

June 18, 2021

Predicting continuous ground reaction forces from accelerometers during uphill and downhill running: A recurrent neural network solution

Ryan S. Alcantara^{1,2}, W. Brent Edwards³, Guillaume Y. Millet⁴, and Alena M. Grabowski¹

¹ Department of Integrative Physiology, University of Colorado Boulder, Boulder, CO, USA

² Department of Bioengineering, Stanford University, Palo Alto, CA, USA

³ Human Performance Laboratory, Faculty of Kinesiology, University of Calgary, Calgary, Canada

⁴ Univ Lyon, UJM-Saint-Etienne, Laboratoire Interuniversitaire de Biologie de la Motricité, EA 7424, F-42023, Saint-Etienne, France

Conflict of Interest Disclosure: None

Correspondence Address:

Ryan S. Alcantara, PhD

354 UCB

Boulder, CO, 80309-0354, USA

Phone: 541-951-7926

Email: ryan.alcantara@colorado.edu

Running Title: Predicting continuous ground reaction forces

Abstract

Ground reaction forces (GRFs) are important for understanding human movement, but their measurement is generally limited to a laboratory. Previous studies used neural networks to predict GRF waveforms during running from wearable device data, but these predictions are limited to the stance phase of level-ground running. We sought to develop a recurrent neural network capable of predicting continuous normal (perpendicular to surface) GRFs across a range of running speeds and slopes from accelerometer data.

19 subjects ran on a force-measuring treadmill at five slopes (0° , $\pm 5^\circ$, $\pm 10^\circ$) and three speeds (2.5, 3.33, 4.17 m/s) per slope with sacral- and shoe-mounted accelerometers. We then trained a recurrent neural network to predict normal GRF waveforms frame-by-frame. The predicted versus measured GRF waveforms had an average \pm SD RMSE of 0.16 ± 0.04 BW and relative RMSE of $6.4 \pm 1.5\%$ across all conditions and subjects.

The recurrent neural network predicted continuous normal GRF waveforms across a range of running speeds and slopes with greater accuracy than neural networks implemented in previous studies. This approach may facilitate predictions of biomechanical variables outside the laboratory in near real-time and improves the accuracy of quantifying and monitoring external loads experienced by the body when running.

Keywords: biomechanics, machine learning, imu, lstm, grf

Word Count: 3979

Introduction

Ground reaction forces (GRFs) are applied to the body when the foot is in contact with the ground and their measurement has facilitated numerous insights into the etiology of running-related injuries¹. However, measurement of GRFs is generally restricted to a laboratory environment. To determine the effects of sport-specific environments on running kinetics and kinematics, previous studies have replicated aspects of an athlete's competitive environment (e.g., running surface, slope) within a laboratory environment²⁻⁴. Alternatively, inertial measurement units (IMUs; wireless wearable devices that measure magnetism, linear acceleration, and angular velocity) have been used to measure athletes' leg joint angles, stride kinematics, and segmental accelerations during competitive events⁵⁻⁷. Although IMUs cannot directly measure GRFs, previous studies have used algorithms to estimate discrete biomechanical variables like peak vertical GRF, ground contact time, vertical impulse, and vertical loading rate⁸⁻¹² from IMU data.

Recently, neural networks have been used to predict GRF waveforms during running¹³⁻¹⁶, from which a variety of discrete variables can be calculated. Although predictions of the entire GRF waveform represent a more versatile outcome compared to predicting a discrete variable, previous studies have used neural network architectures that required waveforms to be normalized to the duration of a step¹³ or stance phase^{14,16}, preventing the calculation of biomechanical variables with a temporal component (e.g., ground contact time, step frequency, vertical impulse, and loading rate). Additionally, previous studies have predicted GRF waveforms only during level-ground running¹³⁻¹⁶, limiting the application to environments that are flat (e.g., level treadmill or athletics track). Road and trail running are internationally popular forms of physical activity^{17,18} and require runners to navigate a variety of uphill and downhill slopes. A method that accurately predicts GRF waveforms from wearable device data across a range of running slopes, while

maintaining the temporal component, could allow researchers, clinicians, and coaches to measure and monitor a variety of kinetic and kinematic variables in outdoor environments.

Long Short-Term Memory (LSTM) networks¹⁹ are a type of recurrent neural network that can overcome the traditional requirement of normalizing GRF waveforms to the duration of stance phase because LSTM networks can recurrently predict smaller, uniform portions of a larger sequence of any length. As such, a sequence of continuous GRF data can be predicted if it can be broken up into uniform portions. For the prediction of a given portion, LSTM networks use information from previous portions, effectively “remembering” the portion’s context and have been used to predict sequential data during natural language processing tasks²⁰. In the field of Biomechanics, LSTM networks have been used to make frame-by-frame predictions of GRF waveforms using motion capture data²¹ and predictions of the center of mass position relative to center of pressure from IMU data during walking²². Developing an LSTM network to predict GRF waveforms from wearable device data would allow researchers to predict GRF waveforms not only during the stance phase, but continuously for the entire duration of a run. IMUs have already been used to longitudinally measure biomechanical variables related to running injury^{5–7,10} and applying an LSTM network to such data could effectively provide a way to indirectly measure continuous GRF waveforms outside of the laboratory at a scale that was previously unattainable.

The purpose of this exploratory study was to develop an LSTM network that could predict continuous normal (perpendicular to running surface) GRF waveforms across a range of running speeds and slopes using data from accelerometers. We sought to develop a network that could predict GRF waveforms with accuracy better than state-of-the-art predictions of time-normalized vertical GRF data during level-ground running using data from multiple IMUs: a root mean square

error (RMSE) of 0.21 BW¹³ and relative RMSE (rRMSE; RMSE normalized to the average range of the compared waveforms; Eq. 1) of 13.92%¹⁴.

Methods

Subjects: We analyzed a pre-existing dataset^{23–25} where 21 subjects ran at a combination of running speeds and slopes. Two subjects were excluded from the current analysis due to equipment data acquisition errors, leaving 19 subjects remaining (10 Male, 9 Female; 29 ± 9 years, 173 ± 9 cm, 68.1 ± 9.9 kg). All subjects provided informed consent and the experimental protocol was approved by the University of Calgary Conjoint Health Research Ethics Board (#REB14-1117).

Experimental Protocol: Following a 5 min warm up at a self-selected speed, each subject completed thirty 30 s trials on a force-measuring treadmill (2000 Hz; Bertec, OH, USA), which included five slopes (0° , $\pm 5^\circ$, $\pm 10^\circ$) at three speeds (2.5, 3.33, 4.17 m/s) per slope, and three step frequencies (preferred and $\pm 10\%$) at 3.33 m/s for each slope. Three custom biaxial accelerometers (2000 Hz) were adhered with tape to subjects during all conditions: two on the right shoe and one on the sacrum. The accelerometers on the shoe were only used to determine the foot strike pattern for each condition using a previously validated method²⁶, which provided the percentage of a trial's foot strikes classified as either rearfoot, midfoot, or forefoot strikes. The biaxial accelerometer on the sacrum was oriented such that it measured vertical and anteroposterior accelerations relative to the accelerometer's local coordinate system (Figure 1).

Data Processing: We analyzed 5 seconds of data (approximately 13 foot-ground contacts) from each trial and downsampled the normal GRF, vertical sacral acceleration, and anteroposterior sacral acceleration to 500 Hz to reduce the computational cost and match the sampling frequency of prior studies^{12,27}. We normalized GRFs to bodyweight (BW) and filtered them using a 4th order

low-pass Butterworth filter with a 30 Hz cut-off. We filtered the sacral acceleration data with a 4th order low-pass Butterworth filter with a 20 Hz cut-off. Preliminary analysis revealed that a 20 Hz cut-off improved prediction accuracy and preserved approximately 89% and 82% of the vertical and anteroposterior signal power, respectively.

Vertical sacral accelerometer data were further processed so that all negative values were replaced with zeros. Preliminary analysis revealed that negative vertical acceleration values occur during the aerial phase and replacing them with zeros helped the LSTM network avoid predictions of negative normal GRFs during the aerial phase. For each condition, we used the 2500 frame (5 s trial @ 500 Hz) sequences of vertical and anteroposterior sacral accelerometer data to predict the simultaneously collected 2500 frame sequence of normal GRFs. The recurrent nature of the LSTM network requires sequential data to be divided into smaller portions that are iteratively used to make predictions. To accomplish this, we divided acceleration data for each trial into overlapping windows with a 6 frame (12 ms) width and padded the beginning and end of each trial's acceleration data with the first and final values, respectively, to ensure the number of windows equaled the number of normal GRF frames (2500) and that the windows were centered on the corresponding frame of the normal GRFs (Figure 1). Pilot testing revealed that a window width of 6 frames was the smallest window we could use without decreasing LSTM network prediction accuracy. Thus, the LSTM network iteratively predicted a single frame of the normal GRF at time t using acceleration data from frames $t-3$ through $t+2$ (Figure 1).

Feature Engineering: A total of 13 features were used as inputs in the LSTM network (Figure 1). We calculated the mean, standard deviation (SD), and range of vertical and anteroposterior acceleration data for each 12 ms window and used them as input features. The use of summary statistics as input features has been shown to maintain neural network accuracy while

benefiting from a reduced computational cost²⁸. These three summary statistics were normalized to a range of 0 – 1 and represent 6 (3 features x 2 acceleration axes) of the 13 input features. The remaining input features were selected due to their effect on running kinetics and kinematics: subject height, body mass, running speed, slope, and percentage of steps classified as either rearfoot, midfoot, or forefoot strikes^{12,24,25,29,30}. We chose not to include step frequency as an input feature, despite the presence of the $\pm 10\%$ preferred step frequency conditions, to increase the variability in the data used to predict GRF waveforms. Doing so theoretically represents a greater challenge for the LSTM network as there is additional variability between trials that is not being explicitly accounted for with an input variable.

Neural Network Architecture: The neural network consisted of a Bidirectional LSTM and a multilayer perceptron (MLP) with three fully connected layers containing 128, 384, and 320 neurons, respectively (Figure 2). The Bidirectional LSTM consists of two LSTM layers where the order of the input sequence is reversed for the second layer. Reversing the sequence for the second LSTM layer allows the network to utilize information from future portions of the sequence just as the first LSTM layer utilizes information from prior portions. The outputs from each LSTM layer are then averaged before being passed along to the MLP. The number and size of the layers were determined using the Hyperband hyperparameter optimization algorithm³¹ on the data of two randomly selected subjects. The LSTM network was trained using a batch size of 32, learning rate of 0.001, and mean square error loss function. Network weights and biases were updated using the adaptive moment estimation (Adam) optimization algorithm at the end of each epoch³² and training lasted a maximum of 1000 epochs or until the mean square error failed to decrease by 0.001 BW after 30 consecutive epochs. The neural network was developed using the Tensorflow (v2.2.0) python library³³.

Network Validation: We assessed the accuracy and generalizability of the network using a Leave-One-Subject-Out (LOSO) cross validation method³⁴. LOSO cross validation is a variation of K-fold cross validation that requires the dataset to be subset by subject, with one subject's data withheld for testing purposes and the rest of the subjects' data used to train the network. This process is repeated until the network has been tested on every subject's data, ultimately providing an ensemble of networks and their respective accuracy metrics. Performing LOSO cross validation can be computationally costly, as the network must be trained and tested a number of times equal to the number of subjects ($n = 19$), but the benefits of this method include the ensemble of accuracy metrics and assurance that a given subject's data are not included in both the training and testing subsets, which can artificially increase the reported accuracy of a network^{35,36}.

In addition to the LOSO cross validation method, we performed a test-train split of one representative subject's data according to slope ($\pm 5^\circ$ trials reserved for testing, 0° and $\pm 10^\circ$ slopes used for training) to test the accuracy of the model when predicting speed-slope combinations that were not present during training. We selected Subject 14 as a representative subject because their RMSE during LOSO cross validation was similar to the average RMSE across all subjects (Figure 3) and their GRF waveforms illustrated an interaction between running slope and normal GRF impact peak magnitude³⁷. This single-subject validation method prioritizes accuracy over generalizability (ability to make accurate predictions for a variety of individuals) and represents a potential circumstance where an LSTM network is trained on data collected from a single athlete prior to their competitive season and later used to predict only that athlete's GRF data from wearable device data during their competitive season.

Prediction error for each trial's GRF waveforms was quantified as the root mean square error (RMSE) and relative RMSE (rRMSE), which is RMSE normalized to the average range of compared waveforms, expressed as a percentage and defined as

$$rRMSE = \frac{RMSE}{0.5 \times \sum_{i=1}^2 (\max(x_i) - \min(x_i))} \times 100, \quad (1)$$

where x_1 and x_2 are the GRF waveforms predicted by the LSTM network and measured by the force-measuring treadmill³⁸. Additionally, we used a threshold of 5% BW to identify stance phase and calculated the active peak of the normal GRF waveform, normal impulse, normal GRF loading rate, contact time, and step frequency from the predicted and measured GRF data. The normal GRF active peak was calculated as the maximum normal GRF value occurring between 40 – 60% of stance phase because the magnitude of the impact peak can exceed the active peak during downhill running and occurs during early stance phase (0 – 30%)^{25,37}. We calculated impulse as the integral of the normal GRF waveform during the stance phase with respect to time, loading rate as the average slope of the normal GRF waveform during the first 25 ms of stance phase³⁹, contact time as the duration when the normal GRF was greater than 5% BW, and step frequency as the number of initial foot-ground contacts per second. We report the mean absolute percent error (MAPE) of these discrete variables for each subject. Data analysis was performed in python (v3.6.9) and R (v4.0.4) using custom libraries^{40–47}.

We enforced two biomechanical boundaries upon the predicted GRF waveforms to ensure that data fell within established biomechanical limits and could be used to calculate the discrete biomechanical variables of interest. First, the predicted GRF waveform had to have an equal number of foot-ground contacts as the GRF waveform measured by the force-measuring treadmill, determined using the same 5% BW threshold. Second, the step frequency over the duration of the

predicted GRF waveform had to be ≤ 4 Hz. We selected these criteria based on previous research of running biomechanics, as thresholds of 5% BW have been previously used to identify the stance phase for the calculation of kinetic or kinematic variables^{12,27} and during uphill and downhill running, step frequency is ≤ 4 Hz^{48,49}. Trials that failed to meet either of these criteria were used to calculate the LSTM network's overall prediction failure rate and removed from subsequent analyses.

Results

We analyzed 529 trials for the present study. The predicted GRF waveforms for 32 trials (6%) failed to meet one or both criteria and were considered failed predictions. Specifically, we identified 22 trials (4%) that required a threshold greater than 5% BW to identify an equal number of steps between predicted and measured GRF waveforms and 10 trials (2%) that had a step frequency greater than 4 Hz. Thus, 94% of GRF waveforms predicted by the LSTM network fell within the imposed biomechanical boundaries.

Leave-One-Subject-Out cross validation revealed that the LSTM network predictions of each subject's normal GRF waveforms had an average \pm SD RMSE of 0.16 ± 0.04 BW (Figure 3) and rRMSE of $6.4 \pm 1.5\%$ compared to GRF waveforms measured by the force-measuring treadmill across all conditions (Table 1). RMSE values were generally lower during slow uphill running (2.5 m/s, $+10^\circ$; 0.13 BW) compared to fast downhill running (4.17 m/s, -10° ; 0.20 BW) (Figures 4 and 5). The MAPE for step frequency was $0.1 \pm 0.1\%$, contact time was $4.9 \pm 4.0\%$, impulse was $6.4 \pm 6.9\%$, normal GRF active peak was $8.5 \pm 8.2\%$, and loading rate was $27.6 \pm 36.1\%$ (Table 2).

The prediction error for one representative subject's (Subject 14) normal GRF waveforms at $\pm 5^\circ$ during single-subject validation was lower than those resulting from the LOSO cross

validation, with an average \pm SD RMSE of 0.08 ± 0.02 BW and rRMSE of $3.3 \pm 0.9\%$. The MAPE of step frequency ($0.1 \pm 0.1\%$), contact time ($3.0 \pm 2.3\%$), impulse ($2.5 \pm 1.9\%$), normal GRF active peak ($2.7 \pm 2.0\%$), and loading rate ($17.6 \pm 16.9\%$) calculated from predicted GRF waveforms were also generally lower than those resulting from LOSO cross validation.

Discussion

We developed a recurrent neural network capable of predicting continuous normal GRF waveforms across a range of running speeds ($2.5 - 4.17$ m/s), slopes ($0^\circ, \pm 5^\circ, \pm 10^\circ$), and step frequencies (preferred, $\pm 10\%$) from accelerometer data. Our findings indicate that an LSTM network given the runner's mass, height, running speed, slope, foot strike pattern, and sacral acceleration can predict normal GRF waveforms across a range of speeds and slopes with an RMSE of $0.12 - 0.20$ BW and rRMSE of $5.4 - 7.3\%$ (Figure 4). For comparison, previous studies report an RMSE of 0.39 ± 0.26 BW¹⁶, an RMSE of 0.21 ± 0.03 BW¹³, and an rRMSE of 13.92% ¹⁴ when using neural networks to predict the stance phase vertical GRF waveform during level-ground running. In contrast to previous studies, the LSTM network does not require preliminary stance phase identification or time normalization, which preserves the temporal component of the predicted GRF waveform. This characteristic of the LSTM network allowed us to calculate stride kinematic variables like step frequency and contact time with a MAPE $< 5\%$. Additionally, the recurrent nature of the LSTM network facilitates frame-by-frame predictions of GRF waveforms and can be used to make predictions over any duration of running. Thus, an LSTM network could be used to quantify changes in normal GRF waveforms over the course of a prolonged run (e.g., a marathon race).

The accuracy of predicted GRF waveforms varied across speeds and slopes, with a combination of faster running speeds and negative slopes producing greater RMSE values than

250 slower running speeds and positive slopes (Figure 4). The greater RMSE values during downhill
251 running may be due to the LSTM network's inability to account for changes in impact peak
252 magnitude across slopes (Figure 5). Previous studies have found that the presence of an impact
253 peak in the normal GRF waveform is subject-specific, affected by changes in running slope, and
254 associated with acceleration of the effective mass of the lower extremity during early stance
255 phase^{25,37,50}. Thus, predictions of normal GRF waveforms across slopes may be further improved
256 by incorporating accelerations measured at the feet or lower legs.

257 We also quantified the accuracy of the LSTM network when trained and tested on data
258 from the same subject. Although not a valid method of determining the LSTM network's
259 generalizability, single-subject validation provides insight into how well a personalized neural
260 network could predict an individual's GRF waveforms for unknown combinations of speed and
261 slope in the future. We found that predicted GRF waveforms of a representative subject (Subject
262 14) during the $\pm 5^\circ$ slope conditions had an average \pm SD RMSE of 0.08 ± 0.02 BW, compared to
263 0.16 ± 0.03 BW during LOSO cross validation. These findings indicate that a subject-specific
264 LSTM network was twice as accurate as the LOSO cross validated LSTM network. A single-
265 subject approach may be particularly beneficial for researchers, coaches, or clinicians who have
266 the resources to train personalized LSTM networks and wish to monitor a specific athlete's
267 biomechanics over the course of a competitive season. For example, an athlete could run at a
268 variety of speeds and slopes while wearing accelerometers during a baseline data collection on a
269 force-measuring treadmill at the start of their competitive season and a personalized LSTM
270 network could be trained on their data. Then, if accelerometer data were collected from an athlete
271 during training runs, their normal GRF waveforms and a variety of discrete values could be
272 predicted and monitored longitudinally.

The MAPE values for step frequency, contact time, impulse, and normal GRF active peak were $\leq 8.5\%$, but the loading rate MAPE was $27.6 \pm 36.1\%$. The lower MAPE values for step frequency, contact time, impulse, and normal GRF active peak indicate that the LSTM network consistently identified the general shape of the GRF waveform and the boundaries of the stance phase despite changes in speed, slope, and step frequency. However, the network did not consistently predict the presence of an impact peak during early stance phase (Figure 5, -10° trial), which affected the predicted slope of the GRF waveform during early stance phase and thus the accuracy of loading rate values. Although the prominence of an impact peak in the normal GRF waveform is affected by foot strike pattern and slope³⁷, two of the inputs for the LSTM network, the decreased accuracy when estimating loading rate may be because we did not include accelerometer data from the lower legs or feet. A previous study found moderate-strong correlations between axial tibial acceleration and vertical GRF impact peak magnitude ($r = 0.76$) and timing ($r = 0.94$) during running⁵¹. We did not include accelerometer data from the shoes as inputs for the LSTM network because the data were not available for both feet.

Recurrent neural networks represent a promising strategy for predicting continuous normal GRFs from wearable devices in outdoor environments. The LSTM network required data from three accelerometers (one on the sacrum and two on the right shoe to determine foot strike pattern), but we also performed a *post-hoc* analysis of prediction accuracy without the foot strike pattern data to quantify the network's accuracy when only using data from one sacral accelerometer. The *post-hoc* analysis revealed that excluding foot strike pattern data slightly increased the average \pm SD RMSE from 0.16 ± 0.04 BW to 0.17 ± 0.05 BW and rRMSE from $6.4 \pm 1.5\%$ to $6.7 \pm 1.7\%$. Excluding foot strike pattern data affected the MAPE of discrete variables by $< 3\%$ (Table 3). These findings indicate that the LSTM network can predict normal GRF waveforms from a single

accelerometer on the sacrum more accurately than neural networks implemented in previous studies ($RMSE = 0.21 - 0.39$ BW, $rRMSE = 13.92\%$), which required data from 3 – 7 wearable devices^{13,14,16}.

Using a recurrent neural network in combination with accelerometers and a global positioning system (GPS) device to obtain speed and slope data could potentially allow runners to receive biomechanical feedback during an outdoor run. Watches with GPS capabilities are commonly used by runners⁵², have been used to provide real-time feedback of step frequency⁵³, and could provide running speed and slope data to the LSTM network to predict GRF waveforms in near-real time⁵⁴. Discrete biomechanical variables could then be calculated from predicted normal GRF waveforms and sent to a clinician, coach, researcher, or the runner themselves. A similar approach has been implemented during outdoor walking and running using an integrated IMU-GPS device placed in a backpack, but it is unclear how accurate or generalizable this approach is as the network was trained and tested on data from three subjects and the reported accuracy metrics were combined for walking and running⁵⁵. To facilitate calculation of GRF-based variables during outdoor running using accelerometers, we have made the LSTM networks, which were trained on all subjects, with and without the need for foot strike data, publicly available alongside a tutorial at [www.github.com/alcantarar/Recurrent GRF Prediction](https://www.github.com/alcantarar/Recurrent_GRF_Prediction).

There are potential limitations to consider alongside our findings. The accelerometers used in the present study were adhered to subjects using tape and a less secure attachment method may introduce movement artefact into the accelerometer data. Previous research suggests that attachment method can affect peak tibial acceleration during running⁵⁶, but the lower leg experiences larger accelerations than the sacrum during running²³ and thus is more sensitive to different attachment methods. Using the LSTM network to predict normal GRF waveforms from

a sacral accelerometer adhered differently than in the present study may affect prediction accuracy, but the 20 Hz low-pass filter we applied to the accelerometer data can potentially mitigate this effect. Additionally, predictions made with the LSTM network may not be generalizable for speeds or slopes that fall outside the range of the training data (2.5 – 4.17 m/s and $\pm 10^\circ$) as biomechanics change when running on steep slopes (e.g. 20 – 40°)^{3,57} or with changes in speed¹². Lastly, the LSTM network was trained on data collected on a stiff force-measuring treadmill and thus accelerometer data collected during running on less stiff surfaces (e.g., grass) may result in greater prediction errors given the effects of surface stiffness on running biomechanics and thus energy absorption^{58,59}.

In conclusion, we developed a recurrent neural network that used accelerometer data to predict continuous normal GRF waveforms across a range of running speeds (2.5 – 4.17 m/s) and slopes (0°, $\pm 5^\circ$, $\pm 10^\circ$) with an average \pm SD RMSE of 0.16 ± 0.04 BW and rRMSE of $6.4 \pm 1.5\%$. Unlike neural networks implemented in prior studies, the recurrent neural network does not require preliminary identification of the stance phase or temporal normalization and allows for near real-time predictions of normal GRF waveforms during running. Accurate predictions of normal GRF waveforms using wearable devices will improve the ability to longitudinally monitor biomechanical variables in non-laboratory environments.

Acknowledgements

We utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University.

Table 1. Mean \pm SD root mean square error (RMSE) and relative RMSE (rRMSE) for normal GRF waveforms predicted by the LSTM network compared to the measured normal GRF waveforms for each subject.

Subject	RMSE [BW]	rRMSE [%]
1	0.11 ± 0.04	4.0 ± 1.1
2	0.11 ± 0.02	4.1 ± 0.7
3	0.12 ± 0.02	4.4 ± 0.9
4	0.12 ± 0.03	4.6 ± 1.2
5	0.12 ± 0.04	4.7 ± 0.9
6	0.12 ± 0.03	5.2 ± 1.4
7	0.13 ± 0.03	5.1 ± 1.2
8	0.15 ± 0.08	5.3 ± 3.3
9	0.15 ± 0.03	5.4 ± 1.0
10	0.15 ± 0.04	5.2 ± 1.3
11	0.15 ± 0.03	5.8 ± 0.8
12	0.16 ± 0.03	6.1 ± 1.1
13	0.16 ± 0.06	5.7 ± 1.5
14	0.17 ± 0.07	6.9 ± 2.5
15	0.17 ± 0.04	7.9 ± 2.0
16	0.19 ± 0.05	6.7 ± 1.7
17	0.21 ± 0.07	7.2 ± 2.5
18	0.29 ± 0.03	14.0 ± 1.7
19	0.31 ± 0.07	13.7 ± 2.7
Mean \pm SD	0.16 ± 0.04	$6.4 \pm 1.5\%$

Table 2. Mean \pm SD discrete variables calculated from normal ground reaction force (GRF) waveforms predicted by the LSTM network (“Predicted”) and normal GRF waveforms measured from the force-measuring treadmill (“Measured”) across all speeds and subjects for each slope.

Slope	Step Frequency [Hz]		Contact Time [ms]		Impulse [BW*s]	
	Predicted	Measured	Predicted	Measured	Predicted	Measured
-10°	3.1 \pm 0.3	3.1 \pm 0.3	215 \pm 26	214 \pm 29	0.33 \pm 0.03	0.33 \pm 0.05
-5°	3.1 \pm 0.3	3.1 \pm 0.3	222 \pm 23	223 \pm 27	0.34 \pm 0.04	0.34 \pm 0.04
0°	3.1 \pm 0.3	3.1 \pm 0.3	228 \pm 24	229 \pm 27	0.33 \pm 0.04	0.34 \pm 0.04
+5°	3.2 \pm 0.3	3.2 \pm 0.3	228 \pm 25	232 \pm 26	0.32 \pm 0.03	0.33 \pm 0.04
+10°	3.4 \pm 0.3	3.4 \pm 0.3	223 \pm 26	227 \pm 26	0.30 \pm 0.04	0.30 \pm 0.04

Slope	Active Peak [BW]		Loading Rate [BW/s]	
	Predicted	Measured	Predicted	Measured
-10°	2.38 \pm 0.24	2.36 \pm 0.40	64.7 \pm 19.9	68.4 \pm 20.6
-5°	2.41 \pm 0.25	2.45 \pm 0.36	54.9 \pm 16.7	60.6 \pm 19.4
0°	2.42 \pm 0.22	2.51 \pm 0.36	42.0 \pm 12.8	49.0 \pm 17.2
+5°	2.36 \pm 0.21	2.40 \pm 0.33	34.8 \pm 8.5	39.3 \pm 13.3
+10°	2.24 \pm 0.22	2.20 \pm 0.29	30.1 \pm 8.3	31.3 \pm 12.7

345
346

Table 3. Error metrics for the predicted waveforms and discrete variables when training the Long Short-Term Memory (LSTM) network with and without foot strike pattern as an input feature.

	LSTM With Foot Strike (Sacral + Right Foot Accelerometers)	LSTM Without Foot Strike (Only Sacral Accelerometer)
GRF Waveform		
RMSE [BW]	0.16 ± 0.04	0.17 ± 0.05
rRMSE	$6.4 \pm 1.5\%$	$6.7 \pm 1.7\%$
Discrete Variables		
MAPE		
Step Frequency	$0.1 \pm 0.1\%$	$0.1 \pm 0.1\%$
Contact Time	$4.9 \pm 4.0\%$	$5.6 \pm 4.5\%$
Impulse	$6.4 \pm 6.9\%$	$6.0 \pm 7.1\%$
Active Peak	$8.5 \pm 8.2\%$	$7.7 \pm 6.3\%$
Loading Rate	$27.6 \pm 36.1\%$	$30.3 \pm 41.6\%$

Note. Root mean square error (RMSE) and relative RMSE (rRMSE) are presented for the predicted normal ground reaction force (GRF) waveforms. Mean absolute percent error (MAPE) values are presented for the discrete variables calculated from normal GRF waveforms predicted by both LSTM networks.

Figure Captions

Figure 1 – Overview of the Long Short-Term Memory (LSTM) network's input features and function. **A)** The LSTM network's input features included body mass, height, running speed, slope, and percentage of a trial's steps classified as rearfoot (RFS), midfoot (MFS), or forefoot (FFS) strikes. **B)** Filtered vertical and anteroposterior sacral acceleration data, relative to the device's local coordinate system. **C)** Accelerometer data were divided into overlapping 6 frame (12 ms) windows, one for each frame of the ground reaction force (GRF) data. The mean, standard deviation (SD), and range of vertical and anteroposterior sacral acceleration values were calculated for each window and used as input features to the LSTM network. For the prediction of a normal GRF value at a given time (t), the respective window of acceleration data begins at $t-3$ and ends at $t+2$. **D)** Normal GRFs were predicted frame-by-frame by the LSTM network using the 13 input features.

Figure 2 – The Long Short-Term Memory (LSTM) network consisted of a Bidirectional LSTM layer with a hyperbolic tangent activation function followed by a multilayer perceptron (MLP) with rectified linear unit activation functions for three hidden layers with 128, 384, and 320 neurons, respectively. The Bidirectional LSTM layer is unraveled to illustrate its recurrent nature and dashed lines signify inputs (x) and outputs (h) at time $t-1$ and $t+1$. A dropout rate of 20% was applied to the input layer of the network and a dropout rate of 40% was applied to the output of the Bidirectional LSTM layer to limit network overfitting. For each prediction of the normal ground reaction force (GRF) at a given time (t), the network received 13 features as inputs (x_t ; Figure 1), passed the output from the Bidirectional LSTM layer (h_t) to the MLP, and predicted a single value (y_t) with a linear activation function in the output layer.

Figure 3 – Ground reaction force waveform prediction error for each subject across all conditions. The average root mean square error (RMSE) across subjects was 0.16 BW (dotted line). Filled circles represent each trial, and the color indicates slope (0° , $\pm 5^\circ$, $\pm 10^\circ$) at three speeds (2.5, 3.33, 4.17 m/s). Open circles represent each subject's average RMSE, horizontal bars are the median RMSE, box plot edges indicate the interquartile range (IQR; 25th and 75th percentile), and the whiskers encompass values that fall within $1.5 \times \text{IQR}$. Subjects are sorted from lowest to highest RMSE.

Figure 4 – The average \pm SD root mean square error (RMSE) of the predicted ground reaction force (GRF) waveforms compared to the GRF waveform measured by the force-measuring treadmill for each condition during leave-one-subject-out (LOSO) cross validation.

Figure 5 – Normal ground reaction force (GRF) waveforms predicted by the recurrent neural network (dashed red lines) and measured by the force-measuring treadmill (solid blue lines) at 3.33 m/s and all slopes (0° , $\pm 5^\circ$, $\pm 10^\circ$) are presented for Subject 14. Subject 14 was selected because they had similar RMSE values (0.17 ± 0.07 BW) as the average across all subjects (0.16 ± 0.04 BW) and their GRF waveforms illustrate an interaction between running slope and normal GRF impact peak magnitude.

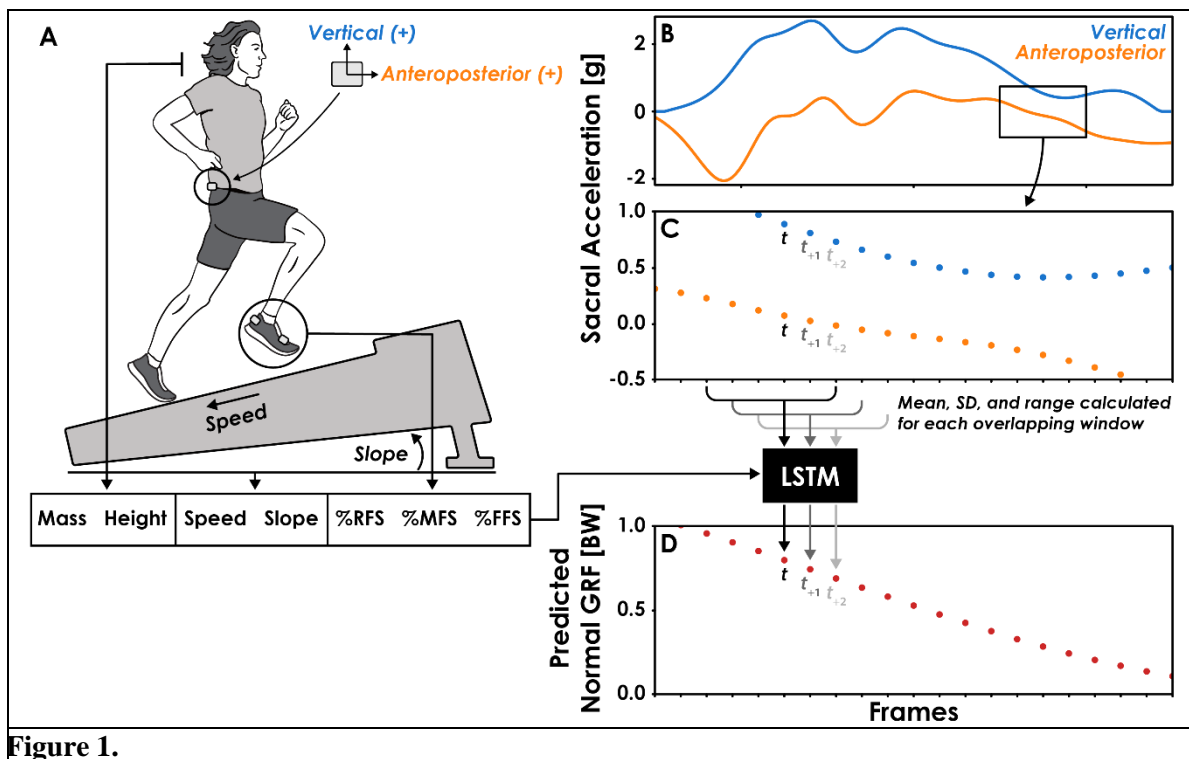


Figure 1.

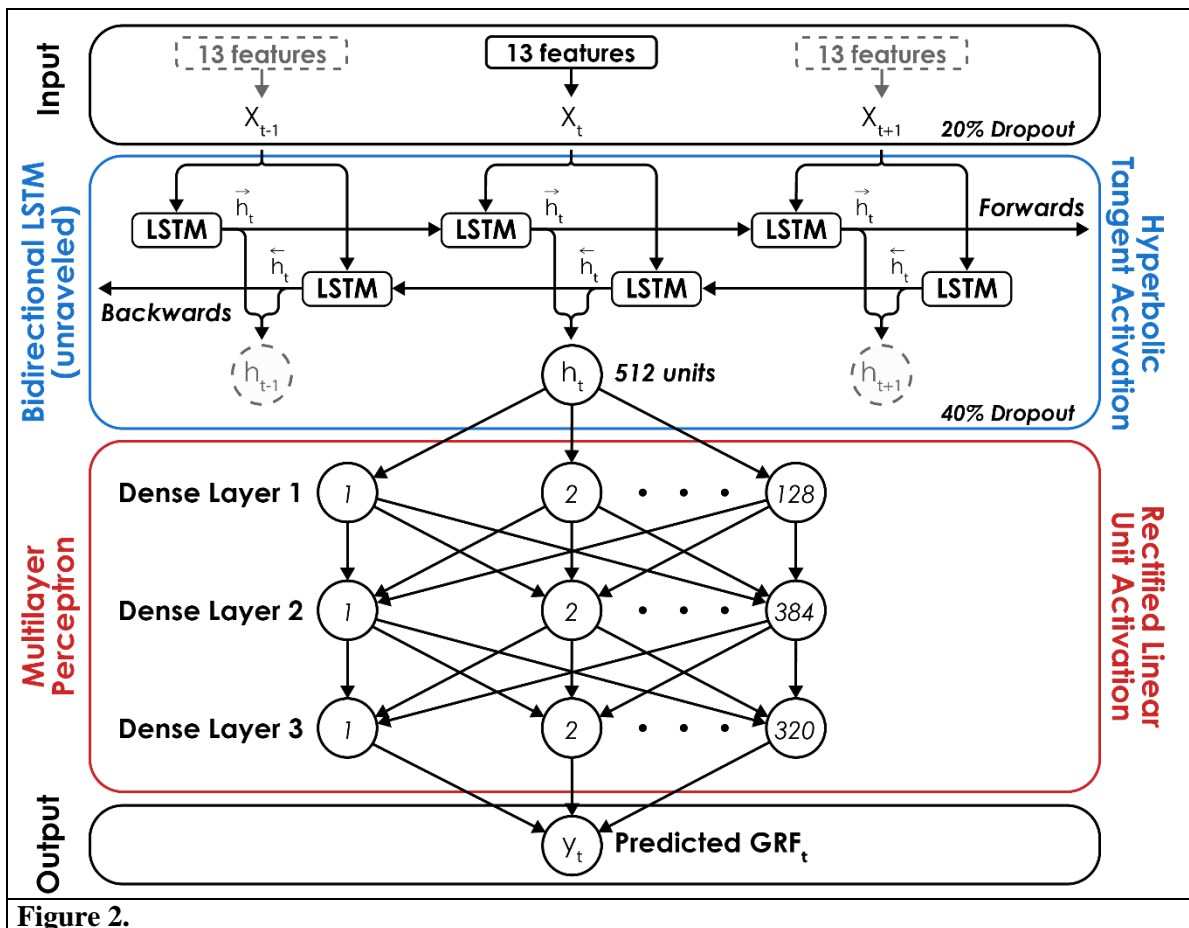


Figure 2.

393

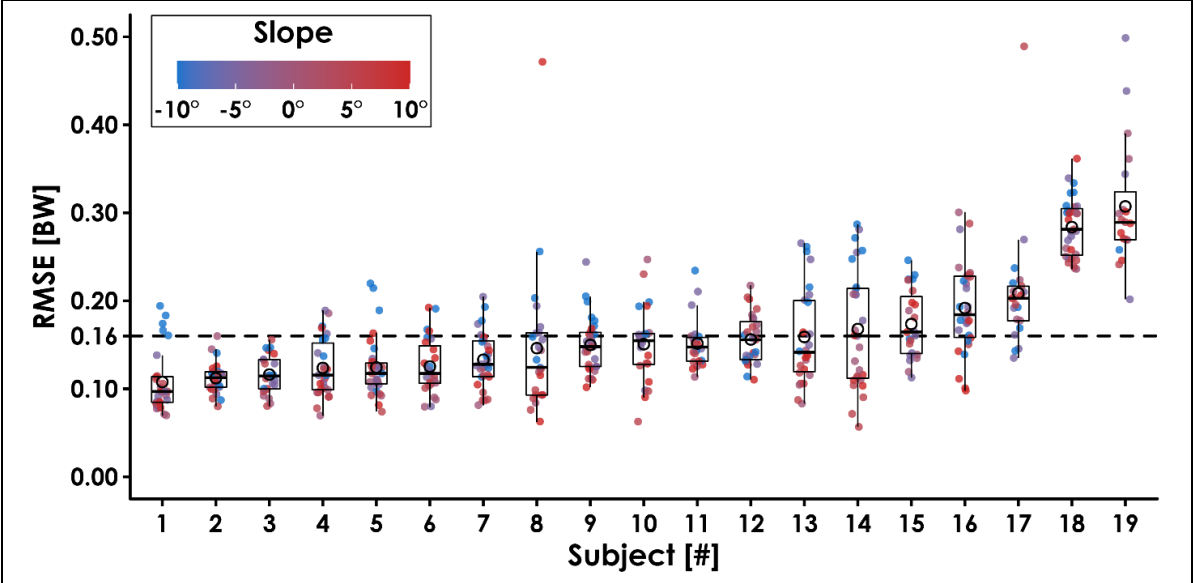


Figure 3.

394

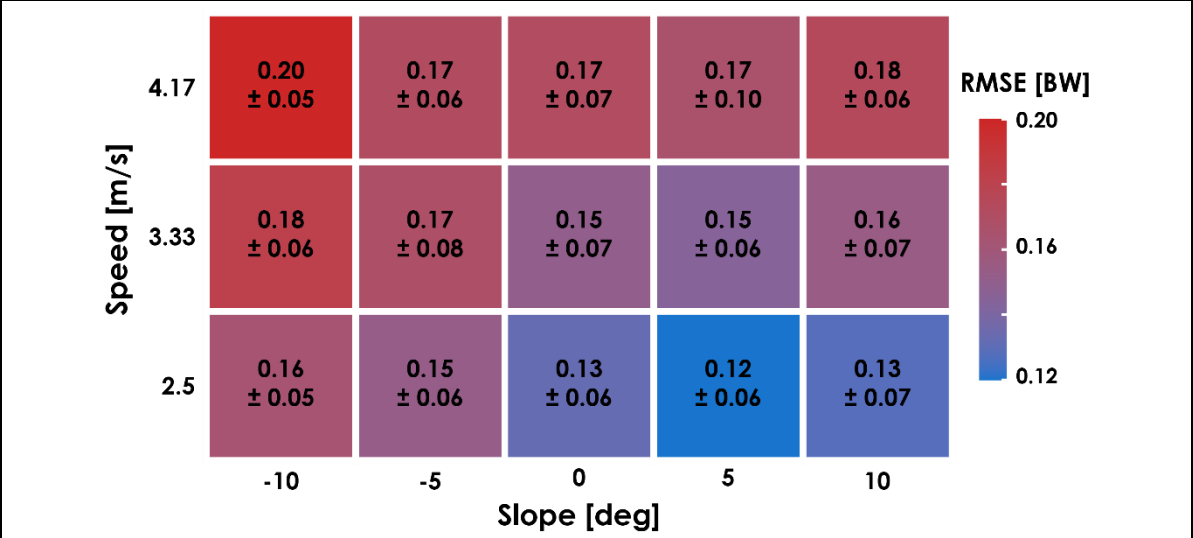


Figure 4.

395

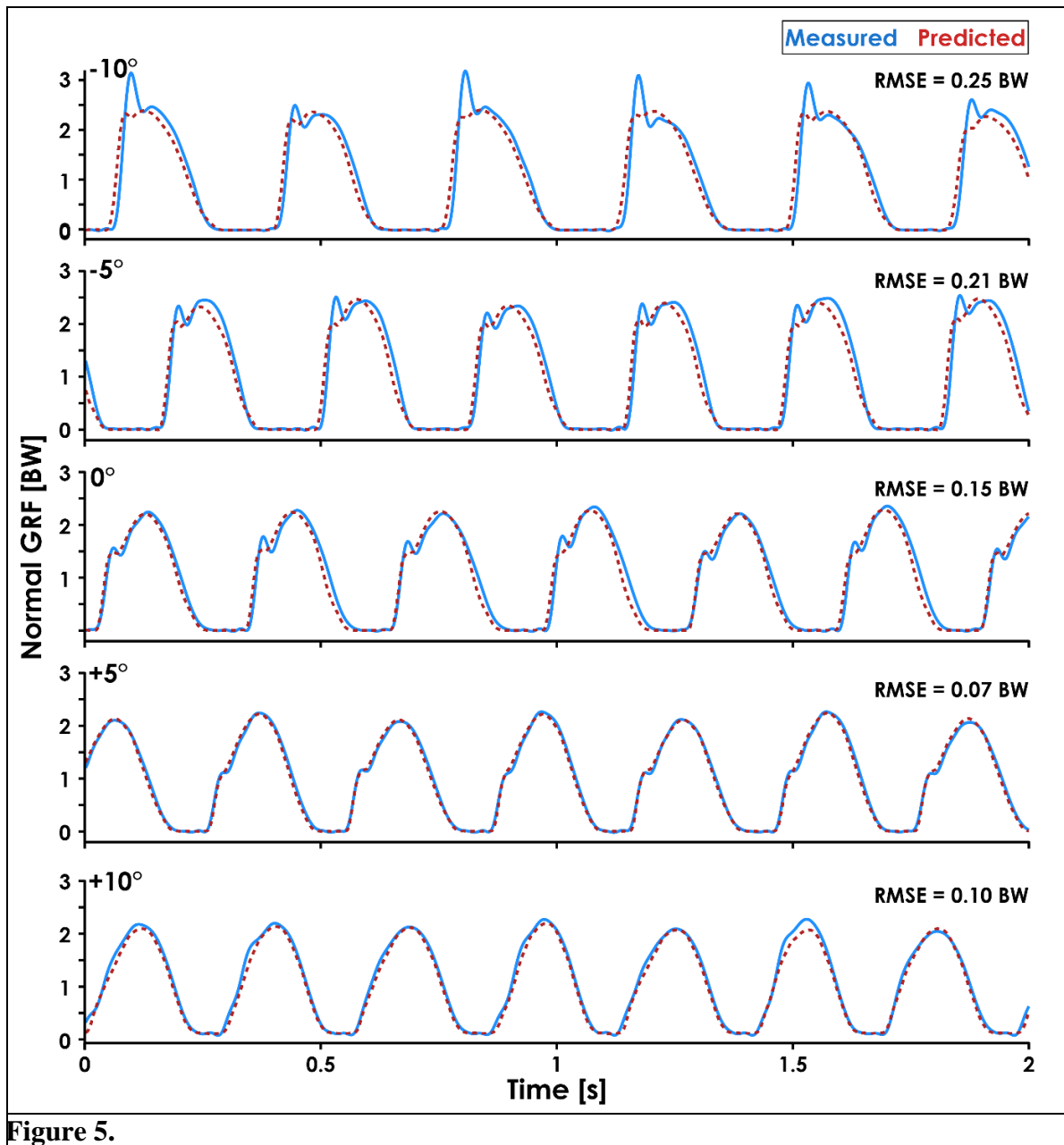


Figure 5.

- 398 1. Ceysens L, Vanelderen R, Barton C, Malliaras P, Dingenen B. Biomechanical risk factors
399 associated with running-related injuries: A systematic review. *Sport Med.* 2019;49(7):1095-
400 1115. doi:10.1007/s40279-019-01110-z
- 401 2. Voloshina AS, Ferris DP. Biomechanics and energetics of running on uneven terrain. *J Exp*
402 *Biol.* 2015;218(5):711-719. doi:10.1242/jeb.106518
- 403 3. Whiting CS, Allen SP, Brill JW, Kram R. Steep (30°) uphill walking vs. running: COM
404 movements, stride kinematics, and leg muscle excitations. *Eur J Appl Physiol.*
405 2020;120(10):2147-2157. doi:10.1007/s00421-020-04437-y
- 406 4. Kipp S, Taboga P, Kram R. Ground reaction forces during steeplechase hurdling and
407 waterjumps. *Sports Biomech.* 2017;16(2):152-165. doi:10.1080/14763141.2016.1212917
- 408 5. Clermont CA, Benson LC, Edwards WB, Hettinga BA, Ferber R. New Considerations for
409 wearable technology data: Changes in running biomechanics during a marathon. *J Appl*
410 *Biomech.* 2019;35(6):401-409. doi:10.1123/jab.2018-0453
- 411 6. Reenalda J, Maartens E, Homan L, Buurke JH. Continuous three dimensional analysis of
412 running mechanics during a marathon by means of inertial magnetic measurement units to
413 objectify changes in running mechanics. *J Biomech.* 2016;49(14):3362-3367.
414 doi:10.1016/j.jbiomech.2016.08.032
- 415 7. Ruder M, Jamison ST, Tenforde A, Mulloy F, Davis IS. Relationship of foot strike pattern
416 and landing impacts during a marathon. *Med Sci Sports Exerc.* 2019;51(10):2073-2079.
417 doi:10.1249/MSS.0000000000002032
- 418 8. Neugebauer JM, Hawkins DA, Beckett L. Estimating youth locomotion ground reaction
419 forces using an accelerometer-based activity monitor. *PLoS One.* 2012;7(10):e48182.
420 doi:10.1371/journal.pone.0048182
- 421 9. Ancillao A, Tedesco S, Barton J, O'Flynn B. Indirect measurement of ground reaction
422 forces and moments by means of wearable inertial sensors: A systematic review. *Sensors*
423 *(Basel).* 2018;18(8). doi:10.3390/s18082564
- 424 10. Kiernan D, Hawkins DA, Manoukian MAC, et al. Accelerometer-based prediction of
425 running injury in National Collegiate Athletic Association track athletes. *J Biomech.*
426 2018;73:201-209. doi:10.1016/j.jbiomech.2018.04.001
- 427 11. Derie R, Robberechts P, Van den Berghe P, et al. Tibial acceleration-based prediction of
428 maximal vertical loading rate during overground running: A machine learning approach.
429 *Front Bioeng Biotechnol.* 2020;8. doi:10.3389/fbioe.2020.00033
- 430 12. Alcantara R, Day E, Hahn M, Grabowski A. Sacral acceleration can predict whole-body
431 kinetics and stride kinematics across running speeds. *PeerJ.* 2021.

13. Dorschky E, Nitschke M, Martindale CF, van den Bogert AJ, Koelewijn AD, Eskofier BM. CNN-based estimation of sagittal plane walking and running biomechanics from measured and simulated inertial sensor data. *Front Bioeng Biotechnol.* 2020;8:604. doi:10.3389/fbioe.2020.00604
14. Johnson WR, Mian A, Robinson MA, Verheul J, Lloyd DG, Alderson J. Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning. *IEEE T Biomed Eng.* 2021;68(1):289-297. doi:10.1109/TBME.2020.3006158
15. Pogson M, Verheul J, Robinson MA, Vanrenterghem J, Lisboa P. A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities. *Med Eng Phys.* 2020;78:82-89. doi:10.1016/j.medengphy.2020.02.002
16. Wouda FJ, Giuberti M, Bellusci G, et al. Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors. *Front Physiol.* 2018;9. doi:10.3389/fphys.2018.00218
17. Running USA. 2019 U.S Running Trends Report. Published online 2019.
18. International Trail Running Association. 2020 Trail Running Infographics. Published online 2020. <https://itra.run/documents/Infographics/EN-ITRA-Trail-Running-Infographics-2020.pdf>
19. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9(8):1735-1789.
20. Wang S, Jiang J. Learning natural language inference with LSTM. *arXiv:151208849 [cs]*. Published online November 10, 2016. Accessed November 22, 2020. <http://arxiv.org/abs/1512.08849>
21. Mundt M, Koeppe A, David S, Bamer F, Potthast W, Markert B. Prediction of ground reaction force and joint moments based on optical motion capture data during gait. *Med Eng Phys.* 2020;86:29-34. doi:10.1016/j.medengphy.2020.10.001
22. Choi A, Jung H, Mun JH. Single inertial sensor-based neural networks to estimate COM-COP inclination angle during walking. *Sensors (Basel).* 2019;19(13). doi:10.3390/s19132974
23. Baggaley M, Vernillo G, Martinez A, et al. Step length and grade effects on energy absorption and impact attenuation in running. *Euro J Sport Sci.* Published online September 24, 2019:1-11. doi:10.1080/17461391.2019.1664639
24. Khassestarash A, Vernillo G, Martinez A, et al. Biomechanics of graded running: Part II—Joint kinematics and kinetics. *Scand J Med Sci Sport.* 2020;30(9):1642-1654. doi:10.1111/sms.13735

25. Vernillo G, Martinez A, Baggaley M, et al. Biomechanics of graded running: Part I - Stride parameters, external forces, muscle activations. *Scand J Med Sci Sport*. 2020;30(9):1632-1641. doi:10.1111/sms.13708
26. Giandolini M, Poupard T, Gimenez P, et al. A simple field method to identify foot strike pattern during running. *J Biomech*. 2014;47(7):1588-1593. doi:10.1016/j.jbiomech.2014.03.002
27. Day EM, Alcantara RS, McGeehan MA, Grabowski AM, Hahn ME. Low-pass filter cutoff frequency affects sacral-mounted inertial measurement unit estimations of peak vertical ground reaction force and contact time during treadmill running. *J Biomech*. Published online February 9, 2021:110323. doi:10.1016/j.jbiomech.2021.110323
28. Figo D, Diniz P, Ferreira D, Cardoso J. Preprocessing techniques for context recognition from accelerometer data. *Pers Ubiquit Comput*. 2010;14:645-662. doi:10.1007/s00779-010-0293-9
29. Almeida MO, Davis IS, Lopes AD. Biomechanical differences of foot-strike patterns during running: A systematic review with meta-analysis. *J Orthop Sports Phys Ther*. 2015;45(10):738-755. doi:10.2519/jospt.2015.6019
30. Vincent HK, Kilgore JE, Chen C, Bruner M, Horodyski M, Vincent KR. Impact of body mass index on biomechanics of recreational runners. *PM&R*. 2020;12(11):1106-1112. doi:https://doi.org/10.1002/pmrj.12335
31. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *J Mach Learn Res*. 2018;18(185):1-52.
32. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv:1412.6980 [cs]*. Published online January 29, 2017. Accessed November 9, 2020. <http://arxiv.org/abs/1412.6980>
33. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467 [cs]*. Published online March 16, 2016. Accessed November 9, 2020. <http://arxiv.org/abs/1603.04467>
34. Halilaj E, Rajagopal A, Fiterau M, Hicks JL, Hastie TJ, Delp SL. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *J Biomech*. 2018;81:1-11. doi:10.1016/j.jbiomech.2018.09.009
35. Chaibub Neto E, Pratap A, Perumal TM, et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *npj Digital Medicine*. 2019;2(1):1-6. doi:10.1038/s41746-019-0178-x
36. Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. *Gigascience*. 2017;6(5). doi:10.1093/gigascience/gix019

- 502 37. Gottschall JS, Kram R. Ground reaction forces during downhill and uphill running. *J*
503 *Biomech.* 2005;38(3):445-452. doi:10.1016/j.jbiomech.2004.04.023
- 504 38. Ren L, Jones RK, Howard D. Whole body inverse dynamics over a complete gait cycle
505 based only on measured kinematics. *J Biomech.* 2008;41(12):2750-2759.
506 doi:10.1016/j.jbiomech.2008.06.001
- 507 39. Yong JR, Silder A, Montgomery KL, Fredericson M, Delp SL. Acute changes in foot strike
508 pattern and cadence affect running parameters associated with tibial stress fractures. *J*
509 *Biomech.* 2018;76:1-7. doi:10.1016/j.jbiomech.2018.05.017
- 510 40. Alcantara RS. Dryft: A Python and MATLAB package to correct drifting ground reaction
511 force signals during treadmill running. *J Open Source Softw.* 2019;4(44):1910.
512 doi:10.21105/joss.01910
- 513 41. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature.*
514 2020;585(7825):357-362. doi:10.1038/s41586-020-2649-2
- 515 42. pandas development team. *Pandas-Dev/Pandas: Pandas.*; 2020.
516 <https://doi.org/10.5281/zenodo.3509134>
- 517 43. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundataion
518 for Statistical Computing; 2020.
- 519 44. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for
520 scientific computing in Python. *Nat Methods.* 2020;17(3):261-272. doi:10.1038/s41592-
521 019-0686-2
- 522 45. Wickham H, François R, Henry L, Müller K. *Dplyr: A Grammar of Data Manipulation.*;
523 2020. <https://CRAN.R-project.org/package=dplyr>
- 524 46. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag; 2009.
525 doi:10.1007/978-0-387-98141-3
- 526 47. Wickham H. *Stringr: Simple, Consistent Wrappers for Common String Operations.*; 2019.
527 <https://CRAN.R-project.org/package=stringr>
- 528 48. Cavagna GA, Mantovani M, Willems PA, Musch G. The resonant step frequency in human
529 running. *Pflügers Arch.* 1997;434(6):678-684. doi:10.1007/s004240050451
- 530 49. Snyder KL, Farley CT. Energetically optimal stride frequency in running: the effects of
531 incline and decline. *J Exp Biol.* 2011;214(Pt 12):2089-2095. doi:10.1242/jeb.053157
- 532 50. McMahon TA, Valiant G, Frederick EC. Groucho running. *J Appl Physiol.*
533 1987;62(6):2326-2337. doi:10.1152/jappl.1987.62.6.2326
- 534 51. Hennig EM, Lafortune MA. Relationships between ground reaction force and tibial bone
535 acceleration parameters. *Int J Sport Biomech.* 1991;7(3):303-309.

- 536 52. Janssen M, Walravens R, Thibaut E, Scheerder J, Brombacher A, Vos S. Understanding
537 different types of recreational runners and how they use running-related technology. *Int J*
538 *Environ Res Public Health*. 2020;17(7). doi:10.3390/ijerph17072276
- 539 53. Willy RW, Buchenic L, Rogacki K, Ackerman J, Schmidt A, Willson JD. In-field gait
540 retraining and mobile monitoring to address running biomechanics associated with tibial
541 stress fracture. *Scand J Med Sci Sport*. 2016;26(2):197-205.
542 doi:https://doi.org/10.1111/sms.12413
- 543 54. Scott MTU, Scott TJ, Kelly VG. The validity and reliability of global positioning systems
544 in team sport: A brief review. *J Strength Cond Res*. 2016;30(5):1470-1490.
545 doi:10.1519/JSC.0000000000001221
- 546 55. Vitali RV, Cain SM, Ojeda LV, et al. Body-worn IMU array reveals effects of load on
547 performance in an outdoor obstacle course. Jan Y-K, ed. *PLoS One*. 2019;14(3):e0214008.
548 doi:10.1371/journal.pone.0214008
- 549 56. Johnson CD, Outerleys J, Tenforde AS, Davis IS. A comparison of attachment methods of
550 skin mounted inertial measurement units on tibial accelerations. *J Biomech*.
551 2020;113:110118. doi:10.1016/j.jbiomech.2020.110118
- 552 57. Giovanelli N, Ortiz ALR, Henninger K, Kram R. Energetics of vertical kilometer foot
553 races; is steeper cheaper? *J Appl Physiol*. 2016;120(3):370-375.
554 doi:10.1152/jappphysiol.00546.2015
- 555 58. Derrick TR, Hamill J, Caldwell GE. Energy absorption of impacts during running at various
556 stride lengths. *Med Sci Sports Exerc*. 1998;30(1):128-135. doi:10.1097/00005768-
557 199801000-00018
- 558 59. Ferris DP, Louie M, Farley CT. Running in the real world: adjusting leg stiffness for
559 different surfaces. *Proc R Soc Lond B*. 1998;265(1400):989-994.
560 doi:10.1098/rspb.1998.0388