closely with force-based contact time ($r = 0.96$). The hybrid estimated peak force with a mean absolute error of 0.090 body weights and R^2 of 0.70, against 0.132 body weights for a speed-only baseline. The calibrated spring-mass model alone reached 0.090 body weights and was as accurate as the best purely data-driven model. When a whole running speed was withheld, the hybrid and ridge regression transferred better than either the spring-mass model or gradient boosting. Loading rate was estimated less well (R^2 0.50). A spring-mass model with two fitted calibration coefficients therefore captures most of the information in kinematic measurements about peak running force, and machine learning should be judged against it.

Keywords: Running Biomechanics, Ground Reaction Force, Spring-Mass Model, Machine Learning, Leave-One-Subject-Out Validation, Explainable Modelling.

1. Introduction

Running is one of the most widely practised forms of physical activity. It needs little equipment, it can be done almost anywhere, and it is used in school physical education, in recreational fitness and in the preparation of athletes for many sports. The same features that make running popular also expose large numbers of people to repetitive loading of the lower limb, and running-related injuries are frequent enough that coaches, clinicians and physical educators routinely ask how much mechanical load a runner absorbs with each step. A systematic review of prospective studies found that only a few biomechanical variables have consistent links with overuse injury in distance runners, and that these links depend on the injury in question (Willwacher *et al.*, 2022). A large secondary analysis of a randomised trial including more than 800 recreational runners showed that spatiotemporal and ground reaction force characteristics carry part of the injury risk (Malisoux *et al.*, 2022).

The vertical ground reaction force is the most common description of external load in running. Its peak, usually between two and three times body weight during the mid-stance of each step, and the rate at

which it rises after initial contact, are reported in most laboratory studies of running mechanics. Whether impact forces directly cause injury has been doubted for both theoretical and empirical reasons (Gruber, 2023). Even so, vertical force and loading rate are still the measures most often used to describe how hard ground reaction force is transmitted through the body, and they separate groups such as recreational and collegiate runners (Evans *et al.,* 2023). Of all the factors that affect these quantities, running speed has the largest effect. Faster running brings a higher cadence, a shorter contact time and greater peak force and loading rate (Leacox *et al.,* 2025); reference values from large samples of recreational runners show that speed alters nearly every spatiotemporal and kinetic variable.

There are practical reasons why researchers now want to measure load away from the laboratory. Coaches and therapists often advise runners to change step rate or foot-strike pattern in order to reduce loading, and systematic reviews have examined what such changes do to injury, performance and biomechanics (Anderson *et al.,* 2022). Recent narrative work has also argued that some common beliefs about heel striking and cadence are not well supported by evidence (Patoz *et al.,* 2026). Whatever the merits of a particular intervention, judging it requires a measure of load that can be collected repeatedly, in the setting where the runner trains. Force plates and instrumented treadmills provide the reference measurement but they are expensive and fixed to a laboratory, and even the laboratory measurement is usually limited to a handful of steps on a single day.

Machine learning has therefore been used to estimate ground reaction forces from signals that are easier to collect, such as plantar pressure (Honert *et al.,* 2022), inertial sensors and joint kinematics. The resulting studies, reviewed in the next section, report useful accuracy, but two features limit how far their results can be trusted for practice. First, the way in which data are divided for testing is not always reported clearly, and when strides or trials from the same person appear in both training and test data, the result reflects familiarity with that person as much as an understanding of running. This sort of leakage is now known to inflate results in machine-learning research in general (Kapoor & Narayanan, 2023). Second, few studies check machine-learning estimates against what well-established mechanics would already predict. Should a simple physical model account for most of the variation in peak force, a complex learner would add little and could obscure the basis of its own predictions.

The present study tackled both issues. An open three-dimensional dataset of treadmill running at three speeds was used. The force signal served only for outcomes and the marker data only for predictors, and every model was tested on runners withheld from all training and tuning. There were three aims. The first was to estimate peak vertical ground reaction force and loading rate in unseen runners from kinematics alone. The second was to compare data-driven learners with a calibrated spring-mass estimate and with a physics-informed hybrid, in which machine learning is used only to correct the spring-mass residual. The third was to see how well each approach generalises to a running speed missing from the training data. It was hypothesised that nonlinear learners would improve on linear and speed-only baselines, and that a hybrid would combine the stability of the physical model with the flexibility of machine learning. The questions concern estimation only; injury prediction is not tested here. The remainder of the paper reviews the relevant literature, describes the data and methods, reports and discusses the results, and closes with conclusions for practice and research.

The study makes four contributions. It provides a participant-independent benchmark for estimating peak vertical force and loading rate from kinematics on an open dataset that others can reuse, with the code needed to reproduce every number. It puts forward the calibrated spring-mass model as an explicit baseline for machine-learning force estimators and shows that the baseline provided competitive predictive accuracy. It proposes a physics-informed hybrid in which a learner corrects the physical estimate rather than replacing it, and tests its behaviour when a running speed is absent from training. Finally, it explains the fitted models in mechanical terms, so that practitioners can see why an estimate is high or low rather than receiving a number from an opaque system.

2. Literature Survey

2.1 Running Mechanics, Running Style and External Load

A useful starting point is the description of running as a sequence of contact and flight phases whose timing reflects the way each runner manages vertical and horizontal motion of the body. A synthesis of the biomechanics of running styles argued that much of the variation between runners can be summarised by a few global measures, especially duty factor (the fraction of the stride spent in contact) and step frequency, and that these measures capture how much

a runner relies on vertical bouncing as opposed to longer ground contact (van Oeveren *et al.,* 2024). A systematic review and meta-analysis linking running biomechanics with running economy reached a related conclusion: spatiotemporal variables and the vertical motion of the body are among the variables most consistently related to the energy cost of running (Van Hooren et *al.,* 2024).

Several recent studies have connected these global measures to external load. Field measurements of cadence and duty factor in recreational runners showed that both can be modulated deliberately and that they map out distinct running styles (Nijs *et al.,* 2023). A laboratory study found that habitual duty factor, rather than stride frequency, was related to loading magnitude (Van der Meulen *et al.,* 2024). For individual runners, duty factor and foot-strike pattern were found to describe different things, so one cannot simply stand in for the other (Patoz *et al.,* 2022). In heel-toe running, foot strike determines how the centre of pressure moves and how severe the impact is (Van den Berghe *et al.,* 2022), and rearfoot strikers with low loading rates differ in particular kinematic features from those with high rates. Slope affects loading rate as well, which in turn relates to oxygen consumption in inclined running (Lemire *et al.,* 2022). Taken together, these studies suggest that timing variables and a small number of kinematic descriptors should carry much of the information needed to estimate external load, which is the premise of the present work.

2.2 The Spring-Mass Model of Running

The spring-mass model treats the runner as a point mass bouncing on a massless linear leg spring. Despite its simplicity, it reproduces the main features of the vertical force curve and gives closed-form relations between contact time, flight time, peak force and leg stiffness. Because these relations need only timing information, the model is widely used in applied research. It has been used to describe mechanical asymmetries during graded treadmill tests (Girard *et al.,* 2022), (Mtibaa *et al.,* 2023), to characterise elite distance runners in matched comparisons across speeds (Mayo-Nieto *et al.,* 2026) and to examine how vertical and leg stiffness change with speed and incline. It has also been applied to follow spring-mass characteristics before and after a 56 km ultramarathon (Burns *et al.,* 2024) and to compare leg stiffness in highly cushioned shoes with a carbon-fibre plate and in traditional shoes (Hata *et al.,* 2022). In most of this work the model is used to derive descriptive variables rather than to estimate force for runners whose force has not been measured, and its accuracy as a stand-alone estimator in unseen runners is rarely reported. Nor has it been used, to the author's knowledge, as a formal baseline for machine-learning estimators of running force.

2.3 Machine Learning for Ground Reaction Force Estimation

The largest body of related work uses inertial measurement units. A long short-term memory network was trained to map data from three inertial sensors to insole-measured force waveforms during fixed-pace running outside the laboratory (Donahue & Hahn, 2023b), and the same group estimated gait events and kinetic waveforms with wearable sensors in an unconstrained environment (Donahue & Hahn, 2023a). Consumer-priced pressure insoles combined with an inertial sensor and deep learning estimated vertical and anteroposterior force waveforms in fifty runners across speeds and gradients (Carter *et al.,* 2024). Linear regression, support-vector regression and a neural network were compared for improving sacral-acceleration estimates of temporal variables and peak vertical force in one hundred runners, and the authors recommended the simplest model because the learners performed alike (Patoz *et al.,* 2025). Convolutional and recurrent architectures have been used to predict vertical force from lower-limb joint angles at five speeds (Chen *et al.,* 2025) and from a single inertial sensor in rearfoot and forefoot strikers (Li *et al.,* 2025). A recent study estimated vertical force from inertial sensors placed at the wrist and sternum, the positions of a smartwatch and a heart-rate strap, and validated the estimates outdoors (Binnekamp *et al.,* 2026).

Other work has broadened the outputs and the methods. Lightweight embedding regression and nearest-neighbour models were compared with recurrent networks and found to be similarly accurate, with personal data reducing error (Song *et al.,* 2025). Braking and propulsion forces have been estimated during overground running in and out of the laboratory, with fine-tuning on a few individual strides (Baker *et al.,* 2025). Other groups have predicted Achilles tendon and patellar ligament loads from inertial data by pairing musculoskeletal simulation with gradient boosting (Rasmussen *et al.,* 2023), and estimated tibial bone load from sensors placed on the tibias and the sternum (van Middelaar & Reenalda, 2025). Three-dimensional kinematics and kinetics have also been reconstructed from virtual inertial data with optimal-control simulation of full-body musculoskeletal models (Nitschke *et al.,*

2024). Work has also covered sparse sensor layouts for inferring treadmill running motion (Pitts *et al.,* 2025) and the reliability of inertial spatiotemporal and kinetic measures in endurance runners (Miqueleiz *et al.,* 2025), and there is now an open multi-modal dataset for benchmarking force estimation from consumer wearables (Ghaffarzadeh *et al.,* 2026).

Two trends stand out in this body of work. One is a gradual shift towards more complex models, from linear regression and shallow networks to recurrent, convolutional and hybrid deep networks, frequently trained on only a few dozen people. The second is variety in what is reported: some studies give errors for whole waveforms, others for peaks or impulses, some express error in body weights, others as percentages or correlation coefficients, and the number of runners held out for testing ranges from one or two to all participants in turn. This variety makes it hard to judge whether added complexity brings added accuracy, and it is one reason why a common, simple reference model would be useful.

2.4 Kinematic Inputs from Markerless and Marker-Based Capture

The present study uses marker trajectories as inputs, which raises the question of whether the same inputs can be obtained in the field. Markerless motion capture has advanced quickly. A low-cost smartphone-based system was validated for lower-extremity kinematics during return-to-sport tasks (Turner *et al.,* 2024), and a recent study reported the validity and reliability of whole-body, joint and muscle kinematics during running from a markerless system. The predictors used here, namely contact and flight times, pelvis vertical excursion, heel position relative to the pelvis and sagittal knee and shank angles, are among the simplest quantities such systems provide. A model built on these inputs is therefore a realistic candidate for transfer to video-based or sensor-based capture.

2.5 Physics-Informed Machine Learning

Physics-informed machine learning combines mechanistic knowledge with data-driven models, either by building physical relations into the model structure or by using a physical model as a baseline whose residual is learned. In biomechanics, adding physical constraints improved the detection of head impacts from wearable sensor data (Raymond *et al.,* 2022), and physics-informed deep learning improved the estimation of muscle force from unlabelled surface electromyography (Ma *et al.,* 2024). Residual learning suits small datasets well. The physical model gives a reasonable estimate even where the learner has seen few examples, and the learner has only to account for what the physics misses. To the author's knowledge, nobody has yet used this form to estimate running ground reaction force from kinematics.

2.6 Validation Design and Data Leakage

In biomechanical datasets each participant usually contributes many strides or trials, and how these are split between training and testing determines what a reported error actually means. Leakage between training and test data has been blamed for irreproducible findings in many fields that use machine learning. Guidance for avoiding it in biological applications is now available (Bernett *et al.,* 2024), and in neuroimaging it was found to raise prediction performance considerably (Rosenblatt *et al.,* 2024). Movement data are highly personal; wrist accelerometry recorded during walking is enough to identify individuals (Koffman *et al.,* 2023). Among the force-estimation studies reviewed above, some held out whole participants, and others found that training on a runner's own data cut the error considerably. Both findings lead to the same conclusion. Accuracy for a known runner differs from accuracy for a new one, and only the latter matters when a tool is given to people who took no part in building it.

2.7 Comparison of Existing Studies and Research Gap

Table 1 summarises representative studies on the estimation of running ground reaction force, using the information reported in their abstracts. The studies differ in sensors, outputs and error measures, so their figures are not directly comparable with one another or with the present study, the table is intended to show design features rather than to rank methods.

Three gaps follow from this review. First, only one of the representative studies compared its learners with a non-learned estimate of peak force, and it found that the learned predictions were no better, none compared learners with a spring-mass model computed from the same timing information, so it is unknown how much of their accuracy comes from machine learning as such. Second, transfer to running speeds outside the training range is rarely tested, although practitioners will often want estimates at paces that were not covered during calibration. Third, the drivers of model predictions are seldom examined in mechanical terms.

Table 1. Representative studies estimating running ground reaction force with machine learning

Study	Participants and inputs	Model	Testing on unseen runners	Physical baseline
Honert *et al.*	18 runners, plantar pressure, speed, slope	Linear models, recurrent network	Reported, subject-specific training reduced error	No
Donahue and Hahn	15 runners, three IMUs	LSTM	Not stated in abstract	No
Carter *et al.*	50 runners, insoles and IMU	LSTM	Leave-one-subject-out	No
Patoz *et al.*	100 runners, sacral IMU	Linear, SVR, neural network	Held-out group of 20 runners	IMU-based estimate used as input
Chen *et al.*	12 runners, joint angles at five speeds	CNN-xLSTM	Not stated in abstract	No
Li *et al.*	15 runners, single IMU	Wavelet and feed-forward networks	Two runners held out	No
Song *et al.*	Runners, sacrum and shank IMUs	SVD regression, KNN, LSTM	Compared other-athlete and same-athlete training	No
Binnekamp *et al.*	11 runners, wrist and sternum IMUs	LSTM	Leave-one-subject-out	No
Present study	39 runners, marker kinematics at three speeds	Linear, SVR, tree ensembles, physics-informed hybrid	Leave-one-runner-out and speed-held-out	Calibrated spring-mass model

The present study responds to all three by reporting a calibrated spring-mass baseline, a physics-informed hybrid and a speed-held-out test, and by interpreting model importance in the light of running mechanics.

3. Materials and Methods

3.1 Dataset and Participants

For this secondary analysis, the public running biomechanics dataset of Fukuchi, Fukuchi and Duarte (Fukuchi *et al.*, 2017) was obtained from its repository (Figshare, DOI 10.6084/m9.figshare.4543435, version 5) on 6 November 2026, together with the updated participant information file. It contains three-dimensional marker trajectories sampled at 150 Hz and treadmill ground reaction forces sampled at 300 Hz for 30 s trials. All 39 runners contributed a trial at 3.5 m/s, and 31 of them also ran at 2.5 and 4.5 m/s, giving 101 participant-speed trials (Figure 1 outlines the analysis). In the original dataset, runners were 34.4 ± 6.9 years old, weighed 70.5 ± 8.9 kg and were 175.8 ± 7.1 cm tall, 38 were men. Sixty trials were classified as rearfoot strike, 21 as midfoot strike and 20 as forefoot strike in the source metadata. No trial was excluded. The original study reported ethical approval and informed consent; the de-identified dataset was made available for secondary analysis.

3.2 Loading Outcomes from the Force Signal

The vertical force was low-pass filtered with a fourth-order zero-lag Butterworth filter at 50 Hz. A stance phase was defined while the force exceeded 50 N, and contacts shorter than 120 ms or longer than 450 ms were discarded as artefacts of crossover steps or noise. For each contact, the peak vertical ground reaction force was divided by body weight (Eq. 1). Loading rate was the mean slope of the rising force between 20% and 80% of its peak, also expressed in body weights per second (Eq. 2). Both outcomes were averaged over all valid contacts of a trial, the median trial contained 84 contacts. Contact time from the force signal was kept as a reference for validating the kinematic event detection but was never used as a predictor.

$$y_{\text{peak}} = \frac{\max F_z(t)}{mg},\ t \in \text{stance} \quad (1)$$

$$LR_{20-80} = \frac{0.6\,F_{\text{peak}}}{mg(t_{80} - t_{20})} \quad (2)$$

Table 2. Participants, trials and loading measures by running speed (mean ± SD)

Speed (m/s)	Runners	Peak vGRF (BW)	Loading rate (BW/s)	Contact time (ms)	Cadence (steps/min)	Rearfoot / midfoot / forefoot
2.5	31	2.33 ± 0.14	26.7 ± 5.7	270 ± 25	161 ± 9	19 / 5 / 7
3.5	39	2.54 ± 0.15	36.2 ± 10.4	230 ± 24	171 ± 10	23 / 10 / 6
4.5	31	2.66 ± 0.20	56.0 ± 21.7	202 ± 32	181 ± 15	18 / 6 / 7

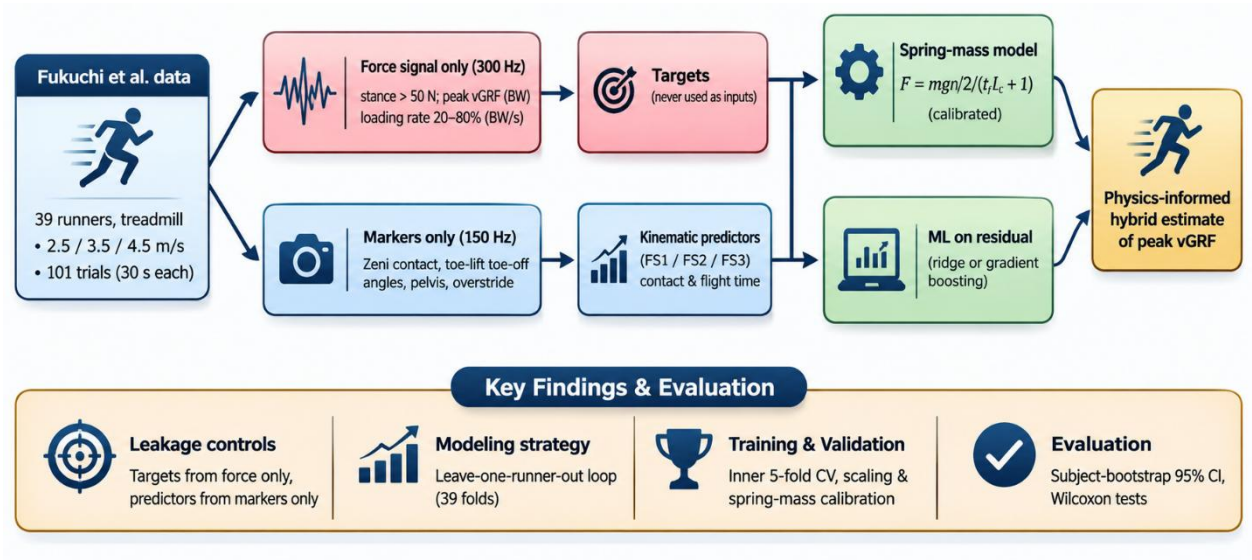


Figure 1. Analysis workflow and leakage controls.

Here F_z is the vertical force, m the body mass, g the gravitational acceleration, and t_{20} and t_{80} the instants at which the rising force first reaches 20% and 80% of its peak.

3.3 Kinematic Event Detection and Predictors

Marker trajectories were low-pass filtered at 12 Hz. Initial contact was taken as the instant of the most anterior position of the heel marker relative to the pelvis centroid, a coordinate-based approach that is widely used for treadmill running. Toe-off was the first frame after the stance minimum at which the mean height of the first and fifth metatarsal markers rose 15 mm. For each leg, the median stride time was computed, and step time T_{step} was half of it. Contact time t_c was the median interval from initial contact to toe-off, and flight time, cadence, duty factor and step length followed from Eqs. 3 to 6, where v is the treadmill speed.

$$t_f = T_{\text{step}} - t_c \quad (3)$$

$$\text{cadence} = \frac{60}{T_{\text{step}}} \quad (4)$$

$$\text{DF} = \frac{t_c}{2\,T_{\text{step}}} \quad (5)$$

$$L_{\text{step}} = v\,T_{\text{step}} \quad (6)$$

Against the force-based reference, kinematic contact time correlated at r = 0.964 with a mean bias of 29 ms (Figure 2). Two further kinematic descriptors were computed at every contact and summarised by their median: overstride, the horizontal distance between heel and pelvis centroid at initial contact, and vertical pelvis excursion, the drop of the pelvis centroid from initial contact to its lowest point during stance. Sagittal knee flexion at initial contact, peak knee flexion in stance, knee range of motion in stance, shank angle at initial contact and foot-strike angle were taken from the joint and segment angles of the marker model.

Three predictor sets were prespecified. FS1 contained speed, body mass, height, age and sex. FS2 added cadence, contact time, foot-strike angle and knee flexion at initial contact, representing what a simple field system might provide. FS3 added flight time, duty factor, step length, peak knee flexion, knee range of motion, shank angle, overstride and vertical pelvis excursion. No predictor was derived from the force signal.

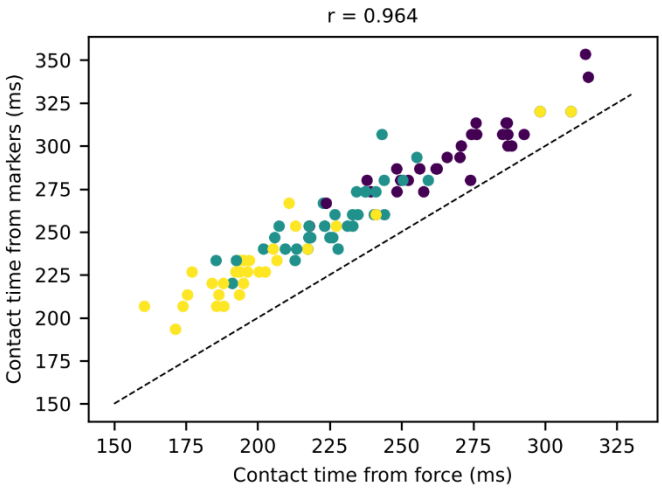


Figure 2. Agreement between contact time detected from markers and from the force signal (101 trials, colour indicates running speed).

3.4 Spring-Mass Estimate

If the vertical force during contact is approximated by a half sine wave, the vertical impulse over a step must equal body weight multiplied by step time. Solving for the peak gives the spring-mass estimate of peak force from contact time and flight time alone (Eq. 7). Expressed in body weights, the estimate depends only on the ratio of flight to contact time.

$$F_{\mathrm{sm}} = mg\,\frac{\pi}{2}\left(\frac{t_f}{t_c} + 1\right) \quad (7)$$

Because kinematic contact time is slightly biased and the half-sine shape is an approximation, the estimate was calibrated linearly on the training runners of each fold (Eq. 8). Fitted on all trials for illustration, the calibration had slope 1.03 and intercept 0.31 body weights, and the uncalibrated estimate correlated with measured peak force at r = 0.84. In every validation fold, only the training runners were used for this calibration.

$$\hat{y}_{\mathrm{sm}} = \beta_0 + \beta_1\,\frac{F_{\mathrm{sm}}}{mg} \quad (8)$$

3.5 Physics-Informed Hybrid

In the hybrid, the calibrated spring-mass estimate stays as the core and a learner deals only with the part that the physical model does not explain. A learner f was trained on the residual r = y − $\hat{y}_{sm}$ of the training runners, with the FS3 predictors and the spring-mass value as inputs, and its output was added back to the spring-mass estimate (Eqs. 9 and 10). Ridge regression and gradient boosting were used as residual learners. If the learner finds no structure in the residual, the hybrid falls back to the physical model, which makes it safer than a purely data-driven model when it is applied to conditions outside the training range.

$$r_i = y_i - \hat{y}_{\mathrm{sm},i} \quad (9)$$

$$\hat{y}_{\mathrm{hyb}} = \hat{y}_{\mathrm{sm}} + f(x, F_{\mathrm{sm}}) \quad (10)$$

3.6 Data-Driven Learners and References

Data-driven learners were ordinary least squares, ridge regression, support-vector regression with a radial basis kernel, random forest and gradient boosting, each run on FS1, FS2 and FS3. Ridge regression minimises the squared error with an L2 penalty on the coefficients (Eq. 11), which stabilises the estimates when predictors are correlated, as running variables usually are. Two simple references were included: the mean outcome of the training trials at the same speed, which shows how much speed alone explains, and the calibrated spring-mass estimate.

$$\beta^* = \arg\min \sum_i (y_i - x_i{}^T\beta)^2 + \lambda \|\beta\|_2^2 \quad (11)$$

3.7 Validation, Error Measures and Statistics

Validation used leave-one-runner-out cross-validation, so all trials of a runner were held out together in 39 outer folds. Hyperparameters (ridge penalty, kernel width and regularisation, forest leaf size, boosting depth, learning rate and number of trees) were chosen by grouped five-fold cross-validation inside the training runners, and scaling and spring-mass calibration were fitted on training runners only. Error was summarised by the mean absolute error (MAE), the root mean square error (RMSE) and the coefficient of

determination (Eqs. 12 to 14), computed over all held-out trials.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \quad (12)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (13)$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (14)$$

Uncertainty was expressed as 95% intervals from 2,000 bootstrap resamples of runners, so that the interval reflects variation between people rather than between trials. Per-runner mean absolute errors were compared between models with the Wilcoxon signed-rank test. Agreement between measured and estimated peak force was examined with a Bland-Altman analysis, in which the limits of agreement are the mean difference plus or minus 1.96 standard deviations of the differences (Eq. 15).

$$\mathrm{LoA} = \bar{d} \pm 1.96\, s_d \quad (15)$$

As a harder secondary task, each speed was withheld in turn and models were trained on the other two speeds, so that estimates at 2.5 and 4.5 m/s required extrapolation and estimates at 3.5 m/s required interpolation. Permutation importance described which inputs the ridge hybrid relied on: each input was shuffled 50 times and the mean increase in MAE was recorded (Eq. 16), where $\mathrm{MAE}_{\pi(j)}$ is the error after permuting input j. Analyses used Python 3.13 with scikit-learn 1.9, SciPy and pandas, and all random seeds were fixed.

$$I_j = \frac{1}{K}\sum_{k=1}^{K}\mathrm{MAE}_{\pi_k(j)} - \mathrm{MAE} \quad (16)$$

4. Results and Discussion

4.1 Event Detection and Cohort Characteristics

Every loading variable changed with running speed (Table 2). Mean peak vertical force went from 2.33 body weights at 2.5 m/s to 2.66 at 4.5 m/s. Over the same range loading rate about doubled, and the spread between runners grew more than threefold. Contact time dropped from 270 to 202 ms, while cadence increased by roughly 20 steps per minute. These changes match the pattern reported for recreational runners across speeds (Malisoux *et al.*, 2023) and confirm that the dataset spans a useful range of loading conditions. Detection of gait events from kinematics performed well. Contact time from markers correlated at r = 0.964 with contact time from the force signal and was on average 29 ms longer (Figure 2), a constant offset that the spring-mass calibration absorbs. Timing variables taken from kinematics thus hold nearly the same information as timing from the force signal, and the rest of the analysis depends on that. Coordinate-based event detection is known to work well on treadmills, and the small constant bias seen here is of the size usually reported when kinematic events are checked against force.

4.2 Peak Vertical Ground Reaction Force in Unseen Runners

Table 3 summarises the main comparison.

Table 3. Leave-one-runner-out estimation of peak vertical ground reaction force (body weights)

Model	MAE	95% CI	RMSE	R^2
Mean of training trials at the same speed	0.132	0.107–0.159	0.165	0.36
Calibrated spring-mass	0.090	0.071–0.110	0.116	0.68
OLS, FS1	0.141	0.116–0.165	0.173	0.30
Ridge, FS1	0.140	0.116–0.163	0.171	0.31
Ridge, FS2	0.102	0.085–0.120	0.126	0.63
Random forest, FS2	0.114	0.098–0.133	0.139	0.54
OLS, FS3	0.096	0.077–0.117	0.121	0.65
Ridge, FS3	0.094	0.076–0.114	0.117	0.68
SVR, FS3	0.104	0.089–0.121	0.123	0.64
Random forest, FS3	0.101	0.084–0.119	0.123	0.64
Gradient boosting, FS3	0.098	0.081–0.117	0.119	0.67
Hybrid: spring-mass + ridge	0.090	0.071–0.111	0.114	0.70
Hybrid: spring-mass + gradient boosting	0.089	0.070–0.110	0.115	0.69

A model that knew only the running speed left an average error of 0.132 body weights, and adding body mass, height, age and sex did not help: ridge regression on FS1 reached 0.140 body weights, slightly worse than speed alone. Anthropometry therefore says little about how hard a runner lands once speed is known, which agrees with reference data showing that body mass explains only a small part of the variation in kinetic variables between recreational runners.

Adding kinematics made a clear difference. Four sagittal variables (FS2) were enough for linear models to bring the error down to about 0.10 body weights, and with the expanded set (FS3) ridge regression reached 0.094 body weights, $R^2 = 0.68$. On no feature set did a nonlinear learner outperform the linear ones. On FS3, random forest and support-vector regression did a little worse than ridge regression, and gradient boosting came only roughly level with it (0.098 body weights). The first hypothesis, that nonlinear learners would improve on linear baselines, was therefore not supported. With 101 trials from 39 runners, there is not enough data for flexible learners to find reliable interactions, and the relation between timing and peak force is close to linear over the observed range. The same pattern was reported in the largest IMU study of peak force, where linear regression, support-vector regression and a neural network performed alike, and lightweight linear methods have matched recurrent networks for force estimation from inertial data.

The calibrated spring-mass model, which uses nothing but kinematic contact and flight times and two fitted constants, reached an MAE of 0.090 body weights ($R^2 = 0.68$). It was as accurate as the best data-driven model, and its per-runner errors did not differ from those of ridge regression on FS3 (Wilcoxon $p = 0.79$) or of gradient boosting ($p = 0.42$). This is the central result of the study. A relation derived from the mechanics of bouncing gait, with no learning beyond a linear calibration, captured essentially all of the information that seventeen kinematic and anthropometric variables offered to purely data-driven models. The spring-mass model has mostly been used to describe running rather than to estimate force in new runners (Meyer et al., 2023), and the present result shows that it deserves the second role as well.

Correcting the spring-mass residual with machine learning gave the lowest errors overall: 0.090 body weights with ridge regression ($R^2 = 0.70$) and 0.089 body weights with gradient boosting. The ridge hybrid was clearly better than the speed baseline ($p = 0.009$), but its advantage over the spring-mass model alone ($p = 0.46$) and over ridge regression without the physical term ($p = 0.13$) was not statistically significant, the hybrid was better than the spring-mass model for 23 of 39 runners. The second hypothesis is therefore supported only in part. The hybrid is never worse than its components and gives the best point estimates, but with this sample the gain over physics alone is small and uncertain. Physics-informed models for head-impact detection and muscle-force estimation reported clear improvements over purely data-driven models. A likely reason for the smaller gain here is that the spring-mass model is already a close description of peak vertical force, leaving little for the learner to correct.

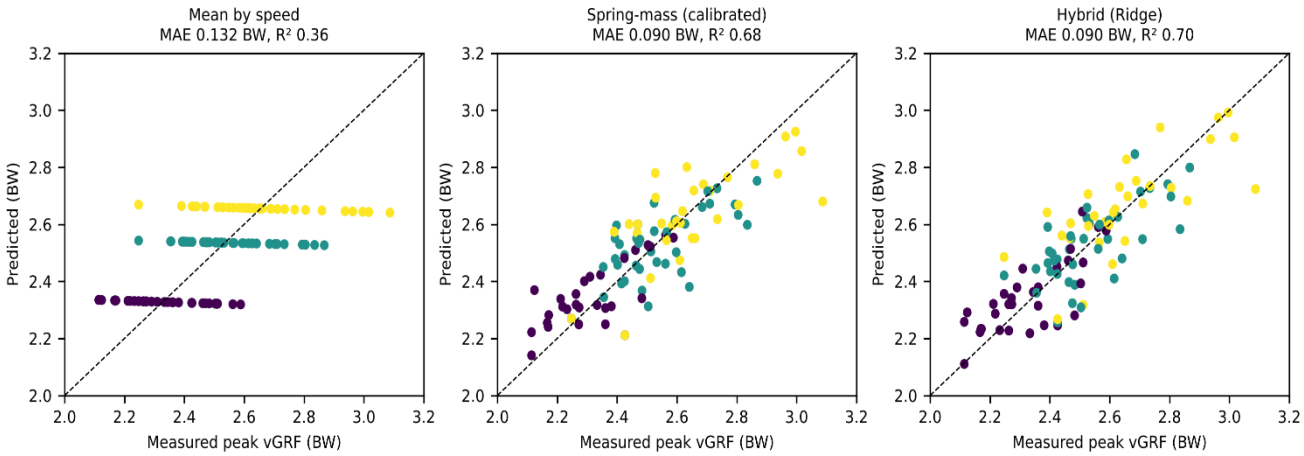


Figure 3. Measured against estimated peak vertical ground reaction force for the speed-only baseline, the calibrated spring-mass model and the spring-mass + ridge hybrid (leave-one-runner-out, colour indicates speed).

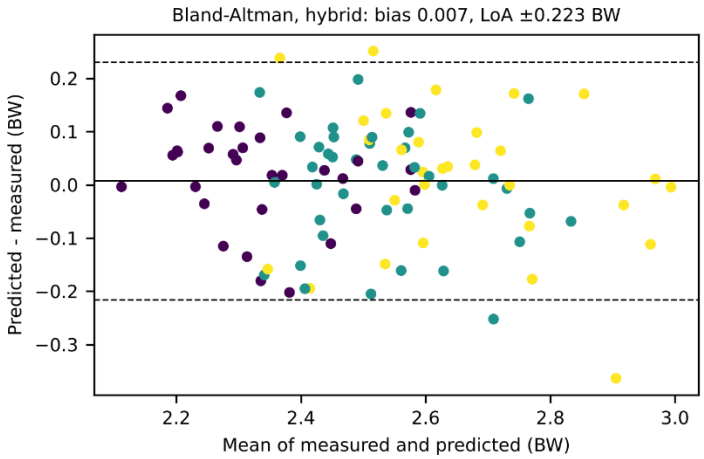


Figure 4. Bland-Altman plot for the spring-mass + ridge hybrid. Solid line, mean difference, dashed lines, 95% limits of agreement.

In practical terms, the hybrid erred by about 3.6% of the typical peak force. The Bland-Altman analysis showed negligible bias (+0.007 body weights) and 95% limits of agreement of ±0.223 body weights (Figure 4). Figure 3 shows that the speed-only model can only assign three values, one per speed, whereas both the spring-mass model and the hybrid follow differences between runners at each speed. The largest single error was an underestimate of 0.36 body weights for a runner whose peak exceeded 3 body weights at 4.5 m/s. Across trials, the hybrid's error was negatively related to the measured peak ($r = -0.41$): unusually high peaks were underestimated and unusually low peaks overestimated. Such regression towards the mean is expected for any estimator trained on few extreme cases.

4.3 Comparison with Existing Studies

The studies summarised in Table 1 used different inputs, outputs and error measures, so their numbers cannot be compared directly with the present results. Many estimated the whole force waveform rather than its peak, several used inertial sensors or pressure insoles rather than kinematics, and error was expressed as relative RMSE, normalised RMSE or the coefficient of determination of waveforms. With those cautions in mind, three contrasts are informative. First, the error obtained here for peak force in unseen runners, approximately 0.09 body weights or under 4% of the peak, is of the same order as the waveform errors reported for leave-one-subject-out designs in Table 1, such as a relative RMSE of 3.2% with insoles and an inertial sensor and an RMSE of 0.10 body weights with wrist and sternum sensors. Second, the largest study in Table 1, with one hundred runners, also found no advantage of a neural network over linear regression, which agrees with the present finding that model complexity buys little. Third, apart from one study in which learned predictions of peak force were no better than the sensor-based estimate they started from, the studies in Table 1 did not report how a non-learned model with the same timing information would have performed. The present results suggest that such a model would often be a strong competitor, and that part of the accuracy credited to deep learning in earlier work may be accuracy that mechanics already provides. The studies that gained a great deal from subject-specific training also make clear that the error for a known runner says little about the error for a new one, and this holds for any learned estimator.

For future studies, this has a direct practical implication. Any learned force estimator ought to be reported alongside a calibrated spring-mass estimate built from the same contact and flight times, ideally on the same held-out runners. Where the learner is not clearly superior, the physical model is the better choice: it is transparent, needs just two constants and remains predictable outside the training range. If the learner is better, the size of its advantage over physics is the honest measure of what machine learning contributes.

4.4 Speed-Specific Error and Transfer to an Unseen Speed

All models made larger errors at higher speeds (Table 4, Figure 5), which matches the wider spread of peak forces at 4.5 m/s.

Table 4. Mean absolute error by running speed and in the speed-held-out transfer task (body weights)

Model	2.5 m/s	3.5 m/s	4.5 m/s	Held-out 2.5	Held-out 3.5	Held-out 4.5	Transfer mean
Speed-only baseline	0.120	0.121	0.158	0.261	0.115	0.229	0.202
Calibrated spring-mass	0.077	0.089	0.105	0.095	0.088	0.102	0.095
Ridge, FS3	0.080	0.093	0.109	0.064	0.069	0.086	0.073
Gradient boosting, FS3	0.081	0.094	0.120	0.099	0.063	0.094	0.086
Hybrid: spring-mass + ridge	0.076	0.089	0.105	0.069	0.065	0.087	0.074

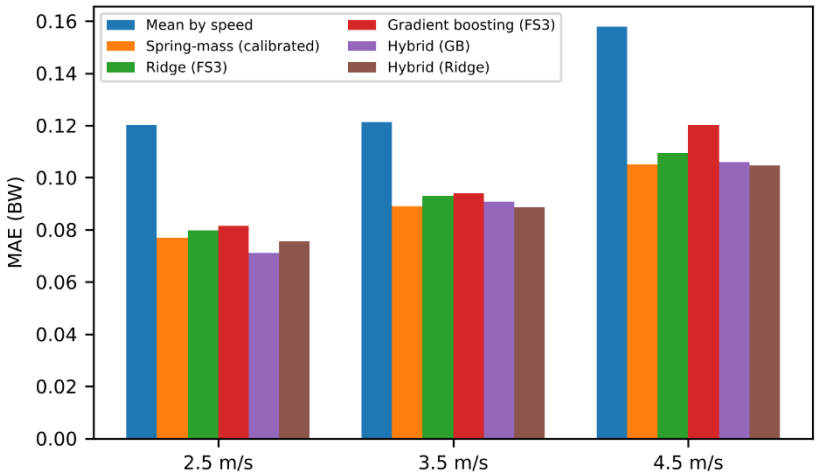


Figure 5. Leave-one-runner-out mean absolute error at each running speed for the principal models.

The ridge hybrid's MAE went up from 0.076 body weights at 2.5 m/s to 0.105 at 4.5 m/s. Relative to the size of the peak, the increase was smaller, because peak force itself rises with speed.

When a whole speed was withheld, differences between model families became clear. The speed-only baseline failed badly at the extremes (0.261 and 0.229 body weights), because it can only repeat the average of the speeds it has seen. Transfer was stable for the spring-mass model (mean 0.095 body weights). This is what one would expect, since its form does not depend on speed, and it fits earlier work in which spring-mass behaviour varied smoothly and predictably with running speed. At the slowest speed, where it was forced to extrapolate, gradient boosting did poorly (0.099 body weights). The best transfer came from ridge regression and the ridge hybrid, whose mean errors were 0.073 and 0.074 body weights. Tree-based learners predict piecewise constant values and cannot extend a trend beyond the range of their training targets. For a coach this is a genuine drawback, because estimates are often wanted at paces outside the calibration range, a slow recovery run, say, or a fast interval session. Linear models and models anchored to physics extend sensibly, and the hybrid combines this property with the small gains in accuracy that the residual learner provides.

4.5 Loading Rate

Loading rate proved much harder to estimate (R^2 at most 0.50). The ridge hybrid again gave the lowest RMSE and the highest R^2, with an MAE of 8.4 body weights per second. Random forest on FS2 had a similar MAE (8.4 body weights per second), against 9.7 for the speed baseline, and none of the differences between learners was statistically significant. The spring-mass estimate carried no information about loading rate beyond speed (9.7 body weights per second). This is expected, because the model describes the peak of a smooth force curve, whereas loading rate depends on how the foot meets the belt during the first few tens of milliseconds of contact. Mean loading rate

also doubled between 2.5 and 4.5 m/s while its spread rose more than threefold (Table 2), so a few runners with steep impact transients dominated the error. Rearfoot strikers with high and low loading rates differ in subtle features of foot and shank motion at contact (Johnson & Davis, 2022), and foot strike governs the centre-of-pressure behaviour that shapes the impact, these details are poorly captured by the sagittal angles used here.

Estimating loading rate well probably needs information that is closer to the impact itself, such as foot or shank acceleration at a high sampling rate, or a model of the impact peak rather than of the whole stance phase. Marker trajectories filtered at 12 Hz smooth out exactly the rapid events that determine loading rate. Loading rate is also sensitive to slope and other conditions, and its link with injury remains debated, so the difficulty of estimating it has limited practical cost. For practitioners, the conclusion is that peak force can be estimated from timing-based kinematics with useful accuracy, whereas loading rate cannot, and claims about loading rate from such inputs should be treated with caution.

Permutation importance for the ridge hybrid ranked vertical pelvis excursion during stance first (0.036 body weights increase in MAE when permuted), followed by overstride and the spring-mass estimate itself (0.023 and 0.023), then duty factor and flight time (Figure 6). All of these describe how stiffly and how far forward the runner lands. The downward excursion of the body during contact is the displacement term of vertical stiffness: for a given peak force, a runner who sinks further is softer, and for a given excursion, a shorter contact implies a higher force. Overstriding places the foot further in front of the centre of mass and lengthens the braking phase. The ranking therefore agrees closely with the mechanics that the spring-mass model encodes, and the residual learner mainly adds information about stiffness that timing alone does not fully capture. Duty factor came fourth. It has been associated with loading magnitude in habitual running styles and sums up the overall running pattern described in syntheses of running biomechanics; for individual runners it carries information that foot-strike pattern does not, which may be why foot-strike angle fell down the ranking once duty factor was in the model.

With timing and excursion already known, joint angles at contact added little, and anthropometric variables almost nothing. Because several inputs are correlated, these importances describe the fitted model rather than causal effects, shuffling one of two correlated inputs understates the importance of both. For explainability, this is an encouraging result. Instead of obscure combinations of inputs, the model relied on variables a biomechanist would pick, and a coach can follow its reasoning easily: a stiff, bouncy landing with short contacts goes with higher peak force.

Several factors probably account for the error that remains. Kinematic contact time differed from force-based contact time by a few milliseconds in individual trials, and because peak force depends on the ratio of flight to contact time, small timing errors translate into noticeable force errors when contact is short, as it is at 4.5 m/s.

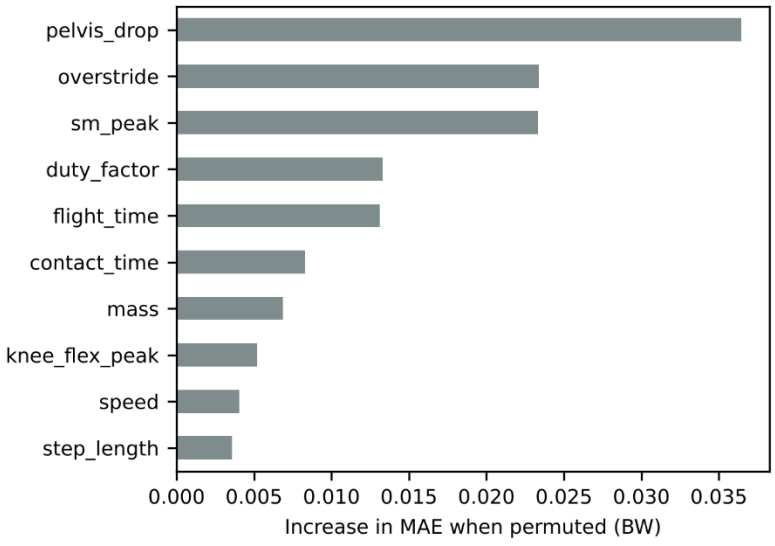


Figure 6. Permutation importance (increase in MAE when an input is shuffled, 50 repeats) for the spring-mass + ridge hybrid.

The half-sine shape assumed by the spring-mass model fits rearfoot strikers with a distinct impact transient less well than midfoot and forefoot strikers, and the dataset contained too few non-rearfoot trials to model this difference reliably. Differences between the left and right legs, which averaging across both legs hides, may also contribute. Each of these points suggests a specific improvement, such as better event detection, a strike-specific force shape or side-specific modelling, rather than a more complex learner.

4.6 Practical Implications

For coaches, physical education teachers and therapists, the main message is that the timing of foot contact already says most of what kinematics can say about peak running force. Any device that records contact and flight time, be it a foot pod, an inertial sensor on the shank or a video system, can give a transparent estimate to within about a tenth of a body weight. Vertical pelvis excursion and overstride can both be seen on ordinary video; adding them improves the estimate a little and makes it easier to explain. Markerless systems can already measure these quantities with acceptable validity during running (Verheul *et al.,* 2026), and inertial spatiotemporal measures are reliable in endurance runners, so field tools can supply the inputs required. Any advice to alter step rate or landing position, however, needs to rest on evidence about what such changes do. More complex learners are worth adopting only if they clearly outperform this reference on runners they were not trained on.

For physical education in schools and universities, the approach offers a simple and safe way to teach running technique. With a phone camera or a cheap sensor, students can measure contact and flight times, work out the spring-mass estimate by hand and observe how speed, step length or landing position change the predicted peak force. Since the model is easy to follow, it can be taught in a lesson on forces and motion as well as used for measurement. Even so, teachers should not use these estimates to judge the injury risk of individual students, for reasons set out below.

Three habits are suggested for researchers. Whenever machine learning is used to estimate running force from timing or kinematic inputs, a spring-mass baseline should be reported too. Validation ought to hold out whole runners and, if possible, whole conditions such as particular speeds or surfaces. Finally, importance analyses should be read in mechanical terms, since this makes mistakes easier to catch and findings easier to believe. The physics-informed hybrid gives a straightforward template: the physical model stays at the core, data are used to refine it, and the model behaves sensibly when data are limited.

4.7 Limitations

Some limitations should be noted. With 39 runners, only one of them a woman, the confidence intervals are wide and subgroup analysis by sex or foot-strike pattern is not possible. All trials were run on one instrumented treadmill, so the estimates may not carry over to overground running, other surfaces, slopes or footwear, transfer from treadmill to overground running has been shown to need individual fine-tuning in comparable work. Open multi-modal datasets with consumer wearables offer a route to test this. Predictors came from laboratory markers rather than wearable sensors or video, and the next step is to repeat the analysis with inertial or markerless data, where timing errors will be larger. Loading outcomes were averaged over each trial, so step-to-step variation and asymmetry between legs were not modelled. The hybrid was tested with only two residual learners and one form of physical model, other formulations, such as models that include horizontal forces, may perform differently. Finally, these are estimates of external load, not predictions of injury, and no claim about injury prevention follows from them.

5. Conclusion

Using only marker-based kinematics and validating on runners withheld from training, peak vertical ground reaction force across three running speeds was estimated with a mean absolute error approximately 0.09 body weights. A calibrated spring-mass model using kinematic contact and flight times performed as well as any purely data-driven learner. The hybrid models achieved similar or marginally lower errors than the calibrated spring-mass baseline; the ridge-corrected hybrid did not improve the reported MAE at three-decimal precision. It also transferred well to an unseen running speed, though its advantage over simpler models was small and not statistically significant. Tree-based learners did not extrapolate well to a speed outside their training range, and loading rate stayed hard to estimate from kinematics. The inputs that mattered most were vertical pelvis excursion, overstride and the spring-mass estimate, which ties the model to well-understood running mechanics. Tools that

estimate running load with machine learning should report a physical baseline of this sort, and future field systems could assess simple timing measures before adding model complexity, provided that transfer to non-laboratory sensors is validated. The approach now needs testing with wearable and video-based inputs, in women and in overground running.

References

Anderson, L.M., Martin, J.F., Barton, C.J., Bonanno, D.R. (2022). What is the effect of changing running step rate on injury, performance and biomechanics? A systematic review and meta-analysis. Sports Medicine - Open, 8(1), 112. [DOI] [PubMed]

Baker, L.M., Weigend, F.C., Swaminathan, K., Kim, D., Chin, A., Lieberman, D.E., Walsh, C.J. (2025). Estimating braking and propulsion forces during overground running in and out of the lab. PLOS ONE, 20(9), e0330042. [DOI] [PubMed]

Bernett, J., Blumenthal, D.B., Grimm, D.G., Haselbeck, F., Joeres, R., Kalinina O.V., List, M. (2024). Guiding questions to avoid data leakage in biological machine learning applications. Nature Methods, 21(8), 1444–1453. [DOI] [PubMed]

Binnekamp, M., Scheltinga, B.L., Wouda, F.J., Reenalda, J. (2026). No extra sensors needed: estimating vertical ground reaction force using IMUs at the smartwatch and heartrate strap location. Sports Biomechanics, 1–17. [DOI] [PubMed]

Burns, G.T., Tam, N., Langerak, N.G., Zernicke, R.F., Lamberts, R.P. (2024). Spring-mass characteristics in runners before and after a 56-km road ultramarathon. Journal of Applied Biomechanics, 40(6), 492–500. [DOI] [PubMed]

Carter, J., Chen, X., Cazzola, D., Trewartha, G., Preatoni, E. (2024). Consumer-priced wearable sensors combined with deep learning can be used to accurately predict ground reaction forces during various treadmill running conditions. PeerJ, 12, e17896. [DOI] [PubMed]

Chen, T., Xu, D., Zhou, Z., Zhou, H., Shao, S., Gu, Y. (2025). Prediction of vertical ground reaction forces under different running speeds: integration of wearable IMU with CNN-xLSTM. Sensors, 25(4), 1249. [DOI] [PubMed]

Donahue, S.R., Hahn, M.E. (2023a). Estimation of gait events and kinetic waveforms with wearable sensors and machine learning when running in an unconstrained environment. Scientific Reports, 13(1), 2339. [DOI] [PubMed]

Donahue, S.R., Hahn, M.E. (2023b). Estimation of ground reaction force waveforms during fixed pace running outside the laboratory. Frontiers in Sports and Active Living, 5, 974186. [DOI] [PubMed]

Evans, R.J., Moffit, T.J., Mitchell, P.K., Pamukoff, D.N. (2023). Injury and performance related biomechanical differences between recreational and collegiate runners. Frontiers in Sports and Active Living, 5, 1268292. [DOI] [PubMed]

Fukuchi, R.K., Fukuchi, C.A., Duarte, M. (2017). A public dataset of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics. PeerJ, 5, e3298. [DOI] [PubMed]

Ghaffarzadeh, P., Chakraborty, D., Aslansefat, K., Dostan, A., Papadopoulos, Y. (2026). A multi-modal dataset for ground reaction force estimation using consumer wearable sensors. Scientific Data, 13(1), 855. [DOI] [PubMed]

Girard, O., Alsenoy, K.V., Li, S.N., Ryu, J.H., Peeling, P. (2022). Constant low-to-moderate mechanical asymmetries during a treadmill graded exercise test. European Journal of Sport Science, 22(4), 530–538. [DOI] [PubMed]

Gruber, A.H. (2023). The 'impacts cause injury' hypothesis: Running in circles or making new strides? Journal of Biomechanics, 156, 111694. [DOI] [PubMed]

Hata, K., Noro, H., Takeshita, T., Yamazaki, Y., Yanagiya, T. (2022). Leg stiffness during running in highly cushioned shoes with a carbon-fiber plate and traditional shoes. Gait & Posture, 95, 9–14. [DOI] [PubMed]

Honert, E.C., Hoitz, F., Blades, S., Nigg, S.R., Nigg, B.M. (2022). Estimating running ground reaction forces from plantar pressure during graded running. Sensors, 22(9), 3338. [DOI] [PubMed]

Johnson, C.D., Davis, I.S. (2022). What differentiates rearfoot strike runners with low and high vertical load rates? Gait & Posture, 96, 149–153. [DOI] [PubMed]

Kapoor, S., Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based

science. Patterns, 4(9), 100804. [DOI] [PubMed]

Koffman, L., Zhang, Y., Harezlak, J., Crainiceanu, C., Leroux, A. (2023). Fingerprinting walking using wrist-worn accelerometers. Gait & Posture, 103, 92–98. [DOI] [PubMed]

Leacox, A., Fashingbauer, L., Ferguson, T., Zajakowski, A., Baum, B., Reinking, M. (2025). The effect of running speed on cadence and running kinetics. International Journal of Sports Physical Therapy, 20(7), 957–963. [DOI] [PubMed]

Lemire, M., Falbriard, M., Aminian, K., Pavlik, E., Millet, G.P., Meyer, F. (2022). Correspondence between values of vertical loading rate and oxygen consumption during inclined running. Sports Medicine - Open, 8(1), 114. [DOI] [PubMed]

Li, S., Pan, J., Wang, D., Yuan, S., Yang, J., Hao, W. (2025). Predicting running vertical ground reaction forces using neural network models based on an IMU sensor. Sensors, 25(13), 3870. [DOI] [PubMed]

Ma, S., Zhang, J., Shi, C., Di, P., Robertson, I.D., Zhang, Z.Q. (2024). Physics-informed deep learning for muscle force prediction with unlabeled sEMG signals. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 1246–1256. [DOI] [PubMed]

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. [DOI] [PubMed]

Malisoux, L., Napier, C., Gette, P., Delattre, N., Theisen, D. (2023). Reference values and determinants of spatiotemporal and kinetic variables in recreational runners. Orthopaedic Journal of Sports Medicine, 11(10), 23259671231204629. [DOI] [PubMed]

Mayo-Nieto, C., Cartón-Llorente, A., Rafales-Perucha, A., Cardiel-Sánchez, S., Jaén-Carrillo, D., Roche-Seruendo, L.E. (2026). The biomechanical signature of elite distance running: A matched Pairs analysis across multiple speeds. Scandinavian Journal of Medicine & Science in Sports, 36(11), e70385. [DOI] [PubMed]

Meyer, F., Falbriard, M., Aminian, K., Millet, G.P. (2023). Vertical and leg stiffness modeling during running: Effect of speed and incline. International Journal of Sports Medicine, 44(9), 673–679. [DOI] [PubMed]

Miqueleiz, U., Aguado-Jimenez, R., Lecumberri, P., Garcia-Tabar, I., Gorostiaga, E.M. (2025). Reliability of inertial measurement unit-based spatiotemporal and kinetic variables in endurance runners during treadmill running. Journal of Human Kinetics, 97, 65–76. [DOI] [PubMed]

Mtibaa, K., Zarrouk, N., Ryu, J.H., Racinais, S., Girard, O. (2023). Mechanical asymmetries remain low-to-moderate during 30 min of self-paced treadmill running. Frontiers in Physiology, 14, 1289172. [DOI] [PubMed]

Nijs, A., Roerdink, M., Beek, P.J. (2023). Exploring running styles in the field through cadence and duty factor modulation. PLoS One, 18(12), e0295423. [DOI] [PubMed]

Nitschke, M., Dorschky, E., Leyendecker, S., Eskofier, B.M., Koelewijn, A.D. (2024). Estimating 3D kinematics and kinetics from virtual inertial sensor data through musculoskeletal movement simulations. Frontiers in Bioengineering and Biotechnology, 12, 1285845. [DOI] [PubMed]

Patoz, A., Gindre, C., Lussiana, T., Breine, B. (2026). The heel-strike fallacy and cadence beliefs in distance running: A narrative review. Sports Medicine - Open, 12(1), 149. [DOI] [PubMed]

Patoz, A., Lussiana, T., Breine, B., Gindre, C., Malatesta, D. (2022). Duty factor and foot-strike pattern do not represent similar running pattern at the individual level. Scientific Reports, 12(1), 13061. [DOI] [PubMed]

Patoz, A., Lussiana, T., Breine, B., Gindre, C., Malatesta, D. (2025). Comparison of different machine learning models to enhance sacral acceleration-based estimations of running stride temporal variables and peak vertical ground reaction force. Sports Biomechanics, 24(4), 825–841. [DOI] [PubMed]

Pitts, M.N., Ebers, M.R., Agresta, C.E., Steele, K.M. (2025). Evaluating sparse inertial measurement

unit configurations for inferring treadmill running motion. Sensors, 25(7), 2105. [DOI] [PubMed]

Rasmussen, J., Skejø, S., Waagepetersen, R.P. (2023). Predicting tissue loads in running from inertial measurement units. Sensors, 23(24), 9836. [DOI] [PubMed]

Raymond, S.J., Cecchi, N.J., Alizadeh, H.V., Callan, A.A., Rice, E., Liu, Y., Zhou, Z., Zeineh, M. and Camarillo, D.B. (2022). Physics-informed machine learning improves detection of head impacts. Annals of Biomedical Engineering, 50(11), 1534–1545. [DOI] [PubMed]

Rosenblatt, M., Tejavibulya, L., Jiang, R., Noble, S., Scheinost, D. (2024). Data leakage inflates prediction performance in connectome-based machine learning models. Nature Communications, 15(1), 1829. [DOI] [PubMed]

Song, B., Paolieri, M., Stewart, H.E., Golubchik, L., McNitt-Gray, J. L., Misra, V., Shah, D. (2025). Estimating ground reaction forces from inertial sensors. IEEE Transactions on Biomedical Engineering, 72(2), 595–608. [DOI] [PubMed]

Turner, J.A., Chaaban, C.R., Padua, D.A. (2024). Validation of OpenCap: A low-cost markerless motion capture system for lower-extremity kinematics during return-to-sport tasks. Journal of Biomechanics, 171, 112200. [DOI] [PubMed]

Van den Berghe, P., Warlop, L., Derie, R., Leman, M., De Clercq, D., Breine, B. (2022). Foot strike determines the center of pressure behavior and affects impact severity in heel-toe running. Journal of Sports Sciences, 40(7), 808–820. [DOI] [PubMed]

Van der Meulen, L., Bonnaerens, S., Caekenberghe, I.V., Clercq, D., Segers, V., Fiers, P. (2024). Habitual running style matters: Duty factor, and not stride frequency, relates to loading magnitude. Journal of Human Kinetics, 94, 37–45. [DOI] [PubMed]

Van Hooren, B., Jukic, I., Cox, M., Frenken, K.G., Bautista, I., Moore, I.S. (2024). The relationship between running biomechanics and running economy: A systematic review and meta-analysis of observational studies. Sports Medicine, 54(5), 1269–1316. [DOI] [PubMed]

Van Middelaar, R.P., Reenalda, J. (2025). Estimating tibial bone load during running using only inertial measurement units on the tibias and sternum: A Centre of Pressure approach. Journal of Biomechanics, 189, 112839. [DOI] [PubMed]

Van Oeveren, B.T., de Ruiter, C.J., Beek, P.J., van Dieën, J.H. (2024). The biomechanics of running and running styles: A synthesis. Sports Biomechanics, 23(4), 516–554. [DOI] [PubMed]

Verheul, J., Hughes, O., Hitchens, L., Atherton, T., Sauter, T., Radwan, A., Pryse, E., Mitchell, A., Burton, S. (2026). Markerless motion capture for running: Validity and reliability of whole-body, joint, and muscle kinematics. Journal of Biomechanics, 195, 113133. [DOI] [PubMed]

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. [DOI] [PubMed]

Author's contribution

Samuel Asante: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. Archibald Danquah-Amoah: Conceptualization, Supervision, Validation, Writing – review & editing. Alex Abrokwa Fosuhene: Formal analysis, Investigation, Visualization, Writing – review & editing. Dominic Akwasi Bonah: Data curation, Validation, Writing – review & editing. All the authors read and approved the final version of this manuscript.

Acknowledgments

The authors thank Fukuchi, Fukuchi and Duarte for making their dataset openly available.

Funding Source

This study received no external funding.

Ethics Statements

This secondary analysis used a public, de-identified dataset collected with ethical approval and informed consent in the original study, the present secondary analysis did not recruit or assess participants. Data are available from Figshare (DOI 10.6084/m9.figshare.4543435) and the analysis code is available from the author. Generative AI tools assisted with drafting and proofreading. The authors reviewed and verified the manuscript's content, including the reported results, and take full responsibility for the published work.

Does this article pass screening for similarity?

Yes

About the License

© The Author(s) 2026. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License.