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

Abstract

Background

Ground reaction forces (GRFs) are important for understanding the biomechanics of human movement but the measurement of GRFs is generally limited to a laboratory environment. Wearable devices like accelerometers have been used to measure biomechanical variables outside the laboratory environment, but they cannot directly measure GRFs. Previous studies have used neural networks to predict the entire GRF waveform during the stance phase from wearable device data, but these networks require normalization of GRFs to the duration of a step or stance phase, resulting in a loss of the GRF waveform's temporal component. Additionally, previous studies have predicted GRF waveforms during level-ground, but not uphill or downhill running. A method of predicting the normal (perpendicular to running surface) GRF waveform using wearable devices across a range of running speeds and slopes, while maintaining the GRF waveform's temporal component, 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 normal GRF waveforms across a range of running speeds and slopes using data from accelerometers located on the sacrum and shoe.

Methods

19 subjects completed 30-s running trials on a force-measuring treadmill at five slopes (0°, ±5°, ±10°) and three speeds (2.5, 3.33, and 4.17 m/s) per slope. One biaxial accelerometer was adhered to the sacrum and two uniaxial accelerometers were adhered to the right shoe during all trials. Accelerometers on the shoe were used to classify foot strike patterns as rearfoot, midfoot, or forefoot, and sacral acceleration data were divided into overlapping 12-ms windows, allowing the neural network to iteratively predict the experimentally-measured normal GRF waveform frame-by-frame. The mean, SD, and range of

accelerometer data for each 12-ms window were included as neural network input features, along with the subject's body mass, height, running speed, slope, and percentage of a trial's steps classified as a rearfoot, midfoot, or forefoot strike. We assessed the accuracy and generalizability of the neural network using leave-one-subject-out cross validation, which provided an ensemble of Root Mean Square Error (RMSE) and relative RMSE (rRMSE) values comparing the normal GRF waveform predicted by the neural network to the normal GRF waveform measured by the force-measuring treadmill. Additionally, we calculated the mean absolute percent error (MAPE) of step frequency, contact time, normal impulse, normal GRF active peak, and loading rate between the predicted and measured GRF waveforms.

Results

The average ± SD RMSE was 0.16 ± 0.04 BW and rRMSE was 6.4 ± 1.5% for neural network predictions of each subject's normal GRF waveform compared to measured GRF waveforms across all conditions. RMSE values were lower during slow uphill running (2.5 m/s, +10°; 0.13 ± 0.07 BW) compared to fast downhill running (4.17 m/s, -10°; 0.20 ± 0.05 BW). The MAPE ± SD for step frequency was 0.1 ± 0.1%, contact time was 4.9 ± 4.0%, normal impulse was 6.4 ± 6.9%, normal GRF active peak was 8.5 ± 8.2%, and loading rate was 27.6 ± 36.1%.

Conclusions

We developed a recurrent neural network that uses accelerometer data to predict the continuous normal GRF waveform across a range of running speeds and slopes. The neural network does not require preliminary identification of the stance phase, maintains the temporal component of the GRF waveform, can be applied to up- and downhill running, and facilitates the prediction of kinetic and kinematic variables outside the laboratory environment. This represents a substantial step towards accurately quantifying and monitoring the external loads experienced by the body when running outdoors.

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, the measurement of GRFs is generally restricted to a laboratory environment. To determine the effects of sport-specific environments on running kinetics and kinematics, previous research studies have replicated aspects of an athlete's training or competitive environment within a laboratory environment. For example, Kipp et al. (Kipp, Taboga and Kram, 2017) measured GRFs during a steeplechase "water jump" by mounting the steeplechase barrier to a force platform and embedding another force platform below the landing surface, Voloshina and Ferris (Voloshina and Ferris, 2015) developed a custom force-measuring treadmill with an uneven running surface to mimic a trail surface, and Whiting et al. (Whiting *et al.*, 2020) measured stride kinematics during uphill running on a custom treadmill inclined to 30° to mimic slopes experienced during trail running. However, measuring GRFs within a laboratory environment may not adequately replicate the conditions (e.g. running surface, non-constant velocity) encountered during training or competition.

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; Clermont *et al.*, 2019; Ruder *et al.*, 2019). However, IMUs cannot directly measure GRFs and require the use of algorithms to estimate discrete variables like peak vertical GRF, ground contact time, vertical impulse, and vertical loading rate (Neugebauer, Hawkins and Beckett, 2012; Ancillao *et al.*, 2018; Kiernan *et al.*, 2018; Derie *et al.*, 2020; Alcantara *et al.*, 2021). Recently, neural networks have been used to predict GRF waveforms during running (Wouda *et al.*, 2018; Dorschky *et al.*, 2020; Pogson *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 data to be time-normalized to the duration of a step (Dorschky *et al.*, 2020) or stance phase (Wouda *et al.*, 2018; Johnson *et al.*, 2021), preventing the calculation of biomechanical variables with a non-normalized temporal component (e.g. ground contact time, step frequency, vertical impulse, and loading rate). Pogson et al. (2020) zero-padded the end of accelerometer data to the length of the longest stance phase to preserve the temporal component, but observed decreased accuracy when predicting the vertical GRF waveform during the stance phase (Root Mean Square Error ≈ 200 N)(Pogson *et al.*, 2020). Additionally, previous studies have only predicted GRF waveforms during level-ground running (Wouda *et al.*, 2018; Dorschky *et al.*, 2020; Pogson *et al.*, 2020; Johnson *et al.*, 2021), limiting the application to environments where no uphill or downhill running occurs (i.e. a 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 running slopes and terrain. A method of predicting GRF waveforms from wearable devices across a range of running slopes, while maintaining the temporal component, could allow researchers, clinicians, and coaches to measure a variety of kinetic and kinematic variables in a range of outdoor environments.

Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) are a type of recurrent neural network that overcome the requirement of the time-normalized inputs like that of the neural networks used in previous research (Wouda *et al.*, 2018; Johnson *et al.*, 2021). LSTM networks overcome

this limitation by recurrently predicting small, uniform portions of a larger sequence of data that can be of variable 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. LSTM networks have been used in this manner to analyze a variety of sequential data, ranging from natural language processing (Wang and Jiang, 2016) to stock market forecasting (Selvin *et al.*, 2017). 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 and Mun, 2019). To our knowledge, LSTM networks have not been used to predict GRF waveforms from wearable device data exclusively during running. Developing such a network would allow researchers to predict the GRF waveform not only during the stance phase, but continuously for the entire duration of a run. Considering that IMUs have already been used to longitudinally measure biomechanical variables related to running injury (Reenalda *et al.*, 2016; Kiernan *et al.*, 2018; Clermont *et al.*, 2019; Ruder *et al.*, 2019), applying an LSTM network to such data would effectively provide a way to indirectly collect continuous GRF waveforms in environments and quantities currently impossible to accomplish.

The purpose of this exploratory study was to develop an LSTM network that could predict the normal (perpendicular to running surface) GRF waveform across a range of running speeds and slopes using data from accelerometers. Given the novelty of this approach, we sought to develop a network that could predict the GRF waveform 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).

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 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, and 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: one on the sacrum and two on the right shoe. The accelerometers on the shoe were 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 a rearfoot, midfoot, or forefoot strike. Only the vertical and anteroposterior axes relative to the local coordinate system of the sacral accelerometer and the normal (perpendicular to the treadmill surface) GRFs were used in the

current analysis. We chose to exclude the shoe accelerometer data for predictions of GRF waveforms because the data were not available for both feet.

Data Processing

We analyzed 5 seconds of data from each trial (approximately 13 foot-ground contacts) 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 (Alcantara *et al.*, 2021; Day *et al.*, 2021). The GRFs were normalized to bodyweight (BW) and filtered using a 4th order low-pass Butterworth filter with a cut-off frequency of 30 Hz and the sacral acceleration data were filtered with a 4th order low-pass Butterworth filter with a cut-off frequency of 20 Hz. Pilot testing revealed that a 20 Hz cut-off frequency improved prediction accuracy and preserved approximately 89% and 82% of the vertical and anteroposterior signal power, respectively.

The vertical accelerometer data were further processed so that all negative values were replaced with zeros. Pilot testing revealed that negative vertical acceleration values primarily 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 @ 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 long sequences to be divided into smaller sequences that are iteratively used to make predictions. To accomplish this, we divided the acceleration data for each trial into overlapping windows with a width of 6 frames (12 ms) and padded the beginning and end of the acceleration data with the nearest value to ensure that the number of windows was equal to the number of frames in the normal GRFs (2500) and that the windows were centered on the corresponding frame in 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).

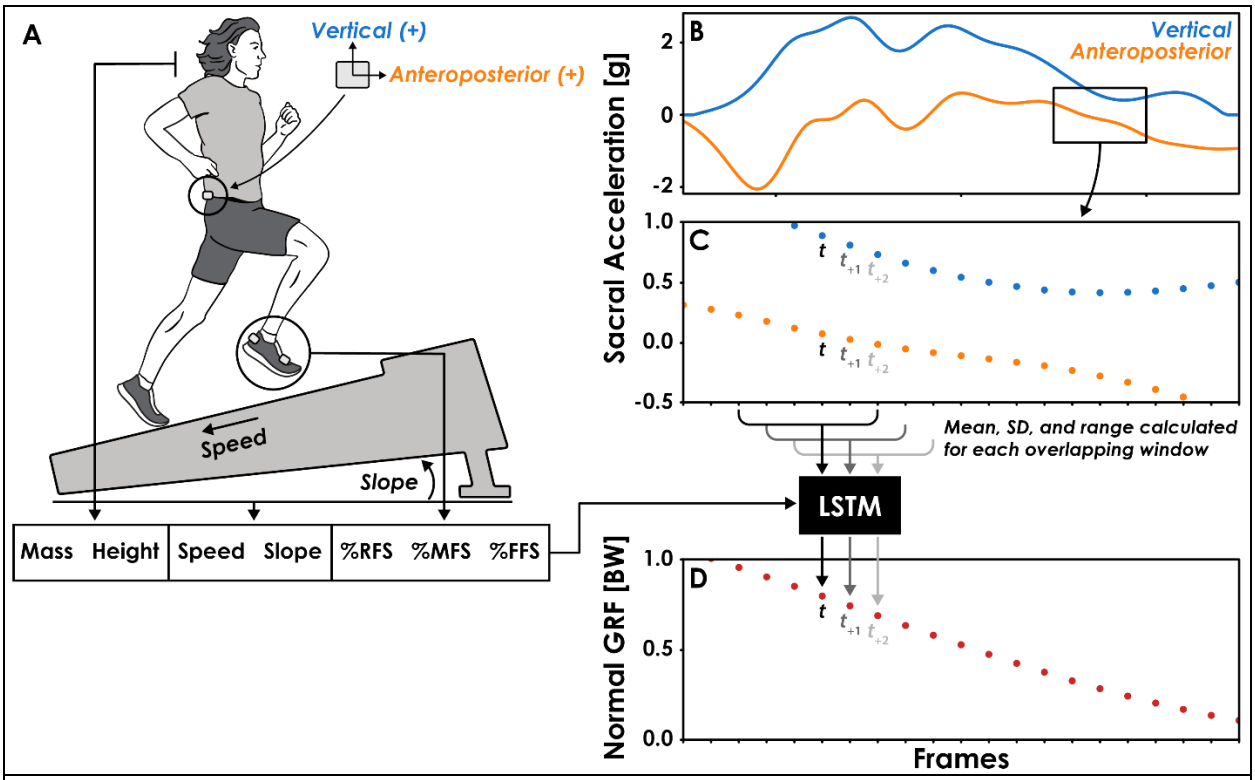

Figure 1. Overview of the Long Short-Term Memory (LSTM) network's input features. A) The LSTM network's input features included body mass, height, running speed, slope, and percentage of a trial's steps classified as a rearfoot (RFS), midfoot (MFS), or forefoot (FFS) strike. Foot strike percentages were calculated from data collected by accelerometers on the right foot. **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 were 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.

Feature Engineering

A total of 13 features were used as inputs to 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 a rearfoot, midfoot, or forefoot strike (Almeida, Davis and 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 accelerometer data used to predict the GRF waveforms. Doing

so represents a greater challenge for the LSTM network as there is known variability between trials that is not being explicitly accounted for as an input variable.

Neural Network Architecture

The neural network consisted of a Bidirectional LSTM and 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 layers are then averaged before being passed along to the MLP. The number and size of the layers in this network 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 and 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 set aside 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. These metrics illustrate how accurate the network is when applied to data from many different individuals. Performing LOSO cross validation is 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 subject's data is not included in both the training and testing subsets, which can artificially increase the reported accuracy of a network since it was tested on data from a subject included during training (Saeb *et al.*, 2017; Chaibub Neto *et al.*, 2019).

In addition to the LOSO cross validation method, we performed a test-train split according to slope (±5° trials reserved for testing, 0° and ±10° slopes used for training) for one representative subject's data to test the accuracy of a model when predicting speed-slope combinations that were not present during training. We selected subject 14 as a representative subject because their RMSE was similar to the average RMSE across all subjects and their GRF waveforms illustrated the interaction between running slope and normal GRF impact peak magnitude (Gottschall and Kram, 2005). This single-subject validation method prioritizes accuracy over generalizability (the ability to make accurate predictions for a variety of athletes) and represents the potential circumstance of training the LSTM network on data collected from a single athlete before the start of their competitive season and using it to predict the same athlete's GRF data during their competitive season.

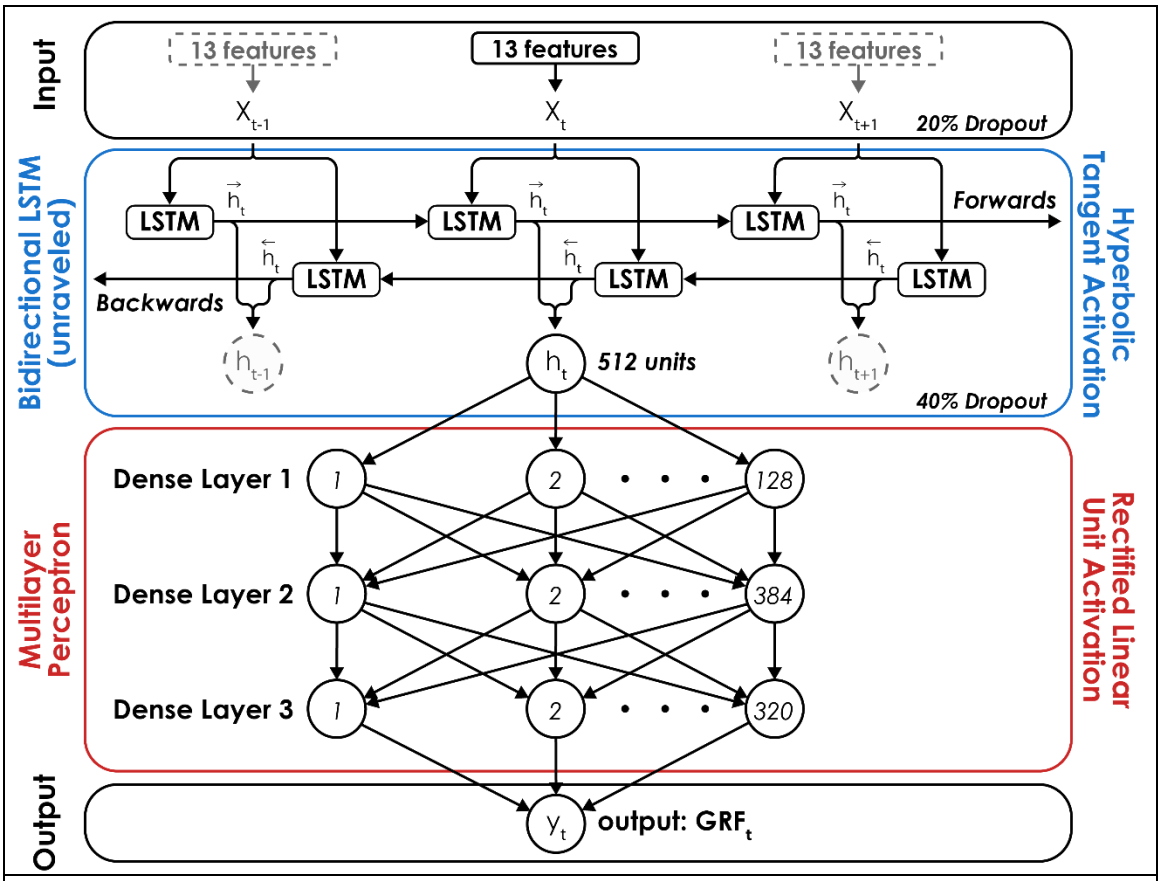

Figure 2. Neural Network Architecture. The Long Short-Term Memory (LSTM) network consisted of a Bidirectional LSTM layer with a hyperbolic tangent activation function followed by a multilayer perceptron (MLP) with rectified linear unit activation functions for three hidden layers with 128, 384, and 320 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 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.

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

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

where x_1 and x_2 are the GRF waveforms predicted by the LSTM network and measured by the force-measuring treadmill (Ren, Jones and Howard, 2008). Additionally, we used a threshold of 5% BW to identify the start and end of the 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 the 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 and 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 the 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 and R using custom libraries (Wickham, 2009, 2019; Alcantara, 2019; Harris *et al.*, 2020; pandas development team, 2020; R Core Team, 2020; Virtanen *et al.*, 2020; Wickham *et al.*, 2020).

We applied two constraints to the predicted GRF waveform to ensure that the data fell within established biomechanical limits and could be used to calculate 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 a 5% BW threshold. Second, the step frequency over the duration of the predicted GRF waveform had to be ≤ 4 Hz. These criteria represent biomechanical constraints to the network predictions, as thresholds of 5% BW have been previously used to identify the stance phase for the calculation of kinetic or kinematic variables (Alcantara *et al.*, 2021; Day *et al.*, 2021) and step frequency is ≤ 4 Hz during uphill and downhill running (Cavagna *et al.*, 1997; Snyder and Farley, 2011). Trials that failed to meet either of these criteria were used to calculate the LSTM network's prediction failure rate and removed from subsequent analyses.

Results

LSTM Network Prediction Failure Rate

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 classified as failed predictions. Specifically, 22 trials (4%) required a threshold greater than 5% BW to identify an equal number of steps between the predicted and measured GRF waveforms and 10 trials (2%) had a step frequency greater than 4 Hz. Thus, 94% of the GRF waveforms predicted by the LSTM network fell within the imposed biomechanical boundaries.

LOSO Cross Validation

The average ± SD RMSE was 0.16 ± 0.04 BW (Figure 3) and rRMSE was 6.4 ± 1.5% for the LSTM network predictions of each subject's normal GRF waveforms compared to the GRF waveforms measured by the force-measuring treadmill across all conditions (Table 1). RMSE values were 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 & 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).

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%

Table 1. Mean ± SD Root Mean Square Error (RMSE) and relative RMSE (rRMSE) for the normal GRF waveform predicted by the LSTM network compared to the measured normal GRF waveform for each subject.

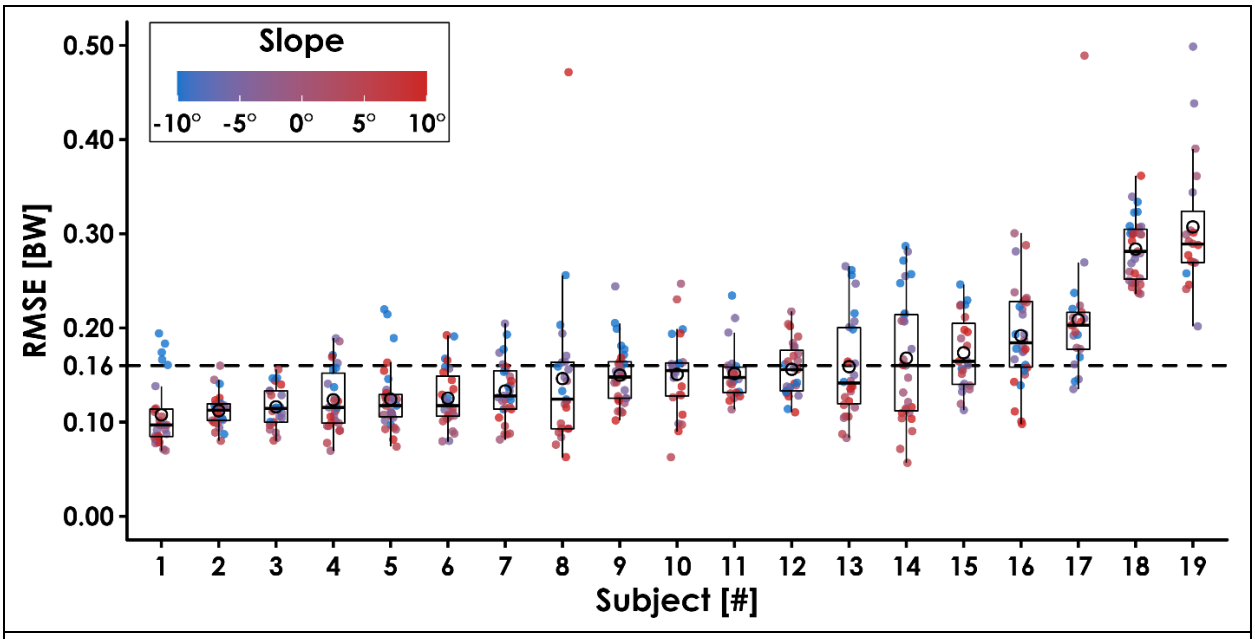

Figure 3. Ground reaction force waveform prediction error for each subject across all conditions. The average Root Mean Square Error (RMSE) across subjects was 0.16 BW (dotted line). Filled circles represent each trial, and the color indicates slope (0°, ±5°, ±10°) at three speeds (2.5, 3.33, and 4.17 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.

Single-Subject Validation

The error metrics calculated during the single-subject validation reflect the accuracy of the LSTM network when tested on the data from a representative subject (Subject 14) during the ±5° slope trials and trained on their remaining trials. The prediction error for the representative subject's normal GRF waveforms at ±5° 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 the GRF waveform predicted by the LSTM network were also generally lower than those resulting from the LOSO cross validation.

	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

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

Discussion

We developed a recurrent neural network capable of predicting the continuous normal GRF waveform 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 given the runner's mass, height, running speed, slope, foot strike pattern, and sacral acceleration can predict the normal GRF waveform 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, recent 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. Previous studies also quantified the accuracy of their networks using the Pearson correlation between the predicted and measured GRF waveforms, but we chose not to include correlation as a primary metric of accuracy. We refrained from calculating the Pearson correlation because it appears to lack the sensitivity of RMSE or rRMSE when comparing GRF waveforms (Ancillao *et al.*, 2018). Previous studies have achieved very strong correlations (≥ 0.90) when predicting the vertical GRF waveform but RMSE and rRMSE values were up to twice as large as the average values achieved in the present study.

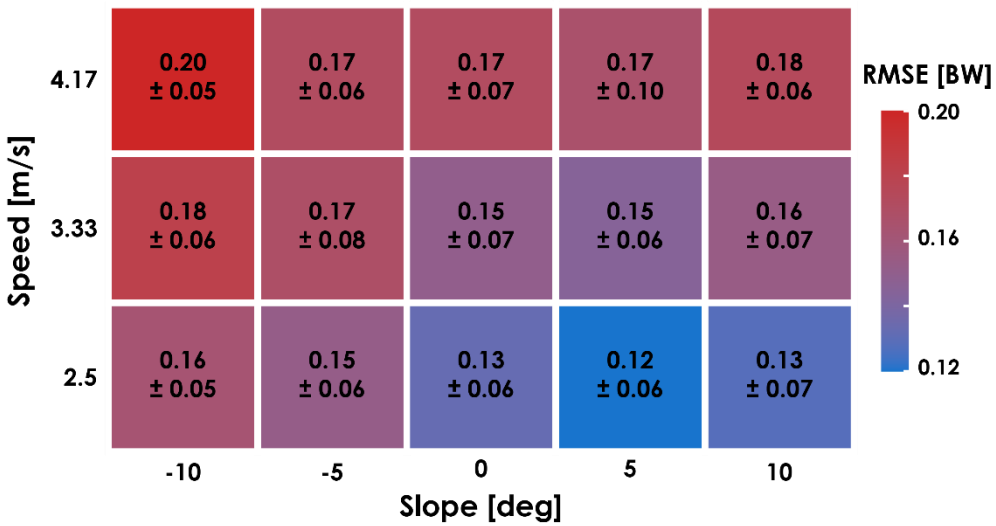

Figure 4. Ground reaction force waveform prediction error for each condition. The average ± SD Root Mean Square Error (RMSE) of the predicted ground reaction force (GRF) waveforms compared to the GRF waveform measured by the force-measuring treadmill for each condition during leave-one-subject-out cross validation.

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 the GRF waveform and can be used to make predictions over any duration of running if the length of the accelerometer signal is divisible by the window width (12 ms). Thus, the LSTM network could be used to quantify changes in the normal GRF waveform over the course of prolonged runs (e.g. a marathon race).

The accuracy of the 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). We suspect that 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 the acceleration of the effective mass of the lower extremity during early stance phase (McMahon, Valiant and Frederick, 1987; Gottschall and Kram, 2005; Vernillo *et al.*, 2020). Thus, predictions of the normal GRF waveform across slopes may be further improved by incorporating accelerations measured at the feet or lower legs.

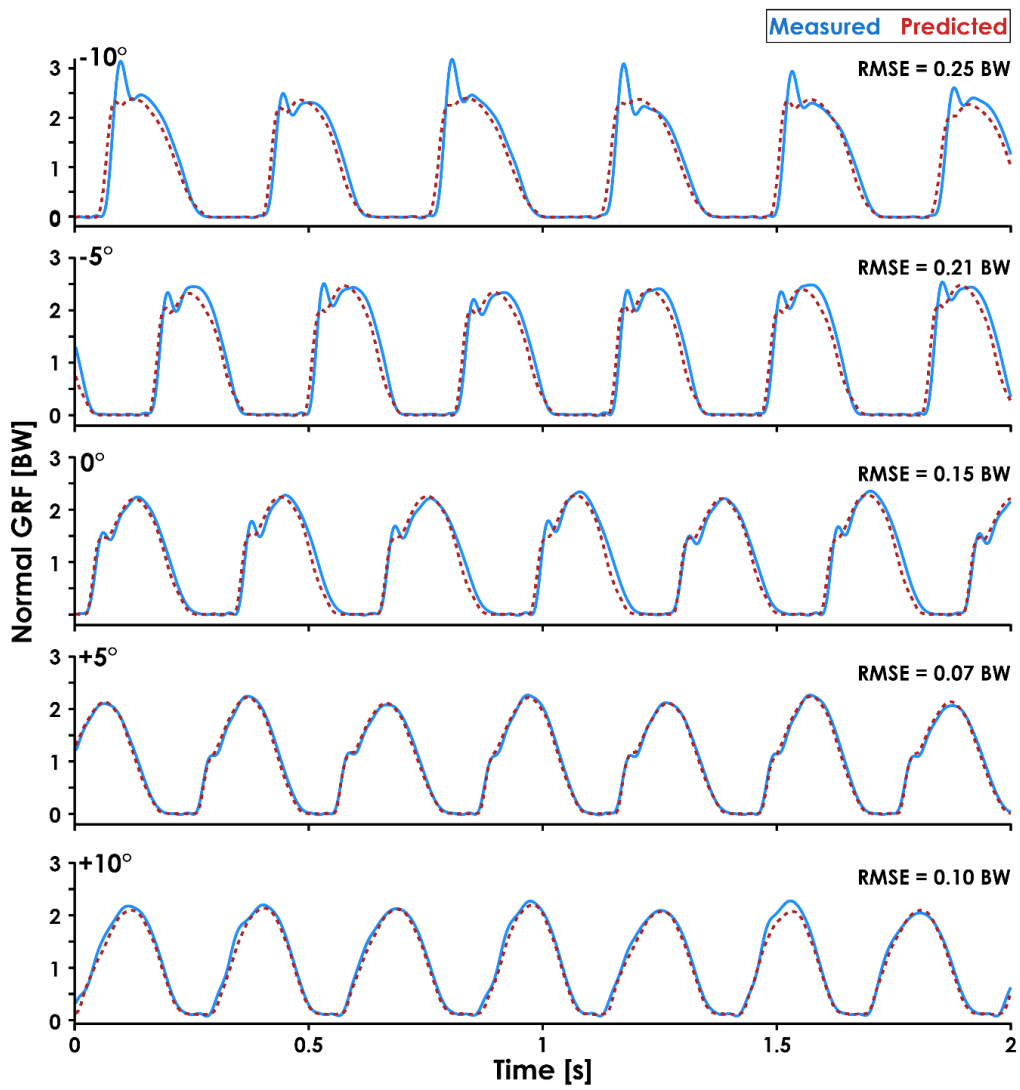

Figure 5. Predicted and measured normal GRF waveforms across slopes for a representative subject. The normal ground reaction force (GRF) waveforms predicted by the recurrent neural network (dashed red lines) and measured by the force-measuring treadmill (solid blue lines) at 3.33 m/s and all slopes (0°, ±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 data illustrate the effect of slope on the presence of an impact peak in the normal GRF waveform.

Single-Subject Validation

We 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 generalizability of a machine learning algorithm, single-subject validation provides insight into how well a personalized neural network could predict an individual's GRF waveform for unknown combinations of speed and slope in the future. We found that predicted GRF waveforms during the ±5° slope conditions had an average ± SD RMSE of 0.08 ± 0.02 BW, indicating that a subject-specific LSTM network was approximately 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 baseline testing 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. Although this approach benefits from improved prediction accuracy, it requires computational resources necessary to train personalized LSTM networks for each athlete.

Discrete Variable Accuracy

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 affects the slope of the GRF waveform during early stance phase and thus the accuracy of the loading rate values. Although the prominence of an impact peak in the normal GRF waveform is affected by foot strike pattern and slope (Gottschall and Kram, 2005), which are two variables that we included as input features 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. The impact peak of the GRF waveform is associated with the acceleration of the effective mass of the lower extremity during early stance phase (McMahon, Valiant and Frederick, 1987; Gottschall and Kram, 2005) and 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 and Lafortune, 1991). We did not include the accelerometer data from the shoes as inputs for the LSTM network because the data were not available for both feet but including data from an accelerometer located on the lower legs or feet as inputs to a neural network may further improve the prediction accuracy of loading rate.

External cumulative loading is generally calculated as the product of a given external load's magnitude (e.g. peak GRF) and the number of steps taken over a given duration of running (Firminger and Edwards, 2016; Kiernan *et al.*, 2018; Backes *et al.*, 2020; Ryan *et al.*, 2020) and has been associated with running-related overuse injuries (Colby *et al.*, 2014; Bertelsen *et al.*, 2017; Kiernan *et al.*, 2018). Previous studies have predicted the peak vertical GRF from pelvis accelerometer data with a MAPE of 4.0 – 8.3% during level-ground running (Neugebauer, Hawkins and Beckett, 2012; Alcantara *et al.*, 2021), but a personalized LSTM network could be used to measure an athlete's external cumulative load with improved accuracy

(MAPE = 2.7 ± 2.0%) across a range of running speeds and slopes. Thus, an LSTM network could be used to monitor an athlete's risk of sustaining a running-related overuse injury over the course of a competitive season.

Considerations for LSTM Network Implementation

The LSTM network represents a promising strategy for predicting continuous normal GRFs from wearable devices in outdoor environments. The LSTM network required three accelerometers (one on the sacrum and two on the right shoe to determine foot strike pattern), but we performed a *post-hoc* analysis of the 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 require data from 3 – 7 wearable devices (Wouda *et al.*, 2018; Dorschky *et al.*, 2020; Johnson *et al.*, 2021).

	LSTM With Foot Strike (Sacrum + Right Foot Accelerometers)	**LSTM Without Foot Strike (Only Sacrum 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%

Table 3. Error metrics for the predicted waveform and discrete variables when training the Long Short-Term Memory (LSTM) network with and without foot strike pattern as input features. Root Mean Square Error (RMSE) and relative RMSE (rRMSE) are presented for the normal ground reaction force (GRF) waveforms predicted by an LSTM network trained with and without foot strike pattern data from the two accelerometers on the right foot. Mean absolute percent error (MAPE) values are presented for the discrete variables calculated from the normal GRF waveforms predicted by both LSTM networks.

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. GPS devices 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 the LSTM network with the running speed and slope data necessary to predict the GRF waveform (Scott, Scott and Kelly, 2016). This could facilitate near real-time predictions of the normal GRF waveform if the accelerometer data can be synchronized with the GPS data and split into portions with near constant

speeds and slopes. Discrete biomechanical variables could then be calculated from the predicted normal GRF waveform 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 generalizable or accurate this approach is as the network was trained and tested on data from the same three subjects and the reported accuracy metrics are combined for walking and running (Vitali *et al.*, 2019). To facilitate the 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.

Limitations

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 signal. Securing an IMU to the lower leg using a commercial strap resulted in peak accelerations that were 1.2 g greater than when securing the IMU with athletic tape (Johnson *et al.*, 2020), but the lower leg experiences larger accelerations than the sacrum during running (Baggaley *et al.*, 2019) and may be more sensitive to different attachment methods. Using the LSTM network to predict normal GRF waveforms from accelerometers adhered differently than in the present study may decrease 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°) as biomechanics change when running on steep slopes (e.g. 20 – 40°) (Giovanelli *et al.*, 2016; Whiting *et al.*, 2020). 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 and Caldwell, 1998; Ferris, Louie and Farley, 1998).

Conclusions

We developed a Long Short-Term Memory (LSTM) neural network that used accelerometer data to predict the continuous normal GRF waveform 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 LSTM network does not require preliminary identification of the stance phase or temporal normalization and allows for near real-time predictions of the normal GRF waveform during running. Accurate predictions of the normal GRF waveform using wearable devices will improve our ability to longitudinally monitor biomechanical variables in non-laboratory environments.

Acknowledgements

This work 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.

References

Abadi, M. *et al.* (2016) 'TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems', *arXiv:1603.04467 [cs]*. Available at: http://arxiv.org/abs/1603.04467 (Accessed: 9 November 2020).

Alcantara, R. (2019) 'Dryft: A Python and MATLAB package to correct drifting ground reaction force signals during treadmill running', *Journal of Open Source Software*, 4(44), p. 1910. doi: 10.21105/joss.01910.

Alcantara, R. *et al.* (2021) 'Sacral acceleration can predict whole-body kinetics and stride kinematics across running speeds', *PeerJ*, In Press.

Almeida, M. O., Davis, I. S. and Lopes, A. D. (2015) 'Biomechanical Differences of Foot-Strike Patterns During Running: A Systematic Review With Meta-analysis', *Journal of Orthopaedic & Sports Physical Therapy*, 45(10), pp. 738–755. doi: 10.2519/jospt.2015.6019.

Ancillao, A. *et al.* (2018) 'Indirect Measurement of Ground Reaction Forces and Moments by Means of Wearable Inertial Sensors: A Systematic Review', *Sensors (Basel, Switzerland)*, 18(8). doi: 10.3390/s18082564.

Backes, A. *et al.* (2020) 'Predicting cumulative load during running using field-based measures', *Scandinavian Journal of Medicine & Science in Sports*, 00, pp. 1–9. doi: 10.1111/sms.13796.

Baggaley, M. *et al.* (2019) 'Step length and grade effects on energy absorption and impact attenuation in running', *European Journal of Sport Science*, pp. 1–11. doi: 10.1080/17461391.2019.1664639.

Bertelsen, M. L. *et al.* (2017) 'A framework for the etiology of running-related injuries', *Scandinavian Journal of Medicine & Science in Sports*, 27(11), pp. 1170–1180. doi: https://doi.org/10.1111/sms.12883.

Cavagna, G. A. *et al.* (1997) 'The resonant step frequency in human running', *Pflügers Archiv*, 434(6), pp. 678–684. doi: 10.1007/s004240050451.

Ceyssens, L. *et al.* (2019) 'Biomechanical Risk Factors Associated with Running-Related Injuries: A Systematic Review', *Sports Medicine*, 49(7), pp. 1095–1115. doi: 10.1007/s40279-019-01110-z.

Chaibub Neto, E. *et al.* (2019) 'Detecting the impact of subject characteristics on machine learning-based diagnostic applications', *npj Digital Medicine*, 2(1), pp. 1–6. doi: 10.1038/s41746-019-0178-x.

Choi, A., Jung, H. and Mun, J. H. (2019) 'Single Inertial Sensor-Based Neural Networks to Estimate COM-COP Inclination Angle During Walking', *Sensors (Basel, Switzerland)*, 19(13). doi: 10.3390/s19132974.

Clermont, C. A. *et al.* (2019) 'New Considerations for Wearable Technology Data: Changes in Running Biomechanics During a Marathon', *Journal of Applied Biomechanics*, 35(6), pp. 401–409. doi: 10.1123/jab.2018-0453.

Colby, M. J. *et al.* (2014) 'Accelerometer and GPS-Derived Running Loads and Injury Risk in Elite Australian Footballers', *The Journal of Strength & Conditioning Research*, 28(8), pp. 2244–2252. doi: 10.1519/JSC.0000000000000362.

Day, E. M. *et al.* (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*, p. 110323. doi: 10.1016/j.jbiomech.2021.110323.

Derie, R. *et al.* (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, T. R., Hamill, J. and Caldwell, G. E. (1998) 'Energy absorption of impacts during running at various stride lengths', *Medicine and Science in Sports and Exercise*, 30(1), pp. 128–135. doi: 10.1097/00005768-199801000-00018.

Dorschky, E. *et al.* (2020) 'CNN-Based Estimation of Sagittal Plane Walking and Running Biomechanics From Measured and Simulated Inertial Sensor Data', *Frontiers in Bioengineering and Biotechnology*, 8, p. 604. doi: 10.3389/fbioe.2020.00604.

Ferris, D. P., Louie, M. and Farley, C. T. (1998) 'Running in the real world: adjusting leg stiffness for different surfaces', *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1400), pp. 989–994. doi: 10.1098/rspb.1998.0388.

Figo, D. *et al.* (2010) 'Preprocessing techniques for context recognition from accelerometer data', *Personal and Ubiquitous Computing*, 14, pp. 645–662. doi: 10.1007/s00779-010-0293-9.

Firminger, C. R. and Edwards, W. B. (2016) 'The influence of minimalist footwear and stride length reduction on lower-extremity running mechanics and cumulative loading.', *Journal of science and medicine in sport*. doi: 10.1016/j.jsams.2016.03.003.

Giandolini, M. *et al.* (2014) 'A simple field method to identify foot strike pattern during running', *Journal of Biomechanics*, 47(7), pp. 1588–1593. doi: 10.1016/j.jbiomech.2014.03.002.

Giovanelli, N. *et al.* (2016) 'Energetics of vertical kilometer foot races; is steeper cheaper?', *Journal of Applied Physiology*, 120(3), pp. 370–375. doi: 10.1152/japplphysiol.00546.2015.

Gottschall, J. S. and Kram, R. (2005) 'Ground reaction forces during downhill and uphill running', *Journal of Biomechanics*, 38(3), pp. 445–452. doi: 10.1016/j.jbiomech.2004.04.023.

Halilaj, E. *et al.* (2018) 'Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities', *Journal of biomechanics*, 81, pp. 1–11. doi: 10.1016/j.jbiomech.2018.09.009.

Harris, C. R. *et al.* (2020) 'Array programming with NumPy', *Nature*, 585(7825), pp. 357–362. doi: 10.1038/s41586-020-2649-2.

Hennig, E. M. and Lafortune, M. A. (1991) 'Relationships Between Ground Reaction Force and Tibial Bone Acceleration Parameters', *International Journal of Sport Biomechanics*, 7(3), pp. 303–309.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long Short-Term Memory', *Neural Computation*, 9(8), pp. 1735–1789.

International Trail Running Association (2020) ‘2020 Trail Running Infographics’. Available at: https://itra.run/documents/Infographics/EN-ITRA-Trail-Running-Infographics-2020.pdf.

Janssen, M. *et al.* (2020) ‘Understanding Different Types of Recreational Runners and How They Use Running-Related Technology’, *International Journal of Environmental Research and Public Health*, 17(7). doi: 10.3390/ijerph17072276.

Johnson, C. D. *et al.* (2020) ‘A comparison of attachment methods of skin mounted inertial measurement units on tibial accelerations’, *Journal of Biomechanics*, 113, p. 110118. doi: 10.1016/j.jbiomech.2020.110118.

Johnson, W. R. *et al.* (2021) ‘Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning’, *IEEE Transactions on Biomedical Engineering*, 68(1), pp. 289–297. doi: 10.1109/TBME.2020.3006158.

Khassetarash, A. *et al.* (2020) ‘Biomechanics of graded running: Part II—Joint kinematics and kinetics’, *Scandinavian Journal of Medicine & Science in Sports*, 30(9), pp. 1642–1654. doi: 10.1111/sms.13735.

Kiernan, D. *et al.* (2018) ‘Accelerometer-based prediction of running injury in National Collegiate Athletic Association track athletes’, *Journal of Biomechanics*, 73, pp. 201–209. doi: 10.1016/j.jbiomech.2018.04.001.

Kingma, D. P. and Ba, J. (2017) ‘Adam: A Method for Stochastic Optimization’, *arXiv:1412.6980 [cs]*. Available at: http://arxiv.org/abs/1412.6980 (Accessed: 9 November 2020).

Kipp, S., Taboga, P. and Kram, R. (2017) ‘Ground reaction forces during steeplechase hurdling and waterjumps’, *Sports Biomechanics*, 16(2), pp. 152–165. doi: 10.1080/14763141.2016.1212917.

Li, L. *et al.* (2018) ‘Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization’, *Journal of Machine Learning Research*, 18(185), pp. 1–52.

McMahon, T. A., Valiant, G. and Frederick, E. C. (1987) ‘Groucho running’, *Journal of Applied Physiology*, 62(6), pp. 2326–2337. doi: 10.1152/jappl.1987.62.6.2326.

Mundt, M. *et al.* (2020) ‘Prediction of ground reaction force and joint moments based on optical motion capture data during gait’, *Medical Engineering & Physics*, 86, pp. 29–34. doi: 10.1016/j.medengphy.2020.10.001.

Neugebauer, J. M., Hawkins, D. A. and Beckett, L. (2012) ‘Estimating Youth Locomotion Ground Reaction Forces Using an Accelerometer-Based Activity Monitor’, *PLOS ONE*, 7(10), p. e48182. doi: 10.1371/journal.pone.0048182.

pandas development team (2020) *pandas-dev/pandas: Pandas*. Available at: https://doi.org/10.5281/zenodo.3509134.

Pogson, M. *et al.* (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, pp. 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. *et al.* (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(14), pp. 3362–3367. doi: 10.1016/j.jbiomech.2016.08.032.

Ren, L., Jones, R. K. and Howard, D. (2008) 'Whole body inverse dynamics over a complete gait cycle based only on measured kinematics', *Journal of Biomechanics*, 41(12), pp. 2750–2759. doi: 10.1016/j.jbiomech.2008.06.001.

Ruder, M. *et al.* (2019) 'Relationship of Foot Strike Pattern and Landing Impacts during a Marathon', *Medicine and Science in Sports and Exercise*, 51(10), pp. 2073–2079. doi: 10.1249/MSS.0000000000002032.

Running USA (2019) '2019 U.S Running Trends Report'.

Ryan, M. R. *et al.* (2020) 'Comparison of different measures to monitor week-to-week changes in training load in high school runners', *International Journal of Sports Science & Coaching*, p. 174795412097030. doi: 10.1177/1747954120970305.

Saeb, S. *et al.* (2017) 'The need to approximate the use-case in clinical machine learning', *GigaScience*, 6(5). doi: 10.1093/gigascience/gix019.

Scott, M. T. U., Scott, T. J. and Kelly, V. G. (2016) 'The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review', *The Journal of Strength & Conditioning Research*, 30(5), pp. 1470–1490. doi: 10.1519/JSC.0000000000001221.

Selvin, S. *et al.* (2017) 'Stock price prediction using LSTM, RNN and CNN-sliding window model', in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1643–1647. doi: 10.1109/ICACCI.2017.8126078.

Snyder, K. L. and Farley, C. T. (2011) 'Energetically optimal stride frequency in running: the effects of incline and decline', *The Journal of Experimental Biology*, 214(Pt 12), pp. 2089–2095. doi: 10.1242/jeb.053157.

Vernillo, G. *et al.* (2020) 'Biomechanics of graded running: Part I - Stride parameters, external forces, muscle activations', *Scandinavian Journal of Medicine & Science in Sports*, 30(9), pp. 1632–1641. doi: 10.1111/sms.13708.

Vincent, H. K. *et al.* (2020) 'Impact of Body Mass Index on Biomechanics of Recreational Runners', *PM&R*, 12(11), pp. 1106–1112. doi: https://doi.org/10.1002/pmrj.12335.

Virtanen, P. *et al.* (2020) 'SciPy 1.0: fundamental algorithms for scientific computing in Python', *Nature Methods*, 17(3), pp. 261–272. doi: 10.1038/s41592-019-0686-2.

Vitali, R. V. *et al.* (2019) 'Body-worn IMU array reveals effects of load on performance in an outdoor obstacle course', *PLOS ONE*. Edited by Y.-K. Jan, 14(3), p. e0214008. doi: 10.1371/journal.pone.0214008.

Voloshina, A. S. and Ferris, D. P. (2015) 'Biomechanics and energetics of running on uneven terrain', *Journal of Experimental Biology*, 218(5), pp. 711–719. doi: 10.1242/jeb.106518.

Wang, S. and Jiang, J. (2016) 'Learning Natural Language Inference with LSTM', *arXiv:1512.08849 [cs]*. Available at: http://arxiv.org/abs/1512.08849 (Accessed: 22 November 2020).

Whiting, C. S. *et al.* (2020) 'Steep (30°) uphill walking vs. running: COM movements, stride kinematics, and leg muscle excitations', *European Journal of Applied Physiology*, 120(10), pp. 2147–2157. doi: 10.1007/s00421-020-04437-y.

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag (Use R!). doi: 10.1007/978-0-387-98141-3.

Wickham, H. (2019) *stringr: Simple, Consistent Wrappers for Common String Operations*. Available at: https://CRAN.R-project.org/package=stringr.

Wickham, H. *et al.* (2020) *dplyr: A Grammar of Data Manipulation*. Available at: https://CRAN.R-project.org/package=dplyr.

Willy, R. W. *et al.* (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(2), pp. 197–205. doi: https://doi.org/10.1111/sms.12413.

Wouda, F. J. *et al.* (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, J. R. *et al.* (2018) 'Acute changes in foot strike pattern and cadence affect running parameters associated with tibial stress fractures', *Journal of Biomechanics*, 76, pp. 1–7. doi: 10.1016/j.jbiomech.2018.05.017.