ImpactEar: Cross Activity Ground Reaction Force Estimation using Earable IMUs

JAKE STUCHBURY-WASS*†, University of Cambridge, United Kingdom
MATHIAS CILIBERTO†, University of Cambridge, United Kingdom
QIANG YANG, University of Cambridge, United Kingdom
KAYLA-JADE BUTKOW, University of Cambridge, United Kingdom
YANG LIU, Florida State University, United States
TOBIAS RÖDDIGER, IPAI Foundation gGmbH, Germany
EZIO PREATONI, University of Bath, United Kingdom
DONG MA, University of Cambridge, United Kingdom
CECILIA MASCOLO, University of Cambridge, United Kingdom

Ground Reaction Forces (GRFs) are the forces exerted between the foot and the ground during movement. They are key biomechanical indicators for gait monitoring, injury recovery tracking, and sporting performance. Traditional force plates provide gold-standard accuracy but are expensive and confined to laboratories, while leg- or foot-mounted devices require specialised hardware and sacrifice user comfort. Recent work utilising commodity devices such as smartwatches and earphones offer greater accessibility, yet wrist sensors suffer from arm-swing artifacts and existing earable studies remain activity-specific, limited to either walking or running (and never both) as GRF patterns differ drastically across locomotion types. In this paper, we propose ImpactEar, the first system to estimate complete GRF curves across multiple activities using only a pair of earable inertial measurement units (IMUs). Our core insight is to treat the head as a proxy for the body's center of mass and leverage bilateral earable IMUs to capture both translational and rotational dynamics. Combined with temporal context modeling, this enables a cross-activity mapping from head motion to GRF curves with a single model. ImpactEar employs a lightweight encoder–decoder network that reconstructs left–right GRF profiles across walking, running, and jumping in real time. The evaluation on 30 participants shows that ImpactEar achieves 8.6% normalized root mean square error (NRMSE) and 9.5 ms latency on a mobile phone streaming earable IMU data, outperforming state-of-the-art single-activity baselines and remaining robust under practical conditions, such as reduced sampling rates, single earbuds, different ground surfaces, music playback and head and facial motions. We further demonstrate the power of a cross activity model over a range of different downstream applications including gait asymmetry analysis, jump power assessment, and ballet motion evaluation. Finally, we release ImpactEarDS, the first public earable–GRF dataset with synchronised IMU and force-plate ground truth, paving the way for ubiquitous human kinetics and biomechanics research.

CCS Concepts: • **Human-centered computing** → *Ubiquitous and mobile computing design and evaluation methods.*

*Corresponding author.
†Equal Contribution.

Authors' Contact Information: Jake Stuchbury-Wass, (Corresponding Author), University of Cambridge, United Kingdom, js2372@cl.cam.ac.uk; Mathias Ciliberto, University of Cambridge, United Kingdom, mc2514@cl.cam.ac.uk; Qiang Yang, University of Cambridge, United Kingdom, qiang.yang@cl.cam.ac.uk; Kayla-Jade Butkow, University of Cambridge, United Kingdom, kjb85@cl.cam.ac.uk; Yang Liu, Florida State University, United States, yl25r@fsu.edu; Tobias Röddiger, IPAI Foundation gGmbH, Germany, tobias.roeddiger@ipai-foundation.ai; Ezio Preatoni, University of Bath, United Kingdom, e.preatoni@bath.ac.uk; Dong Ma, University of Cambridge, United Kingdom, dongma@smu.edu.sg; Cecilia Mascolo, University of Cambridge, United Kingdom, cm542@cl.cam.ac.uk.

ACM 2474-9567/2026/9-ART152
https://doi.org/10.1145/3831649

Additional Key Words and Phrases: IMU, Ground reaction forces, Earable devices, Dataset

ACM Reference Format:
Jake Stuchbury-Wass, Mathias Ciliberto, Qiang Yang, Kayla-Jade Butkow, Yang Liu, Tobias Röddiger, Ezio Preatoni, Dong Ma, and Cecilia Mascolo. 2026. ImpactEar: Cross Activity Ground Reaction Force Estimation using Earable IMUs. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 3, Article 152 (September 2026), 25 pages. https://doi.org/10.1145/3831649

1 Introduction

Ground reaction forces (GRFs) describe the force between the foot and the ground and are fundamental to the understanding of human kinetics. Represented as force–time curves, they characterise the magnitude and timing of impacts, revealing lower-limb loading, balance, and propulsion, serving as critical biomechanical parameters for musculoskeletal health, performance optimisation, injury prevention, and rehabilitation monitoring [38, 55]. Estimating GRF across multiple activities, such as walking, running, and jumping, provides a comprehensive understanding of how different locomotion patterns contribute to lower-limb loading [53] and performance in various exercises [1] and rehabilitation scenarios [56].

Traditionally, force plates, instrumented treadmills, or instrumented walkways offer gold-standard accuracy for GRFs estimation, but they are expensive and confined to laboratory settings [55]. To extend GRF estimation beyond the lab, researchers have explored leg- or foot-mounted devices [2, 20] and instrumented insoles [9] to estimate GRF parameters [4, 9]. However, most studies use specialised devices, yielding low accessibility and user acceptance. Recently, commodity wearables such as smartwatches and earphones have attracted attention for their accessibility and integrated Inertial Measurement Units (IMUs) sensors to estimate gait parameters [4, 23] and GRF [50]. However, wrist sensors are dominated by arm swing which make gait sensing difficult [30], while existing earable studies estimate GRF under a single activity (either walking or running), showing limited generalisation to different locomotion patterns [50]. This leads us to a fundamental research question: *Can we estimate GRFs across different activities using only a pair of earable IMUs?*

In this paper, we introduce ImpactEar, the first system that can predict GRFs from all impacts across multiple activities including walking, running, and jumping using earables. As shown in Figure 2(a), our key insight is to treat the head as a proxy for the body's centre of mass, a location commonly used for GRF prediction [2], and to leverage bilateral earable IMUs to capture translational dynamics of the head and estimate correlated GRF with deep learning. However, achieving cross-activity GRF estimation from earables is non-trivial and introduces three technical challenges. First, the morphology and dynamics of GRF curves differ drastically across walking, running, and jumping, varying in amplitude, duration, and frequency (Figure 3). This cross-activity variability makes it difficult for a single model to capture shared force patterns without overfitting to specific locomotion types. Second, head-mounted IMUs record not only body translation but also neck rotations and postural changes, while similar sensor readings can correspond to distinct physical states, such as standing still versus being airborne, posing ambiguity in force estimation. Third, earbuds and mobile phones paired to them are resource constrained: this demands that the models are both lightweight and energy efficient.

To address these challenges, ImpactEar introduces a unified modelling framework that generalises across activities, disentangles complex head dynamics, and remains lightweight for real-time operation. Specifically, to overcome cross-activity variability, ImpactEar standardises all GRF data into a normalised representation and employs a shared encoder–decoder architecture that learns common force patterns across walking, running, and jumping rather than activity-specific models. To handle head dynamics complexity, the system fuses bilateral accelerometer and gyroscope signals to separate translational and rotational motion, while incorporating temporal context windows to disambiguate similar IMU readings occurring during standing, airborne, or impact states. To meet resource constraints of earbuds and the phones they are paired with, ImpactEar adopts a lightweight convolutional autoencoder optimised for low latency and energy efficiency: our implementation efficiently

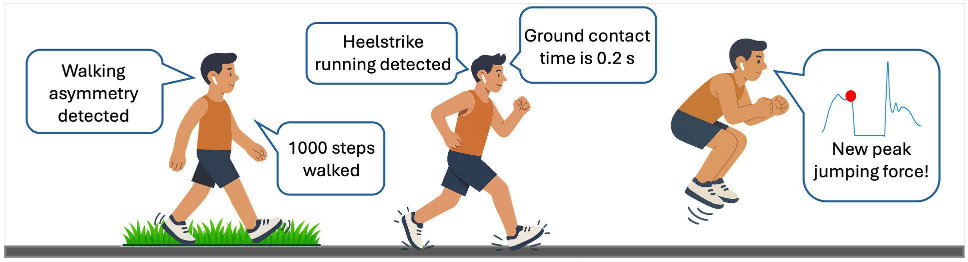


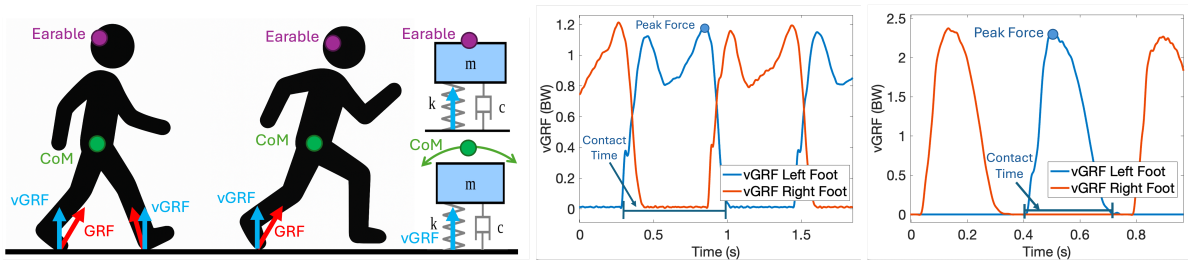
Fig. 1. Illustration of a user undergoing different activities that can be tracked with ImpactEar along with useful downstream applications derived from the estimated GRF curves.

streams the data from the earbuds to the phone and runs inference there. We evaluate ImpactEar on data collected from 30 participants performing walking, running, and jumping, with synchronised force-plate ground truth across different surfaces. The result shows that ImpactEar achieves an average Leave-One-Subject-Out (LOSO) NRMSE of 8.6% and 9.5 ms latency, with comparable performance to state-of-the-art single-activity related works. We also show the feasibility of earable on-device deployment of ImpactEar using quantisation methods. In addition, we investigate real-world considerations, including lower IMU sampling rates comparable to commercial devices, single-channel conditions, and varied ground surfaces, music playback, and head and facial motions demonstrating the robustness of ImpactEar under practical deployment constraints. Beyond numerical accuracy, we demonstrate the system's unique versatility through important downstream applications such as gait asymmetry analysis, jump power estimation, running footstrike type detection, and ballet motion assessment, a selection of these are illustrated in Figure 1 along with the corresponding activity. While these are the downstream tasks selected to showcase ImpactEar, there are many more applications of GRF curves, such as advanced gait analysis [19] or sports coaching that ImpactEar can potentially support. Finally, we release *ImpactEarDS*, the first public dataset that combines bilateral earable IMUs with 3D GRF, moment, and centre-of-pressure measurements, paving the way for future research on ubiquitous and cross-activity biomechanical analysis.

To summarise, this work presents the following contributions:

- We present ImpactEar, the first system that can continuously estimates complete vertical ground reaction force (vGRF) curves across multiple activities using only a pair of IMU-equipped earbuds.
- We propose a modelling framework that treats the head as a proxy for the body's centre of mass and fuses bilateral translational and rotational dynamics within temporal context windows to enable across-activity GRF estimation using a single model with a single modality.
- We comprehensively evaluate ImpactEar on data from 30 participants performing walking, running, and jumping, achieving an average LOSO NRMSE of 8.6% and 9.5 ms latency, outperforming state-of-the-art single-activity baselines. We further assess its robustness under real-world conditions, including single earbud, reduced sampling rates, and different ground surfaces.
- We demonstrate the versatility of ImpactEar through several important biomechanical use cases, including gait metric analysis, gait asymmetry estimation, footstrike type detection, jump peak force estimation, and ballet motion assessment.
- We release ImpactEarDS, the first public dataset that combines bilateral earable IMUs with synchronised 3D GRF, moment, and centre-of-pressure measurements, enabling reproducible and future research on earable-based biomechanics.

The following sections of this paper include a primer to explain the mechanics of GRFs, how they can be modelled at the head as well as the challenges associated with this. We then review related work showing the research gap for cross activity GRF monitoring as well as earable GRF monitoring. We also show the lack of



(a) (b) (c)

Fig. 2. Visualisation of the GRF and vGRF on a person while walking (a, left) and running (a, right). Examples of a vGRF curve while a user is walking (b) and running (c) for both left (blue) and right (orange) feet, normalised in the body-weight unit.

relevant datasets and motivating the collection of ImpactEarDS. We then describe how we designed ImpactEar to overcome these challenges and address these research gaps. Finally, we present the dataset and full evaluation results and downstream tasks followed by a discussion of the results and implications of our findings.

2 Background and Challenges

This primer is included to explain how GRFs are produced, why they're important and how we can model them from inertial data at the head, as well as the challenges associated with this.

2.1 Ground Reaction Forces

Ground reaction forces (GRFs) occur when there is contact between a foot and the ground and these forces are then distributed through the body [55]. The GRF curve gives useful insights into gait analysis for both health and sporting applications, and can predict loading on the body throughout any activity, e.g., jump landings [11]. These forces act at a location called the centre of pressure (CoP). For example, when a person is standing stationary, this force will be split between their two feet at roughly half their body weight per foot. While this static scenario is of little interest, GRFs change when dynamic motion happens, which can be in any activity such as walking, running and jumping. Additionally, GRFs can vary highly from 0 to 5 times body weight or more in certain dynamic scenarios [11]. While the GRF is a 3D force, the most commonly studied component is the vertical ground reaction force (vGRF) with examples shown in Figure 2(b) for walking and Figure 2(c) for running. vGRF presents the largest magnitudes because it also includes the majority of the body's weight, thus it represents most of the loading on the body. When assessing GRFs in the literature, it is common to normalise the forces that are directly measured in Newtons into the unit of Body Weight (BW) by dividing the measurement by the user's body weight. Similarly, ImpactEar estimates GRF in BW units. Some example use cases for GRF curves include gait metrics such as peak forces as annotated in Figure 2(b) and (c). However, sometimes it is important for more detailed gait assessments to be performed by analysing the whole GRF curve [19].

2.2 Sensing Ground Reaction Forces From The Head

While measuring the GRF at the head may at first seem counterintuitive, head motion tracks the body's centre of mass (CoM) motion well. Other works often use the sacrum, a close approximation to the human CoM [37], as a location for an inertial sensor for estimating GRF [2]. Therefore, it can be extended that the head could be a sensible location for GRF assessment. More specifically, many biomechanical models describe relationships between different parts of the human body [7]. One representative example is the relationship between a person's centre of mass and the centre of pressure, from which a model linking the acceleration at the CoM to the GRF can be derived. This is often in the form of a damped mass spring model with the motion of an inverted pendulum [35],

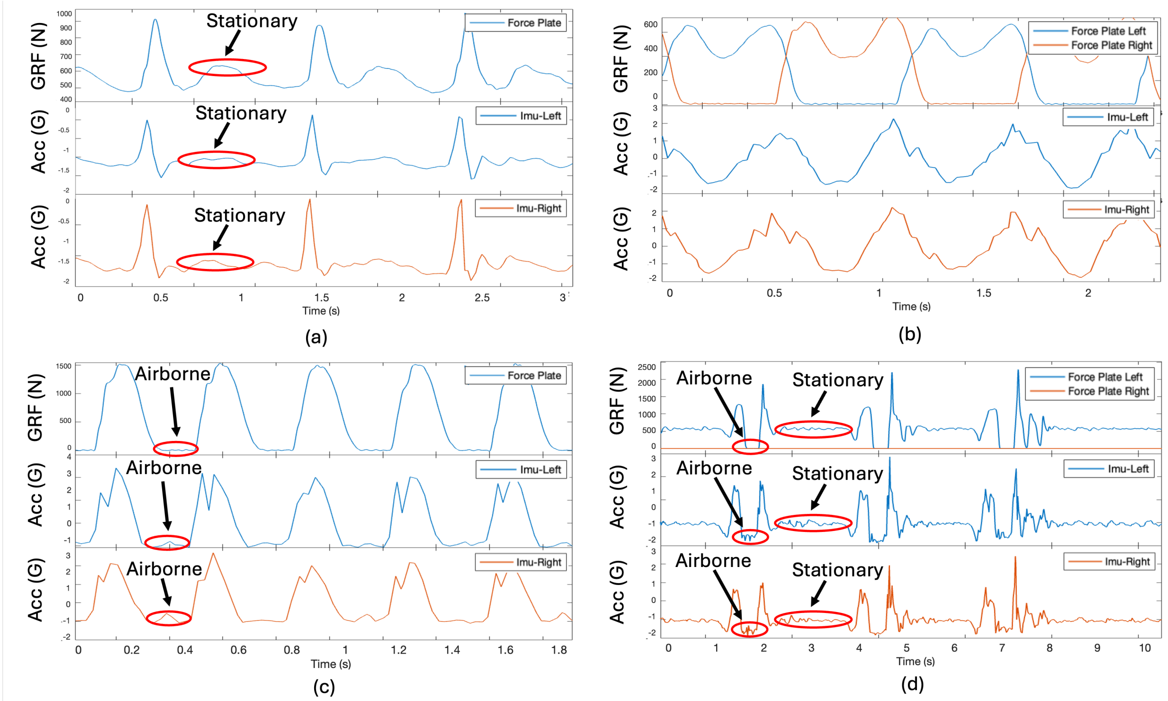


Fig. 3. Examples of GRFs (top) coupled with the earable accelerometer (Acc) data from the left (middle) and right (bottom) devices for (a) footstamps, (b) walking, (c) running, (d) and jumping, showing similar signal morphologies between these two modalities for the cases of (a), (c) and (d) and different morphology for (b) when both legs are interacting with the ground.

as illustrated in Figure 2(a). We could extend this with the body and neck acting as another damped mass spring system from the motion of the CoM to that at the head. Therefore, the dynamics captured by head-mounted measurements can reflect the interactions between the body and the ground, allowing GRFs to be inferred from head motion as shown in Figure 2(a).

2.3 Challenges

To better understand how head-mounted IMU signals relate to GRFs, Figure 3 visualises synchronised GRF and bilateral earable IMU recordings across four representative activities: footstamps, walking, running, and jumping. In Figure 3(a, c, d), footstamps, running, and jumping—the IMU waveforms follow the GRF profiles closely, showing that foot impacts propagate effectively to the head. In contrast, Figure 3(b) illustrates that during walking, the IMU correlates more strongly with the net GRF (the sum of both feet) rather than each individual foot. This observation reveals the **first challenge**: across different locomotion types, GRF magnitudes, durations, and waveform shapes vary drastically, and the head predominantly reflects whole-body motion rather than foot-specific forces. This makes it difficult for a single model to disentangle left–right contributions and to generalise across walking, running, and jumping dynamics. The **second challenge** comes from the complexity of head dynamics. As seen in Figure 3(a, c, d), similar IMU readings (around -1 g) can occur when the user is either standing still stationary (a, d) or airborne while running or jumping (c, d), creating ambiguity between weight-support and free-fall states. Additionally, for all activities, voluntary head rotations and tilts generate angular velocities that can overshadow gait-induced motion, further complicating the mapping from head kinematics to GRF signals. The **third challenge** arises from the limited computational and memory resources available on earbuds and their paired mobile phones. Such constraints requires models that are both lightweight and energy-efficient while maintaining low latency for GRF estimation.

In Section 4, we will introduce our system, specifically designed to address these three challenges, that can generalise across activities, disentangle noisy head motion, and operate efficiently under hardware limitations.

3 Related Work

3.1 Estimating Ground Reaction Forces

Ground reaction forces can only be accurately measured using force plates, which can be embedded in the floor, a walk way or a treadmill which are confined to laboratory settings. Works to estimate ground reaction forces are dominated by wearable systems and are split by activity. Cameras have also been used to estimate GRFs in both walking [16] and jumping activities [33]. However, camera based systems lack the real world flexibility and cause privacy concerns. Therefore, some approaches utilise wearables with IMU sensors to estimate GRF. These sensors are typically located on the lower leg or lower back near the centre of mass of a human. For walking, studies have used a single device [26] as well as multiple devices [20, 32, 57] for GRF estimation. During running, works also use such devices for GRF estimation [2]. There are also works estimating various jumping [28, 31, 33, 39] and landing forces [51]. As well as inertial sensing wearables, insoles instrumented with pressure sensors have been used for GRF estimation [14, 24, 27], by fusing inertial sensing with the pressure sensing [9]. While these methods can show good results for estimation accuracy, they all rely on specialist devices that are not commonly owned by consumers. Smart watches are a commodity wearable form factor, however, there is no research on vGRF curve reconstruction except for one work that estimates peak force from wrist IMU [17]. This lack of full vGRF reconstruction from the wrist may be due to noise in the wrist IMU signal that comes with typical gait patterns and swinging of the arms [30].

3.2 Earables for Assessing Gait and GRF

Earables are an emerging wearable form-factor for sensing to complement their entertainment functions with a number of recent works focusing on human activity recognition and gait [22, 43]. Burgos et al. [8] used the IMU to classify between walking and running with high accuracy, showing there are clear differences at the head between the two gait styles. Additionally, a small number of works have used an earable device to estimate ground reaction force information. Attalah et al. explore correlations between peak GRF and the peak earable IMU signal with a linear regression [4], which shows the promise of using motion and vibrations captured at the head to infer GRF information. However, the simple model and single feature limit regression accuracy, with R^2 parameters reported between 0.3-0.5. While a user is running, Hu et al. [21] use in-ear audio to classify the heel strike type of the user, showing again that the head captures information about the ground impact. But this is limited to a 3-class classification rather than a more complex regression task. Similarly, EarWalk [25] assesses the foot posture in a neutral, toe in or toe out position using an ear worn IMU, showing again a 3-class problem that relates the head to ground impacts. These works show that the head can be used to infer gait patterns and information from the foot ground contact, but all of them estimate coarse-grained information about the ground impact in constrained scenarios and for single activity only.

WalkEar [50] is the first to show that vertical ground reaction force (vGRF) curves, along with key measures like peak force, can be estimated from earable IMUs during walking. However, the approach is limited in practice: it focuses only on walking and depends on custom signal processing that first extracts gait parameters and then estimates vGRF within walking-specific segments. Motion2Press [40] uses IMU sensors on the body, including a head mounted IMU, to estimate pressure distribution from instrumented insoles using a neural network. This work also estimates GRF through integration of the pressure data. However, literature [9] shows that GRFs from

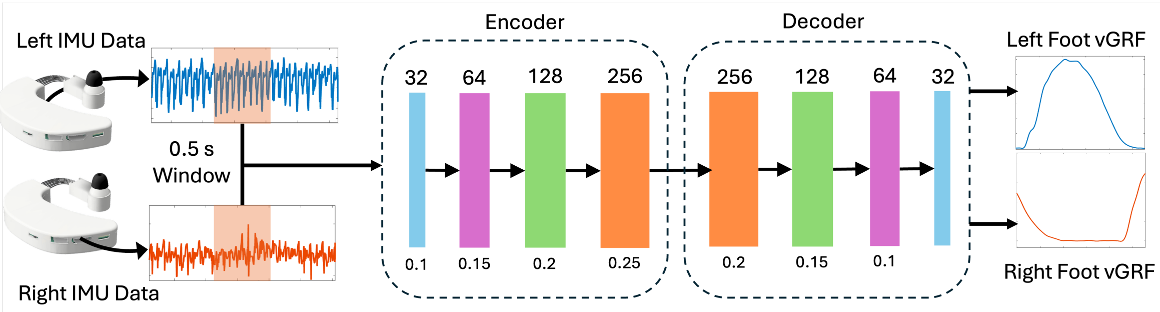


Fig. 4. ImpactEar system architecture showing input windowed IMU data, the Autoencoder including layer size and dropout rate, then the two channels of reconstructed vGRF curves.

insole data is not accurate and remains a challenging problem in biomechanics, with errors ranging from 0.8-8.8%. Additionally, Motion2Press [40] presents some discontinuities in the ground truth insole GRF, which are not possible in human locomotion.

To summarise, related works focus on tackling individual activities that a model is aware of, or only validating on single activity. We see a research gap where a ubiquitous consumer device can be used across scenarios, essentially establishing a force plate at the head building a complete picture of loading on the body continuously across different activities.

3.3 GRF Datasets

Datasets containing GRF data allow detailed studies on gait and movement. There are some publicly available datasets that include GRF information for the study of gait, for example GaitRec [19] with extensive GRF and clinical labels but not including wearable data. There are also numerous datasets containing IMU information for walking [34], running [15] and general human activity recognition [13] but these do not contain GRF information. A dataset from Carter et al. [10] collects running data from 50 participants and contains GRFs and 6 IMU locations, however, these don't include the head. Other research creates synthetic IMU data from camera recordings [47] or simulations [49], however, these approaches may lack the dynamics of a real-life wearable IMU. This review shows a gap in publicly available datasets that contain IMU, specifically from the head, along with GRF data. The contribution of which could be used for novel and reproducible research in earable estimation of human kinetics.

4 ImpactEar

4.1 System Overview

The goal of ImpactEar is to continuously estimate the GRF of each foot using only bilateral earable IMU data. Figure 4 shows the end-to-end architecture of ImpactEar. The system takes synchronised 6-axis IMU signals from the left and right earbuds as input and reconstructs the corresponding left and right vGRF curves in real time. Each input sample corresponds to a 0.5 s sliding window of IMU data containing both accelerometer and gyroscope channels. The model follows an encoder–decoder design that maps short-term head kinematics to continuous vGRF curves. This symmetric design allows the network to learn both global patterns of impact propagation and fine-grained differences between the two sides of the body.

4.2 Data Pre-processing.

We collected earable IMU and GRF data from multiple locomotion experiments, including walking, running, and jumping sessions, which will be described in detail in Section 5.1. The IMU data is streamed from the earable devices at 50 Hz. All IMU channels are then low-pass filtered to the same cutoff frequency as the ground-truth GRF signals. The filtered data from the left and right earbuds, each containing six IMU channels (three-axis accelerometer and three-axis gyroscope), are concatenated into a tensor of size 1×12, allowing convolution across both device axes and sensing modalities. The input is then segmented into 0.5 s windows with a 0.25 s overlap, resulting in an input tensor of size $25 \times 1 \times 12$ for each sample. The 0.5 s window length was selected because most force-plate events (e.g., two steps during running) occur within this duration, and it keeps the model compact for real-time inference.

Due to the limited number of samples for some high-intensity activities (e.g., jumping), a larger overlap of 0.45 s was used to augment these samples and balance the activity distribution. This trade-off allows the model to observe sufficient temporal context while maintaining a small parameter count. To enable a single model to handle multiple activities with distinct force magnitudes and durations, all GRF signals are normalised to body weight (BW) and with the 0.5 s window length chosen to enable a large enough portion of the activity to be seen for context recognition while still keeping model parameters low. This normalisation step standardises the representation across walking, running, and jumping, ensuring that the model learns activity-agnostic mappings rather than activity-specific amplitude scales.

4.3 Model Design

ImpactEar's model is designed to address the two remaining challenges: complex head dynamics and real-time deployment constraints. To capture the full range of head motion, the network jointly processes accelerometer and gyroscope signals from both earbuds, allowing it to distinguish translational motion caused by ground impacts from rotational motion arising from head tilts or neck movements. To enable real-time inference on resource-limited mobile devices, ImpactEar adopts a lightweight encoder–decoder architecture composed of compact dense layers, achieving low latency while maintaining accuracy. Benefiting from the diverse multi-activity dataset used for training, the model generalises across walking, running, and jumping, learning an across-activity mapping between head kinematics and ground reaction forces.

Model Architecture. The ImpactEar model is based on an encoder-decoder architecture displayed in Fig. 4. We adopted a lightweight model architecture for real-time GRF estimation. The model employs a 2D convolutional autoencoder architecture tailored for time-series data, where the input is a multichannel 2D signal representing windowed acceleration data from the ear-worn IMUs, and the output is the corresponding vertical GRF curves over time. The encoder consists of four sequential 2D convolutional layers with progressively increasing channel depths (32, 64, 128, 256). Each convolution is followed by batch normalisation, ReLU activation, and dropout for regularisation. Kernel sizes are set to 11, 5, and 3, with padding to preserve temporal resolution. Dropout rates increase gradually from 0.1 to 0.25 across layers to reduce overfitting. The decoder mirrors the encoder with transposed convolutions, halving the feature channels at each step and reconstructing the original temporal resolution. It ends with a final transposed convolution that outputs 2 channels GRF, one for the left and one for the right foot.

Training and Postprocessing. The model is trained to minimise the L1 loss between predicted and ground truth GRF curves:

$$\mathcal{L}_{\text{train}} = \|\hat{\mathbf{F}} - \mathbf{F}\|_1$$

where $\hat{\mathbf{F}}$ is the predicted force curve and $\mathbf{F}$ is the ground truth from the force plate or treadmill. The model was optimised with the Adam optimiser using batch sizes of 16 and trained for 200 epochs. The model predictions are then brought together into a timeseries by combining each 0.5 s window and taking into account the overlap in prediction windows which are sample-wise averaged. This reconstructed timeseries then has a Butterworth 5th order lowpass filter with a cutoff frequency of 10 Hz applied to smooth any edge effects of combining windows from separate model window inferences.

Evaluation metric. The normalised root mean square error (NRMSE) (sometimes called rRMSE) are commonly used in the literature for GRF predictions so it also allows comparison to related works. We used this as the primary validation metric to assess prediction quality and the formula for evaluation NRMSE is given here:

$$\text{NRMSE} = \frac{\text{RMSE}}{\text{Range}(\mathbf{F})} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mathbf{F}}_i - \mathbf{F}_i\right)^2}}{\max(\mathbf{F}) - \min(\mathbf{F})}$$

Overall, this architecture is lightweight, robust to overfitting due to layered dropout, and able to generalise well across multiple locomotion tasks including walking, running, and jumping.

5 Evaluation

5.1 ImpactEarDS Dataset

To obtain sufficient data for model training and testing across different activities, we collected ImpactEarDS from 30 participants. ImpactEarDS is the first publicly available dataset to contain a pair of earable IMUs along with gold standard 3D Ground Reaction Force, Moment and Centre of Pressure measures taken from force plates and an instrumented treadmill. ImpactEarDS is downloadable from Kaggle[1], with an MIT license. This study had ethical approval from the institution's ethics committee and all participants gave written informed consent. Ethical considerations included not causing the participants too much exertion during running experiments, hence participants were asked to self select running speeds and were able to test them. All participants in the study were healthy with no diagnosed gait impairments. The demographics and anthropometrics of the participants are reported in Table 1. Participants were excluded if they reported any musculoskeletal injury or lower-limb surgery within the preceding six months, any neurological condition affecting gait, chronic pain or joint disorder (e.g. osteoarthritis) that could alter locomotion patterns, use of assistive devices for mobility, or inability to perform the required activities (walking, running, stair ascent/descent) without discomfort.

Figure 5 illustrates the experimental setup. Each participant walked and ran on a *Bertec* instrumented treadmill [5] which recorded the GRF data sampled at 1000 Hz with the setup shown in Figure 5.a. Participants also wore the open source OpenEarable V1.4 [45] (Figure 5.d), which contained a Bosch BMX160 9-axis IMU sampled at 50 Hz. The device weighed 15 g overall and was secured with a headband to prevent it from being lost in the treadmill. The footwear of the participants was not controlled but they were advised to wear shoes comfortable for running; participants arriving in flat shoes that would cause pain in running were given trainers to wear. The data was synchronised using multiple firm foot stamps at the start of each recording. Each participant completed 2 walks, at 3 kph and 5 kph of 4 minutes each followed by 3 runs of 4 minutes each, every participant ran at 10 kph but the other two speeds were self selected, for example, 9, 10, 11 kph. An exact breakdown of running speeds for each participant can be seen in Table 2 where results are given in the *Treadmill - Run* columns only for speeds ran by that participant. The participants also performed over ground walking and running validations in a room instrumented with 4 floor mounted force plates produced by *Kistler Group* [29] sampled at 1000 Hz. This

[1]Dataset Repository: https://kaggle.com/datasets/c35c07d6389fa011e6fa22f8db64cbb89854f83df260fb17e857c6799afe3848

Table 1. Details of participant demographics and anthropometrics (mean ± standard deviation).

Metric	Age	Weight (kg)	Height (cm)	BMI	Foot length (cm)
Male (18)	25.8 ± 6.9	79.2 ± 13.6	181.4 ± 7.9	24.0 ± 3.7	27.7 ± 1.2
Female (12)	31.8 ± 11.3	65.7 ± 7.3	170.1 ± 4.8	22.7 ± 2.5	24.6 ± 1.2
Total (30)	28.2 ± 9.2	73.8 ± 13.2	176.2 ± 8.8	23.5 ± 3.3	26.4 ± 1.9

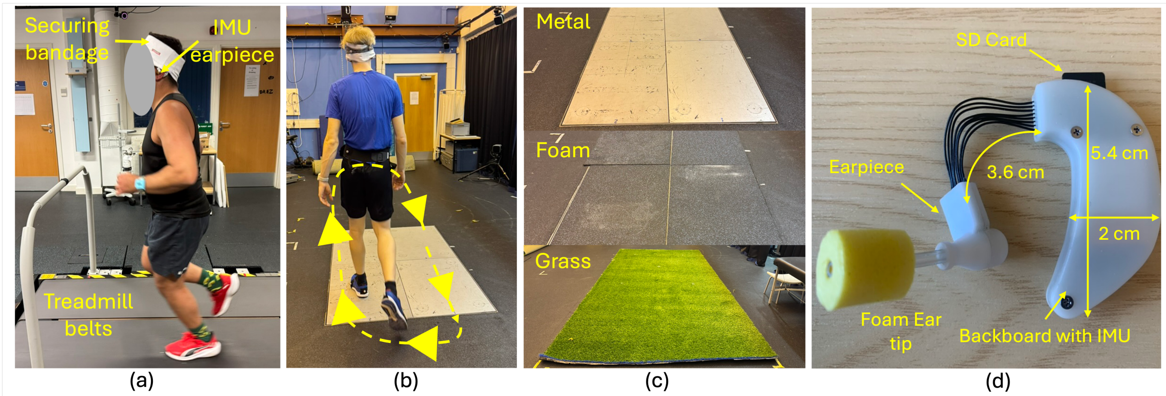


Fig. 5. (a) Photographs of a participant on the instrumented treadmill running, (b) a participant on the force plates while walking with the path annotated, (c) the different surface types used on the force plates, (d) and the OpenEarable 1.4 device used for data collection.

included experiments where every participant would walk and run laps on the force plates at self selected paces with only the rough path assigned as shown in Figure 5.b.

All of the participants also undertook a series of jumping exercises on the force plates to collect a wide range of impacts. The jumping protocol is as follows:

- Five counter-movement jumps (CMJ).
- Five CMJs with swinging arms.
- Six front jumps with swinging arms.
- Five standing single leg jumps (both for left and right legs).
- 60 seconds exertion protocol of CMJs.

This protocol was adapted from [6] with guidelines on measurement of mechanical power and athlete performance during jumps. The jumps were not timed and participants were allowed breaks between individual jumps and between sets of jumping activities due to the strenuous nature of the exercises which created variability in the length of the jumping activity across the participants. The time gaps, while varying, are accounted for by the thresholding method used on the force plates.

Figure 6 summarises the amount of data in ImpactEarDS shown in blue bars, totalling over 6.5 hours. All ground truth data from the force plates and instrumented treadmill was re-processed into left and right foot GRF rather than the original format from different numbers of force plates, this allows the data to be fully standardised in its most useful format and streamline model training. Figure 6 the amount of data seen by ImpactEar which is higher due to upsampling. This is achieved by having window overlap which is 0.25 s for treadmill activities (those given with speeds in the figure) and 0.45 s for overground activities (those appended with FP and jumps). This allows the sparser samples taken from the force plates to be balanced with the high sample count treadmill activities.

5.2 Overall Performance

We evaluated ImpactEar in a Leave-One-Subject-Out (LOSO) scenario, training the autoencoder on 29 users and testing on the remaining one, this was repeated across all 30 participants and calculating Normalised Root Mean Square Error (NRMSE) in each case. Figure 7 presents the results of this evaluation by showing the mean NRMSE across all users in walking, running and for the protocols executed on the force plates. The NRMSE is 8.61%

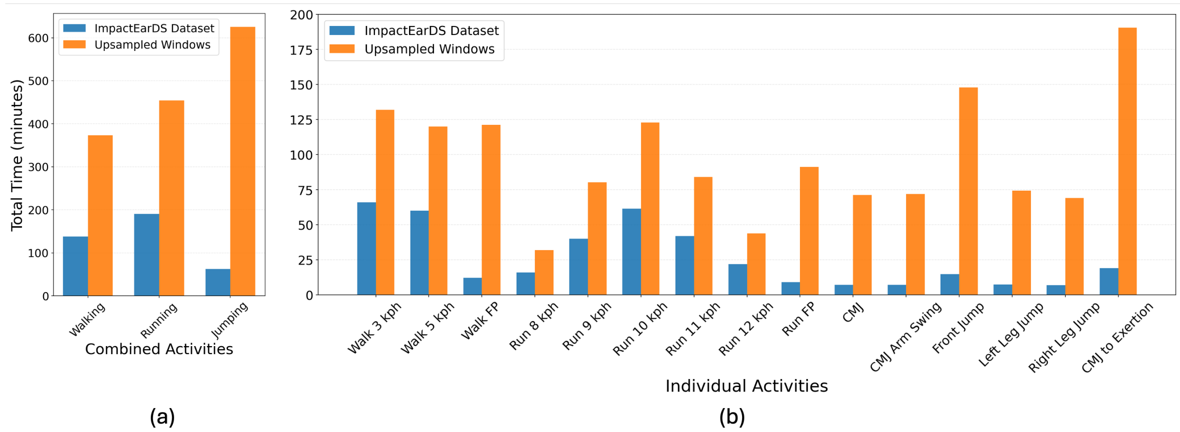


Fig. 6. Overall ImpactEarDS data quantity in minutes for for each activity group (a) and each sub-activity class (b). Raw dataset quantity is given in blue and the amount of data seen by ImpactEar is given in orange. FP denotes activities done in the over ground setting on force plates.

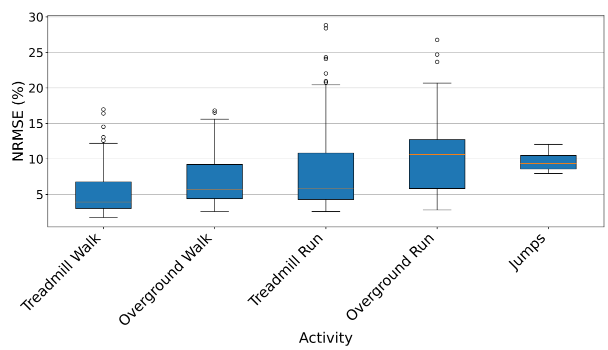
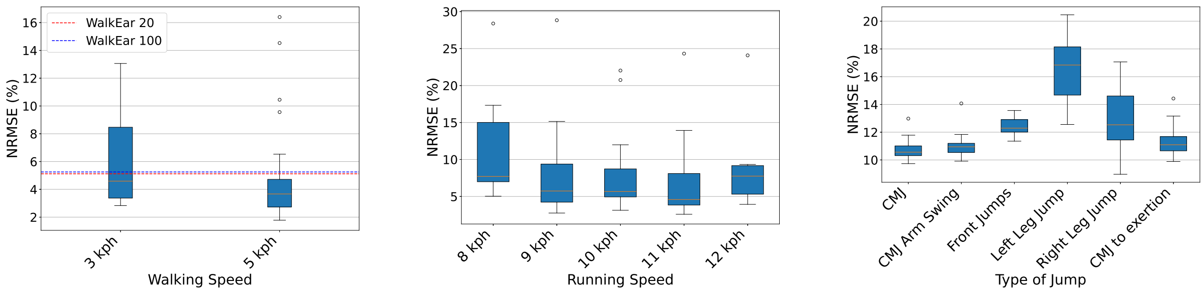


Fig. 7. Boxplot of overall performance given in NRMSE for comparison across the five activities tested.

on average across activities, with a minimum of 5.2% for 5 kph walking. This could be due to the regularity in walking as participants were not tiring thus changing their gait like in the running experiments. In addition, the system performs well across multiple protocols, with an average NRMSE of 6.34% for walking at different speeds and in the two conditions (treadmill and force plates), 9.58% for running on treadmill and force plates, and a 9.58% for jumping.

5.3 Impact of Different Users

Table 2 breaks down the performance for each participant on each activity they participated in. It shows that the NRMSE variation across all the participants ranges from 4.54% to 16.00% with an overall mean of 8.61% and standard deviation of 3.91% across all activities. The range of errors is fairly large across participants and may be explained by individual differences in gait and jumping as well as different amounts of head motion among participants. To further explain this we can look at variations in participants within each activity. Some participants experience large errors in some activities then lower errors in a related activity, for example P03 had a NRMSE of 28.83% in the 9 kph run but a NRMSE of near 10% in the other two runs, this could have been caused by a loose device creating noise in the data.



(a) (b) (c)

Fig. 8. NRMSE breakdown within each activity performed at different speeds or protocols. Specifically, (a) treadmill walking speeds annotated with WalkEar [50] earable IMU related work performance for the two models presented, (b) treadmill running speeds and (c) the different jump types.

For walking it can be seen in the *Treadmill - Walk* columns that the slower walk had slightly higher errors than the faster walk. We assume that this is caused by having weaker GRFs while walking more slowly, as less force is required to propel oneself forwards which would lead to a more difficult prediction as the GRFs have a smaller range and there is less vibration transmitted to the head. Similarly for running, the slower speeds show higher errors, seen in Figure 8(b), across participants which is assumed to also be caused by weaker forces at lower speeds. While examining the *Treadmill - Run* columns in the table, it can be seen that the participants who opted for slower running speeds consistently had higher overall errors. For jumping the range of errors is smaller being from 8.00% to 12.05% with a smaller standard deviation of 1.13%. The lower error range and deviation may be due to the fact that all the jumps followed the same demonstrated protocol. A participant with a strong athletic background produced a peak jumping force of 1.9 BW compared to the dataset average of 1.5 BW, this shows the model has adapted well to a wide range of peak jump forces as, despite having the highest relative jumping forces, this participant had a NRMSE of 8.55% which was below the mean error.

Effect of participant demographics: To assess whether participant demographics affected vGRF reconstruction accuracy, we computed Spearman rank correlations between per-participant NRMSE and each continuous demographic variable (height, weight, and shoe size), and compared NRMSE between male and female participants using a Mann-Whitney U test. These non-parametric tests were chosen as they do not assume normality and are robust to the modest sample size (N = 30). No significant associations were found at the $\alpha = 0.05$ level (height: $\rho = 0.14$, $p = 0.46$; weight: $\rho = -0.09$, $p = 0.64$; shoe size: $\rho = 0.18$, $p = 0.34$; gender: $U = 89$, $p = 0.52$, male median NRMSE = 8.4%, IQR = 6.1−10.7%, female median NRMSE = 8.9%, IQR = 5.6−9.8%). Given the sample size, these tests are powered to detect only moderate-to-strong effects ($|\rho| \geq 0.36$), and weaker associations may emerge with a larger cohort.

5.4 Impact of Different Speeds

In Section 5.2 we presented overall results for grouped activities, in this section we break down performance *within* each activity by looking at the effect of walking and running speed on the NRMSE. The distribution of each speed within the dataset and within model training can be seen in Figure 6(b). Figure 8(a) and (b) shows the performance of ImpactEar for different walking speeds, running speeds respectively. It can be seen that slower speeds typically result in worse performance. This is expected as slower movements typically produce weaker forces as well as vibrations and motion transmitted to the head, giving a reduced estimation performance. Additionally, performance worsen only for specific participants and specific protocols, for which the availability of data is reduced. For example, running at 8 kph was performed only by 8 users and running at 12 kph by 10, reducing the amount of data for training. For the running experiments, due to the pre-processing of the data,

Table 2. Details of per participant and per activity NRMSE. For each participant, the last three columns show the mean error across the same type of activity (walk, run and jumps). The grouped mean values show the mean across the activities both on the treadmill and overground. The last column presents the mean values for each participant to allow comparison between the 30 participants. The final row shows the mean for each protocol.

	Treadmill - Walk		Treadmill - Run					Force Plates			Grouped Mean Values			Participant
Participant	3 kph	5 kph	8 kph	9 kph	10 kph	11 kph	12 kph	Walk	Run	Jumps	Walk	Run	Jumps	Mean
P00	13.41	16.97	-	19.58	13.89	13.45	-	11.44	27.99	11.26	13.94	18.73	11.26	**16.00**
P01	3.42	1.84	-	-	7.28	3.87	-	2.83	4.01	8.55	2.70	5.05	8.55	**4.54**
P02	5.50	2.77	-	3.49	5.31	13.93	-	6.55	10.02	11.34	4.94	8.19	11.34	**7.36**
P03	6.84	4.54	8.09	28.83	10.96	-	-	16.52	11.80	8.10	9.30	14.92	8.10	**11.96**
P04	12.57	4.43	-	14.56	11.99	7.49	-	8.47	9.16	10.43	8.49	10.80	10.43	**9.89**
P05	3.37	16.40	-	-	22.03	3.81	4.18	15.60	20.68	10.62	11.79	12.67	10.62	**12.08**
P06	4.59	4.72	-	-	5.12	7.83	5.60	5.12	2.81	9.30	4.81	5.34	9.30	**5.64**
P07	2.98	2.06	-	3.21	3.13	3.64	-	4.43	13.96	8.00	3.16	5.99	8.00	**5.18**
P08	12.20	9.56	-	-	10.46	24.32	9.01	12.48	23.66	9.38	11.41	16.86	9.38	**13.88**
P09	3.41	2.40	-	15.15	4.56	2.87	-	2.64	2.91	9.31	2.82	6.37	9.31	**5.41**
P10	7.18	3.67	-	-	8.15	8.37	9.33	3.76	10.12	10.68	4.87	8.99	10.68	**7.66**
P11	3.21	1.80	-	5.74	3.91	3.20	-	8.95	12.29	10.82	4.65	6.28	10.82	**6.24**
P12	2.98	2.94	-	4.23	3.56	4.47	-	6.56	5.73	9.03	4.16	4.50	9.03	**4.94**
P13	2.98	2.18	-	-	6.08	4.34	24.08	12.43	26.76	12.05	5.86	15.32	12.05	**11.36**
P14	5.88	3.79	-	-	11.77	11.12	5.02	4.66	4.70	9.56	4.78	8.15	9.56	**7.06**
P15	3.01	3.53	-	-	3.22	2.59	7.75	4.74	6.89	10.63	3.76	5.11	10.63	**5.29**
P16	8.47	6.53	-	4.29	4.70	8.53	-	3.23	3.95	9.10	6.08	5.37	9.10	**6.10**
P17	3.42	3.10	5.03	4.75	5.46	-	-	9.04	11.00	8.59	5.19	6.56	8.59	**6.30**
P18	3.21	1.78	-	-	3.25	4.59	3.93	4.32	11.75	8.53	3.10	5.88	8.53	**5.17**
P19	3.90	2.73	-	4.32	20.75	3.98	-	4.89	11.03	9.52	3.84	10.02	9.52	**7.64**
P20	10.04	3.59	-	-	5.90	5.03	3.79	9.76	18.04	9.50	7.80	8.19	9.50	**8.21**
P21	3.86	2.23	-	2.85	6.44	5.23	-	6.05	6.62	8.19	4.05	5.28	8.19	**5.18**
P22	2.83	3.65	-	-	5.02	3.54	8.27	5.16	11.75	7.96	3.88	7.14	7.96	**6.02**
P23	10.65	14.53	7.36	9.38	5.35	-	-	4.15	10.26	8.23	9.78	8.09	8.23	**8.74**
P24	-	3.98	-	2.76	5.18	-	-	5.43	5.68	9.90	4.70	4.54	9.90	**5.49**
P25	4.07	6.22	28.39	7.76	5.10	-	-	16.81	24.69	11.88	9.03	16.48	11.88	**13.11**
P26	12.20	3.96	6.87	6.53	5.88	-	-	6.71	5.88	9.48	7.63	6.29	9.48	**7.19**
P27	6.16	4.30	8.06	6.29	6.46	-	-	2.62	12.30	9.07	4.36	8.28	9.07	**6.91**
P28	13.06	10.45	17.33	14.16	11.35	-	-	11.91	18.25	9.09	11.81	15.28	9.09	**13.20**
P29	9.38	5.48	20.95	20.43	16.70	-	-	6.27	25.52	9.45	7.04	20.90	9.45	**14.27**
Mean	**6.37**	**5.20**	**12.76**	**9.38**	**7.96**	**6.96**	**8.10**	**7.45**	**12.34**	**9.58**	**6.34**	**9.58**	**9.58**	**8.61**

lower speeds might have affected the quality of the ground truth. This is because slower speeds reduce the flight time or air time between running steps, this increases the possibility that the user touched the treadmill with both feet simultaneously. Due to the experimental setup the running protocol can only have one foot in contact with the treadmill at a time or the ground truth data can be compromised. So effect could contribute to worse results at slower running speeds.

5.5 Impact of Different Jump Types

To assess the impact of each jump type we break down the performance for each type of jump. The jumps included in the protocol were five counter-movement jumps (CMJ), five CMJ with arm swings, six front jumps with both legs, five left leg only jumps followed by five right leg only jumps and finally CMJ until exertion or upto 60 s. As well as these there were also significant portions of the data where the user was stationary during the jumping session while tired from the CMJs or changing jump types. The distribution of each jump type within the dataset and within model training can be seen in Figure 6(b). The results per jump are given in Figure 8(c) as a boxplot. The NRMSE results in this figure are higher than the jumping presented in other parts of the paper as these exclude all stationary parts. It can be seen that for each CMJ type the performance is better, we expect this is due to the relative amount of CMJ jump data in the protocol and hence in the training data. The front jumps also show a similar but slightly worse performance, this could be due to them being similar to the CMJ but the participant lands a distance infront of their starting position rather than the same place which makes the jump less stable and have different dynamics to the CMJ. The single leg jumps have the worse performance, this could be partly due to the difficulty of performing the jumps which led to participants being very unstable creating unique force profiles that the model had very little samples for. These jumps also have added complexity in only involving a single leg so the dual foot GRF output should only have data from a single channel and zeros on the other channel.

5.6 Baseline and State of the Art Comparison

To put the NRMSE results into further context we evaluate two simple baselines then compare the results to state-of-the-art (SOTA) literature. The baselines consist of a prediction purely on the mean ground truth value for each activity which establishes a worse case scenario for model performance. The second baseline is a linear regression model that takes resultant IMU data from a 25 sample window as an input and predicts the next GRF sample. The third baseline is a deep learning recurrent LSTM model. This baseline was taken from recent literature predicting vGRF from insole data, Carter et al. [9], and was adapted for use on the 12 IMU channels coming from both earbuds. In detail, the model has a bidirectional LSTM layer followed by three linear layers to make the vGRF prediction. Hyperparameters including layer size and training parameters were taken from the related work as they were optimised there.

Table 3 shows the results for the baselines and SOTA comparisons showing that ImpactEar strongly outperforms the Ground truth and Linear regression baselines showing that it learns the GRF morphologies. The first two baselines are trained per-activity and even with this prior knowledge ImpactEar, which is trained across activities, still strongly outperforms them.

For the LSTM baseline we see similar performance for both walking and running and the LSTM under performing ImpactEar on the jumping activities. The LSTM model (~2.2M) has around an order of magnitude more parameters than the CNN architecture used in ImpactEar (~270k), this may have over fitted on the data and struggled to reconstruct the more varied jumps compared to the more repetitive walking and running vGRF waveforms. Additionally, a drawback of a recurrent based architecture is related to its deployability to a microcontroller. The primary drawback is latency due to recurrent models needing series computations to calculate the LSTM state vector compared to parallelisable computation of a convolutional architecture [41]. Finally, a recent study found an order of magnitude increase in latency of an LSTM model vs a CNN model for on-device experiments [46].

To describe the SOTA comparisons in more detail, the only work that uses earables that we can compare ImpactEar to is WalkEar [50], which evaluates earables for GRF prediction in a walking scenario only. Figure 8.a includes a blue and red horizontal lines representing the two models proposed by WalkEar, it can be seen both lines lie above the mean NRMSE for ImpactEar showing that it outperforms this related work on the walking activity whilst also predicting GRF in other scenarios. This result can also be seen in Table 3.

For other comparisons we look at non-earable devices including sacrum, lower leg and insole based systems. Shahabpoor [48] estimate waking vGRF using different IMU locations at the chest, sacrum and foot achieving NRMSE between 4-8%. This shows ImpactEar is within these bounds for sensors at more favourable body locations. For running two works are assessed, the first, by Alcantara et al. [2] uses both sacrum and shoe mounted IMU achieve an average NRMSE of 6.4% in running tasks. The second by Carter et al. uses a more advanced fusion of pressure sensing insoles and foot mounted IMUs, achieved an NRMSE average of 3.2% varying from 0.8-8.8% across 50 participants on treadmill running vGRF reconstruction. This compares to ImpactEar having NRMSE average of 8.44% and varying from 2.59%-28.39%, although as seen in Table 2 (Section 5.3) the errors above 20% were outliers for each participant compared to their other runs. While ImpactEar performance is worse than the smart insole based methods, these systems are not commonly owned by consumers. No jumping GRF

Table 3. Baseline and Related Work performances compared to ImpactEar. The walking and running are given for treadmill data only for better comparison with other methods.

	Method	IMU location/s	Activity (NRMSE %) Walk	Run	Jump
Baselines	Ground Truth Mean	Head	38.3	42.6	40.1
	Linear Regression	Head	27.4	25.3	30.1
	LSTM model	Head	6.1	8.3	12.1
Related Works	WalkEar [50]	Head	5.1	-	-
	Shahabpoor et al. [48]	Chest, Sacrum & Foot	4-8	-	-
	Alcantra et al. [2]	Sacrum & Foot	-	6.4	-
	Carter et al. [9]	Foot IMU & Pressure Insoles	-	3.2	-
Ours	**ImpactEar**	Head	**5.8**	**8.5**	**9.6**

curve predictions could be found for comparison but further SOTA comparisons are given in Section 5.9.1 and Section 5.9.2 which discuss downstream applications of ImpactEar to gait metric and jumping metric estimations.

5.7 Power and Latency

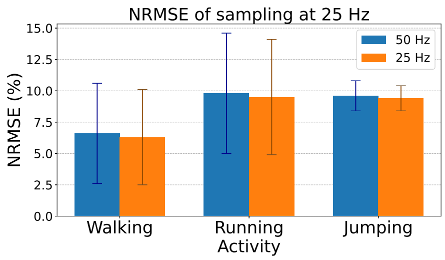
ImpactEar was designed to be lightweight and stream data from the earbuds to the paired mobile phone. To evaluate this we developed an Android app to perform model inference and stream IMU data from the OpenEarable v1.4 via Bluetooth 4. The model was transferred to the PyTorch Lite framework to run on a mobile phone. To evaluate the performance we installed the app on a Google Pixel 9 phone and assessed the average inference latency and power consumption. The experiments showed a latency of **9.5 ms** giving fast real-time performance on 0.5 s windows including the 0.1 s overlap. The power analysis showed that **14 μW** were consumed from the combination of the radio and CPU and memory while running ImpactEar. This measurement excluded the screen as ImpactEar can operate in the background with the screen off. The setup details and code for the ImpactEar mobile implementation will be made available to the community.

Further details on the ImpactEar model relevant to on-device deployment include a parameter count of 270k, a model parameter memory of 1.03 MB; and estimated FLOPs of 4.28E8 using a 50 sample window. If quantised with an INT8 scheme, as is commonly used for on device machine learning [18] leading to only small to moderate losses in performance, then the memory footprint would be reduced to 0.26 MB. This could enable real-time on device inference on an OpenEarable v2.0 device[44] for example, which has 1 MB of flash memory so would be able to hold the model, inference and sensor data handling using the quantised version of ImpactEar.

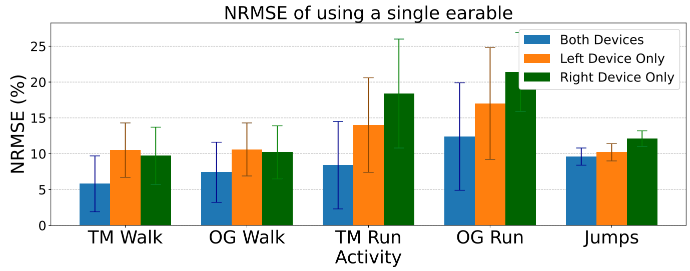
5.8 Real-life Considerations

In order to assess the robustness of ImpactEar we undertook a variety of experiments to simulate real-life challenges that may occur outside of the laboratory. This section details the results of each of these experiments.

5.8.1 Impact of Ground Surfaces: The results presented so far are from data collected on the standard surface of the treadmill belt and metal force plates. While these surfaces can represent some real-world scenarios, they are both relatively hard surfaces. To evaluate the system on other surfaces representative of real-life scenarios, two further surfaces of artificial grass and foam pads, shown in Figure 5(c), were tested in the overground walk, run and jumping activities. Table 4 shows the results in this experiment. It can be seen that the softer surfaces perform worse than the hard metal surface. This may be due to a softer surface creating a weaker vibration



(a)



(b)

Fig. 9. Barcharts recording results of experiments on different potential real-world conditions applied to ImpactEar with error bars corresponding to standard deviation of the NRMSE across participants. All activities on both the treadmill (TM) and over ground (OG) are listed. (a) Comparing performance using a sample rate of 25 Hz, which is the same as the Apple AirPods Pro, to the dataset sampling rate of 50 Hz. (b) Comparing the performance when losing either the left or right earable device compared to having both devices as a baseline.

to the head as the impacts are dampened. Additionally, the softer surfaces were only collected on a subset of participants and activities, which reduces the amount of data from softer surfaces seen by the model.

5.8.2 Lower IMU Sampling Rate: To better assess the feasibility of running our framework on a consumer device with a known IMU streaming rate (for example Apple Airpods Pro [3] has a streaming rate of 25 Hz), we assess the performance of ImpactEar at this lower sample rate, and compare these to the ones when run on the the dataset's original sampling rate of 50 Hz. To reduce the dataset's sampling rate, every other sample was discarded. The models were then retrained on this data in the same LOSO scenario as the main evaluation. Figure 9(a) shows the change in NRMSE across activities for this experiment. From Figure 9(a), it can be seen that there is a small decrease in NRMSE from moving to 25 Hz across experiments. Upon inspection, we found that the simpler waveform at 25 Hz allows for an easier vGRF reconstruction task because the ground truth signals are filtered at 15 Hz to eliminate vibrations of the treadmill (a common practice in the field [9]). However, the reduction is marginal and we still choose 50 Hz as it better supports downstream tasks in Section 5.9.

5.8.3 Impact of a Single Earable: While ImpactEar uses a pair of earables, it can be a relatively common scenario where some people like to use single earbud or one earable device becomes unavailable due to battery constraints or communication problems. To assess ImpactEar in this case, we evaluate the model with only one IMU input stream with the other devices data set to zero to mimic an outage or a single sided setup. Figure 9(b) compares the GRF NRMSE results for single and dual device setups. It can be seen that, as expected, in each case having both devices is better than only a single device but that the performance is still close in the walk settings. It can also be seen that the right device only setting shows worse performance than the left device only setting on the treadmill run and overground run. This is an artifact of our modeling framework, due to the model fitting more strongly to features from the left channel that is given first in the input array or a looser fitting right device making the IMU signals less coupled to the head.

Table 4. NRMSE for the activities performed on different ground surfaces.

	NRMSE (%)		
Surface	Walking	Running	Jumping
Metal	5.8	8.5	9.6
Foam	9.8	7.3	9.9
Grass	11.6	10.0	8.8

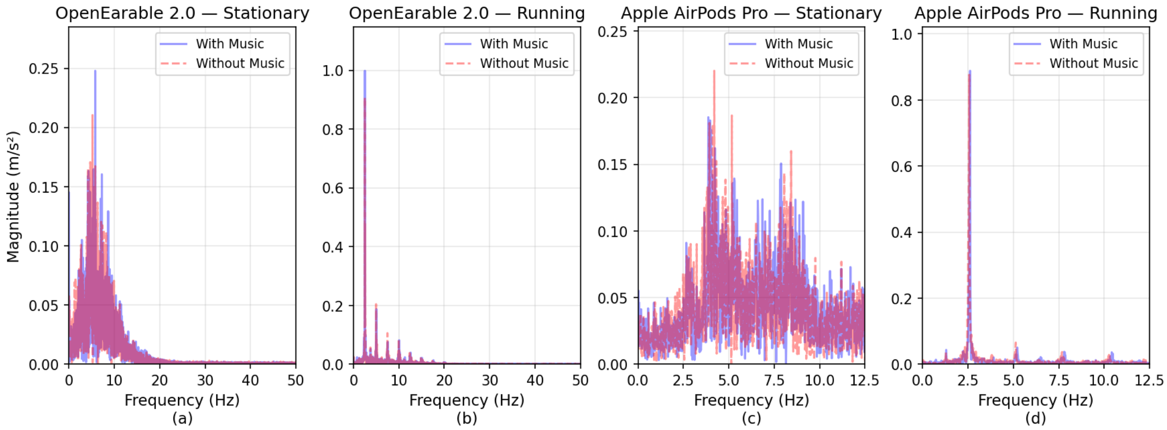


Fig. 10. Comparison of spectra for two different consumer-like earable devices with and without music playback. Note a difference in x-axis scale due to different sampling rates used on the two devices.

5.8.4 Music Playback. In this section we detail two experiments used to assess the impact of music playback on the system and potential future evolutions of the system implemented on different devices. The first is with the OpenEarable v1.4 which assesses the impact of music playback on our system with vGRF ground truth. In the second we use two devices more typical of consumer device earable form factors, i.e. the device is all in one package such as in the Apple Airpods. For these devices we compare the frequency spectrum of the IMU with and without music playback during stationary and running scenarios.

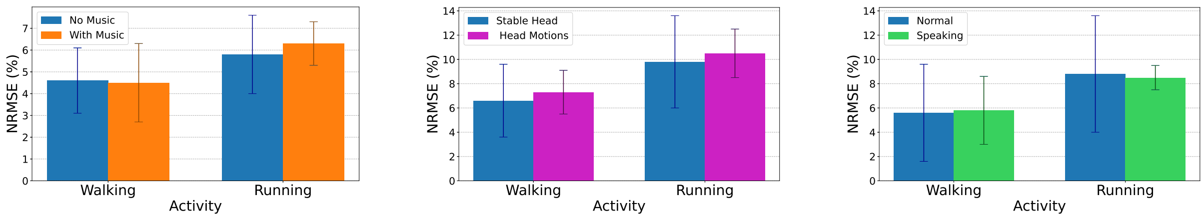
ImpactEar performance with music playback:

In this experiment we had participants perform walking and running on the treadmill with music playback on the OpenEarable v1.4 device. We then compare the accuracy of ImpactEar on vGRF reconstruction in both the music playback and silence conditions for that subset of participants. The results in Fig. 11(a) show that the music playback creates no significant changes in vGRF reconstruction accuracy. This is likely due to relatively small vibrations of the speaker compared to the movements involved with walking and running as well as the distance between the speaker and IMU sensor.

Music effects on IMU frequency spectrum on consumer form-factors:

To more fairly assess if the speaker affects the IMU when devices are packed in a single housing we use both the Apple Airpod Pro Gen 2 [3] and OpenEarable 2.0 [44] devices. To assess the impact of the speaker we play music through both devices in stationary and running scenarios and compare the frequency spectrum to the same scenarios with no music. For the stationary scenario the participant was asked to remain still while in the quiet and music playback scenario. The running scenario uses the same participant at the same speed to create as similar profiles in the running motion as possible between the quiet and music scenarios. The Airpods were sampled at 25 Hz and the OpenEarable was sampled at 100 Hz to try and show higher frequency impact of the music on the IMU.

The spectra from these experiments are shown in Fig. 10 where the IMU data has the resultant taken and is then normalised. It can be seen that for both devices in both scenarios the music has a limited effect on the spectrum, especially in the running scenario where strong peaks and harmonics can be seen caused by the running motion. In the stationary scenario there is more variation but this is unlikely to be significant under the much stronger accelerations present during motion. It is likely that the speaker has more impact at higher frequencies where the music content is playing, however, this isn't recorded by the IMU where we focus on the frequencies relevant for vGRF reconstruction (< 15 Hz).



(a) (b) (c)

Fig. 11. NRMSE performance of ImpactEar in different real world scenarios. Baseline results may differ because of the different inclusion of participants due to data availability so comparisons should be made within each of the figures. (a) The impact of music playback, (b) the impact of head motions and (c) the impact of facial motions caused by speech.

Overall these experiments show that music playback has negligible impact on ImpactEar or general consumer ear-worn IMU sensors.

5.8.5 Impact of Head Motions. To assess the impact of head motions we recorded motion capture data from the participants while they were doing the force plate activities. The markers used for head motion capture can be seen in Fig. 5 (b) present on the grey headband worn by the participant. To make this experiment as natural as possible participants were not instructed to keep their heads still or move them, simply to move naturally. This included some communication with the investigators as well as looking around the laboratory while undertaking the experiments. The experiment looks at the difference in vGRF accuracy between the head motion and off axis labelled windows vs the head still windows that make up the majority of the data.

The results of this experiment are visualised in Fig. 11(b) where it is shown that the head motion and head still conditions yield similar vGRF reconstruction accuracies giving evidence for ImpactEar being robust to head motions and the head being off axis. This may be due to having a pair of earable IMU sensors which can give clear information about head rotations as well as the gyroscope, which has been shown to distinguish between head rotations and gait signals clearly due to the relative difference in radius of rotation [50].

5.8.6 Impact of Facial Motions. To assess the impact of facial motions we assess the windows where the participants were known to be speaking. Speech creates facial movements and contortions that could be seen in the IMU sensor which may affect vGRF reconstruction accuracy, this experiment aims to quantify the effect of these facial movements on ImpactEar. To label the spoken parts we captured audio from the external microphone on the OpenEarable v1.4 during the original collection of ImpactEarDS. This audio data was then passed through a Voice Activity Detector (VAD) to label GRF windows with simultaneous speech and thus facial motions.

The results of this experiment are visualised in Fig. 11(c) where it can be seen that for the speaking and silent conditions vGRF reconstruction accuracy is similar. This gives evidence for the robustness of ImpactEar to facial motions. This may be due to the relatively small magnitude of facial motions compared to movements of the body that propagate to the head during walking or running.

5.9 Downstream Applications

To put the GRF data from ImpactEar further into context we examine some downstream applications. These include: deriving gait and jumping metrics, gait asymmetry analysis on a participant with significant asymmetry, running footstrike analysis and ballet dancing analysis. These are evaluated either against the available ground truth in a LOSO setting or qualitatively in the case of the ballet dancing and compared to relevant SOTA methods where possible.

5.9.1 Gait Metrics: While the GRF curve can already give useful insights, metrics derived from it can be derived to summarise the data. To assess the usefulness of ImpactEar for these scenarios, we calculated ground truth and estimated metrics from the output of ImpactEar. Specifically, the gait metrics assessed are the contact time (i.e., the time the foot is in contact with the ground), peak force (i.e., the maximum force in each gait cycle), and the impulse (i.e., the integral of the curve between two ground contacts). These metrics are illustrated on a vGRF curve in Figure 12 (a).

To ensure that ImpactEar retains the activity agnostic predictions, a simple threshold based algorithm calculates when a foot is not in contact with the floor, i.e., when the GRF value is zero, and uses this to segment the data into ground-foot contact regions. This method can then be applied across any walking and running. All the gait metrics are evaluated using the Mean Absolute Percentage Error (MAPE) to allow comparison with SOTA methods. The formula given below where X_i is a predicted value and $\hat{X}_i$ is a ground truth value for N total samples:

$$\text{MAPE} = 100\% * \frac{1}{N} \sum_{i=1}^{N} \frac{\left(\hat{\mathrm{X}}_i - \mathrm{X}_i\right)}{\hat{\mathrm{X}}_\mathrm{i}}$$

Table 5a shows the error given between the ground truth and ImpactEar estimations. It can be seen that all metrics have a MAPE lower than 10%. The MAPE values for running may be higher than those for walking as the gait profiles can exhibit more variability as participants tire from exertion making the GRFs and therefore the gait parameters more varied. The running contact time may be more accurate than the walking as it is easier to recognise full air time compared to recognising when a single foot leaves the ground like in walking, detecting this toe-off event is known to be challenging from wearable devices [23]. Additionally, for step counting the ImpactEar estimates had a 100% agreement with the ground truth although this was in a laboratory setting rather than in-the-wild.

To put these into context, a related work WalkEar [50] obtains errors of 2.2% and 0.73% for peak force and impulse estimation respectively while walking only and uses earable IMU data in a regression model. Another work by Alcantara et al. [2] uses a leg mounted sensor achieves MAPEs of 8.8%, 6.4% for peak force and impulse estimation respectively, only while a user is running. From these values it can be seen that in this downstream task, ImpactEar predictions are close to values estimated by these specialised systems.

5.9.2 Jump Metrics: To assess athletic performance, we can also use the reconstructed GRF curves to look at the time spent in the air and peak force developed during a jump take off to assess power production as well as the peak landing force to assess the impact. We evaluate the time in the airborne by considering the regions where the GRF is given as zero. The peak forces are evaluated by looking for maximum points either side of the airborn region, one for the peak jumping force and one for the peak landing force. This is illustrated in Figure 12.

We assessed the accuracy of these metrics obtained from the ImpactEar predictions against the ground truth metrics from the force plates. The comparison is again evaluated using MAPE and results are shown in Table 5b. It can be seen that the MAPE is below 10% for all the metrics. To put this in context, related work obtained results from different IMU sensors on the body to estimate peak jumping force [28]. The results were given in MAPE values of 3.9% for the sacrum, 6.6% for the upper back, 15.9% for the chest and 4.7% with a device fusion. This performance comparison shows that ImpactEar allows comparable performance on this downstream task to a specialised system for jumping force prediction. Additionally, Moschina et al. [36] use an earable estimate vertical jumping height which is directly linked to the airborne time calculated in this section. This shows the relevance of the selected downstream applications of ImpactEar to real world tasks desirable to be run on an earable device.

5.9.3 Case Study on Asymmetrical Gait: While analyzing the data, we noticed that one participant ran with significant asymmetry in their gait. This was evident both in the ground truth data and in our IMU data. Overall, the difference between left and right peak forces was on average 7.3% (weaker on the left side) with rises above 10%, this is a significant outlier compared to the other participants who were between 0-3% on average. In a

Table 5. Results of gait metric prediction for both walking and running (a) and jump metric prediction (b). All results are given as a MAPE.

Activity	MAPE (%) Contact Time	Peak Force	Impulse
Walk	5.5	3.2	1.4
Run	4.2	7.9	6.5

a

Activity	MAPE (%) Air Time	Take-off Force	Landing Force
Jump	3.3	6.7	8.9

b

follow-up discussion with the participant, it was revealed that they had a long standing strength asymmetry from childhood sports development, as well as a 1 cm shorter right leg. To assess how well ImpactEar can estimate this asymmetry, we looked at the data from this participant's LOSO evaluation. Specifically, we calculated the symmetry index for the ground truth and the ImpactEar predictions for the peak running forces.

Gait asymmetry can be commonly assessed by calculating the symmetry index [42]. In more detail, symmetry index is a percentage measure of asymmetry with a reference value, often this is the healthy side but here we use the average of the left and right value. Symmetry index is calculated as $SI = \frac{X_{left} - X_{right}}{X_{reference}} * 100\%$, where X_{left} and X_{right} are the left and right foot gait metrics and $X_{reference}$ is the reference value.

The evaluation showed that the ground truth symmetry index was on average **+7.3%** while the value for the ImpactEar prediction was **+9.6%.** This shows that the direction of asymmetry was correctly identified but the magnitude was overestimated by **2.3%**. A related work by Wang et al. [54] uses four lower leg mounted inertial sensors and reported an asymmetry error of 3% showing that ImpactEar is comparable to SOTA methods on asymmetry detection. This shows the promise of using ImpactEar for assessment of gait asymmetries which could be indicative of injuries or used for rehabilitation tracking as asymmetry gradually lessens over time. Additionally, analysis of asymmetry as presented here can also be useful for applications such as balance assessment using GRF curves [32] showing further possible use cases for ImpactEar.

5.9.4 Case Study on Running Footstrike Type: Another use of the vGRF curve can be to assess the running footstrike type [12]. This can be done from the initial loading phase where a peak can be present indicating a heelstrike or no peak indicating a forefoot strike. This difference can be seen in Figure 12 (c) and (d). To assess how well ImpactEar could distinguish these footstrikes a subset of participants were assessed to see whether ImpactEar correctly reconstructed this peak in their running vGRF curves, Two examples are shown in Figure 12 (c) and (d) where the ground truth and ImpactEar estimates are shown, it can be seen that the estimates include the peak in the heelstrike and exclude it in the forefoot strike. Overall, the model always predicted forefoot striking correctly but was sometimes predicting forefoot striking when the ground truth was heelstriking. While related work [21] has performed footstrike estimation using audio from an earable, ImpactEar uses the IMU which has a much smaller power footprint as it is only sampled at 50 Hz compared to 44.1 kHz.

5.9.5 Case Study on Ballet Dancing: To assess the generalisation ability of our framework to unseen activities, we recorded a trained ballet dancer performing a series of dance steps, while wearing a pair of OpenEarable 1.4 devices recording the IMU. While no force plate ground truth was available, we filmed the dance routine and compare the ImpactEar outputs to the timings of the landings and movements in the dance. The IMU and camera were time synchronised with a series of firm foot stamps at the beginning of the recording.

To evaluate how accurately ImpactEar reproduced the GRF from the dance, we assessed at which point each foot was in contact with the ground. Specifically, two examples are shown in Figure 12. The first in subfigure (e) shows the dancer mid air during a jumping spin move, the ImpactEar predicted GRF curve is shown in subfigure (g) showing clear air time with a zero GRF prediction. The other example shown in subfigure (f) depicts the

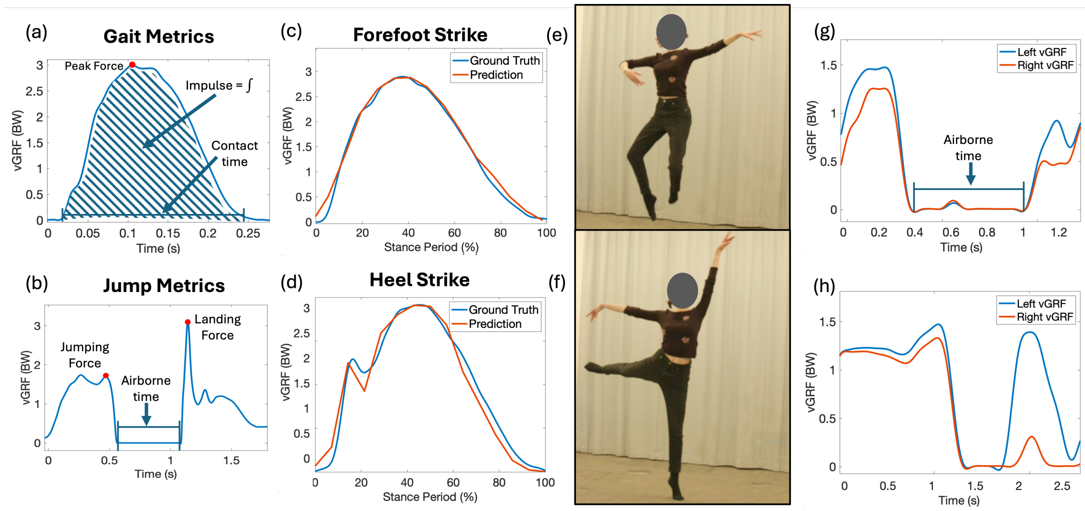


Fig. 12. Illustrations and results from the case studies. (a) shows the gait metrics illustrated on a vGRF curve. (b) shows the jumping metrics on a typical jump and landing GRF curve. (c) and (d) show correctly identified forefoot strike and heelstrike GRF curves including the ground truth (Blue) and ImpactEar estimate (Orange). (e) and (g) show the dancer during a high jumping move and the corresponding GRF curve. (f) and (h) show the dancer during a jump onto a one-legged pivoting move and the corresponding GRF curve.

dancer jumping onto one forefoot with the corresponding predicted GRF curve shown in subfigure (h). Here we can see that a jump and then a single sided impact (left foot as shown in the photograph) is predicted correctly.

6 Discussion and Future Work

In Section 5, particularly Table 3, we show that ImpactEar performs well across all the activities it estimates GRF across. Furthermore, in the downstream tasks, the accuracy of these estimations is also shown to be comparable to other SOTA methods. This demonstrates that one model can be sufficiently accurate at reconstructing GRF curves across different activities.

While our evaluation shows strong results, there are some significant outliers among experiments with NRMSEs over 20% as seen in Figure 8 and Table 2. The cause of the error is likely devices leading to decoupling of the IMU and the head and generating noise. Another possible source of such errors is the variation in device orientation and location by having a different fit per user. This has been shown to have significant impact on GRFs estimation accuracy in walking and jumping scenarios on other IMU locations [28, 52]. Future work could incorporate IMU-gravity alignment as a pre-processing stage for ImpactEar. Other causes of per-participant variation highlighted in Section 5.3 could be caused by movement of the device or the device becoming loose during the experiment which could account for some of the large jumps in error between experiments for some of the participants. This could be solved by using a device that sits in the ear, such as the OpenEarable 2.0 [44] or Apple Airpods rather than behind the ear like the OpenEarable 1.4 [45] where more movement is possible. Additional variations could arise from differences in gait style and jumping patterns between participants where those with an underrepresented style of movement may get lower vGRF reconstruction accuracy. Furthermore, as discussed in Section 5.3 the speed variations chosen by the participants causes a variance in vGRF NRMSE, it is likely that collecting more varied data over more users and speeds could reduce the gap in per-participant accuracy.

In Section 5.9 we presented five different applications of the GRF curves from the ImpactEar predictions. These applications show the wide range of tasks that can be targeted by ImpactEar, however there are many more use cases for GRF curves not explored here such as advanced gait analysis [19] or injury risk assessment [38]. This also leads into real-world use cases for ImpactEar, it could be deployed during sports scenarios where coaches wish to track athlete impact across a variety of different movements that can be present in sports such as soccer. The system could also be used longitudinally to track gait metrics in patients undergoing rehab to track changes in gait profiles over the course of their treatment. In Section 5.9.5 we demonstrate that ImpactEar can make predictions on the unseen ballet dancing activity. However, future work could collect more activity conditions to further expand ImpactEarDS, particularly more static and turning motions as well as stair climbing may prove valuable to capture other sporting and daily motions. This could further improve the generalisation of ImpactEar. As well as this, activity transitions could be collected to assess how well ImpactEar can transfer between contexts as a user moves from walking into running for example.

In Section 5.8.2 we present an experiment showing comparable performance at 25 Hz compared to the standard 50 Hz that ImpactEar uses. We explain this is due to the 15 Hz low pass filter used on the GRF treadmill data thus rendering the ground truth smoother and therefore equally matched by our data at 25 and 50 Hz, following this trend, it is unlikely higher sampling rates would yield better performance.

The dataset we collected during our evaluation, ImpactEarDS, that we are releasing to the research community, has the potential of opening new research avenues, as it contains unique gold standard validation of 3D GRF, moments and centre of pressure. One of the most clear uses would be to enable researchers to validate existing methods that use insoles [40] for GRF or Centre of Pressure estimation, that are not yet validated on force plates.

In Section 5.7 we show results for ImpactEar inference on a mobile phone showing much faster than the required real-time performance as well as low power consumption. In the future, pushing ImpactEar to the earbud itself could allow greater user privacy and less use of the radio cutting battery usage by just transmitting estimated curves rather than IMU data. To enable this transition, ImpactEar was designed to be as lightweight as possible with both a short context window and with a compact CNN autoencoder architecture. Emerging research grade and commercial earable devices such OpenEarable 2.0 [44] with 1 MB of flash memory could enable fully on-device ImpactEar using a quantised model.

7 Conclusion

We presented ImpactEar, which demonstrates for the first time that a single model can accurately predict vGRF across diverse activities without requiring activity context. We show that earbuds with consumer-grade specifications achieve GRF estimation performance comparable to state-of-the-art methods focused on single activities, as well as specialised gait devices. Through a selection of downstream tasks, we demonstrated the broad applicability of ImpactEar's GRF predictions and the extensive range of potential use cases. By developing a lightweight model, ImpactEar achieves real-time performance on mobile phones with data streamed directly from earables, while remaining robust to real-world challenges. Additionally, we release ImpactEarDS to the research community to advance earable research in human motion and kinetics. This work reveals the potential of earables to enable affordable and effective impact monitoring, and demonstrates how the model's cross-activity design effectively provides users *with a portable force plate at the head to measure GRF with no additional devices needed other than commodity earbuds.*

Acknowledgments

This work was supported by Engineering and Physical Sciences Research Council grants EP/Z53447X/1 and EP/S023046/1 for the Engineering and Physical Sciences Research Council Centre for Doctoral Training in Sensor Technologies and Applications as well as EVIDEN/New Atos Life Sciences Centre of Excellence.

References

[1] Inne Aerts, Elke Cumps, EALM Verhagen, Jo Verschueren, and Romain Meeusen. 2013. A systematic review of different jump-landing variables in relation to injuries. *J Sports Med Phys Fitness* 53, 5 (2013), 509–519.

[2] Ryan S Alcantara, W Brent Edwards, Guillaume Y Millet, and Alena M Grabowski. 2022. Predicting continuous ground reaction forces from accelerometers during uphill and downhill running: a recurrent neural network solution. *PeerJ* 10 (2022), e12752. doi:10.7717/peerj.12752

[3] Apple 2025. AirPods Pro (2nd Generation) - Technical Specifications. https://www.apple.com/uk/airpods-pro/specs/.

[4] Louis Atallah, Anatole Wiik, Gareth G Jones, Benny Lo, Justin P Cobb, Andrew Amis, and Guang-Zhong Yang. 2012. Validation of an ear-worn sensor for gait monitoring using a force-plate instrumented treadmill. *Gait & posture* 35, 4 (2012), 674–676. doi:10.1016/j.gaitpost.2011.11.021

[5] Bertec 2025. Instrumented Treadmill. https://www.bertec.com/products/instrumented-treadmills.

[6] Carmelo Bosco, Pekka Luhtanen, and Paavo V Komi. 1983. A simple method for measurement of mechanical power in jumping. *European journal of applied physiology and occupational physiology* 50 (1983), 273–282. doi:10.1007/bf00422166

[7] Heiner Bubb. 2019. Why do we need digital human models? In *DHM and Posturography*. Elsevier, Amsterdam, Netherlands, 7–32. doi:10.1016/b978-0-12-816713-7.00002-7

[8] Clara Piris Burgos, Lea Gärtner, Miguel A González Ballester, Jérôme Noailly, Fabian Stöcker, Martin Schönfelder, Tim Adams, and Simone Tassani. 2020. In-ear accelerometer-based sensor for gait classification. *IEEE Sensors Journal* 20, 21 (2020), 12895–12902. doi:10.1109/jsen.2020.3002589

[9] Josh Carter, Xi Chen, Dario Cazzola, Grant Trewartha, and Ezio Preatoni. 2024. Consumer-priced wearable sensors combined with deep learning can be used to accurately predict ground reaction forces during various treadmill running conditions. *PeerJ* 12 (2024), e17896. doi:10.7717/peerj.17896

[10] Josh Carter, Xi Chen, Dario Cazzola, Grant Trewartha, and Ezio Preatoni. 2024. Full body kinematics and ground reaction forces of fifty heterogeneous runners completing treadmill running at various speeds and gradients. doi:10.15125/BATH-01341 University of Bath Research Data Archive.

[11] Brian Caulfield and Mary Garrett. 2004. Changes in ground reaction force during jump landing in subjects with functional instability of the ankle joint. *Clinical biomechanics* 19, 6 (2004), 617–621. doi:10.1016/j.clinbiomech.2004.03.001

[12] Peter R. Cavanagh and Mario A. Lafortune. 1980. Ground reaction forces in distance running. *Journal of Biomechanics* 13, 5 (Jan. 1980), 397–406. doi:10.1016/0021-9290(80)90033-0

[13] Shing Chan, Yuan Hang, Catherine Tong, Aidan Acquah, Abram Schonfeldt, Jonathan Gershuny, and Aiden Doherty. 2024. CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition. *Scientific Data* 11, 1 (2024), 1135. doi:10.1038/s41597-024-03960-3

[14] Leora A Cramer, Markus A Wimmer, Philip Malloy, Joan A O'Keefe, Christopher B Knowlton, and Christopher Ferrigno. 2022. Validity and reliability of the Insole3 instrumented shoe insole for ground reaction force measurement during walking and running. *Sensors* 22, 6 (2022), 2203. doi:10.3390/s22062203

[15] John J Davis, Marcin Straczkiewicz, Jaroslaw Harezlak, and Allison H Gruber. 2021. CARL: a running recognition algorithm for free-living accelerometer data. *Physiological Measurement* 42, 11 (2021), 115001. doi:10.1088/1361-6579/ac41b8

[16] Ru Feng, Ukadike Christopher Ugbolue, Chen Yang, and Hui Liu. 2025. Estimation of Three-Dimensional Ground Reaction Force and Center of Pressure During Walking Using a Machine-Learning-Based Markerless Motion Capture System. *Bioengineering* 12, 6 (May 2025), 588. doi:10.3390/bioengineering12060588

[17] Cristopher D Garduno Luna, Brian Ruhe, Christopher Ertel, and Daniel Shull. 2025. Modelling peak vertical ground reaction force from smart watch metrics reveals link between higher forces and softer cushioning perception. *Footwear Science* 17, sup1 (2025), S83–S84. doi:10.1080/19424280.2025.2492263

[18] Amir Gholami, Sehoon Kim, Zhen Dong, Zhewei Yao, Michael W Mahoney, and Kurt Keutzer. 2022. A survey of quantization methods for efficient neural network inference. In *Low-power computer vision*. Chapman and Hall/CRC, Boca Raton, FL, USA, 291–326. doi:10.1201/9781003162810-13

[19] Brian Horsak, Djordje Slijepcevic, Anna-Maria Raberger, Caterine Schwab, Marianne Worisch, and Matthias Zeppelzauer. 2020. GaitRec, a large-scale ground reaction force dataset of healthy and impaired gait. *Scientific data* 7, 1 (2020), 143. doi:10.1038/s41597-020-0481-z

[20] Md Sanzid Bin Hossain, Zhishan Guo, and Hwan Choi. 2023. Estimation of lower extremity joint moments and 3d ground reaction forces using imu sensors in multiple walking conditions: A deep learning approach. *IEEE journal of biomedical and health informatics* 27, 6 (2023), 2829–2840. doi:10.1109/jbhi.2023.3262164

[21] Changshuo Hu, Thivya Kandappu, Jake Stuchbury-Wass, Yang Liu, Anthony Tang, Cecilia Mascolo, and Dong Ma. 2024. Detecting Foot Strikes during Running with Earbuds. In *Proceedings of the Workshop on Body-Centric Computing Systems*. Association for Computing Machinery, New York, NY, USA, 35–40. doi:10.1145/3662009.3662023

[22] Changshuo Hu, Qiang Yang, Yang Liu, Tobias Röddiger, Kayla-Jade Butkow, Mathias Ciliberto, Adam Luke Pullin, Jake Stuchbury-Wass, Mahbub Hassan, Cecilia Mascolo, et al. 2026. A survey of earable technology: Trends, tools, and the road ahead. *Proceedings of the ACM*

on Interactive, Mobile, Wearable and Ubiquitous Technologies 10, 2 (2026), 1–41. doi:10.1145/3810217

[23] Delaram Jarchi, Charence Wong, Richard Mark Kwasnicki, Ben Heller, Garry A Tew, and Guang-Zhong Yang. 2014. Gait parameter estimation from a miniaturized ear-worn sensor using singular spectrum analysis and longest common subsequence. *IEEE Transactions on Biomedical Engineering* 61, 4 (2014), 1261–1273. doi:10.1109/tbme.2014.2299772

[24] Eun Som Jeon, Sinjini Mitra, Jisoo Lee, Omik M. Save, Ankita Shukla, Hyunglae Lee, and Pavan Turaga. 2025. Ground Reaction Force Estimation via Time-aware Knowledge Distillation. *IEEE Internet of Things Journal* (2025), 1–1. doi:10.1109/JIOT.2025.3578012 arXiv:2506.10265 [eess].

[25] Nan Jiang, Terence Sim, and Jun Han. 2022. EarWalk: towards walking posture identification using earables. In *Proceedings of the 23rd Annual International Workshop on Mobile Computing Systems and Applications*. Association for Computing Machinery, New York, NY, USA, 35–40. doi:10.1145/3508396.3512883

[26] Xianta Jiang, Christopher Napier, Brett Hannigan, Janice J Eng, and Carlo Menon. 2020. Estimating vertical ground reaction force during walking using a single inertial sensor. *Sensors* 20, 15 (2020), 4345. doi:10.3390/s20154345

[27] Amal Kammoun, Philippe Ravier, and Olivier Buttelli. 2024. Comparison of the Accuracy of Ground Reaction Force Component Estimation between Supervised Machine Learning and Deep Learning Methods Using Pressure Insoles. *Sensors* 24, 16 (Aug. 2024), 5318. doi:10.3390/s24165318

[28] Jordan A. Kerns, Andrew S. Zwart, Pietro S. Perez, Reed D. Gurchiek, and Jeffrey M. McBride. 2023. Effect of IMU location on estimation of vertical ground reaction force during jumping. *Frontiers in Bioengineering and Biotechnology* 11 (March 2023), 1112866. doi:10.3389/fbioe.2023.1112866 Publisher: Frontiers.

[29] Kistler 2025. 3D Force Plates. https://www.kistler.com/GB/en/3d-force-plate/C00000090.

[30] Felix Kluge, Yonatan E Brand, M Encarna Micó-Amigo, Stefano Bertuletti, Ilaria D'Ascanio, Eran Gazit, Tecla Bonci, Cameron Kirk, Arne Küderle, Luca Palmerini, et al. 2024. Real-world gait detection using a wrist-worn inertial sensor: validation study. *JMIR formative research* 8 (2024), e50035. doi:10.2196/50035

[31] Tatsuki Koshio, Naoto Haraguchi, Takayoshi Takahashi, Yuse Hara, and Kazunori Hase. 2024. Estimation of Ground Reaction Forces during Sports Movements by Sensor Fusion from Inertial Measurement Units with 3D Forward Dynamics Model. *Sensors* 24, 9 (April 2024), 2706. doi:10.3390/s24092706

[32] Xiangzhi Liu, Xiangliang Zhang, Bin Zhang, Bin Zhou, Zexia He, and Tao Liu. 2023. An IMU-based ground reaction force estimation method and its application in walking balance assessment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 32 (2023), 223–232. doi:10.1109/tnsre.2023.3347729

[33] Nathan Louis, Jason J Corso, Tylan N Templin, Travis D Eliason, and Daniel P Nicolella. 2022. Learning to estimate external forces of human motion in video. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery, New York, NY, USA, 3540–3548. doi:10.1145/3503161.3548377

[34] Yue Luo, Sarah M Coppola, Philippe C Dixon, Song Li, Jack T Dennerlein, and Boyi Hu. 2020. A database of human gait performance on irregular and uneven surfaces collected by wearable sensors. *Scientific data* 7, 1 (2020), 219. doi:10.1038/s41597-020-0563-y

[35] Michael Mauersberger, Falk Hähnel, Klaus Wolf, Johannes FC Markmiller, Alexander Knorr, Dominik Krumm, and Stephan Odenwald. 2022. Predicting ground reaction forces of human gait using a simple bipedal spring-mass model. *Royal Society Open Science* 9, 7 (2022), 211582. doi:10.1098/rsos.211582

[36] Dennis Moschina, Tobias Röddiger, and Michael Beigl. 2023. Vertical Jump Test Using an Earable Accelerometer. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (Cancun, Quintana Roo, Mexico) *(UbiComp/ISWC '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 324–326. doi:10.1145/3594739.3610669

[37] Christopher Napier, Xianta Jiang, Christopher L MacLean, Carlo Menon, and Michael A Hunt. 2020. The use of a single sacral marker method to approximate the centre of mass trajectory during treadmill running. *Journal of Biomechanics* 108 (2020), 109886. doi:10.1016/j.jbiomech.2020.109886

[38] C Napier, CL MacLean, J Maurer, JE Taunton, and MA Hunt. 2018. Kinetic risk factors of running-related injuries in female recreational runners. *Scandinavian journal of medicine & science in sports* 28, 10 (2018), 2164–2172. doi:10.1111/sms.13228

[39] Pietro Picerno, Valentina Camomilla, and Laura Capranica. 2011. Countermovement jump performance assessment using a wearable 3D inertial measurement unit. *Journal of Sports Sciences* 29, 2 (Jan. 2011), 139–146. doi:10.1080/02640414.2010.523089 Publisher: Routledge.

[40] Junbin Ren, Ruihao Zheng, Wenbo Zhang, Dong She, Yuting Bai, ZhanPeng Jin, and Yang Gao. 2025. Motion2Press: Cross Model Learning from IMU to Plantar Pressure for Gait Analysis. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–33. doi:10.1145/3749499

[41] Nesma M Rezk, Madhura Purnaprajna, Tomas Nordström, and Zain Ul-Abdin. 2020. Recurrent neural networks: An embedded computing perspective. *Ieee Access* 8 (2020), 57967–57996. doi:10.1109/access.2020.2982416

[42] RO Robinson, Walter Herzog, and Benno M Nigg. 1987. Use of force platform variables to quantify the effects of chiropractic manipulation on gait symmetry. *Journal of manipulative and physiological therapeutics* 10, 4 (1987), 172–176.

[43] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. 2022. Sensing with earables: A systematic literature review and taxonomy of phenomena. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 6, 3 (2022), 1–57. doi:10.1145/3550314

[44] Tobias Röddiger, Michael Küttner, Philipp Lepold, Tobias King, Dennis Moschina, Oliver Bagge, Joseph A Paradiso, Christopher Clarke, and Michael Beigl. 2025. Openearable 2.0: Open-source earphone platform for physiological ear sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 1 (2025), 1–33. doi:10.1145/3712069

[45] Tobias Röddiger, Jake Stuchbury-Wass, Mathias Ciliberto, Philipp Lepold, and Michael Beigl. 2024. OpenEarable 1.4: Dual Microphones Earpiece to Capture In-Ear and Outer-Ear Audio Signals. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 930–933. doi:10.1145/3675094.3678483

[46] Bidyut Saha and Riya Samanta. 2026. Rethinking Temporal Models for TinyML: LSTM versus 1D-CNN in Resource-Constrained Devices. *arXiv preprint arXiv:2603.04860* (2026). doi:10.48550/arXiv.2603.04860

[47] Panneer Selvam Santhalingam, Parth Pathak, Huzefa Rangwala, and Jana Kosecka. 2023. Synthetic smartwatch imu data generation from in-the-wild asl videos. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–34. doi:10.1145/3596261

[48] E Shahabpoor and A Pavic. 2018. Estimation of vertical walking ground reaction force in real-life environments using single IMU sensor. *Journal of biomechanics* 79 (2018), 181–190. doi:10.1016/j.jbiomech.2018.08.015

[49] Mohsen Sharifi Renani, Abigail M Eustace, Casey A Myers, and Chadd W Clary. 2021. The use of synthetic IMU signals in the training of deep learning models significantly improves the accuracy of joint kinematic predictions. *Sensors* 21, 17 (2021), 5876. doi:10.3390/s21175876

[50] Jake Stuchbury-Wass, Yang Liu, Kayla-Jade Butkow, Joshua Carter, Mathias Ciliberto, Quing Yang, Ezio Preatoni, Dong Ma, and Cecilia Mascolo. 2025. WalkEar: Holistic Gait Monitoring Using Earables. In *2025 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, Washington, DC, USA, 99–109. doi:10.1109/PerCom64205.2025.00029

[51] Tao Sun, Dongxuan Li, Bingfei Fan, Tian Tan, and Peter B. Shull. 2023. Real-Time Ground Reaction Force and Knee Extension Moment Estimation During Drop Landings Via Modular LSTM Modeling and Wearable IMUs. *IEEE Journal of Biomedical and Health Informatics* 27, 7 (July 2023), 3222–3233. doi:10.1109/JBHI.2023.3268239

[52] Tian Tan, Peter B Shull, Jenifer L Hicks, Scott D Uhlrich, and Akshay S Chaudhari. 2024. Self-Supervised Learning Improves Accuracy and Data Efficiency for IMU-Based Ground Reaction Force Estimation. *IEEE Transactions on Biomedical Engineering* 71, 7 (2024), 2095–2104. doi:10.1109/tbme.2024.3361888

[53] Ryo Tominaga, Yoshimasa Ishii, Takeshi Ueda, and Takashi Kurokawa. 2016. The effects of running speed on ground reaction forces and lower limb kinematics during single-leg stop movement. *The Journal of Strength & Conditioning Research* 30, 5 (2016), 1224–1230. doi:10.1519/jsc.0000000000000286

[54] Lei Wang, Yun Sun, Qingguo Li, and Tao Liu. 2018. Estimation of step length and gait asymmetry using wearable inertial sensors. *IEEE Sensors Journal* 18, 9 (2018), 3844–3851. doi:10.1109/jsen.2018.2815700

[55] David A Winter. 2009. *Biomechanics and motor control of human movement* (4 ed.). John wiley & sons, Hoboken, NJ.

[56] Christian Wolff, Elke Warmerdam, Tim Dahmen, Tim Pohlemann, Philipp Slusallek, and Bergita Ganse. 2025. New parameters based on ground reaction forces for monitoring rehabilitation following tibial fractures and assessment of heavily altered gait. *Sensors* 25, 8 (2025), 2475. doi:10.3390/s25082475

[57] Yeqing Zhu, Di Xia, and Heng Zhang. 2023. Using wearable sensors to estimate vertical ground reaction force based on a transformer. *Applied Sciences* 13, 4 (2023), 2136. doi:10.3390/app13042136