Journal of Biomechanics 198 (2026) 113149

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech

Predicting human gait kinematics and kinetics from a single inertial measurement unit using deep learning and synthetic datasets: A blinded assessment study

Metin Bicer [a,b,c], Andrew TM Phillips [a], Oğuz Faik Seven [d], Matthew Banger [e], Alison H McGregor [e], Luca Modenese [a,f,*]

[a] Department of Civil and Environmental Engineering, Imperial College London, London, UK
[b] Institute of Informatics, Hacettepe University, Ankara, Türkiye
[c] Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
[d] Department of Mechanical Engineering, Gazi University, Ankara, Türkiye
[e] Department of Surgery and Cancer, Imperial College London, London, UK
[f] School of Biomedical Engineering, University of New South Wales, Sydney, Australia

ARTICLE INFO

Keywords:
Gait
Synthetic mocap dataset
Inertial measurement unit
Blinded assessment

ABSTRACT

A single inertial measurement unit (IMU) can be used with neural networks (NNs) to predict joint kinematics and kinetics. Recent studies, however, often employed virtual IMU data computed from marker-based systems to train NN models, without providing realistic performance assessments on real IMU data. To address this, we present a methodology for developing and blindly evaluating NNs predicting lower-limb joint angles and moments from a sacrum-worn IMU.

NN architectures were trained on IMU data simulated from a public marker-based dataset (49 healthy adults), augmented using conditional generative adversarial networks to enhance variability. The developed NNs were blindly tested against a different dataset (seven healthy adults) of real IMU and synchronous marker-based data collected *ad hoc* after the NN development. The two datasets were collected in different labs using different protocols. These NNs were subsequently fine-tuned (retrained) with this dataset and re-evaluated on another real IMU data (three healthy adults) collected after fine-tuning.

The NNs achieved strong predictive performance on virtual IMU data (average root mean squared error (RMSE) of 2.6±1.3° and 0.10±0.05 Nm/kg for joint angles and moments, respectively). However, performance degraded when applied to real IMU data: average RMSE of 4.5±2.0° for joint angles and 0.21±0.14 Nm/kg for moments. Fine-tuning with real IMU data improved model accuracy, recovering RMSEs to 2.6±0.8° and 0.19 ±0.11 Nm/kg for joint angles and moments, respectively.

Overall, our performance metrics were within the reported ranges for systems employing multiple IMU sensors. This work highlights the importance of blinded assessment and fine-tuning for practical biomechanical applications.

1. Introduction

Marker-based measurements and calculations of joint kinematics and kinetics have long been utilised for providing valuable insights into the biomechanics of human motion (Cappozzo, 1984; Cappozzo et al., 2005; Chèze, 2014; Davis et al., 1991). Yet, because these methods depended on extensive marker configurations and specialized equipment, researchers began investigating more practical and accessible alternatives—among them, inertial measurement units (IMUs), which have since shown considerable promise (Fasel et al., 2017; Seel et al., 2014). Kinematics from IMU data has been extracted by traditional processing methods such as complementary filters or Kalman filters (Gui et al., 2015; Sabatini, 2006). These methods, while effective, faced challenges related to sensor fusion, drift and calibration limiting ease of application (Favre et al., 2009; Picerno, 2017; Robert-Lachaine et al., 2017). The main limitation of using traditional methods was that they

* Corresponding author at: Building E26, Office 1001, School of Biomedical Engineering, University of New South Wales, Sydney, NSW 2033, Australia.
E-mail address: l.modenese@unsw.edu.au (L. Modenese).

https://doi.org/10.1016/j.jbiomech.2026.113149
Accepted 5 January 2026
Available online 6 January 2026
0021-9290/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

required an IMU sensor for each segment and a force plate to calculate lower limb kinematics/kinetics. Although multi-IMU approaches usually yield higher accuracy, a single IMU reduces the setup time, cost, chance of placement errors and participant burden (Fortune et al., 2022), but must compensate for missing segment data. To address this, kinematics and kinetics from IMU data have been predicted using deep learning models, such as neural networks (NNs) (Dorschky et al., 2020; Mundt et al., 2021). Predictive deep learning offers the potential to replace multi-sensor IMU setups with a single device by estimating missing data, yet this strategy remains largely underexplored (Argent et al., 2019; Lim et al., 2019; Mantashloo et al., 2023; Sung et al., 2021). Previous work suggests the lower back, on the belt line (Stenger et al., 2025), as an appropriate location for single IMU placement because of easy placement and practical implications (Godfrey et al., 2015; Hickey et al., 2016), such as the sacrum positioned in the middle. Additionally, the sacrum has been demonstrated as a valid proxy for the centre of mass (Napier et al., 2020), whose movement can be considered representative of all segmental movement and forces and directly linked to ground reaction forces (GRFs) through its acceleration (Lim et al., 2019). Using a single sensor not only streamlines the subject preparation process but also enhances the feasibility of deploying IMU technology in diverse environments (García-de-Villa et al., 2023). However, predictive deep learning applications have yet to be validated on large-scale datasets, as their effectiveness strongly depends on access to extensive and diverse training data. In biomechanics, gold-standard motion capture datasets typically consist of marker and GRF measurements, with no inclusion of synchronous IMU recordings. The limited availability of IMU data for training NNs was partially solved by simulating virtual IMU signals from marker data or computed kinematics (Dorschky et al., 2020; Johnson et al., 2020; Mundt et al., 2020b; Rapp et al., 2021; Sharifi Renani et al., 2021). However, these marker-based datasets were still limited due to lack of access to participants and time/financial constraints (Robinson et al., 2021). To increase the size and further enrich the diversity of these datasets, data augmentation techniques have gained interest. Among these, Generative Adversarial Networks (GANs) and conditional GANs (cGANs) offer a powerful framework for producing synthetic datasets that mimic the statistical characteristics of real data, potentially enhancing prediction accuracy when incorporated into the training process (Bicer et al., 2022; Bicer et al., 2024). However, despite their promise, GAN-generated data have also been primarily evaluated using synthetic IMU signals from marker-based motion capture rather than real-world IMU recordings. As a result, their effectiveness in improving NN performance under real-world conditions remains largely unverified, highlighting a critical gap in current validation efforts.

Moreover, regardless of the size of the dataset used for training NNs or the efficacy of data augmentation, the authors are not aware of any model prediction for which the accuracy was assessed using data collected after model development in a different lab by different operators, i.e. completely unseen data that allow a true *blinded* validation of the model to be performed. Additionally, because training and test datasets in studies developing NNs were of the same characteristics being either real or virtual data, there is limited understanding of the transferability of models trained on virtual IMUs applied to real sensor data.

The primary objective of this work was to develop and assess the performance of a deep learning approach, in the form of a NN model, for prediction of healthy human gait kinematics and kinetics from a single IMU placed at sacrum. For this purpose, first we identified the best performing NN architectures through experiments based on publicly available data (Schreiber and Moissenet, 2019), that we augmented using a generative deep learning approach (Bicer et al., 2024). Then we tested the predictions of the resulting deep learning models on a novel, previously unseen dataset, collected *ad hoc* with the purpose of "blindly" testing the trained models. Finally, in a further development iteration, we fine-tuned the models considering also this new dataset in the training and tested them a last time against a separate set of real IMU data, also collected *ad hoc* during this study.

This work advances prior approaches by: (1) providing a true blinded external validation of IMU-based kinematics and kinetics prediction on real sacrum-mounted IMU data collected later in a different laboratory by different operators; (2) quantifying the contribution of cGAN-based synthetic augmentation to accuracy and robustness; (3) conducting a comparative architecture search on a public motion-capture dataset with simulated sacral IMU signals prior to any exposure to real IMU data; and (4) showing that limited fine-tuning on real-IMU trials further improves performance on an independent test set. Together, these steps provide a minimal single-sensor pipeline with evidence for generalisation beyond the development environment.

2. Methods

Dataset curation and processing, NN training (architecture selection and development), validation and fine-tuning methodology were carried out following the framework (Fig. 1) separately for the lower-limb joint angles and moments. The next sections introduce each component of the workflow.

2.1. Datasets

2.1.1. Literature-dataset

The NN training was performed using a publicly available gait dataset (*literature-dataset*) (Schreiber and Moissenet, 2019) containing overground walking data of 49 healthy adults (male: 22/27 females/males, age: 37.4±13.4 years, mass: 72.0±12.0 kg, height: 1.74±0.09 m, approximately 15 trials per participant). Participants walked overground at multiple speed conditions across a wide range (0.18–2.36 m/s) of walking velocities.

In the development of the NN architecture, the literature-dataset was split into two subsets, for training and testing, following the same methodology in Bicer et al. (2024). The obtained datasets were called *train literature-dataset* and *test literature-dataset* with data from 46 and 3 subjects, respectively. Although this test split proportion (6%) was smaller than the usually employed proportion (~15%), it was balanced by the ad hoc dataset collected (see below) to have 17% of all participants in testing NNs.

2.1.2. Synthetic datasets

The *literature-dataset* was augmented using the Multi-cGAN (conditions: gait parameters and subject anthropometrics) and WS-cGAN (condition: walking speed), developed by Bicer et al. (2024). Each cGAN produced ten synthetic trials for every trial in the training dataset. The synthetic datasets underwent the same pre-processing (joint kinematic/kinetic calculations and IMU data simulations) as the literature-dataset. Based on the cGAN model used, these augmented datasets were named as the *Multi-augmented dataset* and *WS-augmented dataset*. Notably, to prevent information leakage, the test literature dataset was never utilized in training the cGANs employed for augmentation.

2.1.3. Collected-dataset

The data collection for this work took place at the Biodynamics Laboratory of Imperial College London, with ethical approval granted by the Imperial College Research Ethics Committee (Reference: 2IC7598), and informed consent obtained from volunteers. 24 motion capture cameras (Vicon Motion Systems Ltd, Oxford, UK) were used to capture marker trajectories, at 200 Hz, that were placed at anatomical landmarks in addition to marker clusters presented in Fig. S1. An IMU sensor (Delsys Inc., Boston, USA), recording data at 200 Hz, was placed at the sacral position as depicted in Fig. S1 and provided raw accelerometer and gyroscope signals in the sensor local frame aligned with the shown axes in Fig. 2. GRF data were acquired using floor-mounted force plates (sampling frequency: 1000 Hz).

The dataset was collected on two separate occasions to allow blinded

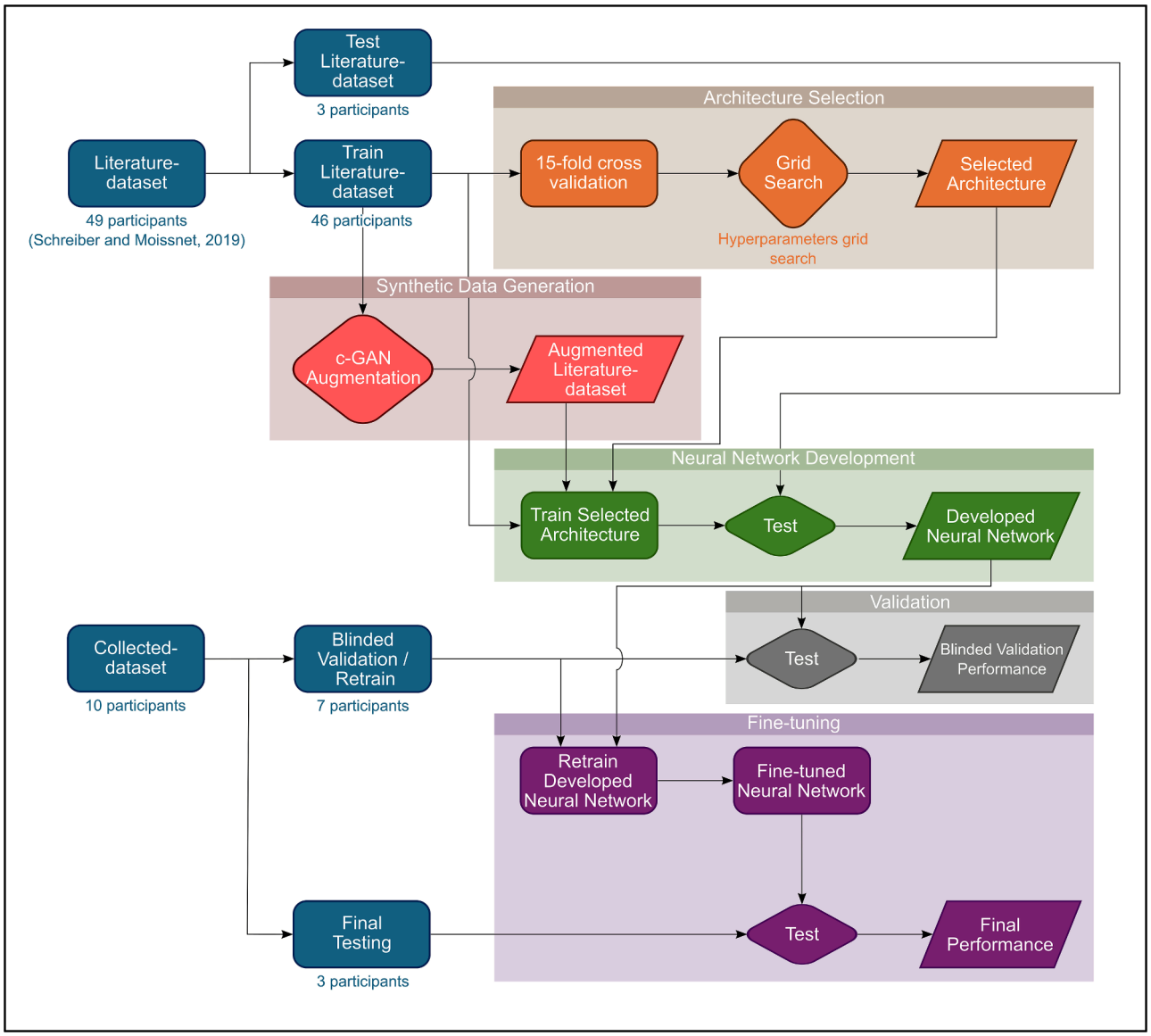


Fig. 1. Framework to train and validate neural network architectures. The same framework was used for joint angle and moment predictions separately. The train literature-dataset was used in 15-fold cross validation strategy in hyperparameter grid search to select the best performing architecture. This dataset was also used in training conditional generative adversarial networks (c-GANs) that augmented the train literature-dataset. The selected architecture was tested using the train literature-dataset and augmented datasets, which led to the developed neural network. This model's performance was tested using the test literature-dataset (unseen in any training level). Next step included blinded evaluation of the developed neural network using data from 7 participants in the collected-dataset. Finally, this data was used to fine-tune (retrain) the developed neural network, which was further blindly tested using data from 3 participants.

evaluation of the developed models. Ten healthy individuals (age: 27.5 ±3.4 years, mass: 77.0±9.9 kg, height: 1.76±0.09 m) performed over-ground walking trials at a self-selected speed wearing comfortable training shoes, with seven participants (six males and one female) collected initially to validate the accuracy of the NNs developed from literature data, and three participants (all females) collected after six months for evaluating the refined NNs considering the collected dataset. The *collected-dataset* was pre-processed following the same methods described for the literature-dataset.

2.1.4. Data processing

For all trials, marker trajectories and GRFs were processed using a generic musculoskeletal model (Rajagopal et al., 2016) to compute joint angles and moments using the inverse kinematics and inverse dynamics tools of OpenSim (Delp et al., 2007; Seth et al., 2018). These angles and moments served as labels during NN training and testing.

For the literature-dataset and synthetic datasets, IMU data were simulated using Simbody (Sherman et al., 2011), the multibody dynamics engine underlying OpenSim. To simulate IMU data, a virtual IMU sensor (Fig. 2) was positioned at the sacral location, aligning its y-axis with the direction of gravity and its z-axis pointing opposite to the direction of motion, similarly to Mundt et al. (2020b). The virtual sensor orientation was assumed to replicate real IMU placement and the noise in the simulated data, derived from the marker trajectories, to match those of real IMU data (Zimmermann et al., 2018). At each time frame, the dynamic state of the musculoskeletal model was realised using the inverse-kinematics-computed joint angles and the joint angular velocities obtained by differentiation of a cubic spline fitted to the joint angles. The virtual IMU position data were also fitted with cubic splines and then differentiated twice to obtain the sensor's linear acceleration. Subsequently, both the angular velocity and linear acceleration of the virtual IMU were transformed into its local reference frame to ensure consistency with a physical device collecting the same data in its local reference frame.

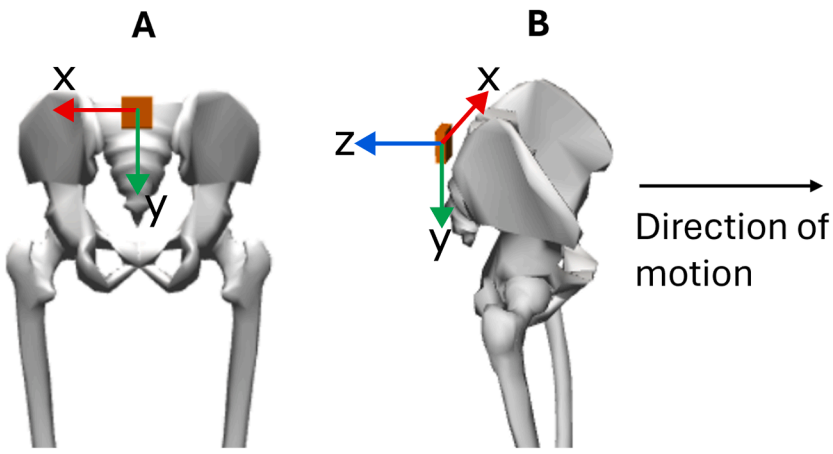


Fig. 2. A virtual IMU sensor is represented with the orange block in rear (A) and side (B) views of the model. The depicted axes in B represent the sensor's local reference frame, with the z-axis (indicated by the blue arrow) directed opposite to the direction of motion, while the y-axis (indicated by the green arrow) aligns with the gravitational vector at the static pose. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Inputs and labels for training neural networks

All NNs were trained using inputs from sacrum-worn IMU data. NNs for kinematic predictions were trained with labels as pelvis (tilt, list, rotation), hip (flexion/extension, abduction/adduction, internal/external rotation), knee (flexion/extension) and ankle (dorsiflexion/plantarflexion) joint angles for the right leg. For kinetic predictions, the labels comprised hip (flexion/extension, abduction/adduction, internal/external rotation), knee (flexion/extension), and ankle (dorsiflexion/plantarflexion) joint moments for the right leg. All joint moments were normalised to subject mass.

2.3. Architecture selection

Feedforward neural networks (FFNNs), Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) were the base candidate architectures, implemented in Pytorch v2.0.0. A hyperparameter grid search was conducted on each network type (Table S1) using the literature-dataset. The hyperparameter search was based on a k-fold cross validation strategy where the train literature-dataset was split into training and validation sets by using k=15, where most folds included three unique validation subjects, while the final fold accommodated an additional one. Data splitting based on subjects ensured a thorough assessment of generalizability for trained architectures, as they were not exposed to data from subjects included in validation sets. The architectures with various hyperparameters (Table S1) were evaluated by average performances across the 15 folds. The architecture and hyperparameters yielding the lowest root-mean squared error (RMSE) averaged over folds, indicating better generalizability on unseen data, were selected for the next phase to develop NNs.

2.4. Neural network development

After selecting the architecture and hyperparameters through the k-fold strategy, two NNs (for joint angles and moments) were separately developed. These NNs were trained from scratch using the train *literature-dataset*, the *Multi-augmented dataset* and the *WS-augmented dataset*, leading to three developed NNs for each task. These NNs were then tested by computing RMSE between their predictions and labels from the test literature-dataset. The networks with the lowest RMSE were designated as the developed NNs and selected for further use to predict the joint angles and moments.

2.5. Validation and fine-tuning of the developed neural network

Validation, fine-tuning and testing of the developed NNs were carried out using the *collected-dataset*. The developed NNs were validated using the first part of the *collected-dataset* (seven subjects), testing their predictions against a dataset collected with a different protocol in a different lab compared to their training dataset. This assessment also evaluated the generalizability of a deep learning model trained using simulated IMU data to real IMU input measurements.

After validation, each developed NN underwent fine-tuning using the first part of the collected-dataset. This fine-tuning process involved continuing the training process from the trained state of the developed NNs (as defined by its weights) at this stage of the workflow (Fig. 1). The fine-tuning was only applied for three training epochs, keeping all hyperparameters of the development phase the same, to prevent overfitting. Since the datasets used for NN development were collected with a specific protocol and included simulated IMU data, the fine-tuning process aimed to enhance training of NNs with real IMU data and integrate protocol-specific adjustments into the developed NNs. Subsequently, the fine-tuned NNs were evaluated using the second part of the collected-dataset.

2.6. Evaluation metrics and statistical comparison

Evaluation metrics for each trained NN were calculated between the predicted and calculated joint angles and moments for each component. These metrics included RMSE, mean absolute error (MAE), correlation coefficient (r) and maximum error expressed as percentage of the range of data (MEP). In addition to scalar error analyses, statistical comparison for each component were conducted using a non-parametric 1D two-tailed paired *t*-test ($\alpha = 0.05$), as implemented in the spm1d package (Pataky et al., 2013). The average percentage of the gait cycle statistically different between predictions and labels was computed.

3. Results

3.1. Architecture selection

The LSTM network exhibited optimal performance in terms of average RMSE and r when predicting joint angles and moments (Fig. S2). The optimal hyperparameter configuration for the LSTM predicting joint angles consisted of a single hidden layer with 512 hidden neurons trained using a batch size of 32 with a learning rate of 0.001. The average RMSE for the joint angle predictions, computed across 15 folds, was determined to be 2.5° (r=0.94) (Table S2). In the prediction of joint kinetics, the LSTM network with 5 hidden layers of 32 hidden neurons trained using a batch size of 32 and a learning rate of 0.001 performed the best. The average RMSE for the joint moment predictions, calculated across 15 folds, was determined to be 0.08 Nm/kg (r=0.92) (Table S2).

3.2. Neural network development

The selected architecture trained using the Multi-augmented dataset (r: 0.96±0.07, RMSE: 2.6±1.3°) outperformed NNs trained with the train literature-dataset and the WS-augmented dataset (Table S3), resulting in average significant differences in only 17±9% of the gait cycle (Fig. 3). In joint moment predictions, minimal differences were observed among predictions of NNs trained with the three datasets (Table S4) (r: 0.93±0.05, RMSE: 0.10±0.05 Nm/kg), thus the Multi-augmented dataset was used to train the developed NN for joint moment predictions. Joint moment predictions differed in 15±8% of the gait cycle (Fig. 4).

3.3. Blind validation and fine-tuning of the developed neural network

Compared to the performance of the developed NNs on the test literature-dataset, their performance on the collected-dataset was degraded for joint angles (r: 0.83±0.20, RMSE: 4.5±2.0°, MAE: 3.8 ±1.6° and MEP: 46±21% in Table 1) and for joint moments (r: 0.86 ±0.13, RMSE: 0.21±0.14 Nm/kg, MAE: 0.15±0.10 Nm/kg, MEP: 41 ±13% in Table 1). On average, 52% and 70% of the gait cycle was significantly different between the predictions and labels for the joint angles (Fig. 5) and moments (Fig. 6), noting that these latter metrics were clearly affected by magnitude offsets.

The developed NNs were fine-tuned, i.e., re-trained starting from the trained weights and biases learnt from the Multi-augmented dataset, using the first part of the collected-dataset. Testing on the second part of the collected-dataset showed better performance of the NNs fine-tuned with the real IMU data (r: 0.85±0.25, RMSE: 2.6±0.8°, MAE: 2.2 ±0.7°, MEP: 36±33% for joint angles in Table 2 and r: 0.88±0.09, RMSE: 0.19±0.11 Nm/kg, MAE: 0.14±0.08 Nm/kg, MEP: 38±13% for joint moments in Table 2) compared to the NNs developed using the simulated IMU data. 44±24% and 63±21% of the gait cycle significantly differed between predictions and labels for the joint angles (Fig. 7) and moments (Fig. 8), respectively.

4. Discussion

In this study, the primary aim was to use predictive deep learning techniques, especially NNs, for the estimation of healthy human gait kinematics and kinetics from a single IMU worn at sacrum. The approach involved the development and validation of NN models trained on simulated IMU data obtained from marker-based measurements, supplemented by synthetic datasets generated using cGANs developed by Bicer et al. (2024). The developed NN models then underwent a blinded

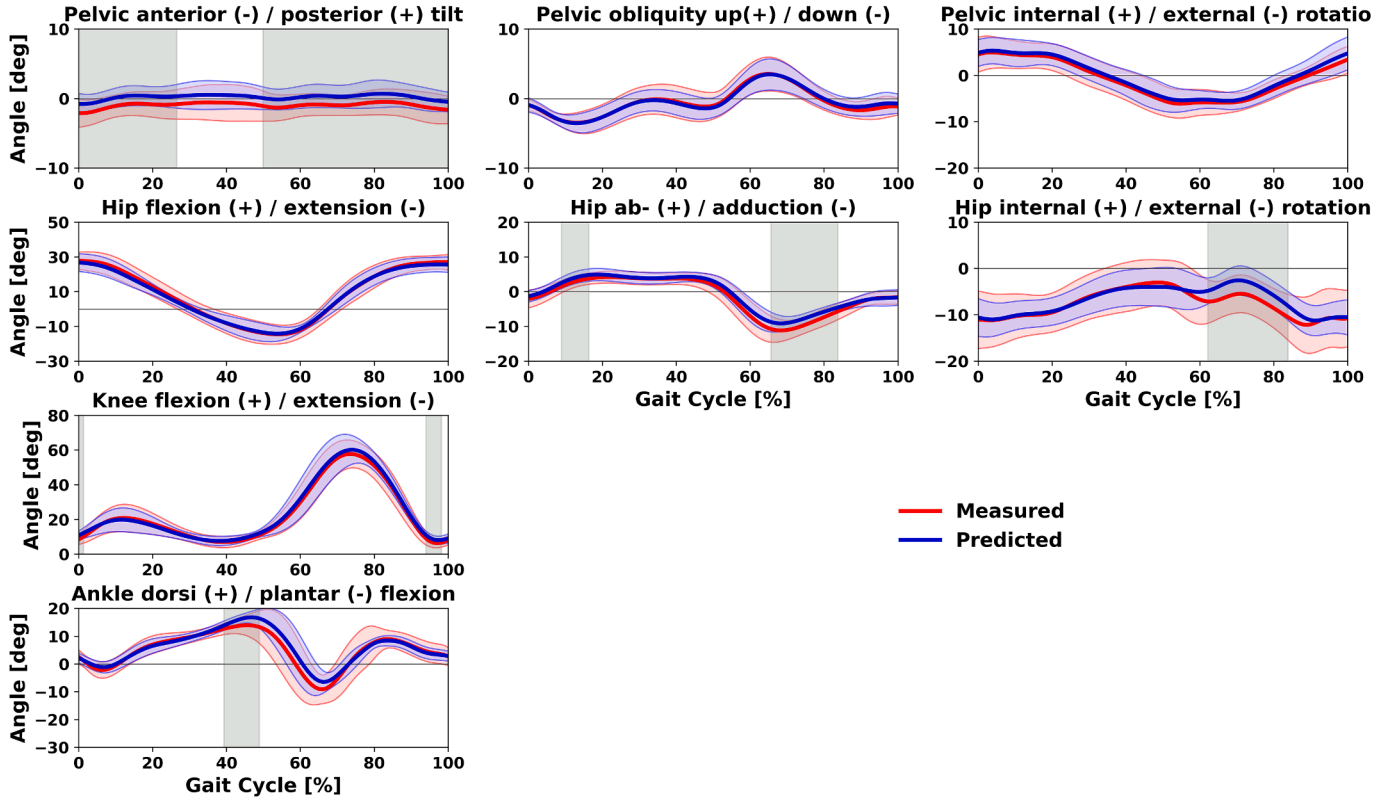


Fig. 3. Measured (red) and predicted (blue) joint angles for the test literature-dataset. Predictions are made by the neural network trained with the Multi-augmented dataset. Grey shaded areas indicates significant differences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

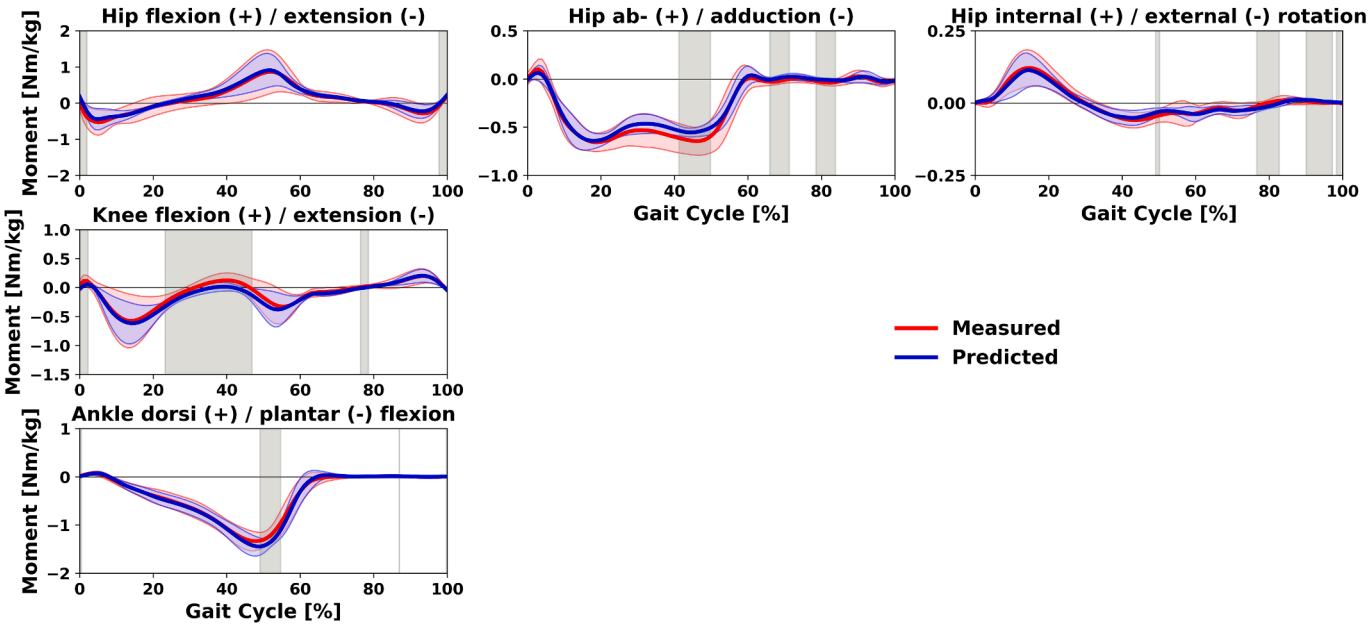


Fig. 4. Measured (red) and predicted (blue) joint moments for the test literature-dataset. Predictions are made by the neural network trained with the Multi-augmented dataset. Grey shaded areas indicates significant differences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Error metrics (r: correlation coefficient, RMSE: root mean squared error, MAE: mean absolute error, MEP: maximum error expressed as percentage of the range of data) for joint angle and moment predictions on the first part of the collected-dataset by the developed neural network trained using the Multi-augmented dataset. Average and standard deviation (SD) across all joint angles or moments are also reported.

	Joint Angle Predictions				Joint Moment Predictions			
	r	RMSE [deg]	MAE [deg]	MEP [%]	r	RMSE [Nm/kg]	MAE [Nm/kg]	MEP [%]
pelvis tilt	0.39	2.2	1.9	92				
pelvis list	0.86	2.0	1.7	40				
pelvis rotation	0.91	3.6	3.3	45				
hip flexion	0.97	5.4	4.6	23	0.93	0.41	0.29	35
hip adduction	0.95	4.3	3.8	42	0.90	0.20	0.16	56
hip rotation	0.71	4.7	4.0	55	0.87	0.03	0.02	36
knee angle	0.96	8.3	6.6	26	0.63	0.21	0.15	52
ankle angle	0.85	5.3	4.3	42	0.95	0.18	0.12	24
Average (SD)	0.83 (0.20)	4.5 (2.0)	3.8 (1.6)	46 (21)	0.86 (0.13)	0.21 (0.14)	0.15 (0.10)	41 (13)

assessment and fine-tuned using real IMU data collected in a different lab with a different protocol with respect to the training dataset.

The results obtained from the NN training experiments demonstrated efficacy in predicting joint angles and moments. Specifically, the LSTM network outperformed the other architectures. Although this finding might appear contradictory to a previous study by Mundt et al. (2021), which found that CNNs were better suited for joint angle predictions and FFNNs for joint moment predictions, differences (data collection/processing protocols and measurement equipment) in training/testing datasets between these two studies make a direct comparisons impractical. Additionally, these findings align with previous research suggesting the effectiveness of LSTM networks in capturing temporal dependencies within sequential data (Mao et al., 2021; Wentz et al., 2022), which is particularly relevant in the context of human movement analysis. In particular, our results confirm the idea that a single IMU on the sacrum is sufficient for predicting lower limb joint kinematics and kinetics with accuracy comparable to previous studies using multiple sensors (Moghadam et al., 2023; Mundt et al., 2020b; Mundt et al., 2020c).

Furthermore, the presented results suggest that the inclusion of synthetic datasets augmented with GANs, particularly the *Multi-augmented dataset*, lead to improved performance in joint angle predictions compared to models trained solely on the literature-dataset, with improvements up to 35% and 25% for pelvis list and hip flexion respectively (Table S3). This finding highlighted the potential of data augmentation techniques in enhancing the generalization of predictive NNs for biomechanical analyses.

The performance of the developed NN models degraded noticeably from the test literature-dataset to the collected-dataset, particularly in joint angle predictions (Table 1). The accuracy of the real-IMU-based assessment was 50% inferior (for both joint angles and moment) compared to using simulated IMU data from the markers of the same *collected-dataset*. This result highlights that NNs trained on simulated IMU data can realistically be evaluated for real usage only using experimental IMU data, and not simulated ones like in previous studies (Dorschky et al., 2019; Mundt et al., 2021; Mundt et al., 2020a; Nitschke et al., 2024). The performance degradation was attributed to a variety of sources including scarce NN generalisation from simulated to real IMU recordings and substantial differences in the protocols, marker sets, equipment and measurement noise between the literature-based (including synthetic datasets) and collected-dataset. By considering these differences in our workflow, we presented, for the first time in the literature, a fair evaluation of NNs concerning IMUs and joint angle/moment estimations and provided an expected performance of the developed NNs on a dataset of real IMU recordings. The presented procedure, in other terms, is a genuine form of blind evaluation,

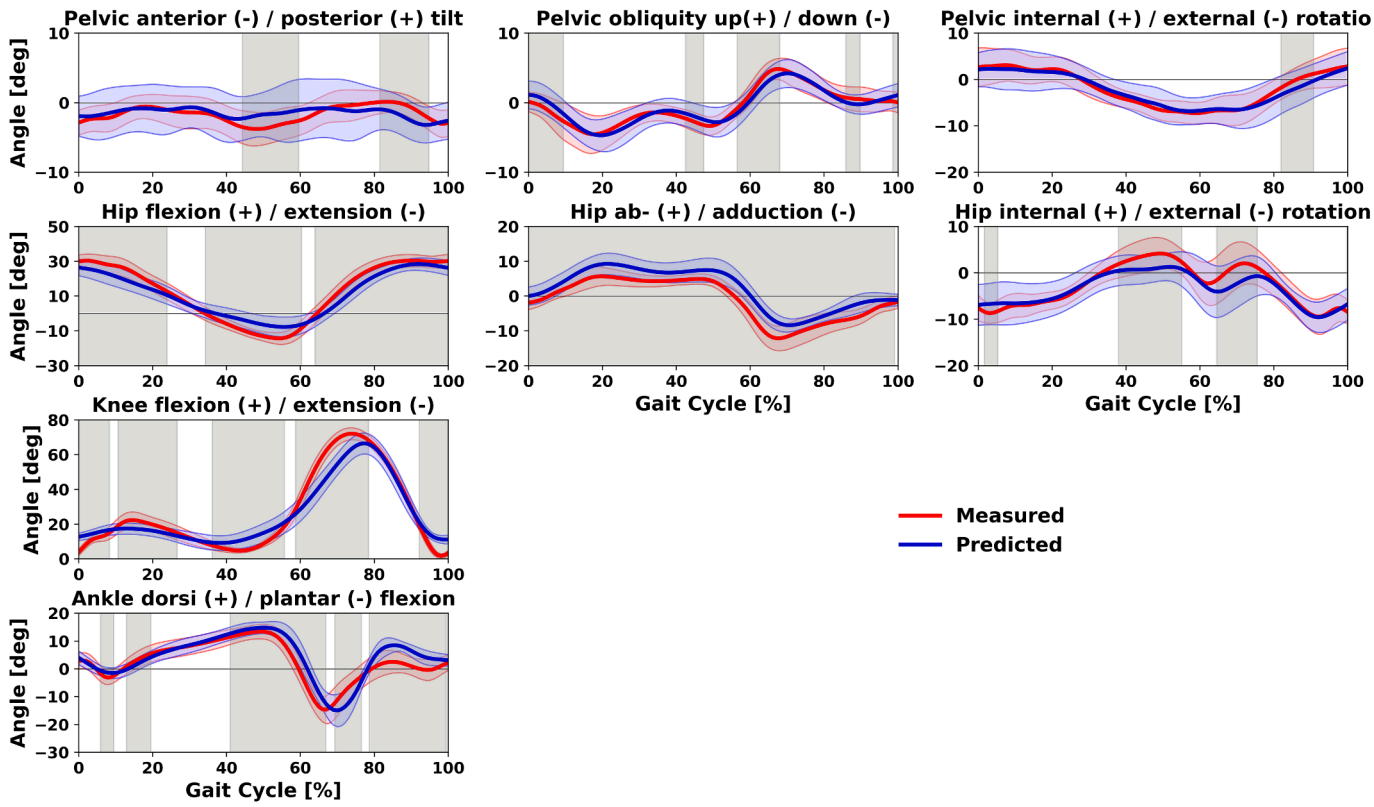


Fig. 5. Measured (red) and predicted (blue) joint angles for the first part of the collected-dataset. Predictions are made by the developed neural network trained with the Multi-augmented dataset. Grey shaded areas indicates significant differences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

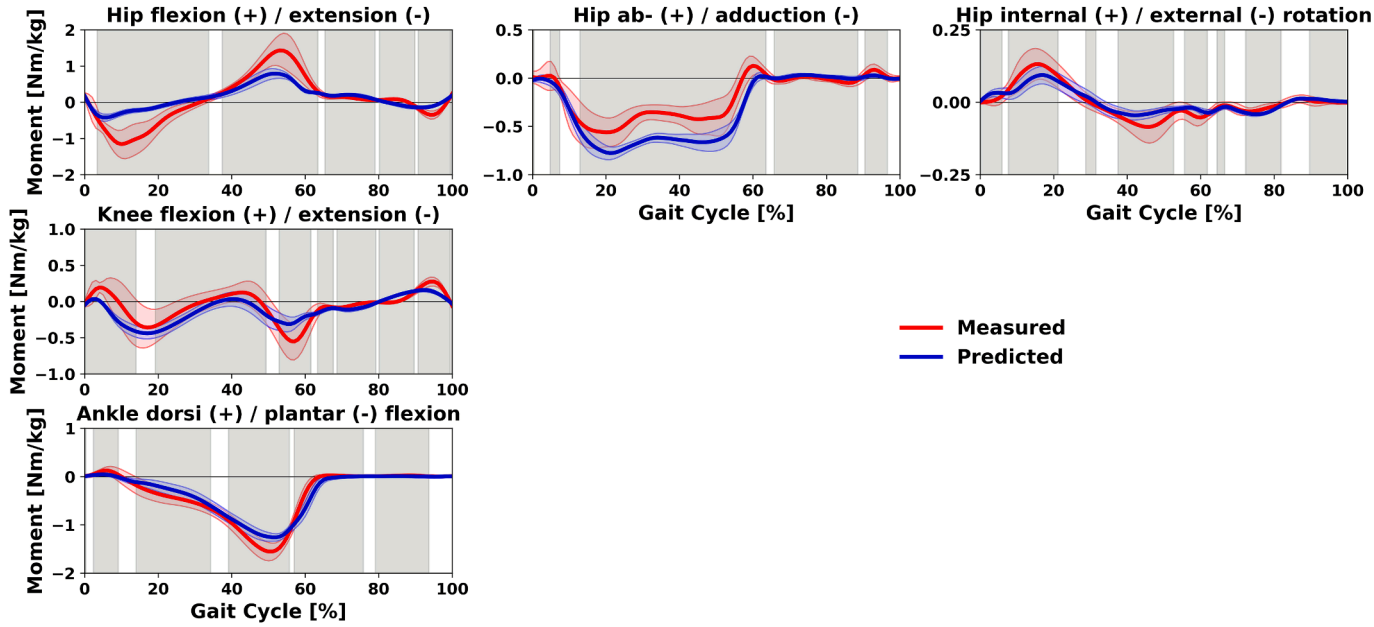


Fig. 6. Measured (red) and predicted (blue) joint moments for the first part of the collected-dataset. Predictions are made by the developed neural network trained with the Multi-augmented dataset. Grey shaded areas indicates significant differences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

providing an unbiased assessment of model performance on previously unseen real-world data, an aspect not addressed in prior studies (Caldas et al., 2017) due to lack of appropriate data for evaluation (García-de-Villa et al., 2023).

The evaluation on the second part of the collected-dataset (Table 2) indicated the importance of fine-tuning NN models using real IMU data to improve their performance and robustness, and to optimise them for specific experimental setups. The performance of the developed and fine-tuned NNs were in the range (RMSE: 2.2–6.5°) shown in literature for using multiple IMUs (Moghadam et al., 2023; Mundt et al., 2020b;

Table 2
Error metrics (r: correlation coefficient, RMSE: root mean squared error, MAE: mean absolute error, MEP: maximum error expressed as percentage of the range of data) for joint angle and moment predictions on the second part of the collected-dataset by the fine-tuned neural network. Average and standard deviation (SD) across all joint angles or moments are also reported.

	Joint Angle Predictions				Joint Moment Predictions			
	r	RMSE [deg]	MAE [deg]	MEP [%]	r	RMSE [Nm/kg]	MAE [Nm/kg]	MEP [%]
pelvis tilt	0.27	1.7	1.4	112				
pelvis list	0.93	1.8	1.5	29				
pelvis rotation	0.97	2.0	1.8	29				
hip flexion	1.00	2.4	2.0	10	0.93	0.33	0.23	50
hip adduction	0.95	2.6	2.2	27	0.87	0.21	0.17	38
hip rotation	0.76	4.3	3.7	52	0.89	0.04	0.03	34
knee angle	0.99	3.1	2.6	9	0.74	0.24	0.16	48
ankle angle	0.96	2.8	2.2	22	0.98	0.13	0.09	18
Average (SD)	0.85 (0.25)	2.6 (0.8)	2.2 (0.7)	36 (33)	0.88 (0.09)	0.19 (0.11)	0.14 (0.08)	38 (13)

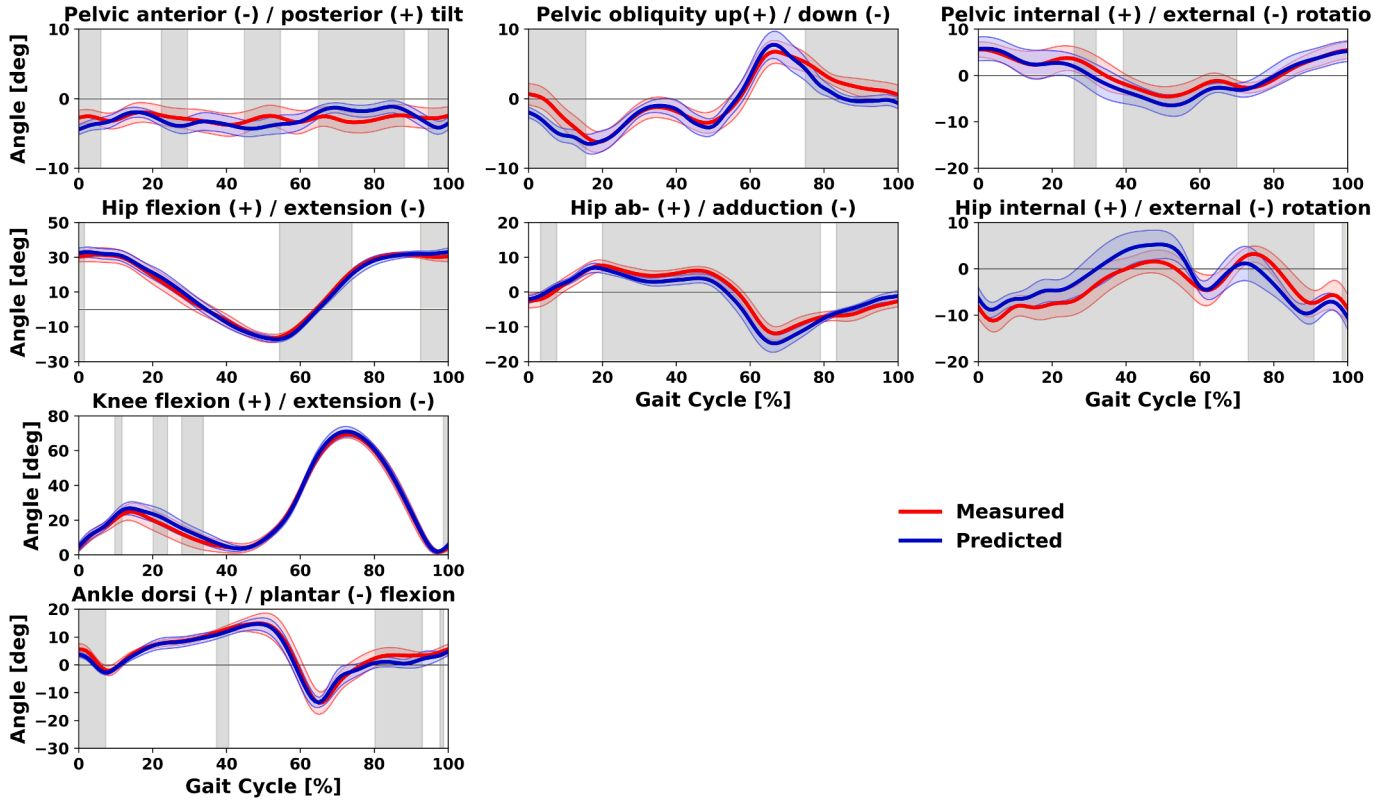


Fig. 7. Measured (red) and predicted (blue) joint angles for the second part of the collected-dataset. Predictions are made by the fine-tuned neural network. Grey shaded areas indicates significant differences. Grey shaded areas indicates significant differences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Mundt et al., 2020c). Additionally, RMSEs for sagittal plane kinematics were lower than those obtained using a single-IMU (RMSEs of 6.2°, 7.0°, and 4.4° for the hip, knee and ankle angles) by Sung et al. (2021). Additionally, the RMSE for kinetics in this study is consistent with previous reports (Moghadam et al., 2023). While statistical differences were observed across a large portion of the gait cycle, the overall waveform similarity remained high, as reflected by strong r values, suggesting that the differences detected by the SPM statistical test were largely due to systematic offsets rather than shape mismatches.

One of the limitations of this study lies in the relatively small size of the collected-dataset used to assess the developed NNs, especially for the final evaluation of the fine-tuned models, which presented all female participants. This may have impacted the generalizability of the presented findings, but also potentially cause an underestimation of the NNs performance in the last assessment. Nevertheless, the number of subjects employed for evaluation was more than those seen in the literature (Dorschky et al., 2020; Mundt et al., 2021; Sharifi Renani et al., 2021; Sung et al., 2021). Another limitation is that the data collection was confined to a laboratory environment, primarily due to the necessity of collecting marker and GRF data to calculate labels (ground-truths), which may influence NNs performance on a dataset collected in outdoor settings. Moreover, we did not consider uncertainties due to the manual positioning of the IMU sensors, e.g. intra- and inter-operator variability, which could be represented in the training data through perturbations of the position and orientation of the marker-based virtual IMUs. Nevertheless, other sources of variability, such as anthropometrics and walking speed, were accounted for in the synthetic datasets through appropriate generation of representative data.

Our blinded, cross-site evaluation demonstrated robust performance within the studied domain (level overground walking in healthy adults); however, cohort- and task-level generalization remain key priorities for

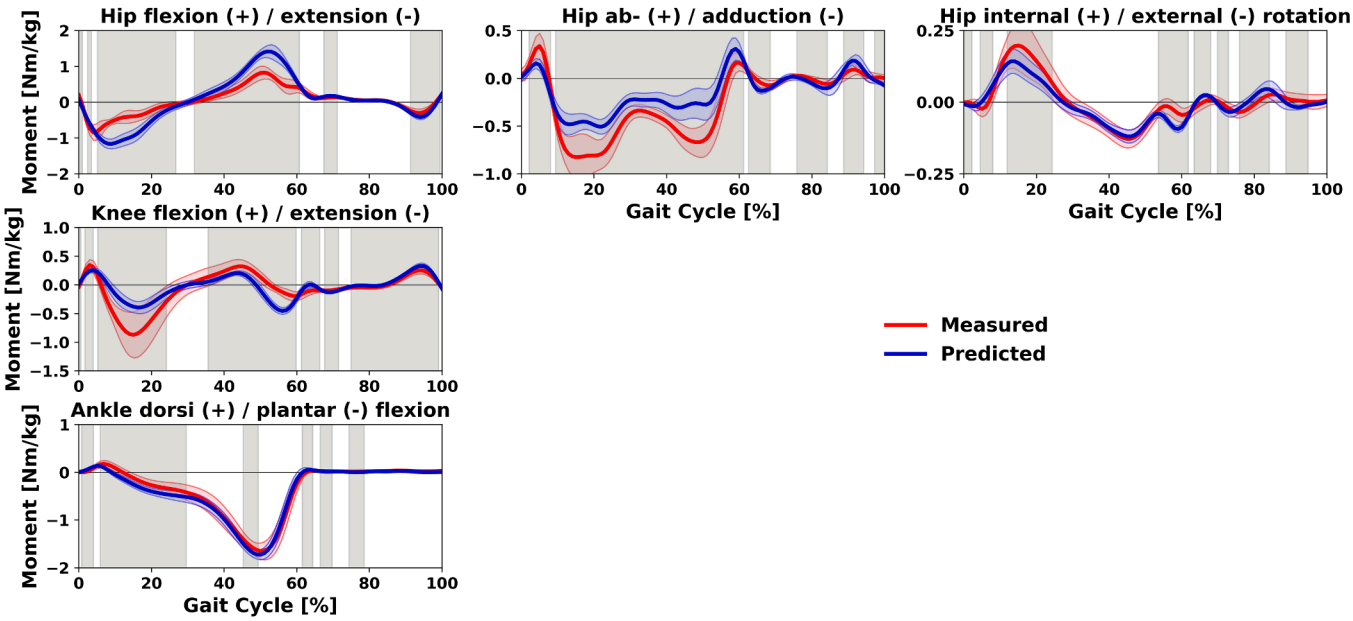


Fig. 8. Measured (red) and Predicted (blue) joint moments for the second part of the collected-dataset. Predictions are made by the fine-tuned neural network. Grey shaded areas indicates significant differences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

future work. Regarding cohort generalizability, our initial datasets comprised exclusively healthy adults, leaving model performance on pathological gait untested. While previous research has cited limited training diversity as a general concern (Marimon et al., 2024; Senanayake et al., 2021), we propose that our results suggest a practical solution: targeted fine-tuning on small, cohort-specific real-sensor datasets could be leveraged to adapt the models and potentially recover accuracy on clinical populations. Concerning task generalizability, the reported metrics do not establish validity for non-walking activities, which is a limitation of the current study. A critical direction for future work is therefore the extension of the proposed pipeline (data augmentation, blinded external testing, limited cohort/task-specific fine-tuning) to additional gait activities, such as stairs or slopes and diverse clinical cohorts.

Overall, this study contributes to the ongoing efforts in advancing IMU-based human movement analysis by demonstrating the potential of predictive deep learning techniques in capturing and predicting complex movement patterns. Achieving this using a single IMU sensor worn on the sacrum reduces the complexity of subject preparation and supports more practical deployment of IMU-based monitoring across different settings. By aiming to bridge the gap between marker-based measurements and IMU-based analysis, the framework lays the foundation for practical and accessible biomechanical assessments.

CRediT authorship contribution statement

Metin Bicer: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Andrew TM Phillips:** Writing – review & editing, Supervision. **Oğuz Faik Seven:** Writing – review & editing, Methodology. **Matthew Banger:** Writing – review & editing, Resources, Data curation. **Alison H McGregor:** Writing – review & editing, Resources. **Luca Modenese:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by a Postgraduate Scholarship granted by the Ministry of National Education of Türkiye to MB and a Scientia Fellowship granted by the University of New South Wales to LM.

Appendix

Codes for model training and evaluation are available at https://github.com/metinbicer/imu_nn_paper. Participant data cannot be publicly released as the study protocol and informed consent agreements do not permit its distribution.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbiomech.2026.113149.

References

Argent, R., Drummond, S., Remus, A., O'Reilly, M., Caulfield, B., 2019. Evaluating the use of machine learning in the assessment of joint angle using a single inertial sensor. Journal of Rehabilitation and Assistive Technologies Engineering 6, 2055668319868544.

Bicer, M., Phillips, A.T.M., Melis, A., McGregor, A.H., Modenese, L., 2022. Generative deep learning applied to biomechanics: a new augmentation technique for motion capture datasets. J. Biomech. 144, 111301.

Bicer, M., Phillips, A.T.M., Melis, A., McGregor, A.H., Modenese, L., 2024. Generative adversarial networks to create synthetic motion capture datasets including subject and gait characteristics. J. Biomech. 177, 112358.

Caldas, R., Mundt, M., Potthast, W., de Lima Neto, F.B., Markert, B., 2017. A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. Gait Posture 57, 204–210.

Cappozzo, A., 1984. Gait analysis methodology. Hum. Mov. Sci. 3, 27–50.

Cappozzo, A., Della Croce, U., Leardini, A., Chiari, L., 2005. Human movement analysis using stereophotogrammetry: Part 1: Theoretical background. Gait Posture 21, 186–196.

Chèze, L., 2014. Kinematic analysis of human movement, Chapters 1&2. John Wiley & Sons.

Davis III, R.B., Ounpuu, S., Tyburski, D., Gage, J.R., 1991. A gait analysis data collection and reduction technique. Hum. Mov. Sci. 10, 575–587.

Delp, S.L., Anderson, F.C., Arnold, A.S., Loan, P., Habib, A., John, C.T., Guendelman, E., Thelen, D.G., 2007. Opensim: Open-source software to create and analyze dynamic simulations of movement. I.E.E.E. Trans. Biomed. Eng. 54, 1940–1950.
Dorschky, E., Nitschke, M., Martindale, C.F., Van den Bogert, A.J., Koelewijn, A.D., Eskofier, B.M., 2020. Cnn-based estimation of sagittal plane walking and running biomechanics from measured and simulated inertial sensor data. Front. Bioeng. Biotechnol. 8, 604.
Dorschky, E., Nitschke, M., Seifer, A.-K., van den Bogert, A.J., Eskofier, B.M., 2019. Estimation of gait kinematics and kinetics from inertial sensor data using optimal control of musculoskeletal models. J. Biomech. 95, 109278.
Fasel, B., Spörri, J., Schütz, P., Lorenzetti, S., Aminian, K., 2017. An inertial sensor-based method for estimating the athlete's relative joint center positions and center of mass kinematics in alpine ski racing. Front. Physiol. 8, 850.
Favre, J., Aissaoui, R., Jolles, B.M., de Guise, J.A., Aminian, K., 2009. Functional calibration procedure for 3d knee joint angle description using inertial sensors. J. Biomech. 42, 2330–2335.
Fortune, E., Cloud-Biebl, B.A., Madansingh, S.I., Ngufor, C.G., Van Straaten, M.G., Goodwin, B.M., Murphree, D.H., Zhao, K.D., Morrow, M.M., 2022. Estimation of manual wheelchair-based activities in the free-living environment using a neural network model with inertial body-worn sensors. J. Electromyogr. Kinesiol. 62, 102337.
García-de-Villa, S., Casillas-Pérez, D., Jiménez-Martín, A., García-Domínguez, J.J., 2023. Inertial sensors for human motion analysis: a comprehensive review. IEEE Trans. Instrum. Meas. 72, 1–39.
Godfrey, A., Del Din, S., Barry, G., Mathers, J., Rochester, L., 2015. Instrumenting gait with an accelerometer: a system and algorithm examination. Med. Eng. Phys. 37, 400–407.
Gui, P., Tang, L., Mukhopadhyay, S., 2015. Mems based imu for tilting measurement: Comparison of complementary and kalman filter based data fusion. In: 2015 IEEE 10th Conference on Industrial Electronics and Applications, pp. 2004–2009.
Hickey, A., Del Din, S., Rochester, L., Godfrey, A., 2016. Detecting free-living steps and walking bouts: Validating an algorithm for macro gait analysis. Physiol. Meas. 38, N1.
Johnson, W.R., Mian, A., Robinson, M.A., Verheul, J., Lloyd, D.G., Alderson, J.A., 2020. Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning. I.E.E.E. Trans. Biomed. Eng. 68, 289–297.
Lim, H., Kim, B., Park, S., 2019. Prediction of lower limb kinetics and kinematics during walking by a single imu on the lower back using machine learning. Sensors 20, 130.
Mantashloo, Z., Abbasi, A., Tazji, M.K., Pedram, M.M., 2023. Lower body kinematics estimation during walking using an accelerometer. J. Biomech. 151, 111548.
Mao, G., Wang, M., Liu, J., Wang, Z., Wang, K., Meng, Y., Zhong, R., Wang, H., Li, Y., 2021. Comprehensive comparison of artificial neural networks and long short-term memory networks for rainfall-runoff simulation. Physics and Chemistry of the Earth, Parts A/b/c 123, 103026.
Marimon, X., Mengual, I., López-de-Celis, C., Portela, A., Rodríguez-Sanz, J., Herráez, I. A., Pérez-Bellmunt, A., 2024. Kinematic analysis of human gait in healthy young adults using imu sensors: Exploring relevant machine learning features for clinical applications. Bioengineering 11, 105.
Moghadam, S.M., Yeung, T., Choisne, J., 2023. A comparison of machine learning models' accuracy in predicting lower-limb joints' kinematics, kinetics, and muscle forces from wearable sensors. Sci. Rep. 13, 5046.
Mundt, M., Johnson, W.R., Potthast, W., Markert, B., Mian, A., Alderson, J., 2021. A comparison of three neural network approaches for estimating joint angles and moments from inertial measurement units. Sensors 21, 4535.
Mundt, M., Koeppe, A., Bamer, F., David, S., Markert, B., 2020a. Artificial neural networks in motion analysis—applications of unsupervised and heuristic feature selection techniques. Sensors 20, 4581.
Mundt, M., Koeppe, A., David, S., Witter, T., Bamer, F., Potthast, W., Markert, B., 2020b. Estimation of gait mechanics based on simulated and measured imu data using an artificial neural network. Front. Bioeng. Biotechnol. 8, 41.
Mundt, M., Thomsen, W., Witter, T., Koeppe, A., David, S., Bamer, F., Potthast, W., Markert, B., 2020c. Prediction of lower limb joint angles and moments during gait using artificial neural networks. Med. Biol. Eng. Compu. 58, 211–225.
Napier, C., Jiang, X., MacLean, C.L., Menon, C., Hunt, M.A., 2020. The use of a single sacral marker method to approximate the centre of mass trajectory during treadmill running. J. Biomech. 108, 109886.
Nitschke, M., Dorschky, E., Leyendecker, S., Eskofier, B.M., Koelewijn, A.D., 2024. Estimating 3d kinematics and kinetics from virtual inertial sensor data through musculoskeletal movement simulations. Front. Bioeng. Biotechnol. 12, 1285845.
Pataky, T.C., Robinson, M.A., Vanrenterghem, J., 2013. Vector field statistical analysis of kinematic and force trajectories. J. Biomech. 46, 2394–2401.
Picerno, P., 2017. 25 years of lower limb joint kinematics by using inertial and magnetic sensors: a review of methodological approaches. Gait Posture 51, 239–246.
Rajagopal, A., Dembia, C.L., DeMers, M.S., Delp, D.D., Hicks, J.L., Delp, S.L., 2016. Full-body musculoskeletal model for muscle-driven simulation of human gait. I.E.E.E. Trans. Biomed. Eng. 63, 2068–2079.
Rapp, E., Shin, S., Thomsen, W., Ferber, R., Halilaj, E., 2021. Estimation of kinematics from inertial measurement units using a combined deep learning and optimization framework. J. Biomech. 116, 110229.
Robert-Lachaine, X., Mecheri, H., Larue, C., Plamondon, A., 2017. Accuracy and repeatability of single-pose calibration of inertial measurement units for whole-body motion analysis. Gait Posture 54, 80–86.
Robinson, M.A., Vanrenterghem, J., Pataky, T.C., 2021. Sample size estimation for biomechanical waveforms: current practice, recommendations and a comparison to discrete power analysis. J. Biomech. 122, 110451.
Sabatini, A.M., 2006. Quaternion-based extended kalman filter for determining orientation by inertial and magnetic sensing. IEEE Trans. Biomed. Eng. 53, 1346–1356.
Schreiber, C., Moissenet, F., 2019. A multimodal dataset of human gait at different walking speeds established on injury-free adult participants. Sci. Data 6, 111.
Seel, T., Raisch, J., Schauer, T., 2014. Imu-based joint angle measurement for gait analysis. Sensors 14, 6891–6909.
Senanayake, D., Halgamuge, S., Ackland, D.C., 2021. Real-time conversion of inertial measurement unit data to ankle joint angles using deep neural networks. J. Biomech. 125, 110552.
Seth, Ajay, Hicks, Jennifer L., Uchida, Thomas K., Habib, Ayman, Dembia, Christopher L., Dunne, James J., Ong, Carmichael F., et al., 2018. OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. PLoS Comput. Biol. 14 (7), e1006223.
Sharifi Renani, M., Eustace, A.M., Myers, C.A., Clary, C.W., 2021. The use of synthetic imu signals in the training of deep learning models significantly improves the accuracy of joint kinematic predictions. Sensors 21, 5876.
Sherman, M.A., Seth, A., Delp, S.L., 2011. Simbody: Multibody dynamics for biomedical research. Procedia IUTAM 2, 241–261.
Stenger, R., Hozhabr Pour, H., Teich, J., Hein, A., Fudickar, S., 2025. Gait event detection and gait parameter estimation from a single waist-worn imu sensor. Sensors 25, 6463.
Sung, J., Han, S., Park, H., Cho, H.-M., Hwang, S., Park, J.W., Youn, I., 2021. Prediction of lower extremity multi-joint angles during overground walking by using a single imu with a low frequency based on an lstm recurrent neural network. Sensors 22, 53.
Wentz, V.H., Maciel, J.N., Gimenez Ledesma, J.J., Ando Junior, O.H., 2022. Solar irradiance forecasting to short-term pv power: Accuracy comparison of ann and lstm models. Energies 15, 2457.
Zimmermann, T., Taetz, B., Bleser, G., 2018. Imu-to-segment assignment and orientation alignment for the lower body using deep learning. Sensors 18, 302.