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

Ryan S Alcantara [Corresp., 1], **W Brent Edwards** [2], **Guillaume Y Millet** [3], **Alena M Grabowski** [1]

[1] Department of Integrative Physiology, University of Colorado, Boulder, CO, United States

[2] Human Performance Laboratory, Faculty of Kinesiology, University of Calgary, Calgary, Alberta, Canada

[3] Laboratoire Interuniversitaire de Biologie de la Motricité, Université Lyon, UJM-Saint-Etienne, Saint-Etienne, France

Corresponding Author: Ryan S Alcantara
Email address: ryan.alcantara@colorado.edu

Background. Ground reaction forces (GRFs) are important for understanding human movement, but their measurement is generally limited to a laboratory environment. Previous studies have 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. A method of predicting the normal (perpendicular to running surface) GRF waveform using wearable devices across a range of running speeds and slopes could allow researchers and clinicians to predict kinetic and kinematic variables outside the laboratory environment.

Purpose. 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.

Methods. 19 subjects ran on a force-measuring treadmill at five slopes (0°, ±5°, ±10°) 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 ± SD RMSE of 0.16 ± 0.04 BW and relative RMSE of 6.4 ± 1.5% across all conditions and subjects.

Results. 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 accurately of quantifying and monitoring external forces experienced by the body when running.

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

Ryan S. Alcantara[1], W. Brent Edwards[2], Guillaume Y. Millet[3], Alena M. Grabowski[1]

[1] Department of Integrative Physiology, University of Colorado Boulder, Boulder, CO, USA
[2] Human Performance Laboratory, Faculty of Kinesiology, University of Calgary, Calgary, Canada
[3] Laboratoire Interuniversitaire de Biologie de la Motricité, Univ Lyon, UJM-Saint-Etienne, Saint-Etienne, France

Corresponding Author:
Ryan S. Alcantara[1]
Department of Integrative Physiology, University of Colorado Boulder, Boulder, CO, USA
Email address: ryan.alcantara@colorado.edu

Abstract

Background. Ground reaction forces (GRFs) are important for understanding human movement, but their measurement is generally limited to a laboratory environment. Previous studies have 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. A method of predicting the normal (perpendicular to running surface) GRF waveform using wearable devices across a range of running speeds and slopes could allow researchers and clinicians to predict kinetic and kinematic variables outside the laboratory environment.

Purpose. 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.

Methods. 19 subjects ran on a force-measuring treadmill at five slopes (0°, ±5°, ±10°) 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 ± SD RMSE of 0.16 ± 0.04 BW and relative RMSE of 6.4 ± 1.5% across all conditions and subjects.

Results. 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 accurately of quantifying and monitoring external forces experienced by the body when running.

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 injuries (Ceyssens et al., 2019). 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 environment (Voloshina & Ferris, 2015; Kipp, Taboga & Kram, 2017; Whiting et al., 2020). 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 events (Reenalda et al., 2016; Ruder et al., 2019; Clermont et al., 2019). 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 rate (Neugebauer, Hawkins & Beckett, 2012; Kiernan et al., 2018; Ancillao et al., 2018; Derie et al., 2020; Alcantara et al., 2021) from IMU data.

Recently, neural networks have been used to predict GRF waveforms during running (Wouda et al., 2018; Pogson et al., 2020; Dorschky et al., 2020; Johnson et al., 2021), 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 step (Dorschky et al., 2020) or stance phase (Wouda et al., 2018; Johnson et al., 2021). Temporal normalization is typically accomplished by identifying gait events in the GRF waveform and segmenting the neural network's input signals, preventing the calculation of biomechanical variables with a temporal component (e.g., ground contact time, step frequency, vertical impulse, and loading rate) outside the laboratory where the GRF waveform is unavailable. Additionally, previous studies have predicted GRF waveforms only during level-ground running (Wouda et al., 2018; Pogson et al., 2020; Dorschky et al., 2020; Johnson et al., 2021), 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(Running USA, 2019; International Trail Running Association, 2020) 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) networks (Hochreiter & Schmidhuber, 1997) 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 tasks (Wang & Jiang, 2016). In the field of Biomechanics, LSTM networks have been used to make frame-by-frame predictions of GRF waveforms using motion capture data (Mundt et al., 2020) and predictions of the center of mass position relative to center of pressure from IMU data during walking (Choi, Jung & Mun, 2019). Developing an LSTM network to predict GRF waveforms from wearable device data would allow researchers to predict GRF waveforms not only during an isolated stance phase, but continuously for multiple steps or the entire duration of a run. IMUs have already been used to longitudinally measure biomechanical variables (Reenalda et al., 2016; Kiernan et al., 2018; Ruder et al., 2019; Clermont et al., 2019) 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. Improving means of remotely measuring biomechanical variables with wearable devices may improve a clinician's ability to identify injury risk factors, monitor rehabilitation, or develop interventions (Gurchiek, Cheney, & McGinnis, 2019).

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 BW (Dorschky et al., 2020) and relative RMSE (rRMSE; RMSE normalized to the average range of the compared waveforms; Eq. 1) of 13.92% (Johnson et al., 2021).

Materials & Methods

Subjects

We analyzed a pre-existing dataset (Baggaley et al., 2019; Khassetarash et al., 2020; Vernillo et al., 2020) 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 written 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°, ±5°, ±10°) at three speeds (2.5, 3.33, 4.17 m/s) per slope, and three step frequencies (preferred and ± 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 method (Giandolini et al., 2014), which provided the percentage of a trial's foot strikes classified as either rearfoot, midfoot, or forefoot strikes. The biaxial accelerometer was placed on the sacrum such that the vertical axis in the accelerometer's local coordinate system was oriented superiorly, but we did not perform a calibration to align accelerometer and treadmill coordinate systems (Figure 1). We did not align the accelerometer and treadmill coordinate systems because an LSTM network can likely learn the transformation between the coordinate systems and requiring this preprocessing calibration may limit the utility of an LSTM network outside a laboratory setting.

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 (Day et al., 2021; Alcantara et al., 2021). We normalized GRFs to bodyweight (BW) and filtered them using a 4^{th} order low-pass Butterworth filter with a 30 Hz cut-off. We filtered the sacral acceleration data with a 4^{th} order low-pass Butterworth filter with a 20 Hz cut-off. Preliminary analysis revealed that a 20 Hz cut-off improved prediction accuracy compared to 5 Hz, 10 Hz, 30 Hz, and no filter 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. Vertical center of mass acceleration is primarily negative during the aerial phase of running (Blickhan 1989) and preliminary analysis revealed that replacing negative vertical sacral accelerometer data 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). Preliminary analysis revealed that a window width of 6 frames was the smallest window we could use without decreasing LSTM network prediction accuracy and we found no improvement in prediction accuracy when using window sizes up to 60 frames (120 ms). 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 cost (Figo et al., 2010). 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 (Almeida, Davis & Lopes, 2015; Khassetarash et al., 2020; Vernillo et al., 2020; Vincent et al., 2020; Alcantara et al., 2021). We chose not to include step frequency as an input feature, despite the presence of the ±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 algorithm (Li et al., 2018) 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 epoch (Kingma & Ba, 2017) 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 library (Abadi et al., 2016).

Network Validation

We assessed the accuracy and generalizability of the network using a Leave-One-Subject-Out (LOSO) cross validation method (Halilaj et al., 2018). 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 (Saeb et al., 2017; Chaibub Neto et al., 2019).

In addition to the LOSO cross validation method, we performed a test-train split of one representative subject's data according to slope (±5° trials reserved for testing, 0° and ±10° 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 magnitude (Gottschall & Kram, 2005). 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

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

where x_1 and x_2 are the GRF waveforms predicted by the LSTM network and measured by the force-measuring treadmill (Ren, Jones & Howard, 2008). 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%) (Gottschall & Kram, 2005; Vernillo et al., 2020). 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 phase (Yong et al., 2018), 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 (Wickham, 2009, 2019; Alcantara, 2019;

pandas development team, 2020; R Core Team, 2020; Wickham et al., 2020; Virtanen et al., 2020; Harris et al., 2020).

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 (Day et al., 2021; Alcantara et al., 2021) and during uphill and downhill running, step frequency is ≤ 4 Hz (Cavagna et al., 1997; Snyder & Farley, 2011). 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 ± SD RMSE of 0.16 ± 0.04 BW (Figure 3) and rRMSE of 6.4 ± 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°; 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 ± 0.1%, contact time was 4.9 ± 4.0%, impulse was 6.4 ± 6.9%, normal GRF active peak was 8.5 ± 8.2%, and loading rate was 27.6 ± 36.1% (Table 2).

The prediction error for one representative subject's (Subject 14) normal GRF waveforms at ±5° during single-subject validation was lower than those resulting from the LOSO cross validation, with an average ± SD RMSE of 0.08 ± 0.02 BW and rRMSE of 3.3 ± 0.9%. The MAPE of step frequency (0.1 ± 0.1 %), contact time (3.0 ± 2.3%), impulse (2.5 ± 1.9%), normal GRF active peak (2.7 ± 2.0%), and loading rate (17.6 ± 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°, ±5°, ±10°), and step frequencies (preferred, ±10%) from accelerometer data. Our findings indicate that an LSTM network with the runner's mass, height, running speed, slope, foot strike pattern, and sacral acceleration as input features 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 BW (Wouda et al., 2018), an RMSE of 0.21 ± 0.03 BW (Dorschky et al., 2020), and an rRMSE of 13.92% (Johnson et al., 2021) 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 slower running speeds and positive slopes (Figure 4). The greater RMSE values during downhill running may be due to the LSTM network's inability to account for changes in impact peak magnitude across slopes (Figure 5). Previous studies have found that the presence of an impact peak in the normal GRF waveform is subject-specific, affected by changes in running slope, and associated with acceleration of the effective mass of the lower extremity during early stance phase (McMahon, Valiant & Frederick, 1987; Gottschall & Kram, 2005; Vernillo et al., 2020). Thus, predictions of normal GRF waveforms across slopes may be further improved by incorporating accelerations measured at the feet or lower legs.

We also quantified the accuracy of the LSTM network when trained and tested on data from the same subject. Although not a valid method of determining the LSTM network's generalizability, single-subject validation provides insight into how well a personalized neural network could predict an individual's GRF waveforms for unknown combinations of speed and slope in the future. We found that predicted GRF waveforms of a representative subject (Subject 14) during the ±5° slope conditions had an average ± SD RMSE of 0.08 ± 0.02 BW, compared to 0.16 ± 0.03 BW during LOSO cross validation. These findings indicate that a subject-specific LSTM network was twice as accurate as the LOSO cross validated LSTM network. A single-subject approach may be particularly beneficial for researchers, coaches, or clinicians who have the resources to train personalized LSTM networks and wish to monitor a specific athlete's

biomechanics over the course of a competitive season. For example, an athlete could run at a variety of speeds and slopes while wearing accelerometers during a baseline data collection on a force-measuring treadmill at the start of their competitive season and a personalized LSTM network could be trained on their data. Then, if accelerometer data were collected from an athlete during training runs, their normal GRF waveforms and a variety of discrete values could be predicted and monitored longitudinally.

The MAPE values for step frequency, contact time, impulse, and normal GRF active peak were ≤ 8.5%, but the loading rate MAPE was 27.6 ± 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 slope (Gottschall & Kram, 2005), 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 extremities and impact accelerations are attenuated at the sacrum compared to the tibia (Baggaley et al. 2019). 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 running (Hennig & Lafortune, 1991), and future research aimed at improving the accuracy of loading rate estimates should include tibial acceleration as an input feature. 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 ± SD RMSE from 0.16 ± 0.04 BW to 0.17 ± 0.05 BW and rRMSE from 6.4 ± 1.5% to 6.7 ± 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 (Wouda et al., 2018; Dorschky et al., 2020; Johnson et al., 2021).

We further analyzed the importance of input features to the LSTM network by calculating prediction accuracy after systematically permuting each input feature across the trials for the

representative subject. This process of calculating Permutation Feature Importance (PFI) effectively severs the learned relationship between an input feature for a given trial and the corresponding GRF waveform (Molnar 2019). PFI is calculated as the ratio between the RMSE of the LSTM network with a given feature permuted and the original prediction RMSE of the LSTM network. After 100 permutations for each input feature, we found that the input feature with the highest PFI was vertical acceleration (6.29), followed by anteroposterior acceleration (1.62), foot strike pattern (1.14), slope, (1.13), speed (1.07), height (1.00), and body mass (1.00). These findings indicate that the inclusion of body mass and height did not improve prediction accuracy and that the LSTM network relies most on the acceleration data when predicting the normal GRF waveform across a range of speeds and slopes.

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 runners (Janssen et al., 2020), have been used to provide real-time feedback of step frequency (Willy et al., 2016), and could provide running speed and slope data to the LSTM network to predict GRF waveforms in near-real time (Scott, Scott & Kelly, 2016). 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 running (Davidson et al. 2019). 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 at www.github.com/alcantarar/Recurrent_GRF_Prediction. We have included a tutorial on how to use an LSTM to continuously predict a signal from wearable device data, an approach that may be used to improve a clinician's ability to remotely quantify a patient's GRFs or monitor rehabilitation progress (Gurchiek, Cheney & McGinnis, 2019).

There are potential limitations to consider alongside our findings. The accelerometers used in the present study were biaxial and the inclusion of mediolateral sacrum accelerations may have further improved prediction as the mediolateral behavior of the center of mass is sensitive to changes in running speed (Nilsson et al. 1989) and slope. Additionally, accelerometers 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 running (Johnson et al., 2020), but the lower leg experiences larger accelerations than the sacrum during running (Baggaley et al., 2019) and thus is more sensitive to different attachment methods. However, variations in accelerometer orientation between subjects may have contributed to the range of RMSE values (0.11 – 0.31

BW) during LOSO cross validation (Tan et al. 2019). Additionally, variations in soft tissue movement artefact between the subjects used to train the LSTM network and other populations may introduce prediction error (Peters et al. 2010). 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 ±10°), for different running surfaces, as biomechanics change when running on steep slopes (e.g. 20 – 40°) (Giovanelli et al., 2016; Whiting et al., 2020), for prolonged runs, on variable terrain (Voloshina & Ferris, 2015), with changes in speed (Alcantara et al., 2021), or in response to muscle fatigue (Derrick, Dereu & McLean, 2002). 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 (Derrick, Hamill & Caldwell, 1998; Ferris, Louie & Farley, 1998).

Conclusions

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°, ±5°, ±10°) with an average ± SD RMSE of 0.16 ± 0.04 BW and rRMSE of 6.4 ± 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.

References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mane D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viegas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. 2016. TensorFlow: Large-Scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467 [cs]*.

Alcantara RS. 2019. Dryft: A Python and MATLAB package to correct drifting ground reaction force signals during treadmill running. *Journal of Open Source Software* 4:1910. DOI: 10.21105/joss.01910.

Alcantara RS, Day EM, Hahn ME, Grabowski AM. 2021. Sacral acceleration can predict whole-body kinetics and stride kinematics across running speeds. *PeerJ* 9:e11199. DOI: 10.7717/peerj.11199.

Almeida MO, Davis IS, Lopes AD. 2015. Biomechanical differences of foot-strike patterns during running: A systematic review with meta-analysis. *Journal of Orthopaedic & Sports Physical Therapy* 45:738–755. DOI: 10.2519/jospt.2015.6019.

Ancillao A, Tedesco S, Barton J, O'Flynn B. 2018. Indirect measurement of ground reaction forces and moments by means of wearable inertial sensors: A systematic review. *Sensors (Basel, Switzerland)* 18. DOI: 10.3390/s18082564.

Baggaley M, Vernillo G, Martinez A, Horvais N, Giandolini M, Millet GY, Edwards WB. 2019. Step length and grade effects on energy absorption and impact attenuation in running. *European Journal of Sport Science*:1–11. DOI: 10.1080/17461391.2019.1664639.

Blickhan R. 1989. The spring-mass model for running and hopping. *Journal of Biomechanics* 22:11-12 1217-1227. DOI: 10.1016/0021-9290(89)90224-8.

Cavagna GA, Mantovani M, Willems PA, Musch G. 1997. The resonant step frequency in human running. *Pflügers Archiv* 434:678–684. DOI: 10.1007/s004240050451.

Ceyssens L, Vanelderen R, Barton C, Malliaras P, Dingenen B. 2019. Biomechanical risk factors associated with running-related injuries: A systematic review. *Sports Medicine* 49:1095–1115. DOI: 10.1007/s40279-019-01110-z.

Chaibub Neto E, Pratap A, Perumal TM, Tummalacherla M, Snyder P, Bot BM, Trister AD, Friend SH, Mangravite L, Omberg L. 2019. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *npj Digital Medicine* 2:1–6. DOI: 10.1038/s41746-019-0178-x.

Choi A, Jung H, Mun JH. 2019. Single inertial sensor-based neural networks to estimate COM-COP inclination angle during walking. *Sensors (Basel, Switzerland)* 19. DOI: 10.3390/s19132974.

Clermont CA, Benson LC, Edwards WB, Hettinga BA, Ferber R. 2019. New considerations for wearable technology data: Changes in running biomechanics during a marathon. *Journal of Applied Biomechanics* 35:401–409. DOI: 10.1123/jab.2018-0453.

Davidson P, Virekunnas H, Sharma D, Piché R, Cronin N. 2019. Continuous Analysis of Running Mechanics by Means of an Integrated INS/GPS Device. *Sensors (Basel, Switzerland)* 19:1480. DOI: 10.3390/s19061480.

Day EM, Alcantara RS, McGeehan MA, Grabowski AM, Hahn ME. 2021. Low-pass filter cutoff frequency affects sacral-mounted inertial measurement unit estimations of peak vertical ground reaction force and contact time during treadmill running. *Journal of Biomechanics*:110323. DOI: 10.1016/j.jbiomech.2021.110323.

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

Derrick TR, Dereu D, McLean SP. 2002. Impacts and kinematic adjustments during an exhaustive run. *Medicine and Science in Sports and Exercise.* 34(6):998-1002. DOI: 10.1097/00005768-200206000-00015.

Derrick TR, Hamill J, Caldwell GE. 1998. Energy absorption of impacts during running at various stride lengths. *Medicine and Science in Sports and Exercise* 30:128–135. DOI: 10.1097/00005768-199801000-00018.

Dorschky E, Nitschke M, Martindale CF, van den Bogert AJ, Koelewijn AD, Eskofier BM. 2020. CNN-based estimation of sagittal plane walking and running biomechanics from

measured and simulated inertial sensor data. *Frontiers in Bioengineering and Biotechnology* 8:604. DOI: 10.3389/fbioe.2020.00604.
Ferris DP, Louie M, Farley CT. 1998. Running in the real world: Adjusting leg stiffness for different surfaces. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265:989–994. DOI: 10.1098/rspb.1998.0388.
Figo D, Diniz P, Ferreira D, Cardoso J. 2010. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing* 14:645–662. DOI: 10.1007/s00779-010-0293-9.
Giandolini M, Poupard T, Gimenez P, Horvais N, Millet GY, Morin J-B, Samozino P. 2014. A simple field method to identify foot strike pattern during running. *Journal of Biomechanics* 47:1588–1593. DOI: 10.1016/j.jbiomech.2014.03.002.
Giovanelli N, Ortiz ALR, Henninger K, Kram R. 2016. Energetics of vertical kilometer foot races; is steeper cheaper? *Journal of Applied Physiology* 120:370–375. DOI: 10.1152/japplphysiol.00546.2015.
Gottschall JS, Kram R. 2005. Ground reaction forces during downhill and uphill running. *Journal of Biomechanics* 38:445–452. DOI: 10.1016/j.jbiomech.2004.04.023.
Gurchiek RD, Cheney N, McGinnis RS. 2019. Estimating Biomechanical Time-Series with Wearable Sensors: A Systematic Review of Machine Learning Techniques. *Sensors (Basel, Switzerland)* 19:5227. DOI: 10.3390/s19235227.
Halilaj E, Rajagopal A, Fiterau M, Hicks JL, Hastie TJ, Delp SL. 2018. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *Journal of biomechanics* 81:1–11. DOI: 10.1016/j.jbiomech.2018.09.009.
Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. 2020. Array programming with NumPy. *Nature* 585:357–362. DOI: 10.1038/s41586-020-2649-2.
Hennig EM, Lafortune MA. 1991. Relationships between ground reaction force and tibial bone acceleration parameters. *International Journal of Sport Biomechanics* 7:303–309.
Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Computation* 9:1735–1789.
International Trail Running Association. 2020. 2020 Trail running infographics.
Janssen M, Walravens R, Thibaut E, Scheerder J, Brombacher A, Vos S. 2020. Understanding different types of recreational runners and how they use running-related technology. *International Journal of Environmental Research and Public Health* 17. DOI: 10.3390/ijerph17072276.
Johnson WR, Mian A, Robinson MA, Verheul J, Lloyd DG, Alderson J. 2021. Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning. *IEEE Transactions on Biomedical Engineering* 68:289–297. DOI: 10.1109/TBME.2020.3006158.
Johnson CD, Outerleys J, Tenforde AS, Davis IS. 2020. A comparison of attachment methods of skin mounted inertial measurement units on tibial accelerations. *Journal of Biomechanics* 113:110118. DOI: 10.1016/j.jbiomech.2020.110118.
Khassetarash A, Vernillo G, Martinez A, Baggaley M, Giandolini M, Horvais N, Millet GY, Edwards WB. 2020. Biomechanics of graded running: Part II—Joint kinematics and

kinetics. *Scandinavian Journal of Medicine & Science in Sports* 30:1642–1654. DOI: 10.1111/sms.13735.
Kiernan D, Hawkins DA, Manoukian MAC, McKallip M, Oelsner L, Caskey CF, Coolbaugh CL. 2018. Accelerometer-based prediction of running injury in National Collegiate Athletic Association track athletes. *Journal of Biomechanics* 73:201–209. DOI: 10.1016/j.jbiomech.2018.04.001.
Kingma DP, Ba J. 2017. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*.
Kipp S, Taboga P, Kram R. 2017. Ground reaction forces during steeplechase hurdling and waterjumps. *Sports Biomechanics* 16:152–165. DOI: 10.1080/14763141.2016.1212917.
Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. 2018. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research* 18:1–52.
McMahon TA, Valiant G, Frederick EC. 1987. Groucho running. *Journal of Applied Physiology* 62:2326–2337. DOI: 10.1152/jappl.1987.62.6.2326.
Molnar C. 2019. Permutation Feature Importance. In: *Interpretable machine learning: A guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/. 210-220.*
Mundt M, Koeppe A, David S, Bamer F, Potthast W, Markert B. 2020. Prediction of ground reactio
n force and joint moments based on optical motion capture data during gait. *Medical Engineering & Physics* 86:29–34. DOI: 10.1016/j.medengphy.2020.10.001.
Neugebauer JM, Hawkins DA, Beckett L. 2012. Estimating youth locomotion ground reaction forces using an accelerometer-based activity monitor. *PLOS ONE* 7:e48182. DOI: 10.1371/journal.pone.0048182.
Nilsson J, Thortensson A. 1989. Ground reaction forces at different speeds of human walking and running. *Acta Physiological Scandinavica* 136:217-227. DOI: 10.1111/j.1748-1716.1989.tb08655.x.
pandas development team. 2020. *pandas-dev/pandas: Pandas*.
Peters A, Galna B, Sangeux M, Morris M, Baker R. 2010. Quantification of soft tissue artifact in lower limb human motion analysis: A systematic review. *Gait & Posture* 31:1-8. DOI: 10.1016/j.gaitpost.2009.09.004.
Pogson M, Verheul J, Robinson MA, Vanrenterghem J, Lisboa P. 2020. A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities. *Medical Engineering & Physics* 78:82–89. DOI: 10.1016/j.medengphy.2020.02.002.
R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundataion for Statistical Computing.
Reenalda J, Maartens E, Homan L, Buurke JH (Jaap). 2016. Continuous three dimensional analysis of running mechanics during a marathon by means of inertial magnetic measurement units to objectify changes in running mechanics. *Journal of Biomechanics* 49:3362–3367. DOI: 10.1016/j.jbiomech.2016.08.032.
Ren L, Jones RK, Howard D. 2008. Whole body inverse dynamics over a complete gait cycle based only on measured kinematics. *Journal of Biomechanics* 41:2750–2759. DOI: 10.1016/j.jbiomech.2008.06.001.

Ruder M, Jamison ST, Tenforde A, Mulloy F, Davis IS. 2019. Relationship of foot strike pattern and landing impacts during a marathon. *Medicine and Science in Sports and Exercise* 51:2073–2079. DOI: 10.1249/MSS.0000000000002032.

Running USA. 2019. 2019 U.S Running Trends Report.

Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. 2017. The need to approximate the use-case in clinical machine learning. *GigaScience* 6. DOI: 10.1093/gigascience/gix019.

Scott MTU, Scott TJ, Kelly VG. 2016. The validity and reliability of global positioning systems in team sport: A brief review. *The Journal of Strength & Conditioning Research* 30:1470–1490. DOI: 10.1519/JSC.0000000000001221.

Snyder KL, Farley CT. 2011. Energetically optimal stride frequency in running: the effects of incline and decline. *The Journal of Experimental Biology* 214:2089–2095. DOI: 10.1242/jeb.053157.

Tan T, Chiasson DP, Hu H, Shull PB. 2019. Influence of IMU position and orientation placement errors on ground reaction force estimation. *Journal of Biomechanics* 97:109416. DOI: 10.1016/j.biomech.2019.109416.

Vernillo G, Martinez A, Baggaley M, Khassetarash A, Giandolini M, Horvais N, Edwards WB, Millet GY. 2020. Biomechanics of graded running: Part I - Stride parameters, external forces, muscle activations. *Scandinavian Journal of Medicine & Science in Sports* 30:1632–1641. DOI: 10.1111/sms.13708.

Vincent HK, Kilgore JE, Chen C, Bruner M, Horodyski M, Vincent KR. 2020. Impact of body mass index on biomechanics of recreational runners. *PM&R* 12:1106–1112. DOI: https://doi.org/10.1002/pmrj.12335.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17:261–272. DOI: 10.1038/s41592-019-0686-2.

Voloshina AS, Ferris DP. 2015. Biomechanics and energetics of running on uneven terrain. *Journal of Experimental Biology* 218:711–719. DOI: 10.1242/jeb.106518.

Wang S, Jiang J. 2016. Learning natural language inference with LSTM. *arXiv:1512.08849 [cs]*.

Whiting CS, Allen SP, Brill JW, Kram R. 2020. Steep (30°) uphill walking vs. running: COM movements, stride kinematics, and leg muscle excitations. *European Journal of Applied Physiology* 120:2147–2157. DOI: 10.1007/s00421-020-04437-y.

Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. DOI: 10.1007/978-0-387-98141-3.

Wickham H. 2019. *stringr: Simple, Consistent Wrappers for Common String Operations*.

Wickham H, François R, Henry L, Müller K. 2020. *dplyr: A Grammar of Data Manipulation*.

Willy RW, Buchenic L, Rogacki K, Ackerman J, Schmidt A, Willson JD. 2016. In-field gait retraining and mobile monitoring to address running biomechanics associated with tibial stress fracture. *Scandinavian Journal of Medicine & Science in Sports* 26:197–205. DOI: https://doi.org/10.1111/sms.12413.

Wouda FJ, Giuberti M, Bellusci G, Maartens E, Reenalda J, van Beijnum B-JF, Veltink PH. 2018. Estimation of vertical ground reaction forces and sagittal knee kinematics during

running using three inertial sensors. *Frontiers in Physiology* 9. DOI: 10.3389/fphys.2018.00218.

Yong JR, Silder A, Montgomery KL, Fredericson M, Delp SL. 2018. Acute changes in foot strike pattern and cadence affect running parameters associated with tibial stress fractures. *Journal of Biomechanics* 76:1–7. DOI: 10.1016/j.jbiomech.2018.05.017.

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)** Vertical and anteroposterior sacral acceleration data were divided into overlapping 6 frame (12 ms) windows, one for each frame of the normal 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}. **(C)** Normal GRFs were predicted frame-by-frame by the LSTM network using the 13 input features.

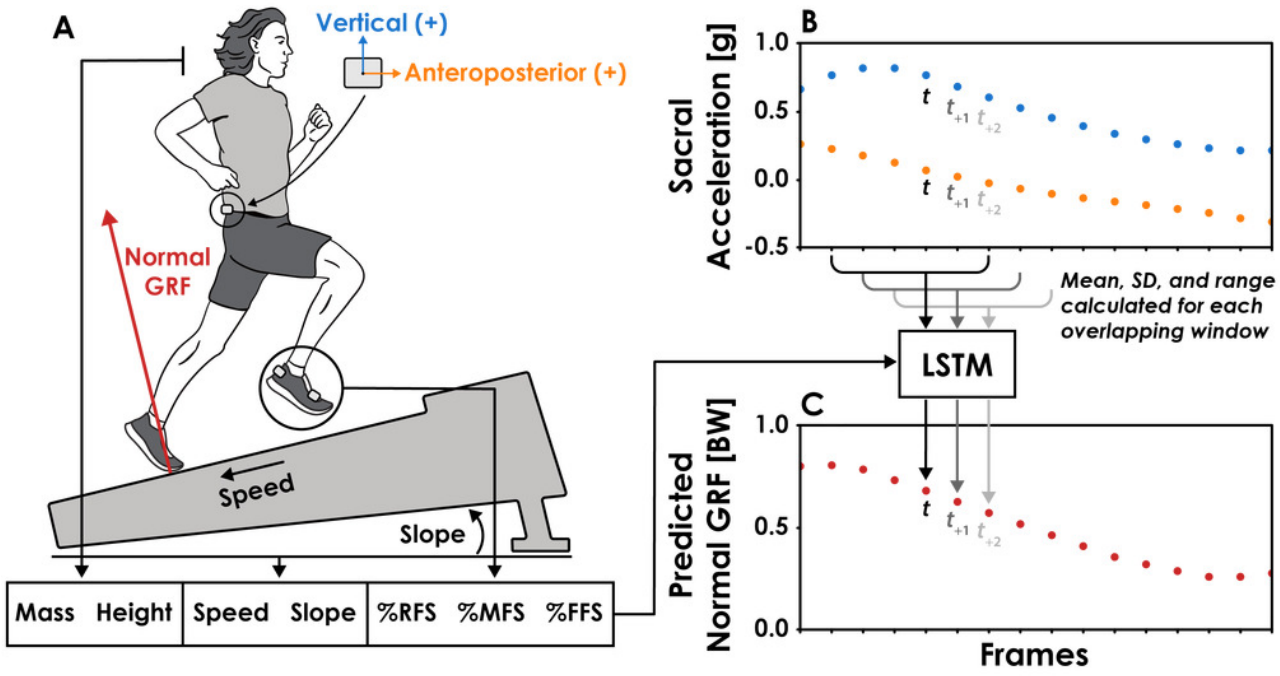

Figure 2

Neural Network Architecture

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.

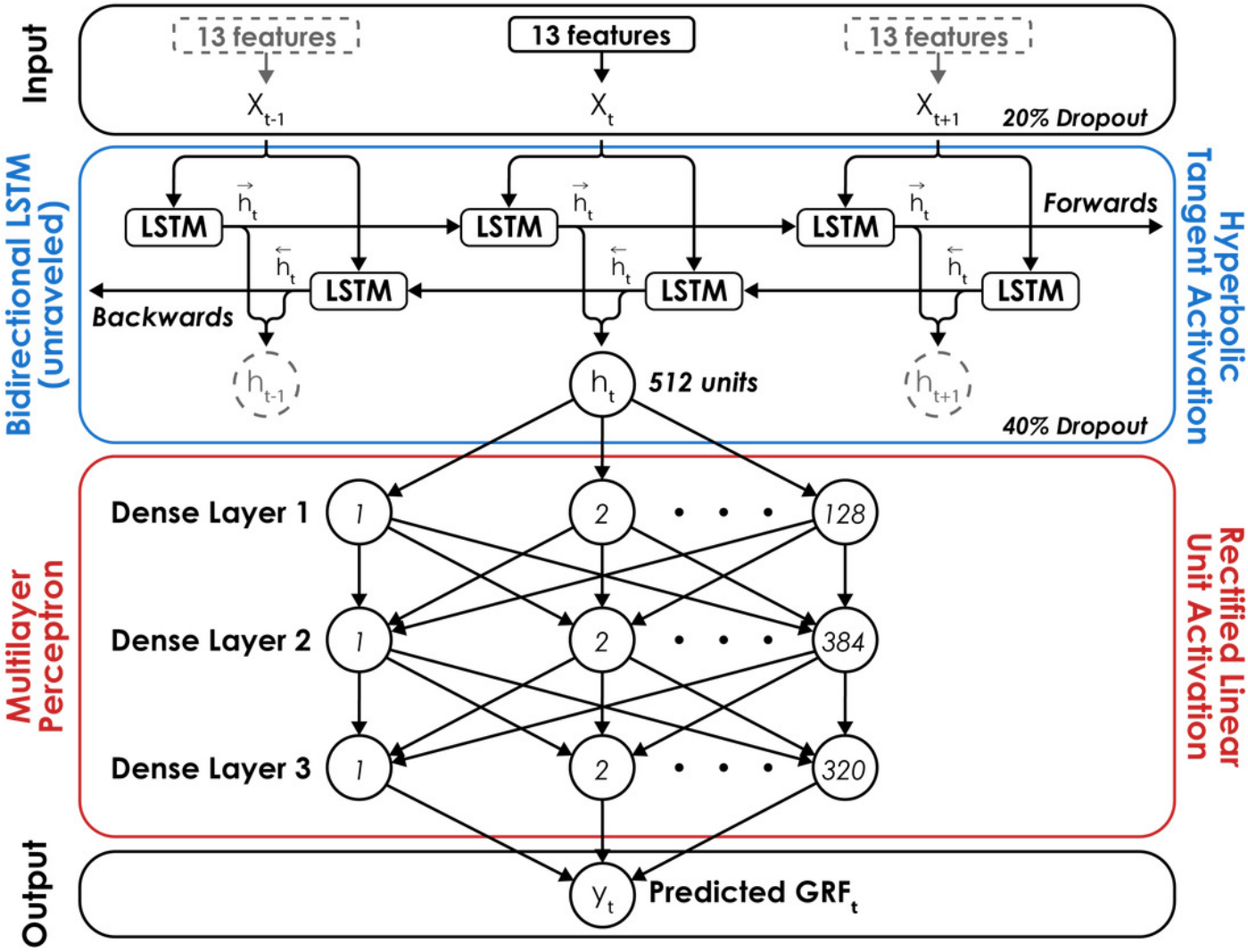
Input
13 features
13 features
13 features
X_{t-1}
X_t
X_{t+1}
20% Dropout
Bidirectional LSTM (unraveled)
LSTM
Forwards
Backwards
h_{t-1}
h_t
512 units
h_{t+1}
40% Dropout
Hyperbolic Tangent Activation
Multilayer Perceptron
Dense Layer 1
Dense Layer 2
Dense Layer 3
1
2
128
384
320
Rectified Linear Unit Activation
Output
y_t
Predicted GRF_t

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°, ±5°, ±10°) 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*IQR. Subjects are sorted from lowest to highest RMSE.

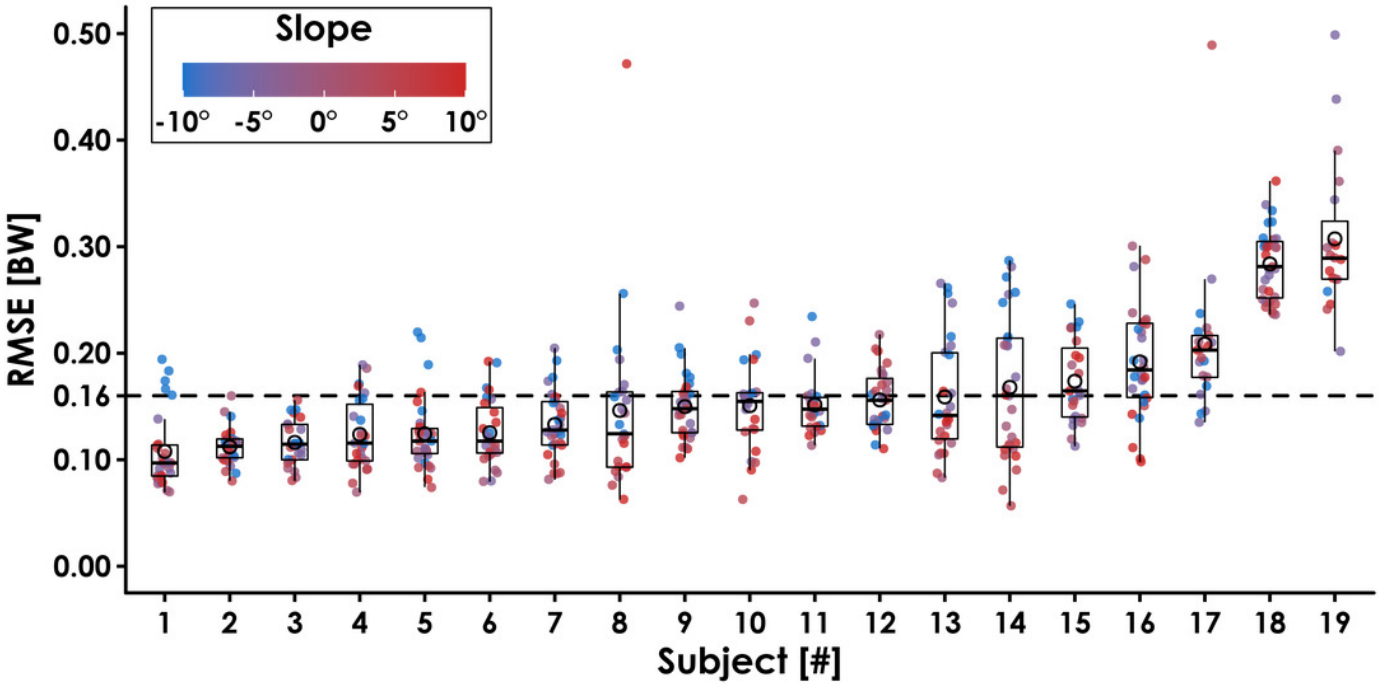

Figure 4

Ground reaction force waveform prediction error for each condition

The average ± 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.

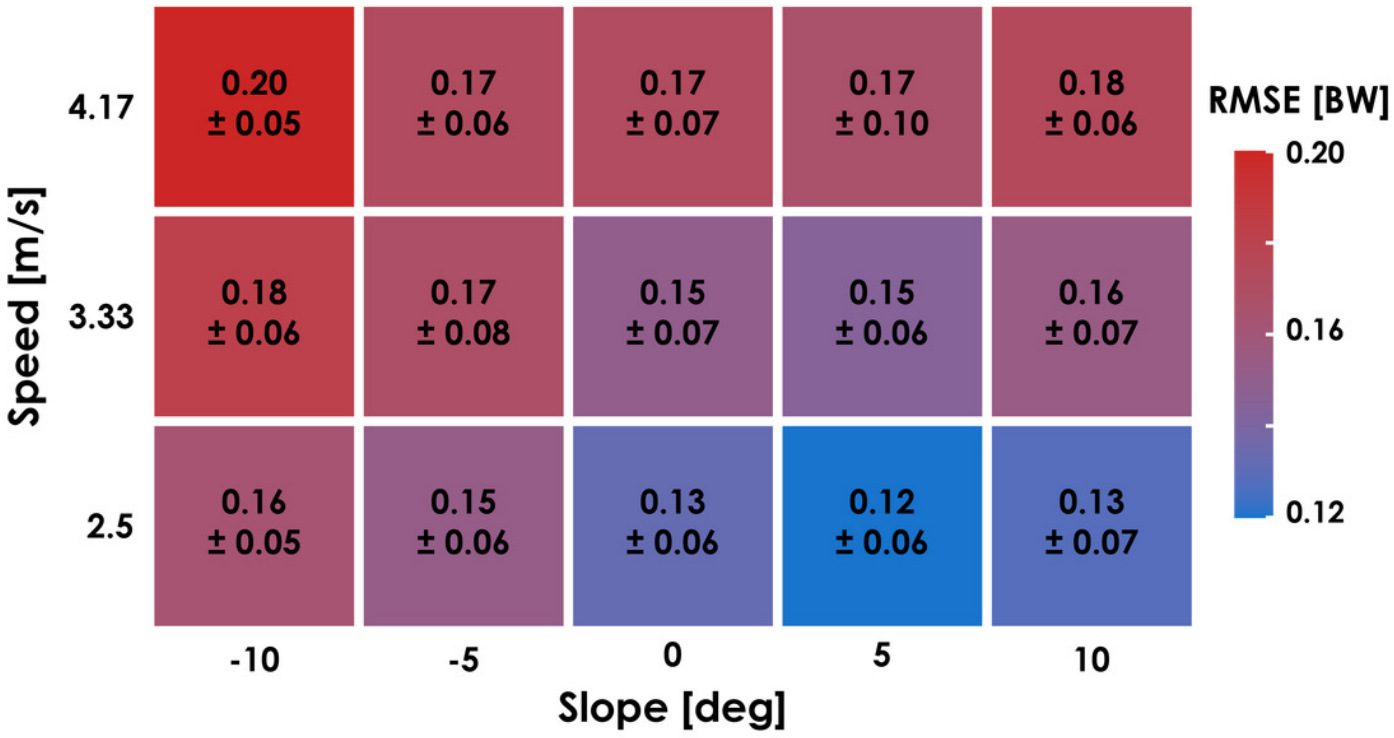

Figure 5

Predicted and measured normal GRF waveforms across slopes for a representative subject

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°, ±5°, ±10°) 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.

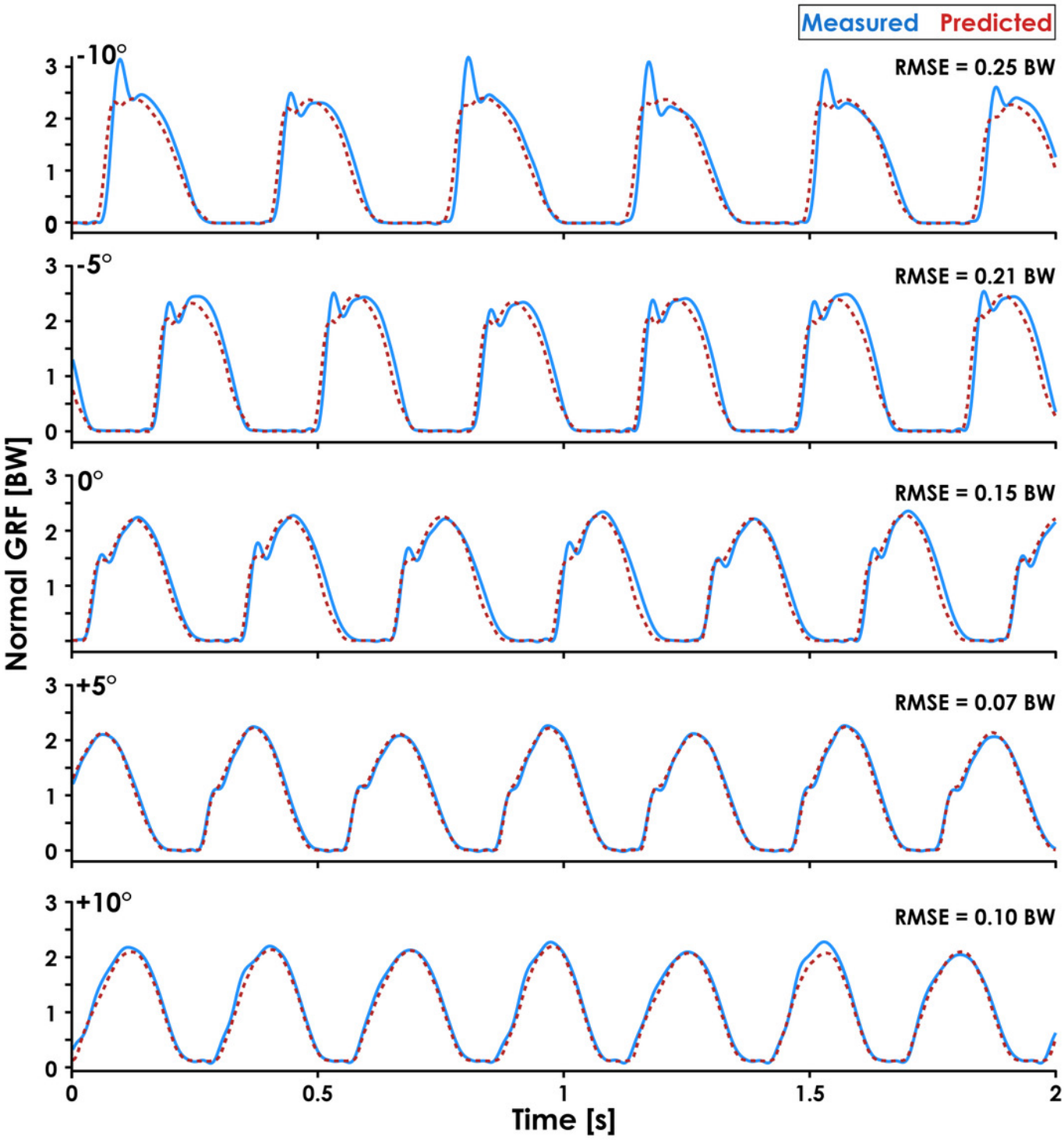
Measured
Predicted
-10°
RMSE = 0.25 BW
-5°
RMSE = 0.21 BW
0°
RMSE = 0.15 BW
+5°
RMSE = 0.07 BW
+10°
RMSE = 0.10 BW
Normal GRF [BW]
Time [s]
0
0.5
1
1.5
2

Table 1(on next page)

Mean ± 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 ± SD	0.16 ± 0.04	6.4 ± 1.5%

1

Table 2(on next page)

Mean ± SD of the discrete biomechanical variables.

Values were 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.

	Step Frequency [Hz]		Contact Time [ms]		Impulse [BW*s]		Active Peak [BW]		Loading Rate [BW/s]	
Slope	Predicted	Measured	Predicted	Measured	Predicted	Measured	Predicted	Measured	Predicted	Measured
-10°	3.1 ± 0.3	3.1 ± 0.3	215 ± 26	214 ± 29	0.33 ± 0.03	0.33 ± 0.05	2.38 ± 0.24	2.36 ± 0.40	64.7 ± 19.9	68.4 ± 20.6
-5°	3.1 ± 0.3	3.1 ± 0.3	222 ± 23	223 ± 27	0.34 ± 0.04	0.34 ± 0.04	2.41 ± 0.25	2.45 ± 0.36	54.9 ± 16.7	60.6 ± 19.4
0°	3.1 ± 0.3	3.1 ± 0.3	228 ± 24	229 ± 27	0.33 ± 0.04	0.34 ± 0.04	2.42 ± 0.22	2.51 ± 0.36	42.0 ± 12.8	49.0 ± 17.2
+5°	3.2 ± 0.3	3.2 ± 0.3	228 ± 25	232 ± 26	0.32 ± 0.03	0.33 ± 0.04	2.36 ± 0.21	2.40 ± 0.33	34.8 ± 8.5	39.3 ± 13.3
+10°	3.4 ± 0.3	3.4 ± 0.3	223 ± 26	227 ± 26	0.30 ± 0.04	0.30 ± 0.04	2.24 ± 0.22	2.20 ± 0.29	30.1 ± 8.3	31.3 ± 12.7

1

Table 3(on next page)

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.

Root mean square error (RMSE) and relative RMSE (rRMSE) are presented for 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.

	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 ± 1.5%	6.7 ± 1.7%
Discrete Variables MAPE		
Step Frequency	0.1 ± 0.1%	0.1 ± 0.1%
Contact Time	4.9 ± 4.0%	5.6 ± 4.5%
Impulse	6.4 ± 6.9%	6.0 ± 7.1%
Active Peak	8.5 ± 8.2%	7.7 ± 6.3%
Loading Rate	27.6 ± 36.1%	30.3 ± 41.6%

1