Article

Deep Metric Learning for Scalable Gait-Based Person Re-Identification Using Force Platform Data

Kayne A. Duncanson [1,*], Simon Thwaites [1], David Booth [1], Gary Hanly [2], William S. P. Robertson [3], Ehsan Abbasnejad [4] and Dominic Thewlis [1]

[1] Adelaide Medical School, The University of Adelaide, Adelaide, SA 5000, Australia; simon.thwaites@adelaide.edu.au (S.T.); david.booth@adelaide.edu.au (D.B.); dominic.thewlis@adelaide.edu.au (D.T.)
[2] Defence Science and Technology Group, Department of Defence, Adelaide, SA 5000, Australia; gary.hanly@defence.gov.au
[3] School of Mechanical Engineering, The University of Adelaide, Adelaide, SA 5000, Australia; will.robertson@adelaide.edu.au
[4] Australian Institute for Machine Learning, The University of Adelaide, Adelaide, SA 5000, Australia; ehsan.abbasnejad@adelaide.edu.au
* Correspondence: kayne.duncanson@adelaide.edu.au

Abstract: Walking gait data acquired with force platforms may be used for person re-identification (re-ID) in various authentication, surveillance, and forensics applications. Current force platform-based re-ID systems classify a fixed set of identities (IDs), which presents a problem when IDs are added or removed from the database. We formulated force platform-based re-ID as a deep metric learning (DML) task, whereby a deep neural network learns a feature representation that can be compared between inputs using a distance metric. The force platform dataset used in this study is one of the largest and the most comprehensive of its kind, containing 193 IDs with significant variations in clothing, footwear, walking speed, and time between trials. Several DML model architectures were evaluated in a challenging setting where none of the IDs were seen during training (i.e., zero-shot re-ID) and there was only one prior sample per ID to compare with each query sample. The best architecture was 85% accurate in this setting, though an analysis of changes in walking speed and footwear between measurement instances revealed that accuracy was 28% higher on same-speed, same-footwear comparisons, compared to cross-speed, cross-footwear comparisons. These results demonstrate the potential of DML algorithms for zero-shot re-ID using force platform data, and highlight challenging cases.

Keywords: gait recognition; biometric; ground reaction force; deep learning; zero-shot re-ID; center of pressure; time series; gait analysis; force plate; classification

1. Introduction

Gait is a characteristic pattern of movements employed to move from one place to another [1]. Human bipedal walking gait has been proposed as a cue to identity, i.e., a biometric, that can be exploited for various authentication, surveillance, and forensics applications [2–4]. The attractiveness of gait as a biometric is that it can be measured unobtrusively and discreetly [3]. Based on early medical and psychological studies on the characterization of gait, it has been said to be unique within individuals, consistent over time, and difficult to obfuscate [5,6]. Therefore, walking gait may be used to control access to secure areas (e.g., within government premises, banks, and research laboratories), detect threats to national security in confined public spaces (e.g., airports and sporting venues), or identify movement of criminals retrospectively. It may be particularly useful in dynamic environments, where a face may be partially or totally occluded by other individuals, or covered (e.g., by a face mask), and other structural features are unobtainable.

Re-identification (re-ID) is the process of matching biometrics over time to re-establish identity. The re-ID task has previously been attempted using a variety of sensing modalities (e.g., video cameras, inertial sensors, pressure sensors, force platforms, and continuous wave radars) that generally acquire data on either body motion or applied forces during gait [7]. In the vast majority of systems, video cameras are utilized to acquire data pertaining to two-dimensional (2D) body appearance and motion. However, such approaches are not universally applicable, due to limitations related to body appearance (e.g., clothing, objects, and camera angle) and visibility (e.g., lighting, occlusion, and background noise) [8–11]. Thus, it is necessary to investigate the viability of complementary sensors that can provide biometric data in settings where vision-based systems are inapplicable or undependable.

Force platforms are a natural complement to video cameras because they circumvent the aforementioned limitations and are a rich source of kinetic data (i.e., data pertaining to the forces that drive motion). These devices measure ground reaction forces (GRFs) along three orthogonal axes (F_x—mediolateral, F_y—anteroposterior, and F_z—vertical) and corresponding ground reaction moments (GRMs) about each axis (M_x, M_y, and M_z). The coordinates of the GRF vector origin, referred to as the center of pressure (COP) coordinates (C_x and C_y), are calculated from directional components of the GRFs and GRMs [12]. The directional components of each parameter result in an eight-channel time series that can be acquired during both left and right stance phases of the gait cycle if two or more force platforms are embedded into a ground surface. For reference, the gait cycle is the interval between two successive occurrences of the same gait event, such as left heel contact. The stance phase is when a given foot is in contact with the ground and comprises approximately 60 % of the gait cycle. Within the stance phase, there are four functional sub-phases: loading response, mid-stance, terminal stance, and pre-swing. A detailed description of these sub-phases can be found elsewhere (e.g., [13]).

Despite the apparent prospect of force platforms as sensors for the re-ID task, there have only been a handful of studies in the field [6,14–20]. All of the recent works implemented machine learning classification models that were trained and evaluated at 'many-shot' (or subject-dependent [21]) re-ID, whereby the training set contains 'many' samples from the identities (IDs) that are encountered in the test set. If one of these models were to be deployed, they could only re-ID those who were included in the training set. This is a significant limitation given that IDs may be introduced or removed from the database upon deployment in many applications. In contrast to this, is the more challenging and general 'zero-shot' (or subject-independent [21]) re-ID, whereby the training set contains 'zero' samples from the IDs that are encountered in the test set (i.e., the training and test sets contain discrete IDs). Zero-shot re-ID can be achieved using deep metric learning (DML); that is, the utilization of a deep neural network (DNN) to learn a feature representation that, when compared between samples using a distance metric, is similar between samples from the same category (i.e., the same ID) and dissimilar between samples from different categories (i.e., different IDs) [22,23] (Figure 1). Using this *matching* approach, any number of IDs, not seen during training, can be re-identified at test time, provided that there is at least one prior sample. Thus, the implementation of DML would increase the scalability of force platform-based re-ID systems.

Previous studies on force platform-based re-ID also provide an incomplete picture on the effects of task and environmental constraints that are known to alter gait (Figure 2). Models have generally been trained and tested on datasets with limited or no variation in clothing, footwear, object carriage, walking speed, and time between trials [6]. The few exceptions to this have provided preliminary evidence that, on datasets with fewer than 80 IDs, these constraints can considerably reduce rank-1 accuracy (referred to herein as accuracy) using classification models (except clothing which has been assumed to have negligible effects) [16,24]. Whether or not these effects scale to larger datasets and apply to DML models remains unknown. Thus, the contributions of this work are as follows:

- We utilized a large force platform dataset that was purpose built for re-ID of persons and, thus, had the most complex set of walking conditions. The dataset contained

5587 walking trials from 193 IDs, with both intra- and inter-individual variations in clothing, footwear, and walking speed, as well as inter-individual variations in time between trials. A public version of the dataset, named ForceID A, contains data from 184 IDs, who consented to their data being published online (see data availability statement).

- To provide scope for future DML model design, we evaluated several different baseline DNN architectures at zero-shot re-ID. Two-layer fully connected neural networks (FCNNs) slightly outperformed more complex architectures over seven-fold cross validation, achieving 85 % accuracy in our challenging evaluation setting, where there was only one prior sample per ID available to compare with each query sample.
- We analyzed the combined effects of changes in walking speed and footwear between measurement instances on re-ID performance. Accuracy across all models on same-speed, same-footwear comparisons (the easiest) was 28 % higher than accuracy on cross-speed, cross-footwear comparisons (the hardest). The code repository for this study can be found at https://github.com/kayneduncanson1/ForceID-Study-1.
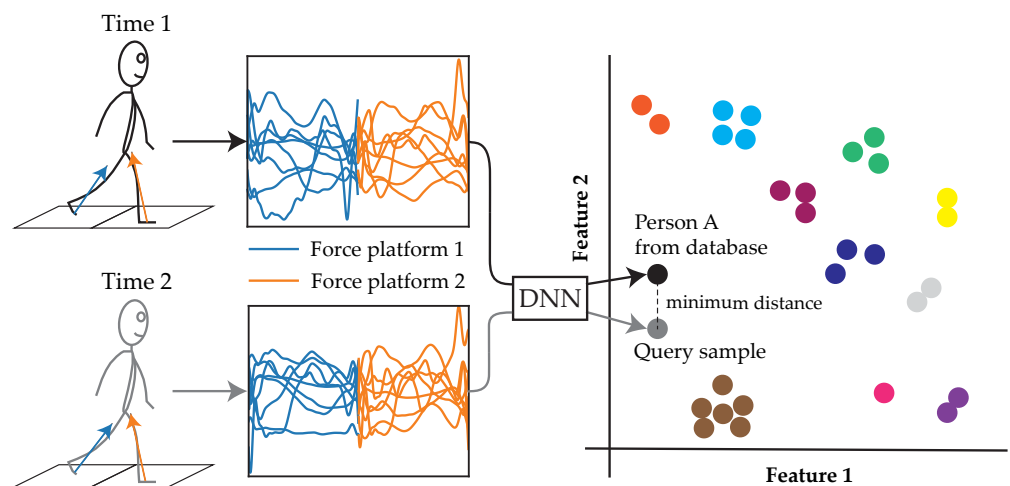
Figure 1. Workflow of person re-identification (re-ID) using a using a deep metric learning model. An unknown identity (ID) can be re-identified by extracting their gait features using a deep neural network (DNN) and comparing them to those of other samples in the database using a distance metric. Following this approach, the same trained model can be used to re-identify any number of IDs who were not seen during training. Unlabeled colors each represent a different ID. The feature space is depicted as two-dimensional for simplicity.
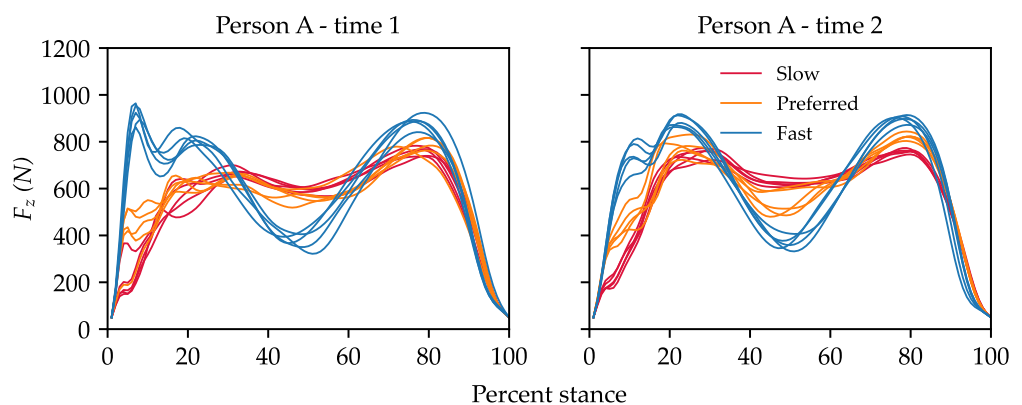
Figure 2. An example of the vertical ground reaction force (F_z) with changes in walking speed (colors) and footwear (**left** vs. **right**) between measurement instances. **Left**—women's ankle boots (flat sole); **right**—ballet flats.

2. Related Work

2.1. Vision-Based re-ID

The state of vision-based systems for gait-based person re-ID has been reported extensively in the literature [21]. The standard benchmark datasets in this space—namely, CASIA-B, OU-ISIR, and OU-MVLP—contain 2D video footage acquired in controlled settings (i.e., settings with sufficient lighting and no occlusions or background noise) and under controlled walking conditions (e.g., limited or no variation in walking speed, clothing, or object carriage between measurement instances) [8,25,26]. Most of the current state-of-the-art models are DML models that rely on some variant of the triplet loss function (described in detail in Section 3.1), either in isolation or in combination with another distance metric loss function [27–33]. These models obtained 93–98% accuracy in the 'Normal' walking condition on CASIA-B (zero-shot, 50 IDs) and the subset of these models that were also implemented on OU-MVLP obtained 63–89% accuracy (zero-shot, 5154 IDs) [28–30,32,33]. The subset of these models that were implemented on OU-ISIR obtained 99–100% accuracy (many-shot, 4007 IDs) [27,33]. In all of the aforementioned cases, walking was conducted at preferred speed in matched clothing without any carried object. When either a carry bag or coat were introduced between measurement instances for the same 50 IDs from CASIA-B, accuracy dropped to 90–94% and 72–82%, respectively [27–32]. These findings suggest that vision-based re-ID systems perform extremely well in tightly controlled conditions, yet have limited robustness to the inclusion of walking conditions that are commonly encountered in natural settings.

The sensitivity of vision-based systems to walking conditions is illustrated by the recent work of Zhu et al. [34]. The authors acquired a video dataset (**G**ait **RE**cognition in the **W**ild (GREW) dataset) of 26,345 IDs walking in diverse public environments with variations in camera angle, lighting, background, walking speed, clothing, object carriage, footwear, and walking direction, as well as the presence of occlusions to the camera field of view. Six models (three DML models and three classification models) that were state-of-the-art at the time on standard benchmark datasets were evaluated at zero-shot re-ID on a test set of 6000 IDs from this dataset. Four out of six of the models almost completely failed (accuracy: PoseGait [35]—0%, GaitGraph [36]—1%, GEINet [37]—7%, and TS-CNN [9]—14%). The other two performed considerably better (GaitPart [30]—44% and GaitSet [38]—46%), yet still performed much worse than they did on the standard benchmark datasets (e.g., accuracy on CASIA-B: GaitPart—79–96% and GaitSet—70–95% depending on the condition). Accordingly, the authors concluded that there is still much room for improvement in the performance of vision-based systems.

2.2. Force Platform-Based re-ID

All studies on force platform-based re-ID have been done in the many-shot setting using classification models. Connor et al. [6] provided an overview of early studies that utilized simple machine learning models (e.g., k-Nearest Neighbour (kNN), Hidden Markov model, and Support Vector Machine). Most recent works have used multiple components of force platform data as inputs to DNN models [18–20]. In studies where data was acquired from fewer than 60 IDs in a single session under tightly controlled walking conditions, 88–99% accuracy was obtained using FCNN models and 99%+ accuracy was obtained using one-dimensional (1D) convolutional neural network (CNN) models [18,20]. Another study utilized an 1D convolutional long short-term memory neural network (CLSTMNN) on a dataset of 79 IDs with intra- and inter-individual variations in clothing and walking speed, as well as inter-individual variations in footwear and time between trials (3–14 days, depending on the ID) [19]. The model obtained 86% validation accuracy when trained on samples from session one and then validated on samples from session two. Instead of using DNNs, Derlatka and Borowska [39] proposed an ensemble of simple machine learning classification models. This system obtained 100 % accuracy on 322 IDs, who completed a single session of walking at preferred speed in personal footwear. Even though there are numerous sources of variation between these studies, that make comparisons difficult

(e.g., different datasets, input representations, model designs, training and evaluation protocols, and implementation details), they collectively highlight the potential of force platform-based re-ID.

3. Materials and Methods

3.1. Problem Formulation and Loss Function

Given a query gait sample $\mathbf{x}_q$ and a set of N prior samples $\mathbf{x}_i$ with corresponding ID labels l_i, $\{(\mathbf{x}_i, l_i)\}_{i=1}^N$, the task is to predict the query ID label l_q. Note that the set is not fixed and new samples and IDs may be added at later times. Under a DML framework, this is done by learning parameters $\boldsymbol{\theta}$ for a DNN $f_{\boldsymbol{\theta}}$ such that, for all i,

$$\left\| f_{\boldsymbol{\theta}}(\mathbf{x}_q) - f_{\boldsymbol{\theta}}(\mathbf{x}_r) \right\| < \left\| f_{\boldsymbol{\theta}}(\mathbf{x}_q) - f_{\boldsymbol{\theta}}(\mathbf{x}_i) \right\|, l_q = l_r, l_i \neq l_r, \tag{1}$$

where $\mathbf{x}_r$ is a prior sample with the same ID label as $\mathbf{x}_q$. Upon evaluation time, l_q is assigned as the label corresponding to the sample that is most similar to $\mathbf{x}_q$, which should be l_r. The $\boldsymbol{\theta}$ that satisfy the conditions in Equation (1) are learnt through sequential optimization. In this study, the triplet margin loss function was used to inform optimization [40]. This function minimizes the distance between samples from the same ID and maximizes the distance between samples from different IDs in feature space. It requires a 'triplet', comprising a query sample, referred to as an anchor $\mathbf{x}_a$, a sample from the same ID as the anchor, referred to as a positive $\mathbf{x}_p$, and a sample from a different ID to the anchor, referred to as a negative $\mathbf{x}_n$. The loss $\bar{L}$ for a given triplet $T = (f_{\boldsymbol{\theta}}(\mathbf{x}_a), f_{\boldsymbol{\theta}}(\mathbf{x}_p), f_{\boldsymbol{\theta}}(\mathbf{x}_n))$ is defined as

$$\bar{L}(f_{\boldsymbol{\theta}}(\mathbf{x}_a), f_{\boldsymbol{\theta}}(\mathbf{x}_p), f_{\boldsymbol{\theta}}(\mathbf{x}_n)) = \left[D(f_{\boldsymbol{\theta}}(\mathbf{x}_a) - f_{\boldsymbol{\theta}}(\mathbf{x}_p)) - D(f_{\boldsymbol{\theta}}(\mathbf{x}_a) - f_{\boldsymbol{\theta}}(\mathbf{x}_n)) + \mu \right]_+, \tag{2}$$

where μ is a margin enforced between $\mathbf{x}_p$ and $\mathbf{x}_n$ relative to $\mathbf{x}_a$ and $[\cdot]_+ = \max(\cdot, 0)$. In this study, the DNN $f_{\boldsymbol{\theta}}$ was trained on varying subsets of the dataset detailed in Section 3.2 using a number of DNN architectures detailed in Section 3.4. Training was conducted on partitions of each subset referred to as mini-batches $\mathbf{X}$. The method for sampling $\mathbf{x}_p$ and $\mathbf{x}_n$ for a given $\mathbf{x}_a$ to form T has been shown to influence the loss and, thus, the optimization of $\boldsymbol{\theta}$ [40]. The 'batch hard' sampling method was chosen in this study based on the L_2 distance metric. Namely, for a given $\mathbf{x}_a$, given the set P of candidate $\mathbf{x}_p$ and the set M of candidate $\mathbf{x}_n$ in $\mathbf{X}$, the hardest $\mathbf{x}_p$ (i.e., the $\mathbf{x}_p$ that maximizes L_2 distance when paired with $\mathbf{x}_a$) and the hardest $\mathbf{x}_n$ (i.e., the $\mathbf{x}_n$ that minimizes L_2 distance when paired with $\mathbf{x}_a$) were selected:

$$f_{\boldsymbol{\theta}}(\mathbf{x}_p) = \max_{\bar{p} \in P} \left\| f_{\boldsymbol{\theta}}(\mathbf{x}_a) - f_{\boldsymbol{\theta}}(\mathbf{x}_{\bar{p}}) \right\|_2^2, \tag{3}$$

$$f_{\boldsymbol{\theta}}(\mathbf{x}_n) = \min_{\bar{n} \in M} \left\| f_{\boldsymbol{\theta}}(\mathbf{x}_a) - f_{\boldsymbol{\theta}}(\mathbf{x}_{\bar{n}}) \right\|_2^2. \tag{4}$$

The training loss L_{train} was defined as

$$L_{\text{train}}(\boldsymbol{\theta}; \mathbf{X}) = \frac{1}{A} \sum_{a=1}^{A} \bar{L}(f_{\boldsymbol{\theta}}(\mathbf{x}_a), f_{\boldsymbol{\theta}}(\mathbf{x}_p), f_{\boldsymbol{\theta}}(\mathbf{x}_n)), \tag{5}$$

where A is the total number of triplets in $\mathbf{X}$. Sampling hard triplets from each mini-batch allows models to learn from challenging examples (i.e., similar samples from different IDs) without oversampling outliers [40].

3.2. Force Platform Dataset

3.2.1. Experimental Protocol

The dataset used in this study is one of the largest force platform datasets and it has the most complex set of walking conditions, because it was purpose built for evaluating person re-ID systems. It contains 5587 walking trials (i.e., walks along the length of our laboratory in one direction) from 193 IDs (54 % female; mean $\pm$ std. dev.: age 27 $\pm$ 7 years, mass 71.86 $\pm$ 15.93 kg, and height 1.72 $\pm$ 0.10 m), with 19–30 trials per ID. All participants gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the University of Adelaide Human Research Ethics Committee (H-2018-009). Participants had no known neurological or musculoskeletal disorder that could affect their gait (this simplified the experimental protocol and facilitated the necessary first step of evaluating re-ID systems on healthy individuals). In regard to personal clothing and footwear, participants each attended two sessions separated by 3–14 days depending on the ID (there were 2777 session one trials and 2810 session two trials). At the start of each session, age, sex, mass, height, and footwear type were recorded. Footwear and clothing were also photographed for future reference. Next, participants walked along the length of the laboratory in one direction ($\approx$10 m) five times at each of three self-selected speed categories: preferred, slower than preferred, and faster than preferred. Two in-ground OPT400600-HP force platforms (Advanced Mechanical Technology Inc., Phoenix, AZ, USA) in the center of the laboratory measured GRFs and GRMs during left and right footsteps. These measures, along with calculated COP coordinates, were acquired through Vicon Nexus (Vicon Motion Systems Ltd., Oxford, UK) at 2000 Hz. Of note, all trials in this dataset were complete foot contacts; that is, each foot completely contacted within the area of each force platform, as identified from video footage. In this study, each sample comprised one trial to simulate a single pass through an area of interest during locomotion. ForceID A (the public version of the dataset) contains 5327 trials from 184 participants who consented to their data being published online (see data availability statement).

3.2.2. Dataset Characteristics

Figure 3 provides a visual representation of the dataset used in this study. A comparison of the sets of signals for three random IDs in the dataset provided an example of how signal shape varied both within and between IDs. The overlap between colors for certain components at certain sub-phases during stance suggests that discriminating between IDs in the dataset (e.g., IDs 108 and 125) could be quite challenging and would, thus, require the identification of subtle shape features. In fact, this dataset was prospectively designed to be challenging via the inclusion of several task and environmental constraints that increase intra-individual variability and can be expected in natural settings.

Table 1 provides an overview of key characteristics of ForceID A compared to other large-scale public force platform datasets. It should be noted that almost all studies in this field have utilized private datasets with fewer than 80 IDs [6,17,19,20,24]. Exceptions to this are Horst et al. [18], who used data from 57 IDs that is available in in their public database [41], and Derlatka and Borowska [39], who used data from 322 IDs (that may be from their public dataset [42], though it was not mentioned). This makes ForceID A the second largest dataset to be implemented for force platform-based re-ID and by far the largest dataset of repeated measures over separate days. ForceID A is also the only dataset with intra- and inter-individual variations in footwear, walking speed, and clothing (Table 1). Hence, while it can be used for a variety of purposes, its primary use case is the development of practical force platform-based re-ID systems.
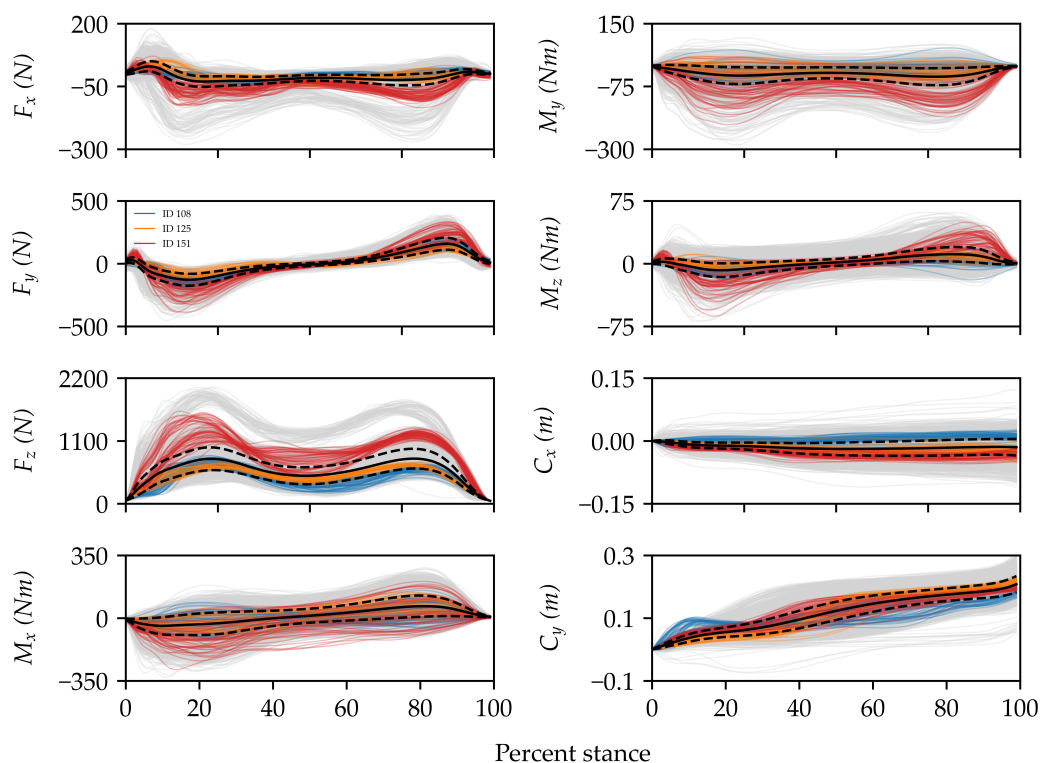
Figure 3. Visualization of all ground reaction forces (F_x—mediolateral, F_y—anteroposterior, and F_z—vertical), ground reaction moments (M_x, M_y, and M_z), and center of pressure coordinates (C_x and C_y) for three random IDs in the dataset (colors) compared to the rest (grey). Solid and dashed black lines are the means and standard deviations across the entire dataset, respectively. For ease of interpretation, the signals have been clipped, filtered, and time normalized to 100 % of the stance phase. Overall, there is a large degree of variability in the dataset, and some components are more variable than others.

Table 1. Overview of key characteristics of public force platform datasets containing healthy individuals. Gutenberg gait database comprises 10 datasets, each used for a separate study. All sessions were conducted on the same day in GaitRec (healthy). * Clothing conditions were not mentioned for datasets other than ForceID A and, thus, were inferred where possible. S = slower than preferred, P = preferred, F = faster than preferred.

Dataset	No. IDs	No. Sessions	Footwear	Walking Speed	Object Carriage	Clothing *
AIST gait database [43]	300	1	Barefoot	P	-	Skin-tight
Gutenberg gait database [41]	350	1 (253 IDs) 2 (47 IDs) 6 (42 IDs) 8 (8 IDs)	Barefoot	P	-	Skin-tight
GaitRec (healthy) [44]	211	1–6	Barefoot and personal	S, P, F	-	Skin-tight
Derlatka & Borowska (2023) [42]	324	1 (294 IDs) 2 (30 IDs)	Personal (semi-controlled)	P	-	Unknown
ForceID A (ours)	184	2 (184 IDs)	Personal	S, P, F	-	Personal

Table 2 shows the distribution of footwear in the private version of the dataset. Of note, 78 participants (≈40%) wore different footwear in session two compared to session one, and 67 participants (≈35%) wore a different *type* of footwear in session two compared to session one. It was assumed that almost all participants wore different footwear to one another. One participant's footwear was unknown in session one because their pictures were corrupted. They were categorized as wearing different footwear between sessions. In terms of walking speed, there were 1892 (≈34%) slower than preferred speed samples, 1900 (≈34%) preferred speed samples, and 1795 (≈32%) faster than preferred speed samples. Even though speed was not measured directly, there was assumed to be inter-individual variation in self-selected speeds within each category [45].

Table 2. Distribution of footwear in the dataset.

Footwear	Session One Count	Session Two Count	Total
Athletic	80	76	156
Flat canvas (slip-on/laced)	45	49	94
Women's ankle boot (flat)	11	14	25
Ballet flat	11	14	25
Men's business	13	11	24
Women's ankle boot (heel)	10	9	19
Men's ankle boot	4	5	9
Sandal	5	4	9
Flip-flop	4	5	9
Loafer	4	4	8
Women's business	2	0	2
Steel capped boot	1	1	2
Five finger	1	1	2
Rubber boot	1	0	1
Unknown	1	0	1

3.2.3. Definition of Training, Validation, and Test Sets

In these experiments, four subsets of the dataset were utilized, each containing a different combination of walking speed and footwear conditions. Each subset was labeled with notation $D_{i,j}$. i corresponds to whether samples from preferred speed only (PS) or all speeds (AS) were included, while j corresponds to whether samples from all IDs (AF), or just those who wore the same footwear between sessions (SF), were included. The subsets were:

- $D_{AS,AF}$—contained samples from all speeds from all IDs (i.e., the entire dataset). This subset allowed same-speed, cross-speed, same-footwear, and cross-footwear comparisons (193 IDs, 5587 samples).
- $D_{AS,SF}$—contained samples from all speeds from IDs who wore the same footwear between sessions. This subset allowed same-speed, cross-speed, and same-footwear comparisons (114 IDs, 3298 samples).
- $D_{PS,AF}$—contained samples from preferred speed from all IDs. This subset allowed same (preferred–preferred) speed, same-footwear, and cross-footwear comparisons (193 IDs, 1900 samples).
- $D_{PS,SF}$—contained samples from preferred speed from IDs who wore the same footwear between sessions. This subset allowed same (preferred–preferred) speed and same-footwear comparisons (114 IDs, 1122 samples).

Each of the data subsets were distributed into training, validation, and test sets. To perform zero-shot re-ID, these sets each contained a distinct group of IDs. To cycle all IDs through validation and test sets, performance was evaluated over seven-fold cross-validation. The distribution of IDs in training, validation, and test sets was approximately 70/15/15% in the first six folds and approximately 80/10/10% in the seventh fold (which contained left-over IDs). This procedure was done once for each subset so that all models were implemented on the same training, validation, and test sets.

3.3. Signal Pre-Processing

As can be seen in Figure 4, the four main steps to the pre-processing method were as follows:

- A 50 N F_z threshold was used to clip all signals to include only the stance phase. For GRFs and GRMs, 20 frames were retained at each end as a buffer. For COP coordinates, an additional 5 % of relative length was excluded at each end to avoid inaccuracies at low force values [46].
- All signals were low-pass filtered (4th order bi-directional Butterworth, cut-off frequency 30 Hz) to minimize high frequency noise. This is common practice for processing time series signals of walking gait kinetics [47]. The precise time points where filtered F_z equaled 50 N were then defined via linear interpolation.
- All signals were time normalized via linear interpolation to time synchronize events within the stance phase and reduce the dimensionality of inputs. A temporal resolution of 300 frames was selected as a conservatively high value, given that a prior study found minimal difference between 100 vs. 1000 frame inputs in a related gait classification task [48].
- Since there were different measurement scales across the eight channels, the features within each channel were standardized to zero mean and unit variance using the means and standard deviations from the training set.
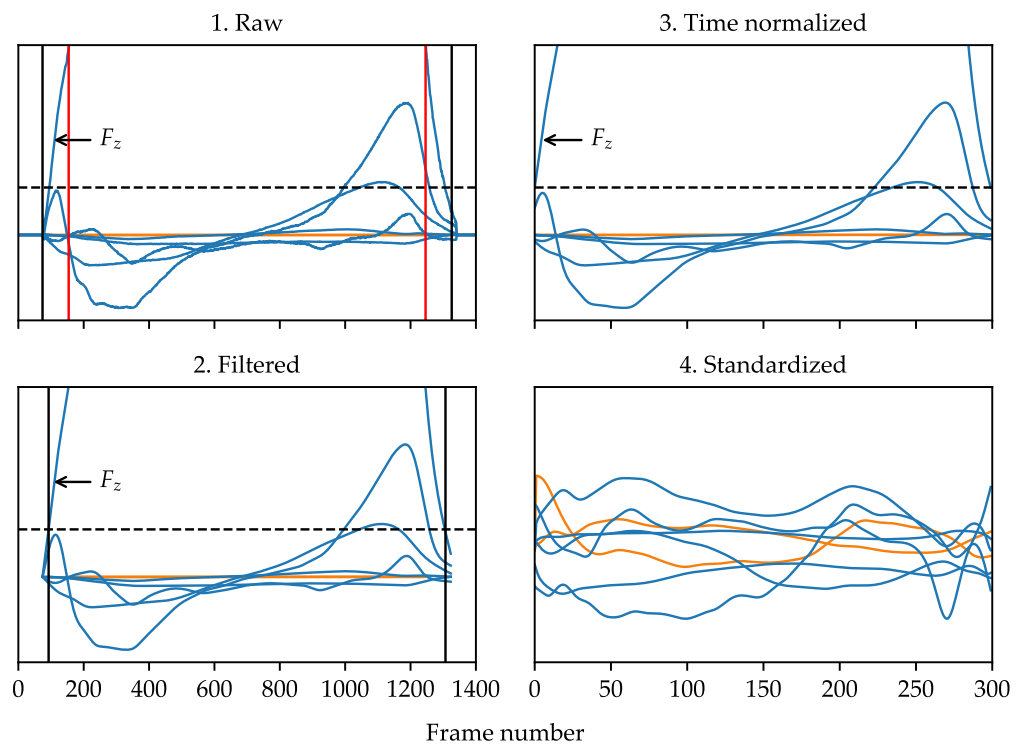
Figure 4. Illustration of the signal pre-processing method for ground reaction forces (GRFs) and ground reaction moments (GRMs) (blue) compared to center of pressure (COP) coordinates (orange). Vertical black lines indicate clip points for GRFs and GRMs. Vertical red lines indicate clip points for COP coordinates. The horizontal dashed line is at 50 N magnitude. The Y-axis limits excluded most of F_z so that other components were easier to see.

3.4. Network Architectures

3.4.1. Overview

To inform future development of DML models for force-platform based re-ID, several baseline DNN architectures of varying type and complexity (in terms of expressive capacity [49]) were implemented (Figure 5):

- FCNN
- CNN
- CLSTMNN
- Convolutional bi-directional long short-term memory neural network (C-Bi-LSTMNN)
- Convolutional transformer neural network (CTNN).

This was the first time a C-Bi-LSTMNN or CTNN had been implemented for force platform-based re-ID. FCNNs, CNNs, and CLSTMNNs had been implemented previously, though comparison between these architectures is difficult due to differences in re-ID system design between studies (e.g., dataset characteristics, loss function, optimizer, and implementation details [18–20,50]). Even if conclusions can be drawn from previous work, they may not generalize to DML models in the zero-shot setting. Hence, in this study, each of the architectures were implemented using the same overall framework to facilitate fair comparison. Ablation experiments were also conducted to determine if input feature selection was warranted under the proposed framework. Each architecture was adapted to accommodate a single channel input and implemented with an individual component of force platform data.
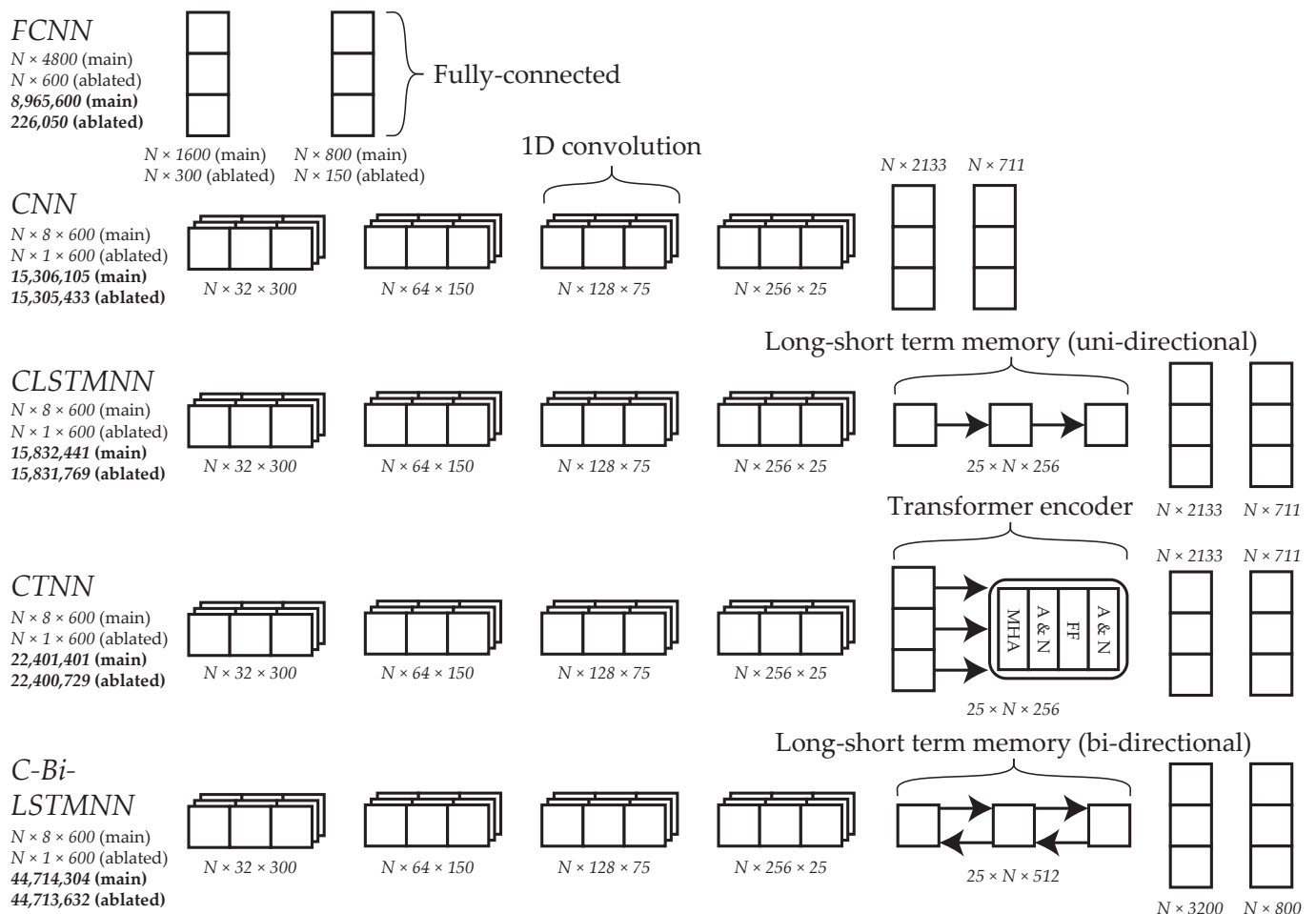
Figure 5. Overview of the DNN architectures used in this study. Shown on the left is the name of each architecture, the shape of its input, and its total number of parameters (top-to-bottom, respectively). The shape notation is number of samples $N \times$ number of features F for the fully connected neural network (FCNN) and number of samples $N \times$ number of channels $C \times$ sequence length L for the others. CNN = convolutional neural network, CLSTMNN = convolutional long short-term memory neural network, CTNN = convolutional transformer neural network, C-Bi-LSTMNN = convolutional bi-directional long short-term memory neural network.

3.4.2. Details

To generate inputs for the FCNN, the pre-processed signals from each force platform were flattened into a feature vector and, thus, the inputs were shaped (number of samples $N \times$ number of features F). This architecture contained two fully connected layers separated by batch normalization (BN) and exponential linear unit (ELU) activation, respectively (referred to herein as a fully connected module) [51,52]. To generate inputs for the remaining architectures, the pre-processed signals from each force platform were concatenated lengthwise and, thus, the inputs were shaped ($N \times$ number of channels $C \times$ sequence length $L(= 600)$). These architectures each contained four 1D convolutional layers at the start (referred to herein as a convolutional module) and a fully connected module at the end. For the convolutions, kernel size was set to three and padding was set to one, such that L remained the same. Initially, C was set to 32 for the first convolution and was then increased by a factor of two for each convolution thereafter ($C = 32 \rightarrow 64 \rightarrow 128 \rightarrow 256$). Between convolutions BN, ELU activation, and local average pooling (LAP) occurred, respectively. Kernel size and stride were set to two for the first three LAP operations, such that L was reduced by a factor of two each time ($L = 300 \rightarrow 150 \rightarrow 75$). These hyper-parameters were set to three for the fourth LAP operation, such that L was reduced by a factor of three ($L = 75 \rightarrow 25$). The CNN just contained a convolutional module and a fully connected module. The CLSTMNN and C-Bi-LSTMNN also contained a long short-term memory (LSTM) layer which was uni-directional in the former and bi-directional in the latter (input and hidden sizes were equal). The CTNN had a transformer encoder layer (with four heads and Gaussian Error Linear Unit (GELU) activation [53]) in between the convolutional and fully connected modules.

3.5. Performance Evaluation

The performance metrics used in this study were accuracy and F1 score [54]. When evaluating performance in a given validation or test set, session two samples were used as queries and session one samples were used as priors, constraining the task to inter-session re-ID. The common method for generating a prediction for a given query is to assign the ID label corresponding to the prior that forms the minimum distance across *all* candidate priors. A more challenging method was implemented in this study, whereby the minimum distance was taken across a subset containing the i-th random prior from each ID. This simulated an initial deployment case, wherein everyone passes through an area once, and then a query ID passes through a second time and must be re-identified. With m as the minimum number of samples per session per ID in a given validation or test set, the process was completed m times, each time with a different subset of priors. This made use of the multiple samples per ID and maximized the number of predictions over which the metrics were computed. For a given query $\mathbf{x}_q$ and set of priors $\mathbf{x}_i$, the prediction was correct if $\|f_{\boldsymbol{\theta}}(\mathbf{x}_q) - f_{\boldsymbol{\theta}}(\mathbf{x}_r)\|_2^2 < \|f_{\boldsymbol{\theta}}(\mathbf{x}_q) - f_{\boldsymbol{\theta}}(\mathbf{x}_i)\|_2^2$, where $\mathbf{x}_r$ is a prior with the same ID label as $\mathbf{x}_q$. Accuracy was defined as the percentage of correct predictions over the total number of predictions. The F1 score was computed using binarized predictions and respective ground truth labels with macro-averaging.

3.6. Hyper-Parameters

Different mini-batch sizes ($N_X = \{32, 64, 128, 256, 512\}$) were compared to determine whether altering the number of candidate positives and negatives during training affected convergence and subsequent test performance. Mini-batches were loaded using a weighted random sampler with the weights being the probabilities of sampling from each ID. The other hyper-parameters were not tuned because the purpose of this set of experiments was to present sensible and comparable baselines, rather than a single fine-tuned model. The μ hyper-parameter for the loss function was set to 0.3, based on previous experiments on person re-ID, where performance was very similar across $\mu = \{0.1, 0.2, 0.5\}$ [40]. Weights were optimized using the AMSGRAD optimizer with default hyper-parameter settings ($\beta = (0.9, 0.999)$, $\epsilon = 1 \times 10^{-8}$, weight decay $= 0$). AMSGRAD was chosen because it has

been shown to guarantee convergence by relying on long-term memory of past gradients when updating model parameters [55]. The number of epochs was set to 1000 at a learning rate of 0.0001. To prevent over-fitting the training set, training was terminated if validation loss did not improve (minimum change $= 0$) over 20 consecutive epochs.

4. Results

4.1. Main Experiments

The results presented herein are for mini-batch size 512 as it led to the best performance (Supplementary Material—Section S2). Accuracy and F1 score were roughly equivalent within each condition, so only accuracy is mentioned in-text. Models with the FCNN architecture were the most accurate on all four subsets (85–92%, depending on the subset in question—Table 3). The amount of variation in mean accuracy, according to architecture, was 2.3 % on $D_{AS,AF}$, 2.7 % on $D_{AS,SF}$, 5.8 % on $D_{PS,AF}$, and 4.2 % on $D_{PS,SF}$. Regardless of the architecture, accuracy was lower on subsets with all IDs compared to subsets with only those who wore the same footwear between sessions (mean difference $= 5.8$ %). In all but one case, accuracy was higher on subsets with samples from all speeds compared to subsets with samples from preferred speed only (mean difference $= 0.97$ %).

Table 3. Accuracy (A) and F1 score (F) of each deep neural network architecture on each data subset (D) over seven-fold cross-validation (provided as mean and min–max range over test sets with the best results in bold). The results for the fully connected neural network (FCNN) architecture on ForceID A are also given as a benchmark. CNN = convolutional neural network, CLSTMNN = convolutional long short-term memeory neural network, CTNN = convolutional transformer neural network, C-Bi-LSTMNN = convolutional bi-directional long short-term memory neural network, AS = all speeds, PS = preferred speed, AF = all footwear, SF = same footwear (between sessions).

Subset	Architecture	Test Performance	
		A (%)	F
$D_{AS,AF}$	FCNN	**85.42(80.76–89.54)**	**0.85(0.80–0.89)**
	CNN	83.12(79.66–87.92)	0.83(0.79–0.87)
	CLSTMNN	83.75(79.18–87.45)	0.83(0.78–0.87)
	CTNN	85.31(79.69–87.97)	0.85(0.79–0.88)
	C-Bi-LSTMNN	85.33(80.31–88.74)	0.85(0.79–0.88)
$D_{AS,SF}$	FCNN	**91.88(85.26–97.18)**	**0.92(0.85–0.97)**
	CNN	89.23(81.92–96.32)	0.89(0.82–0.96)
	CLSTMNN	89.71(82.64–96.45)	0.90(0.82–0.96)
	CTNN	91.10(83.59–97.88)	0.91(0.83–0.98)
	C-Bi-LSTMNN	89.82(81.84–96.36)	0.90(0.82–0.96)
$D_{PS,AF}$	FCNN	**86.77(81.68–90.42)**	**0.86(0.80–0.90)**
	CNN	80.98(76.94–87.74)	0.79(0.74–0.88)
	CLSTMNN	82.34(77.63–88.51)	0.81(0.76–0.88)
	CTNN	82.84(78.02–88.89)	0.82(0.76–0.89)
	C-Bi-LSTMNN	85.12(78.66–89.66)	0.84(0.77–0.89)
$D_{PS,SF}$	FCNN	**91.74(86.03–97.43)**	**0.91(0.85–0.97)**
	CNN	88.91(79.04–95.96)	0.88(0.78–0.96)
	CLSTMNN	88.65(80.88–95.96)	0.88(0.80–0.96)
	CTNN	89.99(77.21–95.59)	0.90(0.77–0.96)
	C-Bi-LSTMNN	87.59(79.78–96.32)	0.87(0.79–0.96)
Force ID A	FCNN	85.57(81.04–90.90)	0.85(0.81–0.91)

The number of correct and incorrect predictions across all models for different walking speeds and footwear comparisons were tallied to determine if there were specific comparisons that generally limited re-ID accuracy on $D_{AS,AF}$. As can be seen in Table 4, accuracy across all models ranged from 68–96%, depending on the type of comparison. Accuracy on cross-footwear comparisons was, on average, 14% lower than that on same-footwear comparisons. Accuracy on cross-speed comparisons was, on average, 8.7% lower than that on same-speed comparisons. Finally, the accuracy on preferred–preferred speed comparisons within $D_{AS,AF}$ was 6.1% higher than the accuracy of preferred–preferred speed comparisons within $D_{PS,AF}$.

The accuracy across all models was calculated on a per-ID basis, and it was found that 44% of all incorrect predictions were attributed to just 20 of the 193 IDs. In this group, the mean per-ID accuracy was 45% (range: 3.0–73%) (Supplementary Material—Table S7). Of the IDs, 85% wore different footwear between sessions and 75% wore a different *type* of footwear between sessions. Of those who wore a different type of footwear between sessions, 75% had a notable difference in heel height (Supplementary Material—Figure S8). When comparing this group to $D_{AS,AF}$ as a whole, the proportions of athletic and flat canvas shoes were 18% and 7% lower, respectively, whereas the proportions of women's ankle boots (heeled), men's business shoes, and women's business shoes were 15%, 14%, and 4% higher, respectively. Finally, the proportion of females in this group was 11% higher (65% female vs. 54% female in $D_{AS,AF}$).

Table 4. Accuracy (A) across all test sets within $D_{AS,AF}$ for specific between-session speed and footwear comparisons. The highest A is in bold.

Footwear	Walking Speed	Walking Speed Comparison	No. Predictions	A (%)
Same	Same	F—F	6760	**95.72**
		S—S	6985	95.26
		P—P	7205	94.13
	Cross	S—P	14,100	92.05
		P—F	13,600	89.49
		S—F	13,600	81.33
Cross	Same	F—F	4345	83.22
		S—S	4830	82.92
		P—P	4940	82.15
	Cross	S—P	9795	78.25
		P—F	9275	71.54
		S—F	9295	68.33
All			104,730	84.45

4.2. Ablations

Across all models, the eight component input led to the highest mean accuracy on all but $D_{PS,SF}$ (Table 5). Using the eight component input, mean accuracy was higher on subsets with samples from all speed categories, compared to subsets with preferred speed samples only. Conversely, using an individual component as the input, mean accuracy was lower on subsets with samples from all speed categories compared to subsets with preferred speed samples only. These relationships held when examining each architecture independently, except in FCNN models (where, for a given input, mean accuracy was nearly always higher on subsets with samples from all speed categories compared to subsets with samples from preferred speed only) and CNN models (where no clear relationship was observed).

Table 5. Accuracy (A) across all models with different inputs (provided as mean and min–max range over test sets with the best results in bold).

Input Component/s	A (%)			
	$D_{AS,AF}$	$D_{AS,SF}$	$D_{PS,AF}$	$D_{PS,SF}$
C_x	32.03(20.96–46.25)	38.27(22.80–52.21)	38.22(26.29–54.79)	41.86(27.21–57.35)
M_z	33.54(22.05–44.67)	38.73(26.00–50.75)	39.67(29.31–51.29)	42.26(21.88–61.40)
M_y	38.39(25.22–56.63)	41.91(20.62–60.67)	39.22(19.40–53.62)	37.30(19.85–56.25)
M_x	48.27(36.82–60.03)	56.74(43.18–68.85)	51.55(29.53–70.39)	48.47(25.37–74.48)
C_y	46.94(34.89–58.25)	65.66(47.63–75.32)	54.93(47.20–64.44)	72.88(59.38–87.13)
F_x	53.90(35.78–68.29)	60.81(44.57–75.95)	65.72(45.69–77.30)	61.78(42.28–85.94)
F_y	56.12(38.42–65.52)	70.41(55.29–77.78)	69.29(55.56–81.91)	81.26(63.60–91.15)
F_z	76.62(66.27–86.53)	85.75(73.86–93.18)	83.11(72.63–94.41)	**91.72(77.21–99.76)**
All	**84.59(79.18–89.54)**	**90.35(81.84–97.88)**	**83.61(76.94–90.42)**	89.38(77.21–97.43)

5. Discussion

In this work, DML was applied for zero-shot re-ID using a large force platform dataset that was purpose built for evaluating person re-ID systems and, thus, had the most complex set of walking conditions. Several baseline DNN architectures of varying types and complexities were implemented on four different data subsets to provide scope for future DNN design and to determine the effects of walking speed and footwear on re-ID performance. The results should be interpreted in the context of the multiple challenges introduced in the evaluation protocol. Namely, re-ID performance was evaluated in the zero-shot setting, with the additional constraints that each query sample had to be from session two and could only be compared with one random prior sample per ID from session one to generate a prediction. To illustrate the effect of constraining the number of priors, mean test accuracy of the FCNN models over seven-fold cross-validation (six sets of 29 IDs and one set of 19 IDs) was 85 % (Table 3) with the constraint applied, yet 94 % without the constraint applied (Supplementary Material—Figure S6 'All'). As this constraint was not applied in previous studies, the value of 94 % (referred to herein as the present benchmark) was used to draw comparisons.

When compared to other studies on force platform-based re-ID, the present benchmark was considerably higher than the previous benchmark at many-shot re-ID conducted across measurement sessions (86 % on 79 IDs), yet lower than the previous benchmarks for many-shot re-ID conducted within a single session (99%+) [18–20,39]. When compared to vision-based re-ID systems, the present benchmark was notably higher than the accuracy of the current state-of-the-art system on CASIA-B under the most comparable conditions (82 %) [27]. Namely, this accuracy was obtained on a similar number of IDs (50) in the condition where clothing differed between measurement instances via the addition of a coat. This provides early evidence that force platform-based re-ID systems could be competitive with video-based re-ID systems. However, caution should be exercised when comparing results between this study and others—particularly those in vision-based re-ID—because there are numerous potential sources of variation, other than those discussed here (e.g., different datasets, input representations, model designs, training and evaluation protocols, and implementation details).

The first main finding of this work was that models with the simplest architecture (i.e., FCNN models) performed the best on all four subsets. All models achieved above 97 % accuracy during training, suggesting a similar level of fit in training sets. However, FCNN models required more epochs (on average) to reach this level of accuracy compared to others (approximate difference: 26–37 on $D_{AS,AF}$, 28–34 on $D_{AS,SF}$, 11–31 on $D_{PS,AF}$, and 9–33 on $D_{PS,SF}$). The fact that FCNN models learnt more slowly than other models (particularly on subsets with more samples) may suggest the discovery of more generic features. Thus, the utility of more complex DNN architectures, such as those implemented in recent works, is called into question when working with datasets containing just a few

thousand samples (even when such datasets contain intra- and inter-individual variations in clothing, walking speed, and footwear, as in $D_{AS,AF}$) [18–20].

The second main finding of this work was that footwear was the main impediment to accuracy across all models on $D_{AS,AF}$. A preliminary investigation into the impact of footwear by Derlatka suggested that changes in footwear between measurement instances complicates force platform-based re-ID using a classification framework [24]. The present results suggest that cross-footwear comparisons also complicate re-ID using DML models on much larger datasets (114–193 IDs, compared to 40). The degree of difficulty appears to scale according to the amount of difference in footwear. This was evidenced by the fact that 85 % of those who were most commonly mistaken wore different footwear between sessions, and 75 % wore different *types* of footwear between sessions. Changes in heel height and sole stiffness, in particular, may pose a challenge, based on the prevalence of changes in heel height and the increased proportion of (presumably) hard-soled footwear (e.g., ankle boots and business shoes) compared to soft-soled footwear (e.g., athletic shoes) in the commonly mistaken group. In 2017, Derlatka compared athletic shoes to high heels and found that the accuracy of a kNN classifier only dropped considerably when high heels were encountered during testing but not training [17]. Thus, the problem posed by footwear type may have been two-fold: shoes that were less represented in the dataset also had characteristics, such as heels, that appear to have complicated re-ID.

The breakdown of accuracy by walking speed revealed why accuracy was slightly higher on $D_{AS,AF}$, compared to $D_{PS,AF}$. Including samples from faster and slower than preferred speed introduced cross-speed comparisons, which had a negative effect on overall accuracy. However, this was offset by a 6.1 % increase in accuracy on preferred–preferred speed comparisons, as well as the inclusion of fast–fast and slow–slow speed comparisons, which were relatively easy. In line with this, Moustakidis, Theocharis, and Giakas found negligible change in classification accuracy on a test set of preferred speed samples when their training set contained samples from all speed categories, compared to just preferred speed samples [16]. These findings suggest that increasing the number of training samples (distributed largely evenly across IDs) can augment test performance, even if the test samples are from a different speed category.

Ablation studies revealed that the eight component input was generally only best on subsets with all speed categories (i.e., $D_{AS,AF}$ and $D_{AS,SF}$). Perhaps this is because the additional information encoded in eight components (compared to one) was only necessary when task complexity increased due to the introduction of cross-speed comparisons. Another possible explanation is that the additional information could only be extracted effectively when the number of samples reached a certain threshold somewhere between 1900 (as for $D_{PS,AF}$) and 3298 (as for $D_{AS,SF}$). In the case of FCNN models, the fact that the eight component input was best on all four subsets suggests that the single channel architecture contained too few parameters to extract meaningful information. As such, the optimal number of parameters to fit the current dataset for zero-shot re-ID could be between that of the single channel FCNN (226,050) and the more complex architectures ($\approx$15,305,000). Up until now, force platform-based re-ID systems have included just a few select components of force platform data (either in their entirety or a reduced set of features thereof), based on their presumed utility for the task [6]. The results from this study suggest that it is more effective to include all components when implementing DNNs on relatively large and complex datasets.

This study was not without limitations. In terms of demographics and experimental protocol, individuals with gait disorders were excluded, despite representing approximately 30 % of the elderly population (60+ years) and presumably a smaller, yet considerable, proportion of the adult population [56]. Data were acquired over a short-term of 3–14 days. Certain applications may require longer term re-ID which has been shown to be more difficult using both video and force platform data [24,57]. Furthermore, object carriage conditions were not included, yet objects alter effective body mass and the location of the center of mass. This causes adaptations in gait that could complicate the re-ID task [58,59].

These conditions simplified the experimental protocol and facilitated investigation into the baseline utility of DML models for force platform-based re-ID. In terms of the dataset, walking trials were only included if each foot landed completely within the sensing area of each force platform. Enforcing this task constraint would be inconvenient in authentication applications because users would likely have to alter stride timing which might alter force patterns [60]. It would be impossible in surveillance applications, because the goal is to have subjects unaware of the system. Finally, in $D_{AS,AF}$ and $D_{AS,SF}$, samples were distributed approximately equally among speed categories to facilitate fair comparison; however, in practice, it would be relatively rare for an individual to walk significantly slower or faster than their preferred speed.

Future work should investigate strategies to improve performance on cross-speed and cross-footwear comparisons without compromising performance elsewhere. Based on the findings from this study, a promising avenue to explore is increasing training set size. It would be valuable to determine the independent effects of training set size (i.e., total number of samples) and sample distribution (between IDs or conditions such as walking speeds). With respect to sample distribution, one recommendation is to include slower and faster than preferred speed samples sparingly to better represent the speed distribution expected in most applications. Other avenues to explore include the following: time-frequency representations; alternative loss functions; more refined DNN architectures; and more advanced training methods (e.g., transfer learning and data augmentation). The independent and combined evaluation of these strategies would provide a clearer picture of the upper limit of performance. On top of this, future work should quantify the effects of gait disorders, object carriage, and partial foot contacts on both short and long-term re-ID, particularly in the challenging zero-shot setting.

6. Conclusions

Force platforms are promising sensors for gait-based person re-ID because they can be applied in settings with limited or no visibility and are a rich source of kinetic information. The purpose of this work was to evaluate several baseline DNN architectures for zero-shot re-ID using a DML framework. This was done on four different subsets of a large, purpose built force platform dataset that contained significant variations in task and environmental constraints. Two-layer FCNN models slightly outperformed models with more complex architectures, achieving 85 % accuracy on the entire dataset in a challenging evaluation setting. A breakdown of performance across all models revealed that re-ID accuracy declined according to the amount of change in individuals' walking speed and footwear between measurement instances. Overall, this work confirmed the baseline utility of DML models for scalable gait-based person re-ID using force platform data.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of The University of Adelaide (protocol code H-2018-009, approved 11 January, 2018) for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available in figshare at https://doi.org/10.25909/14482980.v6. Accessed on 29 September 2022.

Conflicts of Interest: Authors from the Defence Science and Technology Group assisted in the design of the study, interpretation of data, writing of the manuscript, and the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
ID	Identity
re-ID	Re-identification
1D	One-dimensional
2D	Two-dimensional
DML	Deep metric learning
GRF	Ground reaction force
GRM	Ground reaction moment
COP	Center of pressure
F_x	Mediolateral ground reaction force
F_y	Anteroposterior ground reaction force
F_z	Vertical ground reaction force
M_x	Moment about mediolateral axis
M_y	Moment about anteroposterior axis
M_z	Moment about vertical axis
C_x	Mediolateral center of pressure
C_y	Anteroposterior center of pressure
DNN	Deep neural network
FCNN	Fully connected neural network
CNN	Convolutional neural network
LSTM	Long short-term memory
CLSTMNN	Convolutional long short-term memory neural network
CTNN	Convolutional transformer neural network
C-Bi-LSTMNN	Convolutional bi-directional long short-term memory neural network
GREW	Gait Recognition in the Wild
D	Data subset
AS	All speeds
PS	Preferred speed
AF	All footwear
SF	Same footwear (between sessions)
BN	Batch normalization
ELU	Exponential linear unit
LAP	Local average pooling
GELU	Gaussian error linear unit
A	Accuracy
F	F1 score

References

1. Perry, J.; Davids, J.R. Gait analysis: Normal and pathological function. *J. Pediatr. Orthop.* **1992**, *12*, 815. [CrossRef]
2. Stevenage, S.V.; Nixon, M.S.; Vince, K. Visual Analysis of Gait as a Cue to Identity. *Appl. Cogn. Psychol.* **1999**, *13*, 513–526. [CrossRef]
3. Nixon, M.S.; Carter, J.N.; Cunado, D.; Huang, P.S.; Stevenage, S.V. Automatic Gait Recognition. In *Biometrics*; Jain, A.K., Bolle, R., Pankanti, S., Eds.; Springer: Boston, MA, USA, 1999; pp. 231–249. [CrossRef]

4.   Jain, A.K.; Flynn, P.; Ross, A.A. *Handbook of Biometrics*; Springer: Berlin/Heidelberg, Germany, 2007.

5.   Murray, M.P. Gait as a total pattern of movement: Including a bibliography on gait. *Am. J. Phys. Med. Rehabil.* **1967**, *46*, 290–333.

6.   Connor, P.; Ross, A. Biometric recognition by gait: A survey of modalities and features. *Comput. Vis. Image Underst.* **2018**, *167*, 1–27. [CrossRef]

7.   Wan, C.; Wang, L.; Phoha, V.V. A survey on gait recognition. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–35. [CrossRef]

8.   Yu, S.; Tan, D.; Tan, T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Virtual, 20–24 August 2006; Volume 4, pp. 441–444. [CrossRef]

9.   Wu, Z.; Huang, Y.; Wang, L.; Wang, X.; Tan, T. A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 209–226. [CrossRef] [PubMed]

10. Michelini, A.; Eshraghi, A.; Andrysek, J. Two-dimensional video gait analysis: A systematic review of reliability, validity, and best practice considerations. *Prosthetics Orthot. Int.* **2020**, *44*, 245–262. [CrossRef]

11. Song, C.; Huang, Y.; Wang, W.; Wang, L. CASIA-E: A Large Comprehensive Dataset for Gait Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, 1–16. [CrossRef]

12. Winter, D.A.; Patla, A.E.; Frank, J.S. Assessment of balance control in humans. *Med. Prog. Technol.* **1990**, *16*, 31–51.

13. Whittle, M.W. *Gait Analysis: An Introduction*; Butterworth-Heinemann: Oxford, UK, 2014.

14. Cattin, P.; Zlatnik, D.; Borer, R. Sensor fusion for a biometric system using gait. In Proceedings of the Conference Documentation International Conference on Multisensor Fusion and Integration for Intelligent Systems. MFI 2001 (Cat. No. 01TH8590), Baden-Baden, Germany, 20–22 August 2001; pp. 233–238.

15. Jenkins, J.; Ellis, C. Using ground reaction forces from gait analysis: Body mass as a weak biometric. In Proceedings of the International Conference on Pervasive Computing, Atlanta, GA, USA, 18–20 May 2020; Springer: Berlin/Heidelberg, Germany, 2007; pp. 251–267.

16. Moustakidis, S.P.; Theocharis, J.B.; Giakas, G. Feature selection based on a fuzzy complementary criterion: Application to gait recognition using ground reaction forces. *Comput. Methods Biomech. Biomed. Eng.* **2012**, *15*, 627–644. [CrossRef]

17. Derlatka, M. Human gait recognition based on ground reaction forces in case of sport shoes and high heels. In Proceedings of the Proceedings—2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, Poland, 3–5 July 2017; pp. 247–252. [CrossRef]

18. Horst, F.; Lapuschkin, S.; Samek, W.; Müller, K.R.; Schöllhorn, W. Explaining the unique nature of individual gait patterns with deep learning. *Sci. Rep.* **2019**, *9*, 2391. [CrossRef] [PubMed]

19. Yeomans, J.; Thwaites, S.; Robertson, W.S.P.; Booth, D.; Ng, B.; Thewlis, D. Simulating Time-Series Data for Improved Deep Neural Network Performance. *IEEE Access* **2019**, *7*, 131248–131255. [CrossRef]

20. Terrier, P. Gait Recognition via Deep Learning of the Center-of-Pressure Trajectory. *Appl. Sci.* **2020**, *10*, 774. [CrossRef]

21. Sepas-Moghaddam, A.; Etemad, A. Deep gait recognition: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 264–284. [CrossRef] [PubMed]

22. Wang, Z.; Hu, R.; Liang, C.; Yu, Y.; Jiang, J.; Ye, M.; Chen, J.; Leng, Q. Zero-Shot Person Re-identification via Cross-View Consistency. *IEEE Trans. Multimed.* **2016**, *18*, 260–272. [CrossRef]

23. Kaya, M.; Bilge, H.Ş. Deep metric learning: A survey. *Symmetry* **2019**, *11*, 1066. [CrossRef]

24. Derlatka, M. Time Removed Repeated Trials to Test the Quality of a Human Gait Recognition System. In Proceedings of the International Conference on Computer Information Systems and Industrial Management, Bialystok, Poland, 16–18 October 2020; Springer: Cham, Switzerland, 2020; pp. 15–24.

25. Iwama, H.; Okumura, M.; Makihara, Y.; Yagi, Y. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1511–1521. [CrossRef]

26. Takemura, N.; Makihara, Y.; Muramatsu, D.; Echigo, T.; Yagi, Y. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Trans. Comput. Vis. Appl.* **2018**, *10*, 1–14. [CrossRef]

27. Lin, B.; Zhang, S.; Bao, F. Gait recognition with multiple-temporal-scale 3D convolutional neural network. In Proceedings of the 28th ACM International Conference on Multimedia, New York, NY, USA, 12–16 October 2020; Association for Computing Machinery, Inc.: New York, NY, USA, 2020; pp. 3054–3062. [CrossRef]

28. Hou, S.; Cao, C.; Liu, X.; Huang, Y. Gait lateral network: Learning discriminative and compact representations for gait recognition. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part IX; Springer: Berlin/Heidelberg, Germany, 2020; pp. 382–398.

29. Sepas-Moghaddam, A.; Etemad, A. View-invariant gait recognition with attentive recurrent learning of partial representations. *IEEE Trans. Biom. Behav. Identity Sci.* **2020**, *3*, 124–137. [CrossRef]

30. Fan, C.; Peng, Y.; Cao, C.; Liu, X.; Hou, S.; Chi, J.; Huang, Y.; Li, Q.; He, Z. Gaitpart: Temporal part-based model for gait recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14225–14233.

31. Li, X.; Makihara, Y.; Xu, C.; Yagi, Y.; Yu, S.; Ren, M. End-to-end model-based gait recognition. In Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2020.

32. Sepas-Moghaddam, A.; Ghorbani, S.; Troje, N.F.; Etemad, A. Gait recognition using multi-scale partial representation transformation with capsules. In Proceedings of the 2020 25th international conference on pattern recognition (ICPR), Milan, Italy, 10–15 January 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2021; pp. 8045–8052. [CrossRef]

33. Xu, C.; Makihara, Y.; Li, X.; Yagi, Y.; Lu, J. Cross-view gait recognition using pairwise spatial transformer networks. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 260–274. [CrossRef]

34. Zhu, Z.; Guo, X.; Yang, T.; Huang, J.; Deng, J.; Huang, G.; Du, D.; Lu, J.; Zhou, J. Gait Recognition in the Wild: A Benchmark. In Proceedings of the IEEE/CVF international conference on computer vision, Montreal, BC, Canada, 11–17 October 2021; pp. 14789–14799.

35. Liao, R.; Yu, S.; An, W.; Huang, Y. A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognit.* **2020**, *98*, 107069. [CrossRef]

36. Teepe, T.; Khan, A.; Gilg, J.; Herzog, F.; Hörmann, S.; Rigoll, G. Gaitgraph: Graph convolutional network for skeleton-based gait recognition. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 2314–2318. [CrossRef]

37. Shiraga, K.; Makihara, Y.; Muramatsu, D.; Echigo, T.; Yagi, Y. GEINet: View-invariant gait recognition using a convolutional neural network. In Proceedings of the 2016 International Conference on Biometrics (ICB), Halmstad, Sweden, 13–16 June 2016; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2016; pp. 1–8. [CrossRef]

38. Chao, H.; He, Y.; Zhang, J.; Feng, J. GaitSet: Regarding gait as a set for cross-view gait recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 8126–8133. [CrossRef]

39. Derlatka, M.; Borowska, M. Ensemble of heterogeneous base classifiers for human gait recognition. *Sensors* **2023**, *23*, 508. [CrossRef] [PubMed]

40. Hermans, A.; Beyer, L.; Leibe, B. In Defense of the Triplet Loss for Person Re-Identification. *arXiv* **2017**, arXiv:1703.07737.

41. Horst, F.; Slijepcevic, D.; Simak, M.; Schöllhorn, W.I. Gutenberg Gait Database, a ground reaction force database of level overground walking in healthy individuals. *Sci. Data* **2021**, *8*, 232. [CrossRef] [PubMed]

42. Derlatka, M.; Parfieniuk, M. Real-world measurements of ground reaction forces of normal gait of young adults wearing various footwear. *Sci. Data* **2023**, *10*, 60. [CrossRef] [PubMed]

43. Kobayashi, Y.; Hida, N.; Nakajima, K.; Fujimoto, M.; Mochimaru, M. AIST Gait Database 2019, 2019. Available online: https://unit.aist.go.jp/harc/ExPART/GDB2019_e.html (accessed on 16 March 2023).

44. Horsak, B.; Slijepcevic, D.; Raberger, A.M.; Schwab, C.; Worisch, M.; Zeppelzauer, M. GaitRec, a large-scale ground reaction force dataset of healthy and impaired gait. *Sci. Data* **2020**, *7*, 143. [CrossRef]

45. Kuo, A.D.; Donelan, J.M. Dynamic principles of gait and their clinical implications. *Phys. Ther.* **2010**, *90*, 157–174. [CrossRef]

46. Chockalingam, N.; Giakas, G.; Iossifidou, A. Do strain gauge force platforms need in situ correction? *Gait Posture* **2002**, *16*, 233–237. [CrossRef]

47. Yu, B.; Gabriel, D.; Noble, L.; An, K.N. Estimate of the optimum cutoff frequency for the Butterworth low-pass digital filter. *J. Appl. Biomech.* **1999**, *15*, 318–329. [CrossRef]

48. Burdack, J.; Horst, F.; Giesselbach, S.; Hassan, I.; Daffner, S.; Schöllhorn, W.I. Systematic comparison of the influence of different data preprocessing methods on the performance of gait classifications using machine learning. *Front. Bioeng. Biotechnol.* **2020**, *8*, 260. [CrossRef]

49. Hu, X.; Chu, L.; Pei, J.; Liu, W.; Bian, J. Model complexity of deep learning: A survey. *Knowl. Inf. Syst.* **2021**, *63*, 2585–2619. [CrossRef]

50. Derlatka, M. Human gait recognition based on signals from two force plates. In Proceedings of the International Conference on Artificial Intelligence and Soft Computing, Virtual, 21–23 June 2021; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7268 LNAI, pp. 251–258. [CrossRef]

51. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; PMLR: New York City, NY, USA, 2015; Volume 1, pp. 448–456.

52. Clevert, D.A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv* **2015**, arXiv: 1511.07289.

53. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). *arXiv* **2016**, arXiv: 1606.08415.

54. Chinchor, N. MUC-4 evaluation metrics. In Proceedings of the MUC-4—the 4th Conference on Message Understanding, McLean, Virginia, 16–18 June 1992; Association for Computational Linguistics: McLean, Virginia, 1992; pp. 22–29.

55. Reddi, S.J.; Kale, S.; Kumar, S. On the convergence of Adam and beyond. *arXiv* **2019**, arXiv:1904.09237.

56. Mahlknecht, P.; Kiechl, S.; Bloem, B.R.; Willeit, J.; Scherfler, C.; Gasperi, A.; Rungger, G.; Poewe, W.; Seppi, K. Prevalence and Burden of Gait Disorders in Elderly Men and Women Aged 60–97 Years: A Population-Based Study. *PLoS ONE* **2013**, *8*, e69627. [CrossRef] [PubMed]

57. Matovski, D.S.; Nixon, M.S.; Mahmoodi, S.; Carter, J.N. The effect of time on gait recognition performance. *IEEE Trans. Inf. Forensics Secur.* **2011**, *7*, 543–552. [CrossRef]

58. Ren, L.; Jones, R.K.; Howard, D. Dynamic analysis of load carriage biomechanics during level walking. *J. Biomech.* **2005**, *38*, 853–863. [CrossRef]

59. Zhang, X.A.; Ye, M.; Wang, C.T. Effect of unilateral load carriage on postures and gait symmetry in ground reaction force during walking. *Comput. Methods Biomech. Biomed. Eng.* **2010**, *13*, 339–344. [CrossRef] [PubMed]

60. Rastegarpanah, A.; Scone, T.; Saadat, M.; Rastegarpanah, M.; Taylor, S.J.; Sadeghein, N. Targeting effect on gait parameters in healthy individuals and post-stroke hemiparetic individuals. *J. Rehabil. Assist. Technol. Eng.* **2018**, *5*, 205566831876671. [CrossRef]