# Predicting vertical ground reaction forces from 3D accelerometery using reservoir computers leads to accurate gait event detection

Margit M. Bach[1], Nadia Dominici[*1], Andreas Daffertshofer[*1]

[1]Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Amsterdam Movement Sciences & Institute of Brain and Behavior Amsterdam, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[*]Corresponding authors: n.dominici@vu.nl, a.daffertshofer@vu.nl

## Abstract

Accelerometers are low-cost measurement devices that can readily be used outside the lab. However, determining isolated gait events from accelerometer signals, especially foot-off events during running, is an open problem. We outline a two-step approach where machine learning serves to predict vertical ground reaction forces from accelerometer signals, followed by force-based event detection.

We collected shank accelerometer signals and ground reaction forces from five adults during comfortable walking and running on an instrumented treadmill. We trained one common reservoir computer using segmented data using both walking and running data. Despite being trained on just a small number of strides, this reservoir computer predicted vertical ground reaction forces in continuous gait with excellent quality. The subsequent foot contact and foot off event detection proved highly accurate when compared to the golden standard based on co-registered ground reaction forces. Our proof-of-concept illustrates the capacity of combining accelerometery with machine learning for detecting isolated gait events irrespective of mode of locomotion.

## Introduction

Estimating the presence of a step using mobile devices can be realised with fair accuracy and relative ease [1-5]. Yet, many details of the stepping cycle remain opaque such as foot contact and foot off moments, but also more detailed gait characteristics, such as loading responses in (ambulant) clinical contexts. In most of the current literature on wearables, event estimations are rule-based and often require searching for an area of interest [6,7]. This is true for data from inertial measurement units but also for data derived from only accelerometers. Algorithms are optimised for either walking [8-15] or running [16-23] and vary depending on sensor location and type, and on speed. As it stands, they may not generalise to other settings.

Machine learning approaches may provide welcome alternatives. They have been employed to predict stepping moments and gait phases by extracting different features recorded from inertial measurement units [24-37], 3D marker data [38-41], electromyography [42], pressure sensors [43], and textile sensors [44]. Across the board, though, these approaches required a priori feature extractions and are, hence, potentially jeopardised by selection bias.

Stepping instants can readily be identified using (the shape of) ground reaction forces (GRFs), typically obtained from force plate signals [45,46]. With these, one can specify single/double support and flight phases and, correspondingly, the mode of locomotion, i.e., walking or running. As such it seems obvious to first seek to estimate the GRF's shape from wearable sensors and to subsequently using these predicted waveforms to determine gait events. Also here, machine learning has been successful. GRFs during the stance phase, for instance, has been estimated using only acceleration [47-50], a combination of acceleration and angular velocity [51-54], or marker-based kinematics [55,56]. The GRFs during double stance phase could be estimated via marker-based kinematics [57,58] and the GRFs during the full gait cycle using accelerometers placed on the torso [59]. Yet, these approaches often appeared tailored to the data under study rendering their generalisability questionable, but more importantly, in almost all cases, they only managed to predict GRFs for the stance phase, whereas the swing phase is of great importance when investigating running.

A recent review revealed that the shape of the GRF can most accurately be estimated from accelerometery [60] and another found neural networks as a promising tool to do so [61]. This triggered the idea of estimating vertical GRF waveforms from acceleration signals of the lower extremities via reservoir computers like liquid state machines or, more general, echo state networks (ESNs) [62-64]. ESNs are 'minimal' forms of recurrent neural networks. Thanks to the reservoir's 'complex' structure, they may come with great computational capacities [65,66]. In the absence of feedback, one can train them with a very simple and robust rule: optimising output weights by mere linear regression. This is particularly appealing when considering that typical data sets on gait are fairly limited in size and that any implementation of machine learning in wearable devices should come with low computational costs.

In the following, we conceptually prove that a single reservoir computer can accurately predict vertical GRF waveforms from shank accelerometer signals, which allows for detecting gait events during walking and running with particularly high precision.

## Results

Shank accelerometer data (sampling frequency: 2000/14 Hz ≈ 143 Hz) were obtained from five participants during comfortable walking and running on an instrumented treadmill while simultaneously recording ground reaction forces. A total of 378 strides were included from 18 trials (240 walking strides [11 trials] and 138 running strides [7 trials]). A single reservoir computer was trained to predict continuous vertical ground reaction forces based on shank accelerometer data recorded during locomotion (pooled walking and running strides). Here, we would like to note that we only show results of the right-side accelerometer, as the left-side results were very similar. Given this similarity one may pool data across sides to increase the sample size but, as will become clear, this was not needed and not even desired as we sought to illustrate performance for small sample sizes.

Figure 1 contains a schematic of the pre-processing steps and the implemented machine learning approach.

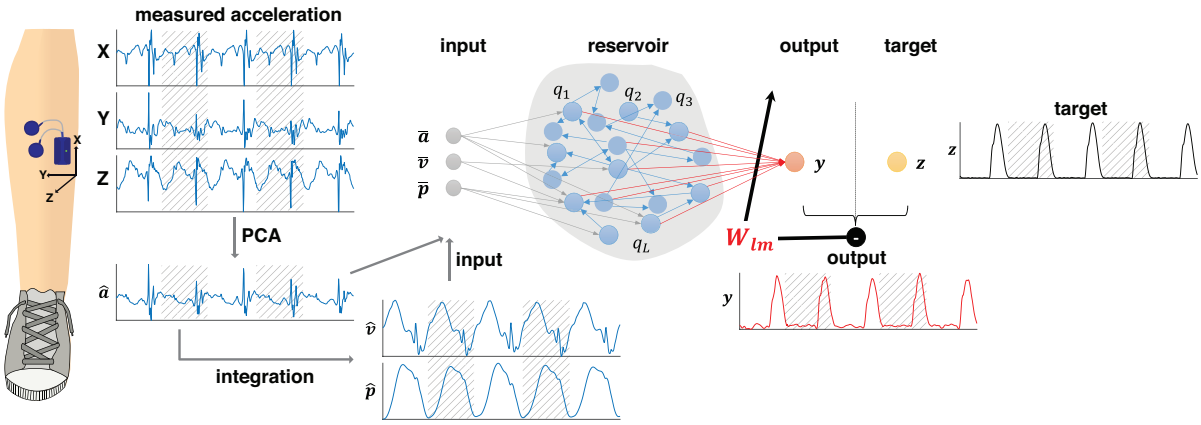


*Figure 1. A reservoir computer was implemented to predict the vertical ground reaction forces. Tri-axial accelerometer data were recorded from the shank. The accelerometer data were re-oriented using a principal component analysis (PCA). The first prinicipal component ($\hat{a}$) was integrated once to obtain the velocity ($\hat{v}$) and position ($\hat{p}$) data. The input $x$ consisting of the normalised accelerometer ($\bar{a}$), velocity ($\bar{v}$), and position ($\bar{p}$) data were subsequently mapped onto the sparsely, randomly connected reservoir $q$. This reservoir generated the output $y$, the predicted normalised vertical ground reaction forces (in red), via output weights $W$. When training, the output was compared to the target $z$, i.e. the measured vertical ground reaction forces (in black). Minimising the difference between generated output and target served to adjust the weights (denoted here as $W_{lm}$). For training, data were segmented into strides, here represented by hatched and unhatched areas. Testing was conducted on continuous data. Data from walking and running were pooled.*

We trained the reservoir computer on stride-segmented data and tested its prediction capacity for continuous data. For each trial, 25% of the continuous data were selected for testing and 25% for validation. The remaining 50% were segmented into strides, pooled across all trials and modes of locomotion, and used for training. Since trials varied substantially in size, we stabilized statistics by always including a maximum number of 15 strides per trial in the training. This split of 50% segmented data used for training, 25% continuous data for validation, and 25% continuous data for testing was repeated 100 times with a random draw each time. The performance of these 100 networks in predicting GRF waveforms exceeded 97%, regardless of mode of locomotion: the coefficients of determination $R^2$ were 0.97 ± 0.01 and the normalised root-mean squared errors $\epsilon$ were 5.2 ± 0.4% (mean ± SD); cf. Figure 2. On average, the subsequently extracted foot contact and foot off events deviated from those based on the measured vertical GRF waveforms by 1.4 samples each. This corresponds to mean absolute errors of 9.8 ± 0.7 ms and 9.8 ± 0.8 ms for foot contact and foot off events, respectively. Here we would like to add that the likewise convincing results when training the network on only walking or on only running are provided as Supplemental Fig. S1, and the results when training and testing on segmented data are provided as Supplemental Fig. S2 and S3.

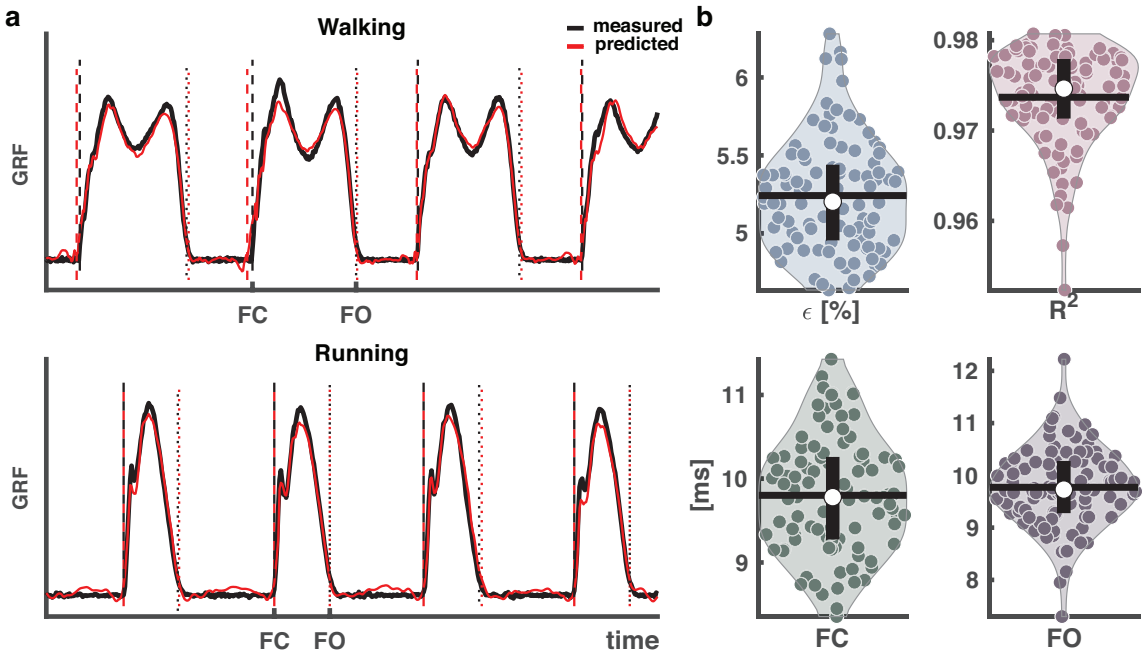


*Figure 2: Output of reservoir computer trained on segmented data and validated and tested on continuous data pooled over conditions.* ***a:*** *Vertical ground reaction force (GRF) waveforms of four randomly selected strides from each condition from a random trial and a random selected training run out of the 100, with the measured vertical GRF waveforms in black and predicted in red. The vertical dashed lines represent the foot contact events (black: measured, red: predicted), the vertical dotted lines the foot off events (black: measured, red: predicted).* ***b****: Normalised root-mean squared error (ϵ), coefficient of determination ($R^2$), mean absolute error of foot contact and foot off. The white dots in the violin-plots illustrate the medians. Black horizontal lines represent the mean and vertical black lines the 1st and 3rd quartiles. Every dot represents one of the 100 training run. Abbreviations: GRF: ground reaction forces, FC: foot contact, FO: foot off.*

To estimate the smallest number of strides needed for a mean reconstruction accuracy above 95%, we changed the size of the training set from 4 to 50% for the total dataset size (again with maximum 15 strides per trial); cf. Figure 3. The size of the validation and test sets were kept fixed at 25% each to guarantee identical accuracy demands. A total of approximately 50 strides, ~30% of the total dataset sufficed to reach $R^2 = 0.95 \pm 0.02$ with $\epsilon = 6.3 \pm 0.7\%$ and a mean absolute error of the foot contact (foot off) of $10.0 \pm 0.7$ ms ($10.1 \pm 0.7$ ms).

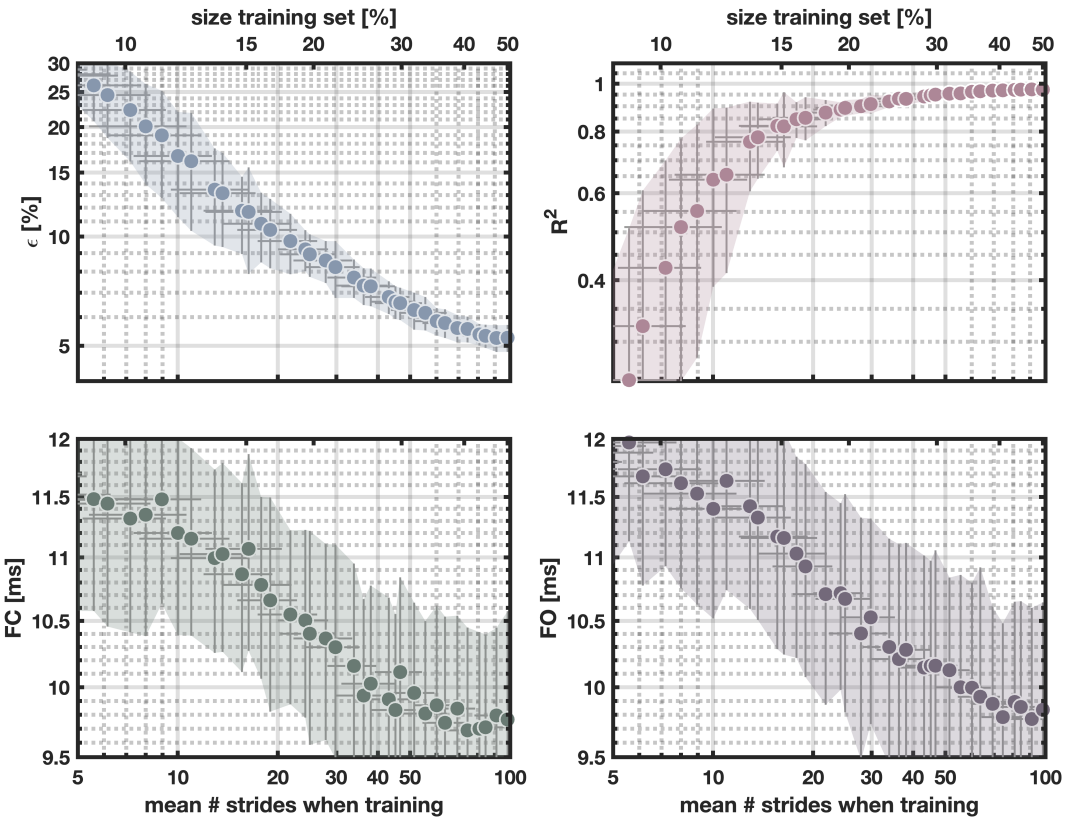


*Figure 3: Output of reservoir computer with training data ranging from 4-50% of the total dataset and validating and testing with 25%, respectively. Upper left panel: normalised root-mean squared error (ϵ), upper right panel: coefficient of determination ($R^2$), lower left panel: mean absolute error of foot contact, and lower right panel: mean absolute error of foot off. Upper x-axes show the percentage of the total dataset used for training and the lower x-axes the corresponding mean number of strides. Each dot represents the mean value for 100 repetitions. The vertical error-bars and colored areas represent the standard deviation of the corresponding measure. The horizontal error-bars represent the standard deviation in the number of strides across the 100 repetitions. Abbreviations: FC: foot contact, FO: foot off.*

Finally, to test whether our approach allows for predicting vertical GRFs in trials that in their entirety were not part of the learning set, we performed a leave-M-out cross-validation. Training was realized in the held in trials using a 75%-25% split of learning and validation (in the training set we used a maximum number of 15 strides per trial). For $M = 5$, the mean $R^2$ reached $0.88 \pm 0.06$, with an error $\epsilon$ of $9 \pm 2\%$. The corresponding mean absolute errors for foot contact and foot off were $10.8 \pm 1.2$ ms and $11.2 \pm 1.6$ ms. For an overview of the results of $M = 1$ to $M = 10$, we refer to Supplemental Fig. S4.

## Discussion

Reservoir computers are a promising tool to predict vertical GRF waveforms based on accelerometer data measured at the shank during walking and running. Accurately predicted GRF waveforms facilitate the detection of gait phases and events. We showed that this 'simple' machine learning approach has excellent prediction accuracy of continuous vertical GRF waveforms independent of the type of locomotion. Put differently, reservoir computers can be used for predicting vertical GRF waveforms for gait of unknown type with excellent performance. This has great potential for uses outside the lab and for collecting large amounts of data. Without a doubt the growing amount of data available for biomechanical analysis in running will greatly drive the field forwards [67]. Machine learning combined with wearable sensors may be the solution to increase the amount of data recorded.

Using machine learning for activity recognition and gait phase recognition based on gait features extracted from biomechanical data, which may be measured with wearable sensors [68,69] has become increasingly popular. The most common techniques to classify gait events or predict GRF waveforms using machine learning are hidden Markov models [26,30,33,34,37,43,70-73], neural networks such as deep neural

networks (more than 1 hidden layer) [25,35,36,44,47,52]; feed-forward neural networks [48,50,53,56-58,74,75]; long short-term models [24,28,38,54,74,75]; convolutional neural networks [29,55]; support vector machines [42,44,76]; (multilayer) perceptron models [28,42,49,51,77], as well as random forest classifiers [36,42,44], K-nearest neighbours [42,54,78], and other types of machine learning using, e.g., Bayesian models [31,32,73,76], Gaussian mixture model [41], and principal component analysis [39,40,51,73]. Echo state networks have the great advantage of low computational costs while still showing excellent performance. They merely require a handful of time series for training and avoid any a priori feature extraction.

There have been several studies conducted on predicting GRFs based on wearables, the vast majority for the stance phase only. Utilising these machine learning techniques properly requires foot contact and foot off to be known. This, however, can only be accomplished by either co-registering footswitches or ground reaction forces, or by implementing a rule-based detection of the gait events. This is exactly what we sought to circumvent. Rule-based detections based on accelerometers during running, searching for a specific peak or valley in some area of interest, are not as precise as an event detection based on vertical ground reaction forces.

We did not restrict the prediction to only the stance phase – we included the entire gait cycle and showed that this worked well on continuous data. We also did not time-normalise the gait cycles and did not impose any other constraints into the timing of the signals. As such, our predictions are robust against variations in speed and stride durations/lengths as well as the type of locomotion. For training the reservoir computer, data were segmented based on the ground truth events, though when these are unknown, the data could be segmented in any way, or they may not be segmented at all; cf. Figure 2. In Supplemental Fig. S2 and S3 online we show the ability of the machine learner when trained, validated, and tested on segmented data. We are convinced that our approach is suitable for lab as well as outdoor use. One very recent study [79] predicted continuous vertical GRFs from trunk accelerations using a long short-term model network with good accuracy during sloped running. The pre-processing involved several filtering and feature extraction steps. By contrast, here we succeeded to reduce the number of pre-processing steps and applied only very weak filtering, i.e., the inputs are by and large the time series (derived from) vertical acceleration.

We validated the use of accelerometer data to estimate the vertical GRF waveform. We used accelerometer data collected by a sensor that could, in fact, also be used for electromyography data collection. The sensors were not aligned in the same way for all participants, which Tan, et al. [80] found can negatively affect the precision of the detection of GRFs using machine learning. However, variability in orientation and position is likely to occur if participants mount their own devices or in large-scale studies. By correcting the orientation of the accelerometers using principal component analysis we circumvent these potential problems and underline the robustness of our method and its applicability in many settings.

Even a small number of strides sufficed to achieve a high reconstruction accuracy. While strides from all participants were pooled for this analysis, it is unlikely that the strides randomly drawn into the training sets were representative for all trials/conditions and that this was also the case for the test sets. We evaluate this via leave-M-out cross-validation, as the test-set should be unknown to the machine [81]. Admittedly, the reconstruction accuracy was not as high, but we consider it still satisfactory.

We would like to note that we did not optimise the machine learner to perform the absolute best it can do. Our primary aim was to show that even in its 'simplest' form, a reservoir computer with regression-based output weights can perform well. Optimising settings will presumably improve performance even further. Adding online feedback will boost the echo state properties and changing the reservoir's connectivity may increase its memory (currently we used a spectral radius of 0.5) [82]. For our proof-of-concept, however, fine-tuning the reservoir might be considered overfitting, which let us decide not to progress along this direction. The most optimal settings will probably depend on the data set under study.

Our accelerometers had a relatively low sampling rate (≈ 143 Hz), which prevents better estimation than 7 ms (i.e., one frame equals 7 ms). An accelerometer with higher sampling frequency will arguably

lead to higher accuracy of the predicted events compared to the ground truth. A sampling frequency of 60-200 Hz is not uncommon when recording kinematics [83-86] and the accuracy is not worse than the accuracy one could obtain using kinematic data. A frequently employed detection algorithm for kinematic gait event detection is the coordinate-based detection algorithm where the distance between the sacrum and foot is used to predict foot contact and foot off events [85, currently cited >800 times]. A review on this and other detection algorithms [86] during running revealed that the coordinate-based detection algorithm has an absolute error of 29 ms for foot contact and 98 ms for foot off (sampling frequency: 200 Hz) whereas the best performing algorithms has an accuracy of 24 ms for foot contact (the foot vertical position [87, >600 citations]) and 6 ms for foot off (the peak knee extension algorithm [88, >600 citations]). For comparison, the best estimation possible with a sampling frequency of 200 Hz is 5 ms, i.e., current algorithms have an accuracy between one and ~ 20 samples. Our approach is comparable to or exceeding this accuracy. Being cheap and easy to collect, being usable outside the lab and for long time-periods are, hence, not the only advantages of accelerometers – they also come with formidable accuracy in step detection when properly combined with reservoir computers.

Our machine learning approach performs well for a dataset comprised of both walking and running data despite a relatively small number of participants and a relatively small number of strides. To investigate its ability for each condition separately, we refer to Supplemental Fig. S1, where we show that training and testing on only one type of locomotion improves the already excellent reconstruction accuracy. The outputs of the reservoir computer can easily be modified to provide other outputs such as other components of the GRF or the centre-of-pressure. By including all three components of the GRF, energetics of the centre-of-mass can be estimated during overground/track running which in turn can provide even more information about the locomotion type. Of course, the energetics will be in arbitrary units given our GRF prediction rely on normalise (z-scored) values. All data for this study were recorded on the treadmill. The next step will be to apply reservoir-based prediction to accelerometer data (or gyroscope data) obtained at other parts of the body, e.g., hip mounted (e.g., activity trackers), arm mounted (e.g., sport watches, smartphones) or head mounted (e.g., augmented/virtual reality glasses) to broaden applicability in daily living contexts and in the clinic. This certainly calls for expanding the current dataset with overground/outdoor locomotion. We expect our findings to be transferable to overground settings. A large meta-analysis suggests that neither vertical ground reaction forces, nor peak tibial accelerations are significantly different between treadmill and overground running [89].

As a final note we would like to recall that our reservoir computer did not include a feedback loop. Adding feedback may allow for not only predicting the GRF accompanying tibial accelerations but eventually also the GRF in forthcoming strides. We trust that future studies will pursue this generalisation as it is beyond the scope of our proof-of-concept study. Given our formidable prediction results, however, we can stress that all the information needed for predicting (vertical) ground reaction forces seems to be present in the (principal component of) accelerometer signals. The use of the latter, hence, provides more opportunities than commonly thought.

Our data and code are made freely available and are ready to use on other datasets and can be extended for use in experiments and clinic without any large hindrances.

## Conclusion

Reservoir computers are an excellent candidate to correctly predict vertical ground reaction force waveforms from accelerometer signals for a small number of participants and strides. The predicted time series can serve to estimate stepping moments with particularly high accuracy. The ease in training procedure, which requires only a (very) limited number of steps and without prior knowledge about the type of locomotion lets us advocate this machine learning approach for future applications in both research and clinic.

## Methods

*Data acquisition*

We included data of five healthy young adults (3 male / 2 female) in the analysis. The participants provided written informed consent in compliance with the Declaration of Helsinki. The experimental design was approved by The Scientific and Ethical Review Board of the Faculty of Behavioural & Movement Sciences, Vrije Universiteit Amsterdam, Netherlands (File number: VCWE-2016-149R1). Data were part of a larger study [90,91]. Participants walked and ran at comfortable speeds on an instrumented dual-belt treadmill (Motek Medical BV, Culemborg, Netherlands) in tied-belt mode wearing their own shoes. They were instructed to step with each foot on a separate belt to be able to record the time series of the ground reaction forces from one leg. Participants familiarised themselves with the setup prior to data collection. Tri-axial accelerometers, built into the probe of the wireless bipolar surface electromyography system (Mini wave plus, Zerowire; Cometa, Bareggio, Italy), were mounted on the right and left tibia, respectively (see Figure 1). Data were sampled at 2000/14 Hz with sensitivity of ± 8g. The vertical GRFs were sampled at 1 kHz but re-sampled to 2000/14 Hz. In total 12 walking trials and 17 running trials were recorded for the five participants. Only consecutive strides absent of artefacts (stepping on the wrong belt) were retained leaving 11 walking and 7 running trials with an average of 21 strides per trial for analysis. One participant did not have any usable running trials but was retained for the walking trials. Participants' mean ± standard deviation age, height, and weight were 24.2 ± 2.5 years, 177.0 ± 5.6 cm, 70.0 ± 8.1 kg, respectively. The recorded speeds were 1.32 ± 0.08 m/s for walking and 2.33 ± 0.07 m/s for running.

*Data pre-processing*

Accelerometer signals $a$ were first 'standardised' to their principal axes using principal component analysis (PCA) [92,93]:

$$a = \left(a_x, a_y, a_z\right) \xrightarrow{\text{PCA}} \hat{a}$$

with $\hat{a}$ along the direction of maximum variance and being the only principal component that was retained. $\hat{a}$ was integrated twice over time (after a bi-directional high pass Butterworth filter with cut-off at 1 Hz, 2nd order) to generate likewise standardised velocities $\hat{v}$ and positional data $\hat{p}$:

$$\hat{v} = \int_0^t \hat{a} \cdot dt \quad \text{and} \quad \hat{p} = \int_0^t \hat{v} \cdot dt$$

Per subset (trial) these signals were normalised [81] by means of

$$\hat{a} \to \bar{a} = \frac{\hat{a}}{\text{range}(\hat{a})} \quad \text{and} \quad \hat{v} \to \bar{v} = \frac{\hat{v}}{\text{range}(\hat{v})} \quad \text{and} \quad \hat{p} \to \bar{p} = \frac{\hat{p}}{\text{range}(\hat{p})}$$

before combining them as three-dimensional input data

$$x = (\bar{a}, \bar{v}, \bar{p}) \in \mathbb{R}^{3\times T}$$

with $T$ indicating the number of samples in time. Vertical ground reaction forces $F_z$ were normalised. With this, the target signal for our machine learner (see below) could be defined as z-score [81]

$$z = \bar{F}_z = \left(F_z - \mu(F_z)\right)/\sigma(F_z) \in \mathbb{R}^{1\times T}$$

with $\mu$ and $\sigma$ denoting the mean and standard deviation over time per stride.

Stepping moments (foot contact and foot off events) were identified based on the measured $F_z$ through mere thresholds: First, the $F_z$ was scaled to a range [0 1], then weakly filtered with a polynomial Savitzky-Golay filter (1st order, 5 samples) [94]. The foot contact was defined as the last point below a threshold (12.5% of the maximum of the data) nearest the ascend of the $F_z$ of the stance phase; similarly, the foot off was defined as the first point crossing the same threshold nearest the descending $F_z$ [46,95].

Data were split according to the defined foot off events for further analysis. We considered 36 samples in time on either side of the foot off as transients when correcting for learning errors in the beginning or end of the data.

*Reservoir computer*

We adopted the leaky ESN implementation by Jaeger and Haas [62], see also [96,97]. In brief, we built a reservoir of $N$ nodes $q = (q_1, q_2, \dots, q_L) \in \mathbb{R}^{N\times T}$ that received an input $x = (x_1, x_2, \dots, x_K) \in \mathbb{R}^{K\times T}$ and generated output $y = (y_1, y_2, \dots, y_M) \in \mathbb{R}^{M\times T}$. During the supervised training mode that output was compared with target $z = (z_1, z_2, \dots, z_M) \in \mathbb{R}^{M\times T}$ by means of the $L_2$-norm[98]; cf. Figure 1.

The reservoir dynamics can be written as

$$dq = \tau^{-1}[-\gamma q + \tanh(Cq + Fx)]dt + d\varepsilon$$

where $C \in \mathbb{R}^{N\times N}$ denotes the connectivity of the reservoir. Here, $C$ was set as a sparse, random matrix specified by a sparseness parameter; its weights were normalised for a given spectral radius (the relative magnitude of the leading eigenvalue of $C$). $F \in \mathbb{R}^{N\times K}$ was set to be a dense matrix allowing for an optional scaling of the input values when mapping them onto the reservoir. The quantity $\varepsilon$ stands for uniformly distributed, uncorrelated noise, i.e., $\varepsilon \in \varepsilon\mathcal{U}(-1,1)$, with $\varepsilon$ being reasonably small. The output is given by

$$y = Wq$$

with $W \in \mathbb{R}^{M\times N}$, which is the matrix of the to-be-learning output weights.

Learning was realised by regression, i.e.

$$\|z - q\| = \|z - Wq\|_2 \to \min \Rightarrow W = Q^{-1}z$$

where $Q = [q_1, \dots, q_T]$ and $(\cdot)^{-1}$ denotes the pseudo-inverse. In the case of multiple time series, i.e. $S$ steps (see below), we defined $Q = \left[q_1^{(1)}, \dots, q_{T_1}^{(1)}; q_1^{(2)}, \dots, q_{T_2}^{(2)}; \dots; q_1^{(S)}, \dots, q_{T_S}^{(S)}\right]$ and accordingly we used $z = \left[z^{(1)}; \dots; z^{(S)}\right]$.

The network parameters were set as follows: $N = 1000$, spectral radius $= 0.5$, $F = [0.1; 0.5]$, $\tau = 1$, $\gamma = 0.5$ and $\varepsilon = 10^{-3}$. The noise was primarily added to minimise the risk of overfitting and we put $\varepsilon = 0$ after learning.

*Stepping moments from the predicted vertical ground reaction forces*

Stepping moment identification of the predicted vertical ground reaction force waveforms was implemented in the same way as for the measured vertical ground reaction forces (see above).

Estimation of gait events such as foot contact and foot off from vertical GRF waveforms are considered the golden standard in movement analysis. An example of the detection algorithm during walking and running can be found in Figure 4. One sample difference between the events based on the measured and predicted vertical GRF waveforms equalled ~7 ms due to the relatively low sampling frequency, common to wearable accelerometers.

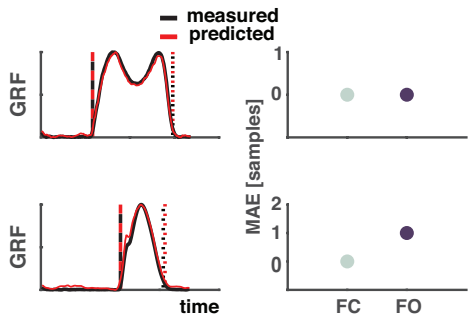


*Figure 4: Example of the estimation of foot contact and foot off events from measured and predicted ground reaction forces. Top: walking, bottom: running. Left side: Measured vertical GRF waveforms are depicted in black and the predicted ones in red. The vertical dashed lines represent foot contact (black: measured, red: predicted) and the dotted lines represent foot off (black: measured, red: predicted). Right side: Differences in samples between events based measured and predicted vertical GRF. One sample equal ~7 ms. Abbreviations: MAE: mean absolute error, FC: Foot contact, FO: Foot off.*

*Statistical evaluation*

The normalised root mean square error $\epsilon$ and the coefficient of determination $R^2$ served for quality assessment of the predicted vertical ground reaction forces. We defined them as follows:

$$\epsilon = \frac{\langle \|z - y\|_2 \rangle}{\mathrm{range}(z)} \quad \text{and} \quad R^2 = 1 - \frac{\langle \|z - y\|_2^2 \rangle}{\langle \|z - \langle z \rangle\|_2^2 \rangle}$$

Prediction of stepping moments were validated using the mean absolute error defined as

$$\frac{1}{t} \sum_{i=1}^{t} \left| E_{\mathrm{target},i} - E_{\mathrm{prediction},i} \right|$$

where, $E_{\mathrm{target},i}$ and $E_{\mathrm{prediction},i}$ refer to target and prediction events $i = 1, \ldots, t$, respectively.

We evaluated the training via cross-validation with 50% of the data segmented and subsequently used for training, 25% continuous data for validation, and the remaining 25% continuous data for testing. We performed 100 iterations with a random draw each time. A continuation rule was used, such that if the $R^2$ of the validation data were all positive, the testing could be employed, and the training was satisfactory. A maximum of 100 repetitions were allocated for validation and in cases where the validation criteria was not satisfied, the training was stopped. In all cases, the number of strides used from each trial during training was reduced to 15 to ensure a balanced design.

Two scenarios served to assess the robustness of the reservoir computer as sketched in Figure 5.

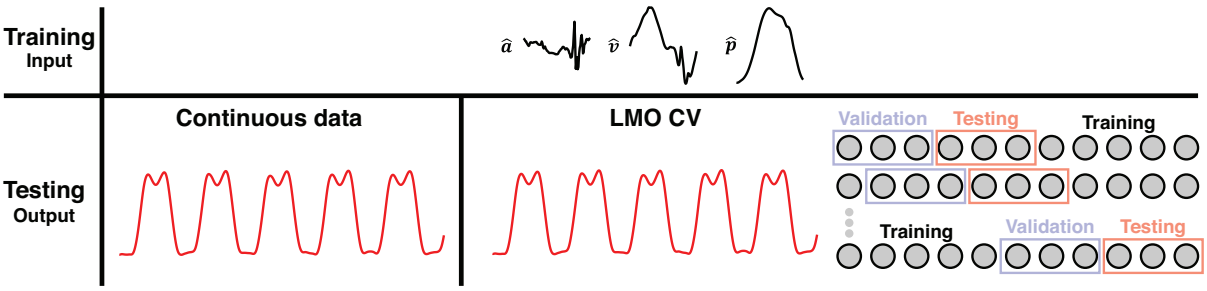


*Figure 5: Schematic of the different testing scenarios used for validating the robustness of the reservoir computer. A random trace of a walking trial is shown here. The input data ($\hat{a}, \hat{v}, \hat{p}$) were first segmented into strides (we considered 36 samples in time on either side of the foot off as transients when correcting for any learning errors in the beginning or end of the data). The first scenario, the training is performed on segmented data and the testing on continuous data. The continuous data (in red) represents the output of the reservoir, the vertical GRF waveforms. Secondly, the training was performed on segmented data, testing was done on continuous, but a leave-M-out cross-validation (LMO CV) was employed (split: 50/25/25% for training, validation, and testing, respectively). The vertical GRF in red represents the output of the trained reservoir during testing. Abbreviations: LMO CV: Leave-M-out cross-validation.*

*Training on segmented data – validating and testing on continuous data*

The applicability of our machine learning approach on continuous data were verified by training the reservoir computer on single strides and subsequent testing on continuous data from each trial (see Figure 2). First, we extracted a random 50% of continuous data from each trial before segmenting the remaining data into strides. The continuous data was split in two so 25% of the data were used for validation and testing respectively. The segmented data were pooled across trials and conditions before being used for training. Training was validated via verifying the mean $R^2$ over the validation set to be positive (see above for definition). Whenever validation did not pass with success, training was repeated using the same subset but other randomly chosen initial conditions (here in all cases validation was passed on first attempt). The entire process was repeated 100 times to allow for statistical evaluation as mentioned earlier.

Additionally, we estimated the minimum amount of data needed to secure a good reconstruction quality ($R^2 > 0.95$), the training data were reduced. We iterated the training 100 times from 4% of the total

dataset to 50% of the total dataset. The validation and test sets remained 25% each for all runs (here the smaller training set sizes required re-learning but eventually validation was passed with success); cf. Figure 3.

*Leave-M-out cross-validation*

To test the machine learner's ability to work as a classifier across participants, it was first trained and validated on a subgroup of trials and then tested on another that was unknown to the machine learner. The cross-validation split was performed based on trials. $M$ trials were held out and the remaining 18-M trials were split 75/25% for training and validation. A total of 100 iterations were performed, with a random draw each iteration. In the main text we report the result for $M = 5$ while the range $M = 1,2 \ldots ,10$ is depicted in Supplemental Fig. S4.

Unless specified otherwise, means and standard deviations are provided.

## Data and code availability

The datasets analysed and the code produced during the current study are available in the GitHub repository, https://github.com/marlow17/PredictingGroundReactionForces.

## References


1 van Oeveren, B. T., de Ruiter, C. J., Beek, P. J., Rispens, S. M. & van Dieen, J. H. An adaptive, real-time cadence algorithm for unconstrained sensor placement. *Med Eng Phys* **52**, 49-58, doi:10.1016/j.medengphy.2017.12.007 (2018).

2 van Oeveren, B. *Running deciphered: The interpretation of running technique from wearable data* PhD thesis, Vrije Universiteit Amsterdam, (2021).

3 de Ruiter, C. J., van Oeveren, B., Francke, A., Zijlstra, P. & van Dieen, J. H. Running speed can be predicted from foot contact time during outdoor over ground running. *PLoS One* **11**, e0163023, doi:10.1371/journal.pone.0163023 (2016).

4 Moe-Nilssen, R. & Helbostad, J. L. Estimation of gait cycle characteristics by trunk accelerometry. *J Biomech* **37**, 121-126, doi:10.1016/s0021-9290(03)00233-1 (2004).

5 Norris, M., Kenny, I. C. & Anderson, R. Comparison of accelerometry stride time calculation methods. *J Biomech* **49**, 3031-3034, doi:10.1016/j.jbiomech.2016.05.029 (2016).

6 Pérez-Ibarra, J. C., Williams, H., Siqueira, A. A. & Krebs, H. I. Real-time identification of impaired gait phases using a single foot-mounted inertial sensor: Review and feasibility study. in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. 1157-1162 (IEEE).

7 Prasanth, H. *et al.* Wearable sensor-based real-time gait detection: A systematic review. *Sensors (Basel)* **21**, doi:10.3390/s21082727 (2021).

8 Trojaniello, D. *et al.* Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: Application to elderly, hemiparetic, parkinsonian and choreic gait. *J Neuroeng Rehabil* **11**, 152, doi:10.1186/1743-0003-11-152 (2014).

9 Rueterbories, J., Spaich, E. G. & Andersen, O. K. Gait event detection for use in fes rehabilitation by radial and tangential foot accelerations. *Med Eng Phys* **36**, 502-508, doi:10.1016/j.medengphy.2013.10.004 (2014).

10 Ben Mansour, K., Rezzoug, N. & Gorce, P. Analysis of several methods and inertial sensors locations to assess gait parameters in able-bodied subjects. *Gait Posture* **42**, 409-414, doi:10.1016/j.gaitpost.2015.05.020 (2015).

11 Greene, B. R. *et al.* An adaptive gyroscope-based algorithm for temporal gait analysis. *Med Biol Eng Comput* **48**, 1251-1260, doi:10.1007/s11517-010-0692-0 (2010).

12 Pacini Panebianco, G., Bisi, M. C., Stagni, R. & Fantozzi, S. Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from imu measurements. *Gait Posture* **66**, 76-82, doi:10.1016/j.gaitpost.2018.08.025 (2018).

13 Selles, R. W., Formanoy, M. A., Bussmann, J. B., Janssens, P. J. & Stam, H. J. Automated estimation of initial and terminal contact timing using accelerometers; development and validation in transtibial amputees and controls. *IEEE Trans Neural Syst Rehabil Eng* **13**, 81-88, doi:10.1109/TNSRE.2004.843176 (2005).

14 Mico-Amigo, M. E. *et al.* A novel accelerometry-based algorithm for the detection of step durations over short episodes of gait in healthy elderly. *J Neuroeng Rehabil* **13**, 38, doi:10.1186/s12984-016-0145-6 (2016).

15 Gurchiek, R. D., Garabed, C. P. & McGinnis, R. S. Gait event detection using a thigh-worn accelerometer. *Gait Posture* **80**, 214-216, doi:10.1016/j.gaitpost.2020.06.004 (2020).

16 Mo, S. & Chow, D. H. K. Accuracy of three methods in gait event detection during overground running. *Gait Posture* **59**, 93-98, doi:10.1016/j.gaitpost.2017.10.009 (2018).

17 Khandelwal, S. & Wickstrom, N. Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the marea gait database. *Gait Posture* **51**, 84-90, doi:10.1016/j.gaitpost.2016.09.023 (2017).

18 Mitschke, C., Heß, T. & Milani, T. Which method detects foot strike in rearfoot and forefoot runners accurately when using an inertial measurement unit? *Applied Sciences* **7**, 959 (2017).

19 Sinclair, J., Hobbs, S. J., Protheroe, L., Edmundson, C. J. & Greenhalgh, A. Determination of gait events using an externally mounted shank accelerometer. *J Appl Biomech* **29**, 118-122, doi:10.1123/jab.29.1.118 (2013).
20 Purcell, B., Channells, J., James, D. & Barrett, R. Use of accelerometers for detecting foot-ground contact time during running. in *BioMEMS and Nanotechnology II.* 603615 (International Society for Optics and Photonics).
21 Lee, J. B., Mellifont, R. B. & Burkett, B. J. The use of a single inertial sensor to identify stride, step, and stance durations of running gait. *Journal of Science and Medicine in Sport* **13**, 270-273 (2010).
22 McGrath, D., Greene, B. R., O'Donovan, K. J. & Caulfield, B. Gyroscope-based assessment of temporal gait parameters during treadmill walking and running. *Sports Engineering* **15**, 207-213, doi:10.1007/s12283-012-0093-8 (2012).
23 Bergamini, E. *et al.* Estimation of temporal parameters during sprint running using a trunk-mounted inertial measurement unit. *J Biomech* **45**, 1123-1126, doi:10.1016/j.jbiomech.2011.12.020 (2012).
24 Tan, H. X., Aung, N. N., Tian, J., Chua, M. C. H. & Yang, Y. O. Time series classification using a modified lstm approach from accelerometer-based data: A comparative study for gait cycle detection. *Gait Posture* **74**, 128-134, doi:10.1016/j.gaitpost.2019.09.007 (2019).
25 Prado, A., Cao, X., Robert, M. T., Gordon, A. M. & Agrawal, S. K. Gait segmentation of data collected by instrumented shoes using a recurrent neural network classifier. *Phys Med Rehabil Clin N Am* **30**, 355-366, doi:10.1016/j.pmr.2018.12.007 (2019).
26 Mannini, A., Genovese, V. & Maria Sabatini, A. Online decoding of hidden markov models for gait event detection using foot-mounted gyroscopes. *IEEE J Biomed Health Inform* **18**, 1122-1130, doi:10.1109/JBHI.2013.2293887 (2014).
27 Mannini, A. & Sabatini, A. M. A hidden markov model-based technique for gait segmentation using a foot-mounted gyroscope. in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 4369-4373 (IEEE).
28 Robberechts, P. *et al.* Predicting gait events from tibial acceleration in rearfoot running: A structured machine learning approach. *Gait Posture* **84**, 87-92, doi:10.1016/j.gaitpost.2020.10.035 (2021).
29 Su, B., Smith, C. & Gutierrez Farewik, E. Gait phase recognition using deep convolutional neural network with inertial measurement units. *Biosensors (Basel)* **10**, doi:10.3390/bios10090109 (2020).
30 Taborri, J., Rossi, S., Palermo, E., Patane, F. & Cappa, P. A novel hmm distributed classifier for the detection of gait phases by means of a wearable inertial sensor network. *Sensors (Basel)* **14**, 16212-16234, doi:10.3390/s140916212 (2014).
31 Martinez-Hernandez, U. & Dehghani-Sanij, A. A. Adaptive bayesian inference system for recognition of walking activities and prediction of gait events using wearable sensors. *Neural Netw* **102**, 107-119, doi:10.1016/j.neunet.2018.02.017 (2018).
32 Meng, L., Martinez-Hernandez, U., Childs, C., Dehghani-Sanij, A. A. & Buis, A. A practical gait feedback method based on wearable inertial sensors for a drop foot assistance device. *IEEE Sensors Journal* **19**, 12235-12243, doi:10.1109/jsen.2019.2938764 (2019).
33 Taborri, J., Scalona, E., Palermo, E., Rossi, S. & Cappa, P. Validation of inter-subject training for hidden markov models applied to gait phase detection in children with cerebral palsy. *Sensors (Basel)* **15**, 24514-24529, doi:10.3390/s150924514 (2015).
34 Abaid, N., Cappa, P., Palermo, E., Petrarca, M. & Porfiri, M. Gait detection in children with and without hemiplegia using single-axis wearable gyroscopes. *PLoS One* **8**, e73152, doi:10.1371/journal.pone.0073152 (2013).
35 Vu, H. T. T. *et al.* Ed-fnn: A new deep learning algorithm to detect percentage of the gait cycle for powered prostheses. *Sensors (Basel)* **18**, doi:10.3390/s18072389 (2018).
36 Yang, J. *et al.* Machine learning based adaptive gait phase estimation using inertial measurement sensors. in *Frontiers in Biomedical Devices.* V001T009A010 (American Society of Mechanical Engineers).
37 Chen, G., Qi, P., Guo, Z. & Yu, H. Gait-event-based synchronization method for gait rehabilitation robots via a bioinspired adaptive oscillator. *IEEE Trans Biomed Eng* **64**, 1345-1356, doi:10.1109/TBME.2016.2604340 (2017).
38 Kidzinski, L., Delp, S. & Schwartz, M. Automatic real-time gait event detection in children using deep neural networks. *PLoS One* **14**, e0211466, doi:10.1371/journal.pone.0211466 (2019).
39 Osis, S. T., Hettinga, B. A. & Ferber, R. Predicting ground contact events for a continuum of gait types: An application of targeted machine learning using principal component analysis. *Gait Posture* **46**, 86-90, doi:10.1016/j.gaitpost.2016.02.021 (2016).
40 Osis, S. T., Hettinga, B. A., Leitch, J. & Ferber, R. Predicting timing of foot strike during running, independent of striking technique, using principal component analysis of joint angles. *J Biomech* **47**, 2786-2789, doi:10.1016/j.jbiomech.2014.06.009 (2014).
41 Aung, M. S. *et al.* Automated detection of instantaneous gait events using time frequency analysis and manifold embedding. *IEEE Trans Neural Syst Rehabil Eng* **21**, 908-916, doi:10.1109/TNSRE.2013.2239313 (2013).
42 Morbidoni, C. *et al.* Machine-learning-based prediction of gait events from emg in cerebral palsy children. *IEEE Trans Neural Syst Rehabil Eng* **29**, 819-830, doi:10.1109/TNSRE.2021.3076366 (2021).
43 Crea, S. *et al.* Development of gait segmentation methods for wearable foot pressure sensors. *Annu Int Conf IEEE Eng Med Biol Soc* **2012**, 5018-5021, doi:10.1109/EMBC.2012.6347120 (2012).
44 Rezaei, A., Ejupi, A., Gholami, M., Ferrone, A. & Menon, C. Preliminary investigation of textile-based strain sensors for the detection of human gait phases using machine learning. in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob).* 563-568 (IEEE).
45 Roerdink, M., Coolen, B. H., Clairbois, B. H., Lamoth, C. J. & Beek, P. J. Online gait event detection using a large force platform embedded in a treadmill. *J Biomech* **41**, 2628-2632, doi:10.1016/j.jbiomech.2008.06.023 (2008).
46 Borghese, N. A., Bianchi, L. & Lacquaniti, F. Kinematic determinants of human locomotion. *J Physiol* **494 ( Pt 3)**, 863-879, doi:10.1113/jphysiol.1996.sp021539 (1996).

47 Davidson, P., Virekunnas, H., Sharma, D., Piche, R. & Cronin, N. Continuous analysis of running mechanics by means of an integrated ins/gps device. *Sensors (Basel)* **19**, doi:10.3390/s19061480 (2019).
48 Ngoh, K. J., Gouwanda, D., Gopalai, A. A. & Chong, Y. Z. Estimation of vertical ground reaction force during running using neural network model and uniaxial accelerometer. *J Biomech* **76**, 269-273, doi:10.1016/j.jbiomech.2018.06.006 (2018).
49 Leporace, G., Batista, L. A. & Nadal, J. Prediction of 3d ground reaction forces during gait based on accelerometer data. *Research on Biomedical Engineering* **34**, 211-216, doi:10.1590/2446-4740.06817 (2018).
50 Lim, H., Kim, B. & Park, S. Prediction of lower limb kinetics and kinematics during walking by a single imu on the lower back using machine learning. *Sensors (Basel)* **20**, doi:10.3390/s20010130 (2019).
51 Pogson, M., Verheul, J., Robinson, M. A., Vanrenterghem, J. & Lisboa, P. A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities. *Med Eng Phys* **78**, 82-89, doi:10.1016/j.medengphy.2020.02.002 (2020).
52 Wouda, F. J. *et al.* Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors. *Front Physiol* **9**, 218, doi:10.3389/fphys.2018.00218 (2018).
53 Lee, M. & Park, S. Estimation of three-dimensional lower limb kinetics data during walking using machine learning from a single imu attached to the sacrum. *Sensors (Basel)* **20**, doi:10.3390/s20216277 (2020).
54 Sharma, D., Davidson, P., Muller, P. & Piche, R. Indirect estimation of vertical ground reaction force from a body-mounted ins/gps using machine learning. *Sensors (Basel)* **21**, doi:10.3390/s21041553 (2021).
55 Johnson, W. R. *et al.* Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning. *IEEE Trans Biomed Eng* **68**, 289-297, doi:10.1109/TBME.2020.3006158 (2021).
56 Komaris, D.-S. *et al.* Predicting three-dimensional ground reaction forces in running by using artificial neural networks and lower body kinematics. *IEEE Access* **7**, 156779-156786, doi:10.1109/access.2019.2949699 (2019).
57 Choi, A., Lee, J.-M. & Mun, J. H. Ground reaction forces predicted by using artificial neural network during asymmetric movements. *International Journal of Precision Engineering and Manufacturing* **14**, 475-483, doi:10.1007/s12541-013-0064-4 (2013).
58 Oh, S. E., Choi, A. & Mun, J. H. Prediction of ground reaction forces during gait based on kinematics and a neural network model. *J Biomech* **46**, 2372-2380, doi:10.1016/j.jbiomech.2013.07.036 (2013).
59 Guo, Y. *et al.* A new proxy measurement algorithm with application to the estimation of vertical ground reaction forces using wearable sensors. *Sensors (Basel)* **17**, doi:10.3390/s17102181 (2017).
60 Horsley, B. J. *et al.* Does site matter? Impact of inertial measurement unit placement on the validity and reliability of stride variables during running: A systematic review and meta-analysis. *Sports Med* **51**, 1449-1489, doi:10.1007/s40279-021-01443-8 (2021).
61 Ancillao, A., Tedesco, S., Barton, J. & O'Flynn, B. Indirect measurement of ground reaction forces and moments by means of wearable inertial sensors: A systematic review. *Sensors (Basel)* **18**, doi:10.3390/s18082564 (2018).
62 Jaeger, H. & Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78-80, doi:10.1126/science.1091277 (2004).
63 Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning*. (MIT press, 2016).
64 Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation* **14**, 2531-2560 (2002).
65 Pathak, J., Hunt, B., Girvan, M., Lu, Z. & Ott, E. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys Rev Lett* **120**, 024102, doi:10.1103/PhysRevLett.120.024102 (2018).
66 Lukoševičius, M. & Jaeger, H. Reservoir computing approaches to recurrent neural network training. *Computer Science Review* **3**, 127-149, doi:10.1016/j.cosrev.2009.03.005 (2009).
67 Novacheck, T. F. The biomechanics of running. *Gait Posture* **7**, 77-95, doi:10.1016/s0966-6362(97)00038-6 (1998).
68 Farrahi, V., Niemela, M., Kangas, M., Korpelainen, R. & Jamsa, T. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait Posture* **68**, 285-299, doi:10.1016/j.gaitpost.2018.12.003 (2019).
69 Figueiredo, J., Santos, C. P. & Moreno, J. C. Automatic recognition of gait patterns in human motor disorders using machine learning: A review. *Med Eng Phys* **53**, 1-12, doi:10.1016/j.medengphy.2017.12.006 (2018).
70 Mannini, A. & Sabatini, A. M. Gait phase detection and discrimination between walking-jogging activities using hidden markov models applied to foot motion data from a gyroscope. *Gait Posture* **36**, 657-661, doi:10.1016/j.gaitpost.2012.06.017 (2012).
71 Chen, G., Salim, V. & Yu, H. A novel gait phase-based control strategy for a portable knee-ankle-foot robot. in *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. 571-576 (IEEE).
72 Guenterberg, E. *et al.* A method for extracting temporal parameters based on hidden markov models in body sensor networks with inertial sensors. *IEEE Trans Inf Technol Biomed* **13**, 1019-1030, doi:10.1109/TITB.2009.2028421 (2009).
73 Yuwono, M., Su, S. W., Guo, Y., Moulton, B. D. & Nguyen, H. T. Unsupervised nonparametric method for gait analysis using a waist-worn inertial sensor. *Applied Soft Computing* **14**, 72-80, doi:10.1016/j.asoc.2013.07.027 (2014).
74 Choi, A., Jung, H., Lee, K. Y., Lee, S. & Mun, J. H. Machine learning approach to predict center of pressure trajectories in a complete gait cycle: A feedforward neural network vs. Lstm network. *Med Biol Eng Comput* **57**, 2693-2703, doi:10.1007/s11517-019-02056-0 (2019).
75 Choi, A., Jung, H. & Mun, J. H. Single inertial sensor-based neural networks to estimate com-cop inclination angle during walking. *Sensors (Basel)* **19**, doi:10.3390/s19132974 (2019).
76 Nutakki, C. *et al.* Classification and kinetic analysis of healthy gait using multiple accelerometer sensors. *Procedia Computer Science* **171**, 395-402 (2020).

77 Mijailovic, N., Gavrilovic, M., Rafajlovic, S., Đuric-Jovicic, M. & Popovic, D. Gait phases recognition from accelerations and ground reaction forces: Application of neural networks. *Telfor Journal* **1**, 34-36 (2009).
78 Chu, K. H., Jiang, X. & Menon, C. Wearable step counting using a force myography-based ankle strap. *J Rehabil Assist Technol Eng* **4**, 2055668317746307, doi:10.1177/2055668317746307 (2017).
79 Alcantara, R. S., Edwards, W. B., Millet, G. Y. & Grabowski, A. M. Predicting continuous ground reaction forces from accelerometers during uphill and downhill running: A recurrent neural network solution. *PeerJ* **10**, e12752, doi:10.7717/peerj.12752 (2022).
80 Tan, T., Chiasson, D. P., Hu, H. & Shull, P. B. Influence of imu position and orientation placement errors on ground reaction force estimation. *J Biomech* **97**, 109416, doi:10.1016/j.jbiomech.2019.109416 (2019).
81 Halilaj, E. *et al.* Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *J Biomech* **81**, 1-11, doi:10.1016/j.jbiomech.2018.09.009 (2018).
82 Yildiz, I. B., Jaeger, H. & Kiebel, S. J. Re-visiting the echo state property. *Neural Netw* **35**, 1-9, doi:10.1016/j.neunet.2012.07.005 (2012).
83 Hreljac, A. & Marshall, R. N. Algorithms to determine event timing during normal walking using kinematic data. *J Biomech* **33**, 783-786, doi:10.1016/s0021-9290(00)00014-2 (2000).
84 O'Connor, C. M., Thorpe, S. K., O'Malley, M. J. & Vaughan, C. L. Automatic detection of gait events using kinematic data. *Gait Posture* **25**, 469-474, doi:10.1016/j.gaitpost.2006.05.016 (2007).
85 Zeni, J. A., Jr., Richards, J. G. & Higginson, J. S. Two simple methods for determining gait events during treadmill and overground walking using kinematic data. *Gait Posture* **27**, 710-714, doi:10.1016/j.gaitpost.2007.07.007 (2008).
86 Fellin, R. E., Rose, W. C., Royer, T. D. & Davis, I. S. Comparison of methods for kinematic identification of footstrike and toe-off during overground and treadmill running. *J Sci Med Sport* **13**, 646-650, doi:10.1016/j.jsams.2010.03.006 (2010).
87 Alton, F., Baldey, L., Caplan, S. & Morrissey, M. C. A kinematic comparison of overground and treadmill walking. *Clin Biomech (Bristol, Avon)* **13**, 434-440, doi:10.1016/s0268-0033(98)00012-6 (1998).
88 Dingwell, J. B., Cusumano, J. P., Cavanagh, P. R. & Sternad, D. Local dynamic stability versus kinematic variability of continuous overground and treadmill walking. *J Biomech Eng* **123**, 27-32, doi:10.1115/1.1336798 (2001).
89 Van Hooren, B. *et al.* Is motorized treadmill running biomechanically comparable to overground running? A systematic review and meta-analysis of cross-over studies. *Sports Med* **50**, 785-813, doi:10.1007/s40279-019-01237-z (2020).
90 Bach, M. M., Daffertshofer, A. & Dominici, N. The development of mature gait patterns in children during walking and running. *Eur J Appl Physiol* **121**, 1073-1085, doi:10.1007/s00421-020-04592-2 (2021).
91 Bach, M. M., Daffertshofer, A. & Dominici, N. Muscle synergies in children walking and running on a treadmill. *Front Hum Neurosci* **15**, 637157, doi:10.3389/fnhum.2021.637157 (2021).
92 Moe-Nilssen, R. A new method for evaluating motor control in gait under real-life environmental conditions. Part 1: The instrument. *Clin Biomech (Bristol, Avon)* **13**, 320-327, doi:10.1016/s0268-0033(98)00089-8 (1998).
93 Rispens, S. M. *et al.* Consistency of gait characteristics as determined from acceleration data collected at different trunk locations. *Gait Posture* **40**, 187-192, doi:10.1016/j.gaitpost.2014.03.182 (2014).
94 Savitzky, A. & Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* **36**, 1627-1639 (1964).
95 Ghoussayni, S., Stevens, C., Durham, S. & Ewins, D. Assessment and validation of a simple automated method for the detection of gait events and intervals. *Gait Posture* **20**, 266-272, doi:10.1016/j.gaitpost.2003.10.001 (2004).
96 Jaeger, H. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report* **148**, 13 (2001).
97 Jaeger, H. *Tutorial on training recurrent neural networks, covering bppt, rtrl, ekf and the" echo state network" approach*. Vol. 5 (GMD-Forschungszentrum Informationstechnik Bonn, 2002).
98 Lukoševičius, M. in *Neural networks: Tricks of the trade: Second edition* (eds Grégoire Montavon, Geneviève B. Orr, & Klaus-Robert Müller) 659-686 (Springer Berlin Heidelberg, 2012).

## Acknowledgements

We would like to acknowledge the participants who participated in the study. We would also like to acknowledge Sjoerd Bruijn and Jaap van Dieën for their valuable inputs to a previous version of the manuscript. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 715945 Learn2Walk) and from the Dutch Organisation for Scientific Research (NWO) VIDI grant (016.156.346 FirSTeps).

## Author contributions

M.B. and N.D. set up the experiment, M.B. collected and pre-processed the data, M.B. and A.D. wrote the code, M.B. prepared all figures, M.B. and A.D. wrote the first draft of the manuscript. All authors commented and provided feedback on the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Figure legends

*Figure 1. A reservoir computer was implemented to predict the vertical ground reaction forces. Tri-axial accelerometer data were recorded from the shank. The accelerometer data were re-oriented using a principal component analysis (PCA). The first prinicipal component ($\hat{a}$) was integrated once to obtain the velocity ($\hat{v}$) and position ($\hat{p}$) data. The input $x$ consisting of the normalised accelerometer ($\bar{a}$), velocity ($\bar{v}$), and position ($\bar{p}$) data were subsequently mapped onto the sparsely, randomly connected reservoir $q$. This reservoir generated the output $y$, the predicted normalised vertical ground reaction forces (in red), via output weights $W$. When training, the output was compared to the target $z$, i.e. the measured vertical ground reaction forces (in black). Minimising the difference between generated output and target served to adjust the weights (denoted here as $W_{lm}$). For training, data were segmented into strides, here represented by hatched and unhatched areas. Testing was conducted on continuous data. Data from walking and running were pooled.*

*Figure 6: Output of reservoir computer trained on segmented data and validated and tested on continuous data pooled over conditions.* ***a:*** *Vertical ground reaction force (GRF) waveforms of four randomly selected strides from each condition from a random trial and a random selected training run out of the 100, with the measured vertical GRF waveforms in black and predicted in red. The vertical dashed lines represent the foot contact events (black: measured, red: predicted), the vertical dotted lines the foot off events (black: measured, red: predicted).* ***b****: Normalised root-mean squared error ($\epsilon$), coefficient of determination ($R^2$), mean absolute error of foot contact and foot off. The white dots in the violin-plots illustrate the medians. Black horizontal lines represent the mean and vertical black lines the 1st and 3rd quartiles. Every dot represents one of the 100 training run. Abbreviations: GRF: ground reaction forces, FC: foot contact, FO: foot off.*

*Figure 3: Output of reservoir computer with training data ranging from 4-50% of the total dataset and validating and testing with 25%, respectively. Upper left panel: normalised root-mean squared error ($\epsilon$), upper right panel: coefficient of determination ($R^2$), lower left panel: mean absolute error of foot contact, and lower right panel: mean absolute error of foot off. Upper x-axes show the percentage of the total dataset used for training and the lower x-axes the corresponding mean number of strides. Each dot represents the mean value for 100 repetitions. The vertical error-bars and colored areas represent the standard deviation of the corresponding measure. The horizontal error-bars represent the standard deviation in the number of strides across the 100 repetitions. Abbreviations: FC: foot contact, FO: foot off.*

*Figure 4: Example of the estimation of foot contact and foot off events from measured and predicted ground reaction forces. Top: walking, bottom: running. Left side: Measured vertical GRF waveforms are depicted in black and the predicted ones in red. The vertical dashed lines represent foot contact (black: measured, red: predicted) and the dotted lines represent foot off (black: measured, red: predicted). Right side: Differences in samples between events based measured and predicted vertical GRF. One sample equal ~7 ms. Abbreviations: MAE: mean absolute error, FC: Foot contact, FO: Foot off.*

*Figure 5: Schematic of the different testing scenarios used for validating the robustness of the reservoir computer. A random trace of a walking trial is shown here. The input data ($\hat{a}, \hat{v}, \hat{p}$) were first segmented into strides (we considered 36 samples in time on either side of the foot off as transients when correcting for any learning errors in the beginning or end of the data). The first scenario, the training is performed on segmented data and the testing on continuous data. The continuous data (in red) represents the output of the reservoir, the vertical GRF waveforms. Secondly, the training was performed on segmented data, testing was done on continuous, but a leave-M-out cross-validation (LMO CV) was employed (split: 50/25/25% for training, validation, and testing, respectively). The vertical GRF in red represents the output of the trained reservoir during testing. Abbreviations: LMO CV: Leave-M-out cross-validation.*