Journal Pre-proofs


# Predicting 3D ground reaction forces across various movement tasks: a convolutional neural network study comparing different inertial measurement unit configurations

Batın Yılmazgün, Jonas Weber, Thorsten Stein, Stefan Sell, Bernd J. Stetter

PII: S0021-9290(25)00400-2
DOI: https://doi.org/10.1016/j.jbiomech.2025.112888
Reference: BM 112888

To appear in: *Journal of Biomechanics*

Accepted Date: 31 July 2025

Please cite this article as: B. Yılmazgün, J. Weber, T. Stein, S. Sell, B.J. Stetter, Predicting 3D ground reaction forces across various movement tasks: a convolutional neural network study comparing different inertial measurement unit configurations, *Journal of Biomechanics* (2025), doi: https://doi.org/10.1016/j.jbiomech.2025.112888

## Article type

Original article

## Title

Predicting 3D Ground Reaction Forces Across Various Movement Tasks: A Convolutional Neural Network Study Comparing Different Inertial Measurement Unit Configurations

Authors

1) **Yılmazgün, Batın[1]***: batin.yilmazguen@kit.edu, M.Sc.

2) Weber, Jonas[1]: weber.jonas.m@gmail.com, M.Sc.

3) Stein, Thorsten[1]: thorsten.stein@kit.edu, Prof. Dr. rer. nat.

4) Sell, Stefan[1,2]: stefan.sell@kit.edu, Prof. Dr. med.

5) Stetter, Bernd, J. [1]: bernd.stetter@kit.edu, Dr.

[1] Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

[2] Joint Center Black Forest, Hospital Neuenbuerg, Neuenbuerg, Germany

*Corresponding author

**Keywords**

machine learning; mobile gait analysis; inertial measurement unit; prediction accuracy; ground reaction force; variability of biomechanical time series

**Word count**

Abstract (289), introduction (495), methods (1206), results (644), discussion (2061)

## Abstract

Ground reaction forces (GRFs) are crucial for understanding movement biomechanics and for assessing the load on the musculoskeletal system. While inertial measurement units (IMUs) are increasingly used for gait analysis in natural environments, they cannot directly capture GRFs. Machine learning can be applied to predict 3D-GRFs based on IMU data. However, previous studies mainly focused on vertical GRF (vGRF) and isolated movement tasks. This study aimed to systematically evaluate the prediction accuracy of convolutional neural networks (CNNs) for 3D-GRFs using IMUs from single and multiple sensor configurations across various movement tasks. 20 healthy participants performed six movement tasks including walking, stair ascent, stair descent, running, a running step turn and a running spin turn at self-selected speeds. CNNs were trained to predict 3D-GRFs on IMU time series data for different configurations (lower body [7 IMUs], single leg [4 IMUs], femur-tibia [2 IMUs], tibia [1 IMU] and pelvis [1 IMU]). Prediction accuracies were assessed based on leave-one-subject-out cross validations using Pearson correlation (r) and relative root mean squared error (relRMSE). Across all tasks, CNNs predicted vGRF most accurately ($r = 0.98$, relRMSE ≤ 7.44%), followed by anterior-posterior GRF ($r \geq 0.92$, relRMSE ≤ 14.24%), with medial-lateral GRF being the least accurate ($r \geq 0.74$, relRMSE ≤ 29.46%). CNNs predicted vGRF consistently across tasks, with similar accuracy for multi IMU (average $r = 0.98$, average relRMSE: 6.06%) and single IMU configurations (average $r = 0.98$, average relRMSE: 6.88%), supporting single IMU configurations for vGRF in practical applications. During cutting maneuvers, the lower body configuration reduces the relRMSE for mlGRF (5.23–12.46%) and apGRF (1.53–3.16%) compared to single IMU configurations. However, for mlGRF and apGRF during cutting tasks, lower body configuration improve accuracy, highlighting a trade-off between simplicity and performance.

## 1. Introduction

Ground reaction force (GRF), typically measured by force plates, is crucial in biomechanical motion analysis, as it represents the force induced by a subject on the ground during movement (Uchida & Delp, 2020). It plays a key role in identifying movement phases (Headon & Curwen, 2001) and analyzing kinetics of movement in clinical (Nazmul Islam Shuzan et al., 2023), therapeutic (Krafft et al., 2017), industrial (Hoogkamer et al., 2018) and training contexts (Huang et al., 2023). In addition, it is an essential input for musculoskeletal modeling (Sylvester et al., 2021).

Observational artefacts, such as the Hawthorne Effect, influence biomechanical time series, highlighting the importance of outdoor research settings as gait parameters differ significantly between controlled and natural environments (Jeon et al., 2023; Schmitt et al., 2021). Researchers recommend focusing on natural environments with less-controlled conditions (Benson et al., 2022). Wearable technologies, particularly inertial measurement units (IMUs), are versatile tools increasingly applied outside laboratory settings (McDevitt et al., 2022).

Since IMUs lack direct kinetic data, machine learning (ML) methods are used to predict kinetic information such as GRFs from kinematics (Lee & Lee, 2022; Stetter & Stein, 2024). Comparisons of ML techniques, such as multilayer perceptron (MLP), convolutional neural networks (CNNs), and long short-term memory networks (LSTM), show that CNNs slightly outperform others. However, the task and dataset remain more critical than the ML architecture (Altai et al., 2023; Mundt et al., 2021). Independent of the ML technique, challenges persist in estimating anterior-posterior GRF (apGRF), medial-lateral GRF (mlGRF), peak vGRF and GRF during double support phases (Ancillao et al., 2018). Furthermore, significant research gaps in biomechanical datasets regarding task variety and comparability of IMU configurations hinder progress (David et al., 2024). While common activities like walking and running are well-studied, broader datasets are needed for diverse tasks (Lee & Lee, 2022). IMU placement and the number of sensors as input for ML used vary across studies. Single sensors on body segments (Jiang et al., 2020; M. Lee & Park, 2020; Lim et al., 2019; Mantashloo et al., 2023) and combinations of multiple sensors (Mundt et al., 2021; Stetter et al., 2020; Wouda et al., 2018) have both been used. For daily use, fewer sensors, such as smartphone-based predictions, may suffice (De Brabandere et al., 2020). For clinical or scientific applications, more sensors might be justified if they improve accuracy. Optimal sensor placement as input for vGRF prediction also varies: the shank (Jiang et al., 2020), top of the shoe (Havashinezhadian et al., 2023) and waist (Pimentel et al., 2024) have been recommended. However, increasing IMU quantity does not necessarily improve prediction accuracy (Moghadam et al., 2023).

The purposes of this study were: (1) to systematically compare the prediction accuracies of five different ML inputs (lower body [7 IMUs], single leg [4 IMUs], femur-tibia [2 IMU], pelvis [1 IMU] and tibia [1 IMU]); and (2) to analyze prediction accuracies across various movement tasks: walking, stair ascent and descent, running, running step turn and spin turn.

## 2. Methods

### 2.1 Participants

A total of 20 healthy sport students (age: 24.61 ± 3.67 years; height: 1.72 ± 0.09 m; body mass: 68.26 ± 11.45 kg; 10 males, 10 females) participated in the study. Sample size was determined based on related studies (n: 17 - 22; Alcantara et al., 2022; Altai et al., 2023; Leporace et al., 2018). Participants were excluded if they had any recent lower body pain, neurological disorders, spinal issues, or conditions affecting their gait. All participants provided written informed consent prior to the measurements. The study was approved by the ethics committee of the Karlsruhe Institute of Technology (KIT).

### 2.2 Experimental protocol and biomechanical data acquisition

All participants visited the biomechanics laboratory once. A total of 16 IMUs (2 feet, 2 shank, 2 femur, 1 pelvis, 1 lower spine, 1 upper spine, 2 upper arm, 2 lower arm, 2 hands, 1 head; 200 Hz, accelerometers ± 200 g, gyroscopes ± 7000°/s; 3D Ultium Motion, Noraxon, USA) were attached to participants. 3D-GRF data were collected using two force plates: one integrated in the ground (1000Hz, AMTI Inc., USA) used for all movement tasks apart from stair climbing (Figure 2) and the other built into the fourth step of a seven-step staircase (Steingrebe et al., 2025) used to collect stair climbing GRFs (1000Hz, custom-made) (Figure 3). Participants completed 1–3 familiarization trials, followed by 10 valid trials (±10% of initial self-selected speed, avoiding both feet on one force plate and proper execution) for each task, including walking, stair ascent, stair descent, running, running step turn (45°±10° turn initiated from the stance limb with external hip rotation), and spin turn (45°±10° turn initiated from the stance limb with internal hip rotation). All tasks were performed at self-selected speeds to ensure natural data acquisition (Padulo et al., 2023). Entry speed was measured using light gates (TAG Heuer, Switzerland) placed just before the force plate over a 2m path (Figure 2). Data collection systems were synchronized via an analog signal from the force plate system (Vicon Lock+ Connectivity) with the IMU system (Noraxon Ultium Receiver).

**2.3 Biomechanical data processing**

Raw 3D-GRF and IMU signals were filtered with a 4$^{th}$ order 15Hz low-pass Butterworth filter (Kristianslund et al., 2012). Ground contact was detected based on vGRF (threshold 20N) (Stetter et al., 2019), and the 3D-GRF was normalized to body weight [N]. IMU measurements, including 3D acceleration (converted to [m/s$^2$]) and 3D angular velocity ([°/s]), were used. The trigger signal segmented the IMU data according to the force plate, and stance phases were determined based on the force plate's ground contact timing. Time series data (GRF, acceleration and angular velocity) were temporally normalized to 101 data points (0-100% of stance phase) using cubic spline interpolation.

**2.4 Machine learning**

For each IMU configuration (Figure 1) one dataset was constructed, containing processed IMU and GRF signal matrices from all stance phases (20 participants x 10 trials x 6 movement tasks). All datasets were divided into training and test sets for model evaluation (Altai et al., 2023). For each iteration, the test set included data from a single participant, while the training set included data from the remaining 19 participants. This process was repeated for all 20 participants, generating 20 unique training and test sets (leave one subject out for cross validation, LOSOCV). Data processing and 3D-array construction was performed using MATLAB 2022b (The MathWorks, USA). ML modeling was conducted in Python (3.11.7), with support from libraries tsai (0.3.8), pandas (2.2.3), numpy (2.0.2), scipy (1.14.1) and matplotlib (3.8.0). The prepared training and test datasets (without the excluded participant) were imported from MATLAB to Python. Five individual CNNs were trained (Figure 1) using filtered and temporally normalized IMU inputs and filtered, bodyweight- and temporally normalized 3D-GRF outputs (see 2.3). The CNNs were trained on data from all six movement tasks in a task-generalized manner to predict all three GRF components jointly, using an established CNN for multivariate timeseries regression (Fauvel et al., 2021). The CNN is structured in two parallel parts: one part applies 2D convolutional filters (Conv 2D) to extract features for each observed variable (e.g., vertical acceleration). It includes Conv 2D (128 filters, kernel size: window size × 1), followed by batch normalization (BN) and rectified linear unit activation (ReLU). Next, 1x1 Conv 2D with ReLU reduces the number of parameters. The filters slide along the time axis (stride=1), and same padding is used to maintain the temporal dimension (T=101), corresponding the stance phase (0-100%). This results in a feature map for each observed variable of size T x D, where D is the number of input variables which depends on the IMU configuration (e.g., D=6 for tibia configuration). The second part, structurally identical to the first, uses 1D convolution filters (Conv 1D) with a kernel size of window size × D to extract and summarize key temporal dependencies and interactions across the IMU time series, producing feature map of size T × 1. The feature maps from both parallel parts are then fused into a single feature map of size T × (D + 1). After the feature maps are combined, the next steps are done in sequence. The fused feature map is processed through a Conv 1D (BN + ReLU, kernel size: window size × (D + 1)) followed by global average pooling to summarize key features. Finally, the prediction is performed with a linear output layer (Altai et al., 2023; Fauvel et al., 2021). The following hyperparameters were used: the adam optimizer was used for its ability to handle noisy, high-dimensional data like 3D acceleration, angular velocity and GRF (Kingma & Ba, 2014); and mean squared error (MSE) was chosen as the loss function to emphasize larger errors. Training included 50 epochs, a batch size of 10, and a 100% window size for extracting full stance phase features (Altai et al., 2023). A cyclical learning rate was applied for improved accuracy and efficiency (Smith, 2017).

**2.5 Statistical analysis**

The predicted 3D-GRFs (pGRF) were evaluated against the reference 3D-GRFs for different IMU configurations, different movement tasks and overall tasks using Pearson's correlation coefficient (r) and the relative Root Mean Squared Error (relRMSE) (Stetter et al., 2019). The relRMSE, derived by dividing RMSE (Eq 1) by the mean peak-to-peak amplitude of pGRF and reference GRF (Eq 2), enables comparison of time series with varying amplitudes.

$$RMSE = \sqrt{\frac{1}{100}\sum_{t=1}^{100}(GRF - pGRF)^2} \qquad \text{(Eq 1)}$$

$$relRMSE = \frac{RMSE}{\frac{1}{2} * ((max(pGRF) - min(pGRF)) + (max(GRF) - min(GRF)))} \quad \text{(Eq 2)}$$

For statistical analysis, the r values were transformed into Fisher-z for averaging and standard deviation calculations, and then retransformed to the original scale. Correlation strength was categorized as weak ($r \leq 0.35$), moderate ($0.35 < r \leq 0.67$), strong ($0.67 < r \leq 0.90$), or excellent ($r > 0.90$) (Karatsidis et al., 2016; Stetter et al., 2020; Taylor, 1990). The significance level was set a priori to $\alpha = 0.05$. Task-specific coefficients of variation (CV) for the reference GRF were calculated to assess GRF variability (Eq 3), using the absolute reference GRF to avoid a small denominator caused by negative and positive values in the ML and AP components (Kowalski et al., 2021).

$$CV = \left(\frac{std(GRF)}{mean(|GRF|)}\right) * 100 \quad \text{(Eq 3)}$$

Average task-specific movement speeds were calculated across all trials of a task and reported as mean and standard deviation (Table 3). All statistical analyses were performed in MATLAB 2022b (The MathWorks, USA).

## 3. Results

### 3.1 Overall performance across IMU configurations

Across all IMU configurations, the vGRF showed the highest accuracy (r: 0.98, relRMSE: 6.02%-7.44%), followed by the apGRF (r: 0.92-0.95, relRMSE: 11.43%-14.24%) and the mlGRF (r: 0.74-0.81, relRMSE: 21.89%-29.46%) (Table 1, Table 2). The pelvis configuration showed the lowest accuracy across all GRF components, corresponding to lower r and upper relRMSE values within the reported ranges. Other configurations showed minimal differences for vGRF (r: 0.98, relRMSE: 6.02%-6.31%) and apGRF (r: 0.95, relRMSE: 11.43%-12.22%) (Table 1, Table 2). For mlGRF, the lower body configuration performed best with lower relRMSE outliers seen in the violin boxplots (Figure 4).

### 3.2 Walking

In walking, the vGRF showed the highest accuracy (r: 0.97-0.98, relRMSE: 5.81%-7.08%), followed by the apGRF (r: 0.96-0.97, relRMSE: 6.65%-9.29%) and the mlGRF (r: 0.66-0.75, relRMSE: 18.38%-23.36%) (Table 1, Table 2). The pelvis configuration showed the lowest accuracy across all GRF components, corresponding to lower r and upper relRMSE values within the reported ranges. The single leg configuration performed best for 3D-GRF prediction, corresponding to higher r and lower relRMSE values within the reported ranges. Average walking speed was 1.38 ± 0.11 m/s, and the reference apGRF (CV: 36.93 ± 28.09%) and vGRF (CV: 9.86 ± 5.59%) showed the lowest CV across tasks (Table 3).

**3.3 Stair ascent**

In stair ascent, the vGRF showed the highest accuracy (r: 0.97, relRMSE: 6.47%-7.24%), followed by the apGRF (r: 0.86-0.88, relRMSE: 14.56%-15.49%) and the mlGRF (r: 0.64-0.67, relRMSE: 23.85%-25.50%) (Table 1, Table2). The accuracy of 3D-GRF prediction varied without a clear best-performing configuration. Average stair ascent speed was 0.53 ± 0.04 m/s, the slowest speed across all tasks, and the reference apGRF (CV: 66.65 ± 34.73%) showed the highest CV (Table 3).

**3.4 Stair descent**

In stair descent, the vGRF showed the highest accuracy (r: 0.97, relRMSE: 5.99%-6.89%), followed by the apGRF (r: 0.88-0.91, relRMSE: 10.93%-13.43%) and the mlGRF (r: 0.60-0.67, relRMSE: 22.18%-26.00%) (Table 1, Table 2). The pelvis configuration showed the lowest accuracy for apGRF and mlGRF, corresponding to lower r and upper relRMSE values within the reported ranges. Average stair descent speed was 0.59 ± 0.07 m/s, the second slowest speed across all tasks, with the reference apGRF (CV: 65.06 ± 33.03%) showing a comparable CV to stair ascent (Table 3).

**3.5 Running**

In running, the vGRF showed the highest accuracy (r: 0.99, relRMSE: 4.55%-6.48%), followed by the apGRF (r: 0.94-0.97, relRMSE: 10.32%-14.40%) and mlGRF (r: 0.46-0.68, relRMSE: 32.15%-49.27%) (Table 1, Table 2). The pelvis configuration showed the lowest accuracy for apGRF and mlGRF, corresponding to lower r and upper relRMSE values within the reported ranges. The tibia configuration was the only one with strong correlations (r ≥ 0.68) for the mlGRF (Table 1). The prediction accuracy for running was strongest for vGRF and weakest for mlGRF. Average running speed was 2.67 ± 0.32 m/s, the fastest speed across all tasks, with the highest CV for the mlGRF (91.82 ± 29.01%) across all tasks (Table 3).

**3.6 Running step turn**

In the running step turn, the vGRF showed the highest accuracy (r: 0.98, relRMSE: 6.39%-8.40%), followed by the apGRF (r: 0.93-0.96, relRMSE: 13.11%-16.27%) and the mlGRF (r: 0.93-0.96, relRMSE: 14.07%-26.53%) (Table 1, Table 2). The lower body and single leg configurations performed best for the mlGRF, showing fewer relRMSE outliers compared to the other IMU configurations (Figure 4). Average running step turn speed was 2.64 ± 0.30 m/s, the second fastest speed across all tasks, with the lowest CV for the mlGRF (39.82 ± 18.59%) across all tasks (Table 3).

**3.7 Running spin turn**

In the running spin turn, the vGRF showed the highest accuracy (r: 0.98-0.99, relRMSE: 6.02%-8.82%), followed by the apGRF (r: 0.93-0.96, relRMSE: 12.45%-16.56%) and the mlGRF (r: 0.85-0.93, relRMSE: 18.65%-25.76%) (Table 1, Table 2). The lower body configuration performed best for the mlGRF, showing fewer relRMSE outliers compared to the other IMU configurations (Figure 4). Average running spin turn speed was 2.58 ± 0.27 m/s, the third fastest speed across all tasks, with the highest CV for the vGRF (25.92 ± 13.99%) across all tasks (Table 3).

## 4. Discussion

This study aimed to systematically evaluate CNN prediction accuracies, trained on different movement tasks, for 3D-GRFs using IMU data from different sensor configurations. The key findings are: first, CNNs predicted the vGRF with the highest accuracy consistently across all movement tasks, regardless of the IMU configuration. Second, a similar pattern was observed for apGRF, though its predictive performance was slightly lower and less consistent than vGRF. Third, in contrast, mlGRF was the most challenging force component to predict across all configurations. Fourth, CNNs trained on lower body and single leg configurations performed noticeably better for mlGRF during cutting maneuvers (running spin turn and step turn) than other configurations. Fifth, CNNs based on the pelvis configuration showed the weakest predictive performance for 3D-GRF across all movement tasks but remained solid for vGRF prediction. Sixth, GRF components with high CV in the training reference data were predicted less accurately by the CNN.

**4.1 Effects of movement task on the different GRF dimensions**

For vGRF, running yielded the highest prediction accuracy (r = 0.99, relRMSE = 4.55%), slightly higher than previous studies (r: 0.96-97, relRMSE: 6.40%-13.92%; Alcantara et al., 2022; Johnson et al., 2021; Scheltinga et al., 2023; Wouda et al., 2018). Walking (r = 0.98, relRMSE = 5.81%) and stair tasks (stair ascent: r = 0.97, relRMSE = 6.47%; stair descent: r = 0.97, relRMSE = 5.99%) also yielded accurate predictions, although slightly inferior to running, that fit well with the results in the literature for walking (e.g., r: 0.95-0.99, relRMSE: 2.32%-7.40%; Hossain et al., 2023; Jiang et al., 2020; M. Lee & Park, 2020; Lim et al., 2019; Mantashloo et al., 2023) and stair ascent and descent (r = 0.98 and relRMSE = 4.51%; Hossain et al., 2023). The running step turn (r = 0.98, relRMSE = 6.39%) and spin turn (r = 0.99, relRMSE = 6.02%) were predicted with accuracies comparable to other tasks (e.g. running) and add to the current literature, as these tasks have not been considered so far. Notably, the variability of the reference vGRF (CV: 9.86%–25.92%) was lowest among all GRF components, which likely contributed to the consistently high prediction accuracy for this force component.

For apGRF, walking had the best prediction accuracy (r = 0.97, relRMSE = 6.65%) followed by running (r = 0.97, relRMSE = 10.32%), consistent with other studies for walking (e.g., r: 0.93-0.99, relRMSE: 2.13%-9.21%; M. Lee & Park, 2020; Lim et al., 2019; Mantashloo et al., 2023) and running (r: 0.91-0.96, relRMSE: 7.80%-17.06%; Johnson et al., 2021; Scheltinga et al., 2023). The predicted apGRF for stair ascent (r = 0.88, relRMSE = 14.56%) and stair descent (r = 0.91, relRMSE = 10.93%) was slightly lower and also showed lower accuracy compared to reported results (r = 0.92, relRMSE = 4.64%; Hossain et al., 2023). Additionally, the running step turn (r = 0.96, relRMSE = 13.11%) and spin turn (r = 0.96, relRMSE = 12.45%) showed good accuracy for apGRF predictions. However, outliers (Figure 4) were likely caused by the dynamic nature and multiplanar execution of the movement, as well as variability in reference apGRF (CV: 42.78%–56.76%). These factors may have led to slightly lower prediction accuracy for a few individual trials. The apGRF showed good predictability but was less accurate than vGRF, likely due to its lower magnitude (Figure 5 – Figure 7) and more variation across tasks and individual techniques in movement execution. However, the CNNs were able to handle this to a large extent.

Our mlGRF prediction for running are worse than those reported in other studies (r: 0.58-0.87, relRMSE = 10.80%-21.56%; Johnson et al., 2021; Scheltinga et al., 2023), but they are consistent in the sense that they also represent the poorest predictions observed within the mlGRF framework. A potential reason for the limited prediction accuracy and outliers is the high variability observed in reference mlGRF. Among all tasks, mlGRF variability was highest for running (CV: 91.81%), which corresponded with relatively low prediction accuracy (r = 0.68, relRMSE = 32.15%). This, combined with low amplitudes (Figure 5) and inter-individual differences (Zinner & Sperlich, 2016), likely contributed to the reduced prediction performance. Additionally, stair ascent (r = 0.67, relRMSE = 23.85) and descent (r = 0.67, relRMSE = 22.18%) showed lower prediction accuracy compared to Hossain et al., 2023 (r = 0.88, relRMSE = 8.16%). In contrast, predicted mlGRF in walking (r = 0.75, relRMSE = 18.38%) showed comparable prediction accuracy to the literature (r: 0.79-0.80, relRMSE: 15.55%-18.50%; Karatsidis et al., 2016; M. Lee & Park, 2020). Tasks with directional changes, like running step turns (r = 0.96, relRMSE = 14.07%) and spin turns (r = 0.93, relRMSE = 18.65%), showed the highest prediction accuracies for mlGRF compared to the other tasks, probably due to their higher amplitudes (Dixon et al., 2014; Figure 7) and lower reference mlGRF variability (CV: 39.82% – 47.86%) compared to the other tasks (Table 3). Consistent force time series were likely easier for the CNN to learn, with the stronger lateral movements during direction changes enhancing IMU signals in these directions and improving predictions. While unique patterns are visible, underestimations explain the strong correlation despite the higher relRMSE, especially in cutting movements (Figure 7). The variability in reference mlGRF can be explained by multiple factors, such as gender differences, leg alignment, step width and foot placement (Racic et al., 2009). Our results suggest that high variability in the targeted model output may reduce prediction accuracy, as seen with mlGRF, which exhibited both the lowest performance and the highest variability. This may also be related to the lower sensitivity of IMUs to lateral motion during straight-line tasks, where most movement occurs in vertical or forward directions. To improve the prediction accuracy of the mlGRF, several options could be explored. Hyperparameter tuning, such as adjusting the window size, batch size, or number of training epochs, could help. While task-specific networks (Shah & Dixon, 2025) might improve mlGRF prediction performance for certain movements, their practical value is limited.

**4.2 Effects of the placement and the number of IMUs on prediction accuracy**

The CNNs predicted vGRF accurately for specific movement tasks, regardless of the IMU configuration used as input. A similar pattern was observed for the apGRF, which was predicted with nearly identical accuracy across all tasks except for cutting movements. In this context, the lower body configuration and single leg configuration showed slightly lower relRMSEs compared to other configurations. However, the pelvis configuration performed with limited accuracy across tasks, with about 1-2% higher relRMSE for vGRF, 2-4% for apGRF and 2-17% for mlGRF. This is different from earlier studies (Pimentel et al., 2024), which found that the pelvis configuration worked best for predicting vGRF using multi-layer perceptron models. The pelvis sensor is often chosen because it tracks the center of mass (CoM), which relates well to external forces (Lim et al., 2019; Pimentel et al., 2024). However, the pelvis sensor is potentially less sensitive to specific changes in movement tasks that require rapid shifts in body posture, such as cutting movements. In contrast, sensors placed closer to the contact points with the ground, like the tibia, provide more accurate and relevant data for such predictions due to their higher sensitivity to these changes. Pelvis-based CNNs tend to underestimate 3D GRFs across tasks (Figure 5 – Figure 7), which may explain their poor performance and the difficulty in predicting accurate peak vGRF reported in the literature (Ancillao et al., 2018). Feature engineering, such as adding jerk or orientation, or alternatively reducing input dimensionality (Barua et al., 2021), could influence the prediction accuracy of the pelvis configuration, or even other configurations. However, both methods remain speculative and were not explored in this study. Earlier research (Moghadam et al., 2023) showed that adding more IMUs does not improve Random Forest predictions of joint moments during gait. Interestingly, CNNs trained on a lower body configuration perform much better at predicting mlGRF during cutting movements and slightly better for walking. Violin plots showed smaller outliers for a lower body configuration in these tasks (Figure 4). Compared to the single leg configuration, the lower body configuration reduces relRMSE by 2–4%, while the pelvis configuration shows a 7–12% higher relRMSE compared to the lower body configuration. We assume that including data from the contralateral leg provides the CNN with more non-redundant information to better generalize the mlGRF. Additionally, the lower body configuration resolves mlGRF underestimations in the running step turn but not in the running spin turn (Figure 7), where strong trunk rotation suggests that a sensor on the thoracic or cervical spine could help.

Placing a single sensor on the tibia and using a Random Forest model has been shown to be the best predictor of vGRF during walking (Jiang et al., 2020). Our results support this sensor configuration, showing that a tibia sensor on the relevant leg gives good accuracy across tasks and is likely comparable to the lower body and single leg configurations. The sensor is close to the contact point while remaining in motion throughout certain parts of the stance phase, meaning that it provides the CNN with relevant information; unlike a foot sensor which stays relatively still during some parts of the stance phase.

**4.3 Limitations**

This study aimed to address certain challenges of laboratory-based research; however, the dataset was collected in a controlled lab environment. Generalizing the results to natural environment remains uncertain, as environmental variability could affect the prediction accuracy. Future studies that predict GRFs on lab-based data should validate whether similar prediction accuracy can be achieved under natural conditions. The data underwent common preprocessing, including filtering, normalization and time alignment. While this reduced variability and ensured consistent input for the CNN, preprocessing may influence prediction accuracy and complicate implementation in real-world scenarios (Jlassi & Dixon, 2024). To maintain consistency, we used the same hyperparameters across all datasets, except for learning rate. However, hyperparameters tailored to specific IMU configurations could yield different results. While the CNN model was selected for its prior success in predicting joint moments (Altai et al., 2023), it cannot be excluded that other models might perform comparably or even better for GRFs. The relatively small sample size ($n = 20$) limits generalizability. However, we used LOSOCV, a stringent validation method. Additionally, comparing IMU configurations is inherently complex. Even with single IMU configurations, sensor placement on different body regions can vary (e.g., proximal or distal positions). Multi-IMU configurations further complicate comparisons due to the vast number of possible configurations. Additionally, multi-IMU configurations involve a trade-off: while they may improve performance in complex tasks like cutting, they also increase computational time, and accumulate placement errors (Tan et al., 2019) across body segments. Moreover, wearing multiple sensors can reduce user comfort and acceptance, limiting practicality for everyday use and making such configurations more suitable for research environments. Future work aiming to maximize GRF prediction accuracy may benefit from combining IMU data with other inputs, like plantar pressure.

**4.4 Conclusion**

Regardless of IMU configuration, our findings show that CNNs can predict vGRF with high accuracy, followed by apGRF and mlGRF, across movement tasks. However, the predictive performance of CNNs may be negatively affected by high variability in biomechanical time series, such as observed for the mlGRF, especially with small datasets. Despite its proximity to the CoM, the pelvis configuration consistently showed limited prediction accuracy for 3D-GRFs, represented by the underestimation of 3D-GRF across all movement tasks. For everyday activities like walking and stair climbing, single-sensor configurations perform comparably to multi-sensor configurations. However, during cutting tasks, lower body or single leg IMU configurations outperform other configurations in predicting the mlGRF. If the focus is exclusively on vGRF prediction, multi-IMU configurations offer no clear advantages. While the tibia configuration performs slightly better, the pelvis configuration is a practical alternative with comparable accuracy for vGRF. Single-sensor solutions are particularly appealing due to their simplicity, cost-effectiveness, energy efficiency, and user-friendliness, making them ideal for mobile health assessment, clinical gait analysis and sports performance monitoring. In scenarios requiring highly accurate predictions of 3D-GRF components (such as sports or real-world joint kinetics calculations), lower body or single leg IMU configurations are suggested based on the findings of this study. Additionally, a tibia sensor alone can provide similar accuracy and offers a practical solution. These insights highlight the importance of adjusting IMU configurations for specific applications: balancing simplicity, practicality and accuracy to meet the needs of clinical, sports and everyday wearable technologies.

## References

Alcantara, R. S., Edwards, W. B., Millet, G. Y., & Grabowski, A. M. (2022). Predicting continuous ground reaction forces from accelerometers during uphill and downhill running: A recurrent neural network solution. *PeerJ*, *10*, e12752. https://doi.org/10.7717/peerj.12752

Altai, Z., Boukhennoufa, I., Zhai, X., Phillips, A., Moran, J., & Liew, B. X. W. (2023). Performance of multiple neural networks in predicting lower limb joint moments using wearable sensors. *Frontiers in Bioengineering and Biotechnology*, *11*, 1215770. https://doi.org/10.3389/fbioe.2023.1215770

Ancillao, A., Tedesco, S., Barton, J., & O'Flynn, B. (2018). Indirect Measurement of Ground Reaction Forces and Moments by Means of Wearable Inertial Sensors: A Systematic Review. *Sensors*, *18*(8), 2564. https://doi.org/10.3390/s18082564

Barua, A., Zakia, U., Menon, C., & Jiang, X. (2021). Deep Learning Techniques in Estimating Ankle Joint Power Using Wearable IMUs. *IEEE Access*, *9*, 83041–83054. https://doi.org/10.1109/ACCESS.2021.3085660

Benson, L. C., Räisänen, A. M., Clermont, C. A., & Ferber, R. (2022). Is This the Real Life, or Is This Just Laboratory? A Scoping Review of IMU-Based Running Gait Analysis. *Sensors*, *22*(5), 1722. https://doi.org/10.3390/s22051722

David, P.-F., David, R.-C., Juan, M. C., & Diego, T. (2024). Human Locomotion Databases. A Systematic Review. *IEEE Journal of Biomedical and Health Informatics*, 1–17. https://doi.org/10.1109/JBHI.2023.3311677

De Brabandere, A., Emmerzaal, J., Timmermans, A., Jonkers, I., Vanwanseele, B., & Davis, J. (2020). A Machine Learning Approach to Estimate Hip and Knee Joint Loading Using a Mobile Phone-Embedded IMU. *Frontiers in Bioengineering and Biotechnology*, *8*, 320. https://doi.org/10.3389/fbioe.2020.00320

Dixon, P. C., Stebbins, J., Theologis, T., & Zavatsky, A. B. (2014). Ground reaction forces and lower-limb joint kinetics of turning gait in typically developing children. *Journal of Biomechanics*, *47*(15), 3726–3733. https://doi.org/10.1016/j.jbiomech.2014.09.011

Fauvel, K., Lin, T., Masson, V., Fromont, É., & Termier, A. (2021). XCM: An Explainable Convolutional Neural Network for Multivariate Time Series Classification. *Mathematics*, *9*(23), 3137. https://doi.org/10.3390/math9233137

Havashinezhadian, S., Chiasson-Poirier, L., Sylvestre, J., & Turcot, K. (2023). Inertial Sensor Location for Ground Reaction Force and Gait Event Detection Using Reservoir Computing in Gait. *International Journal of Environmental Research and Public Health*, *20*(4), 3120. https://doi.org/10.3390/ijerph20043120

Headon, R., & Curwen, R. (2001). Recognizing movements from the ground reaction force. *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, 1–8. https://doi.org/10.1145/971478.971523

Hoogkamer, W., Kipp, S., Frank, J. H., Farina, E. M., Luo, G., & Kram, R. (2018). A Comparison of the Energetic Cost of Running in Marathon Racing Shoes. *Sports Medicine*, *48*(4), 1009–1019. https://doi.org/10.1007/s40279-017-0811-2

Hossain, M. S. B., Guo, Z., & Choi, H. (2023). Estimation of Lower Extremity Joint Moments and 3D Ground Reaction Forces Using IMU Sensors in Multiple Walking Conditions: A Deep Learning Approach. *IEEE Journal of Biomedical and Health Informatics*, *27*(6), 2829–2840. https://doi.org/10.1109/JBHI.2023.3262164

Huang, H., Huang, W.-Y., & Wu, C.-E. (2023). The Effect of Plyometric Training on the Speed, Agility, and Explosive Strength Performance in Elite Athletes. *Applied Sciences*, *13*(6), 3605. https://doi.org/10.3390/app13063605

Jeon, J., Kwon, S., Lee, Y., Hong, J., Yu, J., Kim, J., Kim, S.-G., & Lee, D. (2023). Influence of the Hawthorne effect on spatiotemporal parameters, kinematics, ground reaction force, and the symmetry of the dominant and nondominant lower limbs during gait. *Journal of Biomechanics*, *152*, 111555. https://doi.org/10.1016/j.jbiomech.2023.111555

Jiang, X., Napier, C., Hannigan, B., Eng, J. J., & Menon, C. (2020). Estimating Vertical Ground Reaction Force during Walking Using a Single Inertial Sensor. *Sensors*, *20*(15), 4345. https://doi.org/10.3390/s20154345

Jlassi, O., & Dixon, P. C. (2024). The effect of time normalization and biomechanical signal processing techniques of ground reaction force curves on deep-learning model performance. *Journal of Biomechanics*, *168*, 112116. https://doi.org/10.1016/j.jbiomech.2024.112116

Johnson, W. R., Mian, A., Robinson, M. A., Verheul, J., Lloyd, D. G., & Alderson, J. A. (2021). Multidimensional Ground Reaction Forces and Moments From Wearable Sensor Accelerations via Deep Learning. *IEEE Transactions on Biomedical Engineering*, *68*(1), 289–297. https://doi.org/10.1109/TBME.2020.3006158

Karatsidis, A., Bellusci, G., Schepers, H., De Zee, M., Andersen, M., & Veltink, P. (2016). Estimation of Ground Reaction Forces and Moments During Gait Using Only Inertial Motion Capture. *Sensors*, *17*(1), 75. https://doi.org/10.3390/s17010075

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization* (Version 9). arXiv. https://doi.org/10.48550/ARXIV.1412.6980

Kowalski, E., Catelli, D. S., & Lamontagne, M. (2021). A waveform test for variance inequality, with a comparison of ground reaction force during walking in younger vs. Older adults. *Journal of Biomechanics*, *127*, 110657. https://doi.org/10.1016/j.jbiomech.2021.110657

Krafft, F. C., Stetter, B. J., Stein, T., Ellermann, A., Flechtenmacher, J., Eberle, C., Sell, S., & Potthast, W. (2017). How does functionality proceed in ACL reconstructed subjects? Proceeding of functional performance from pre- to six months post-ACL reconstruction. *PLOS ONE*, *12*(5), e0178430. https://doi.org/10.1371/journal.pone.0178430

Kristianslund, E., Krosshaug, T., & Van Den Bogert, A. J. (2012). Effect of low pass filtering on joint moments from inverse dynamics: Implications for injury prevention. *Journal of Biomechanics*, *45*(4), 666–671. https://doi.org/10.1016/j.jbiomech.2011.12.011

Lee, C. J., & Lee, J. K. (2022). Inertial Motion Capture-Based Wearable Systems for Estimation of Joint Kinetics: A Systematic Review. *Sensors*, *22*(7), 2507. https://doi.org/10.3390/s22072507

Lee, M., & Park, S. (2020). Estimation of Three-Dimensional Lower Limb Kinetics Data during Walking Using Machine Learning from a Single IMU Attached to the Sacrum. *Sensors*, *20*(21), 6277. https://doi.org/10.3390/s20216277

Leporace, G., Batista, L. A., & Nadal, J. (2018). Prediction of 3D ground reaction forces during gait based on accelerometer data. *Research on Biomedical Engineering*, *34*(3), 211–216. https://doi.org/10.1590/2446-4740.06817

Lim, Kim, & Park. (2019). Prediction of Lower Limb Kinetics and Kinematics during Walking by a Single IMU on the Lower Back Using Machine Learning. *Sensors*, *20*(1), 130. https://doi.org/10.3390/s20010130

Mantashloo, Z., Abbasi, A., Tazji, M. K., & Pedram, M. M. (2023). Continuous estimation of kinetic variables while walking using an accelerometer and artificial neural networks. *Measurement Science and Technology*, *34*(9), 095101. https://doi.org/10.1088/1361-6501/acd5f0

McDevitt, S., Hernandez, H., Hicks, J., Lowell, R., Bentahaikt, H., Burch, R., Ball, J., Chander, H., Freeman, C., Taylor, C., & Anderson, B. (2022). Wearables for Biomechanical Performance Optimization and Risk Assessment in Industrial and Sports Applications. *Bioengineering*, *9*(1), 33. https://doi.org/10.3390/bioengineering9010033

Moghadam, S. M., Yeung, T., & Choisne, J. (2023). The Effect of IMU Sensor Location, Number of Features, and Window Size on a Random Forest Model's Accuracy in Predicting Joint Kinematics and Kinetics During Gait. *IEEE Sensors Journal*, *23*(22), 28328–28339. https://doi.org/10.1109/JSEN.2023.3317366

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*(13), 4535. https://doi.org/10.3390/s21134535

Nazmul Islam Shuzan, M., Chowdhury, M. E. H., Bin Ibne Reaz, M., Khandakar, A., Fuad Abir, F., Ahasan Atick Faisal, Md., Hamid Md Ali, S., Bakar, A. A. A., Hossain Chowdhury, M., Mahbub, Z. B., Monir Uddin, M., & Alhatou, M. (2023). Machine learning-based classification of healthy and impaired gaits using 3D-GRF signals. *Biomedical Signal Processing and Control*, *81*, 104448. https://doi.org/10.1016/j.bspc.2022.104448

Padulo, J., Rampichini, S., Borrelli, M., Buono, D. M., Doria, C., & Esposito, F. (2023). Gait Variability at Different Walking Speeds. *Journal of Functional Morphology and Kinesiology*, *8*(4), 158. https://doi.org/10.3390/jfmk8040158

Pimentel, R., Armitano-Lago, C., MacPherson, R., Sathyan, A., Twiddy, J., Peterson, K., Daniele, M., Kiefer, A. W., Lobaton, E., Pietrosimone, B., & Franz, J. R. (2024). Effect of sensor number and location on accelerometry-based vertical ground reaction force estimation during walking. *PLOS Digital Health*, *3*(5), e0000343. https://doi.org/10.1371/journal.pdig.0000343

Racic, V., Pavic, A., & Brownjohn, J. M. W. (2009). Experimental identification and analytical modelling of human walking forces: Literature review. *Journal of Sound and Vibration*, *326*(1–2), 1–49. https://doi.org/10.1016/j.jsv.2009.04.020

Scheltinga, B. L., Kok, J. N., Buurke, J. H., & Reenalda, J. (2023). Estimating 3D ground reaction forces in running using three inertial measurement units. *Frontiers in Sports and Active Living*, *5*, 1176466. https://doi.org/10.3389/fspor.2023.1176466

Schmitt, A. C., Baudendistel, S. T., Lipat, A. L., White, T. A., Raffegeau, T. E., & Hass, C. J. (2021). Walking indoors, outdoors, and on a treadmill: Gait differences in healthy young and older adults. *Gait & Posture*, *90*, 468–474. https://doi.org/10.1016/j.gaitpost.2021.09.197

Shah, V. R., & Dixon, P. C. (2025). Gait Speed and Task Specificity in Predicting Lower-Limb Kinematics: A Deep Learning Approach Using Inertial Sensors. *Mayo Clinic Proceedings: Digital Health*, *3*(1), 100183. https://doi.org/10.1016/j.mcpdig.2024.11.004

Smith, L. N. (2017). Cyclical Learning Rates for Training Neural Networks. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 464–472. https://doi.org/10.1109/WACV.2017.58

Steingrebe, H., Sell, S., Ehmann, H., & Stein, T. (2025). Hip and trunk biomechanics and dynamic balance during steady-state stair walking in people with mild-to-moderate hip osteoarthritis. *Journal of Biomechanics*, *185*, 112709. https://doi.org/10.1016/j.jbiomech.2025.112709

Stetter, B. J., Krafft, F. C., Ringhof, S., Stein, T., & Sell, S. (2020). A Machine Learning and Wearable Sensor Based Approach to Estimate External Knee Flexion and Adduction Moments During Various Locomotion Tasks. *Frontiers in Bioengineering and Biotechnology*, *8*, 9. https://doi.org/10.3389/fbioe.2020.00009

Stetter, B. J., Ringhof, S., Krafft, F. C., Sell, S., & Stein, T. (2019). Estimation of Knee Joint Forces in Sport Movements Using Wearable Sensors and Machine Learning. *Sensors*, *19*(17), 3690. https://doi.org/10.3390/s19173690

Stetter, B. J., & Stein, T. (2024). Machine Learning in Biomechanics: Enhancing Human Movement Analysis. In C. Dindorf, E. Bartaguiz, F. Gassmann, & M. Fröhlich (Hrsg.), *Artificial Intelligence in Sports, Movement, and Health* (S. 139–160). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-67256-9_9

Sylvester, A. D., Lautzenheiser, S. G., & Kramer, P. A. (2021). A review of musculoskeletal modelling of human locomotion. *Interface Focus*, *11*(5), 20200060. https://doi.org/10.1098/rsfs.2020.0060

Tan, T., Chiasson, D. P., Hu, H., & Shull, P. B. (2019). Influence of IMU position and orientation placement errors on ground reaction force estimation. *Journal of Biomechanics*, *97*, 109416. https://doi.org/10.1016/j.jbiomech.2019.109416

Taylor, R. (1990). Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography*, *6*(1), 35–39. https://doi.org/10.1177/875647939000600106

Uchida, T. K., & Delp, S. (2020). *Biomechanics of movement: The science of sports, robotics, and rehabilitation*. The MIT press.

Wouda, F. J., Giuberti, M., Bellusci, G., Maartens, E., Reenalda, J., Van Beijnum, B.-J. F., & Veltink, P. H. (2018). Estimation of Vertical Ground Reaction Forces and Sagittal Knee Kinematics During Running Using Three Inertial Sensors. *Frontiers in Physiology*, *9*, 218. https://doi.org/10.3389/fphys.2018.00218

Zinner, C., & Sperlich, B. (Hrsg.). (2016). *Marathon Running: Physiology, Psychology, Nutrition and Training Aspects*. Springer International Publishing. https://doi.org/10.1007/978-3-319-29728-6

## Data availability statement

The dataset created for this study can be obtained by contacting the corresponding author.

## Ethics statement

The research involving human participants received approval from the Ethics Committee of the Karlsruhe Institute of Technology. Written informed consent was provided by all participants.

## Author contributions

The study was designed by BY, SS, TS, and BJS. BY handled the data collection and analysis. The interpretation and discussion of results involved BY, JW, SS, TS and BJS. BY wrote the initial manuscript, and all authors contributed feedback for the final version.

## Funding

This study received financial assistance from Bauerfeind Innovationscentrum GmbH & Co. KG. The funder had no influence on the study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

## Acknowledgments

We would like to thank Maxine Stiritz for her support throughout the data collection process.

## Conflict of interest

We, Bernd J. Stetter and Stefan Sell, disclose our financial relationship with Bauerfeind AG, consisting of employment and consultancy. We confirm that this financial relationship does not inappropriately influence or bias our professional actions and decisions related to the content of this study. We have conducted this work independently, and we have full control over the content. Additionally, we declare that this financial relationship poses no potential conflicts of interest regarding the research, analysis, or interpretations presented in this manuscript.

| Lower body | Single leg | Femur-tibia | Pelvis | Tibia |
|---|---|---|---|---|

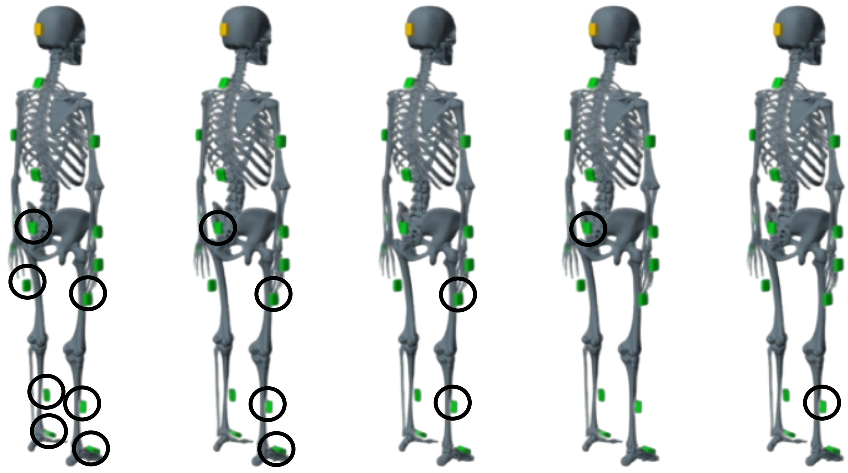

Figure 1: Different IMU configurations from various body regions used as input for the CNN.

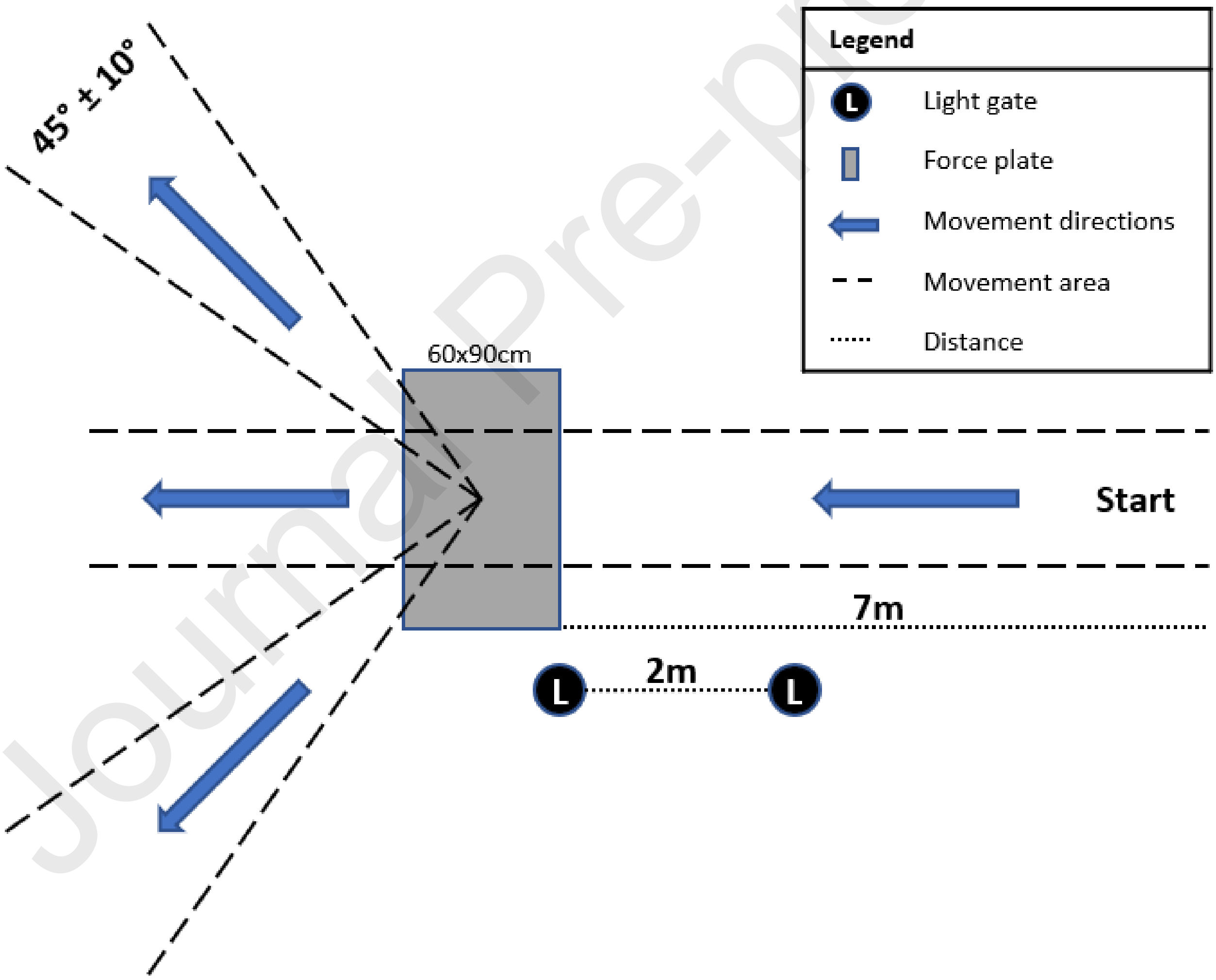


Figure 2: Schematic diagram of the laboratory setup for overground locomotion tasks (walking, running, running spin turn and running step turn).

Figure 3: Picture of the instrumented staircase. Three-dimensional force plates in steps four and five from the bottom, one-dimensional force plates in steps one to three and six. Riser height = 17 cm, tread = 28 cm, inclination angle = 31.26° (Steingrebe et al., 2025).

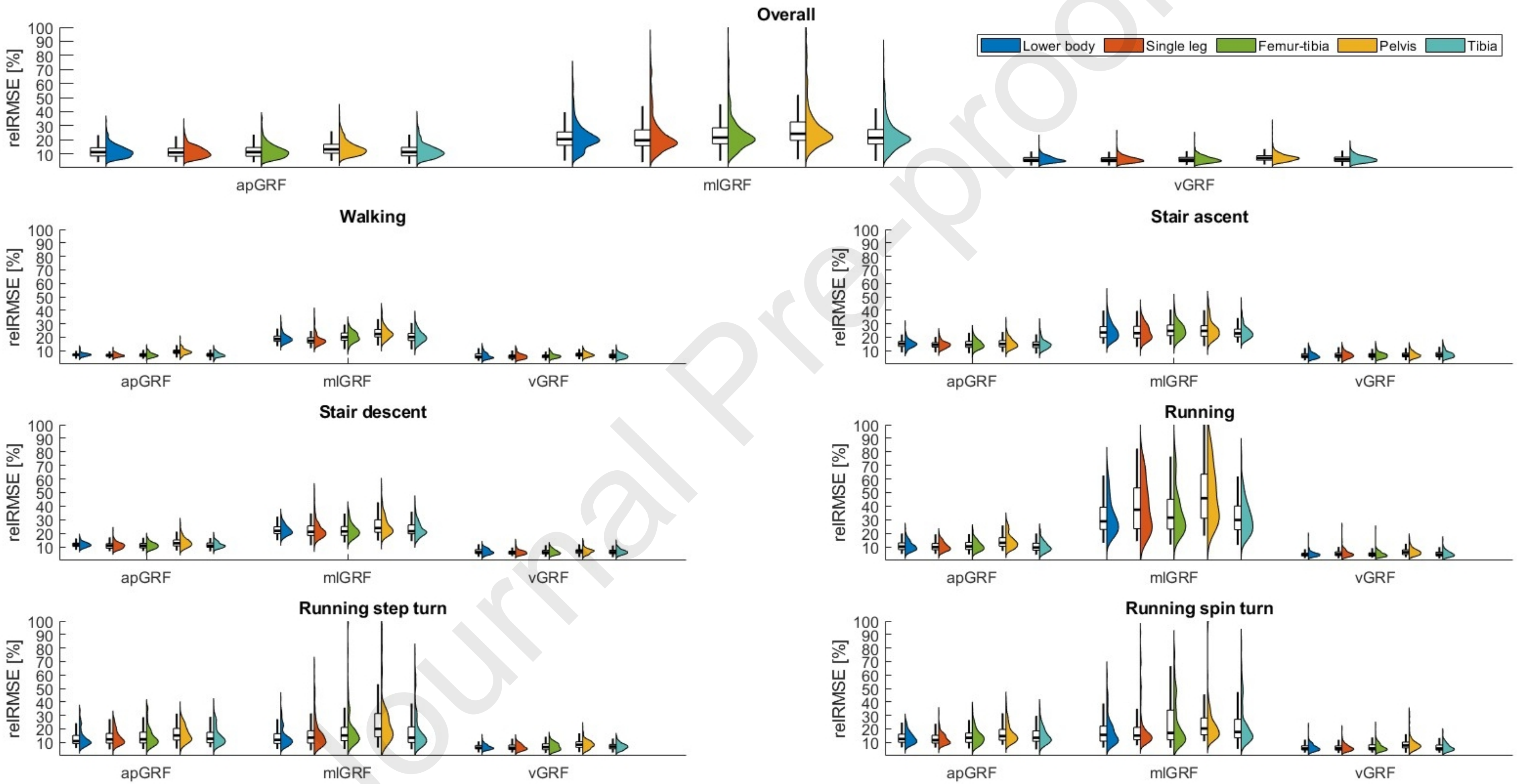

Figure 4: Violin boxplots of the 3D-GRF predictions for the different IMU configurations and movement tasks.

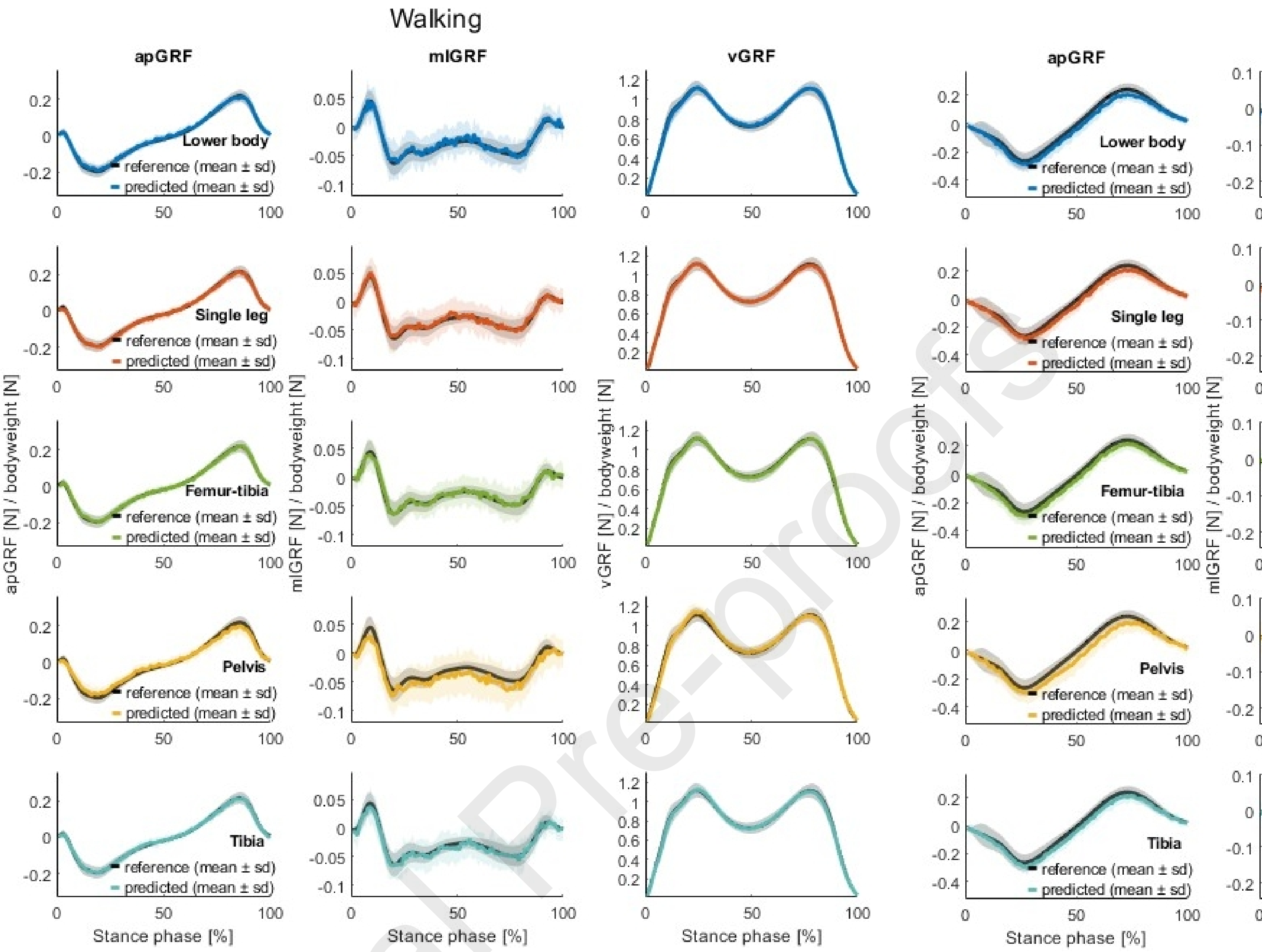


Figure 5: Predicted GRF time series (colored) and reference data (black) for walking and running based on different IMU configurations (from top to bottom). Data are pr

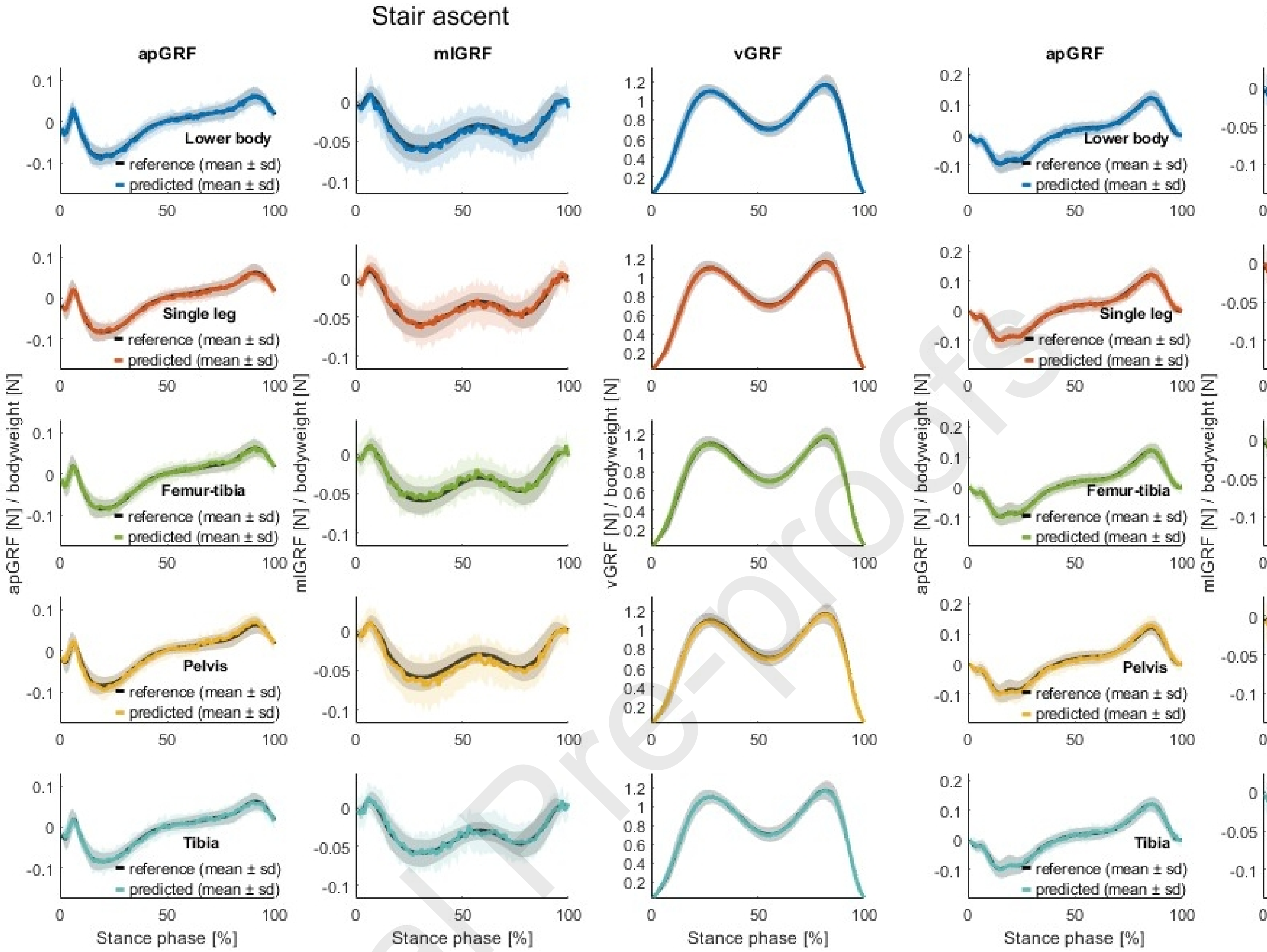


Figure 6: Predicted GRF time series (colored) and reference data (black) for stair ascent and stair descent based on different IMU configurations (from top to the bottom)

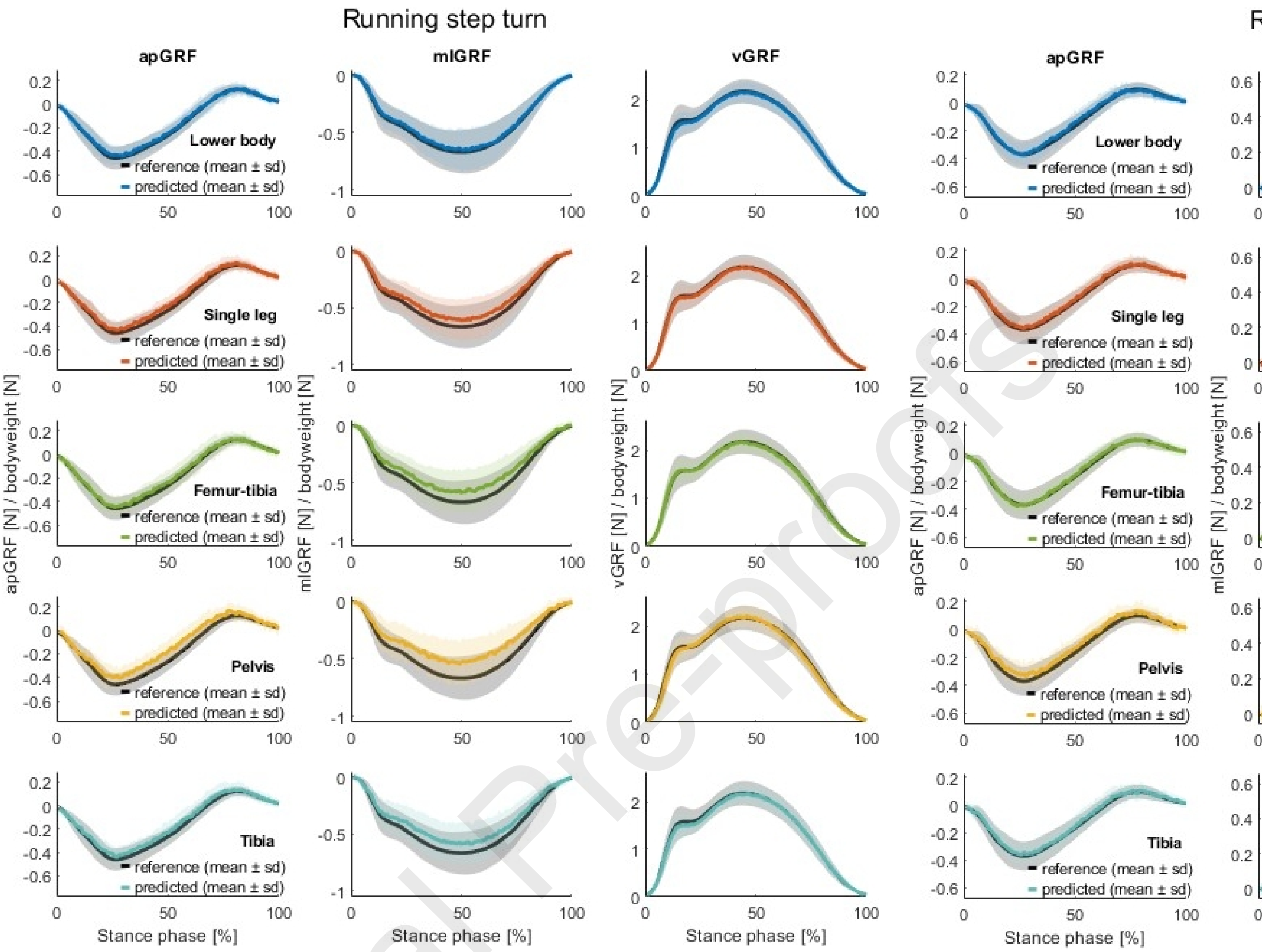


Figure 7: Predicted GRF time series (colored) and reference data (black) for running step turn and running spin turn based on different IMU configurations (from top to b

Table 1: GRF prediction accuracies for different IMU configurations and movement tasks. Accuracy (Pearson's correlation coefficient, r) of predicted GRFs are presented as mean ± standard deviations of the cross validation where one subject was excluded. All correlation coefficients were significant (p < 0.001).

| | | Categorization of the Pearson correlation coefficient | | | | |
|---|---|---|---|---|---|---|
| | | Excellent r > 0.90 | | Strong 0.67 < r ≤ 0.90 | | Moderate 0.35 < r ≤ 0.67 |
| | | **Lower body** | **Single leg** | **Femur-tibia** | **Pelvis** | **Tibia** |
| **Task** | **GRF** | Pearson correlation coefficient [r] | | | | |
| Walking | apGRF | 0.97 ± 0.18 | 0.97 ± 0.14 | 0.97 ± 0.19 | 0.96 ± 0.24 | 0.97 ± 0.22 |
| | mlGRF | 0.69 ± 0.15 | 0.75 ± 0.18 | 0.70 ± 0.20 | 0.66 ± 0.21 | 0.72 ± 0.15 |
| | vGRF | 0.97 ± 0.36 | 0.98 ± 0.30 | 0.97 ± 0.26 | 0.97 ± 0.23 | 0.97 ± 0.30 |
| Stair ascent | apGRF | 0.86 ± 0.20 | 0.88 ± 0.17 | 0.87 ± 0.23 | 0.87 ± 0.22 | 0.87 ± 0.25 |
| | mlGRF | 0.64 ± 0.21 | 0.67 ± 0.24 | 0.64 ± 0.22 | 0.66 ± 0.23 | 0.64 ± 0.19 |
| | vGRF | 0.97 ± 0.35 | 0.97 ± 0.35 | 0.97 ± 0.32 | 0.97 ± 0.31 | 0.97 ± 0.34 |
| Stair descent | apGRF | 0.90 ± 0.16 | 0.91 ± 0.21 | 0.91 ± 0.22 | 0.88 ± 0.29 | 0.90 ± 0.19 |
| | mlGRF | 0.61 ± 0.15 | 0.66 ± 0.23 | 0.67 ± 0.19 | 0.60 ± 0.21 | 0.65 ± 0.18 |
| | vGRF | 0.97 ± 0.30 | 0.97 ± 0.28 | 0.97 ± 0.28 | 0.97 ± 0.32 | 0.97 ± 0.29 |
| Running | apGRF | 0.96 ± 0.23 | 0.96 ± 0.19 | 0.97 ± 0.27 | 0.94 ± 0.28 | 0.97 ± 0.26 |
| | mlGRF | 0.64 ± 0.40 | 0.56 ± 0.41 | 0.64 ± 0.41 | 0.46 ± 0.38 | 0.68 ± 0.35 |
| | vGRF | 0.99 ± 0.35 | 0.99 ± 0.38 | 0.99 ± 0.34 | 0.99 ± 0.32 | 0.99 ± 0.41 |
| Running spin turn | apGRF | 0.96 ± 0.24 | 0.95 ± 0.25 | 0.96 ± 0.30 | 0.93 ± 0.30 | 0.96 ± 0.28 |
| | mlGRF | 0.93 ± 0.34 | 0.91 ± 0.47 | 0.90 ± 0.50 | 0.85 ± 0.56 | 0.91 ± 0.44 |
| | vGRF | 0.99 ± 0.43 | 0.99 ± 0.36 | 0.99 ± 0.45 | 0.98 ± 0.44 | 0.99 ± 0.46 |
| Running step turn | apGRF | 0.96 ± 0.27 | 0.96 ± 0.27 | 0.96 ± 0.34 | 0.93 ± 0.31 | 0.95 ± 0.28 |
| | mlGRF | 0.96 ± 0.28 | 0.96 ± 0.38 | 0.94 ± 0.50 | 0.93 ± 0.49 | 0.95 ± 0.36 |
| | vGRF | 0.98 ± 0.33 | 0.98 ± 0.36 | 0.98 ± 0.39 | 0.98 ± 0.34 | 0.98 ± 0.38 |
| **Overall** | **apGRF** | **0.95 ± 0.38** | **0.95 ± 0.36** | **0.95 ± 0.39** | **0.92 ± 0.34** | **0.95 ± 0.38** |
| | **mlGRF** | **0.81 ± 0.53** | **0.81 ± 0.53** | **0.79 ± 0.51** | **0.74 ± 0.51** | **0.80 ± 0.48** |
| | **vGRF** | **0.98 ± 0.44** | **0.98 ± 0.41** | **0.98 ± 0.43** | **0.98 ± 0.39** | **0.98 ± 0.45** |

Table 2: GRF prediction accuracies for different IMU configurations and movement tasks. Accuracy relative to the root mean squared error (relRMSE) of predicted GRFs are presented as mean ± standard deviations of the cross validation.

| | | Ranges of relRMSE | | | | |
|---|---|---|---|---|---|---|
| | | 0% < relRMSE ≤ 10% | 10% < relRMSE ≤ 20% | 20% < relRMSE ≤ 30% | relRMSE > 30% | |
| | | **Lower body** | **Single leg** | **Femur-tibia** | **Pelvis** | **Tibia** |
| **Task** | **GRF** | Relative Root Mean Squared Error [%] | | | | |
| Walking | apGRF | 7.15 ± 1.50 | 6.65 ± 1.24 | 6.86 ± 1.70 | 9.29 ± 2.18 | 6.95 ± 1.67 |
| | mlGRF | 19.32 ± 3.33 | 18.38 ± 4.37 | 20.40 ± 3.83 | 23.36 ± 5.00 | 20.39 ± 4.23 |
| | vGRF | 6.16 ± 2.23 | 5.81 ± 1.88 | 6.02 ± 1.59 | 7.08 ± 1.62 | 6.03 ± 1.71 |
| Stair ascent | apGRF | 15.22 ± 3.06 | 14.56 ± 2.65 | 14.81 ± 3.22 | 15.49 ± 3.61 | 14.93 ± 3.75 |
| | mlGRF | 24.60 ± 6.31 | 23.91 ± 5.77 | 25.30 ± 5.71 | 25.5 ± 7.10 | 23.85 ± 5.16 |
| | vGRF | 6.47 ± 2.30 | 6.68 ± 2.26 | 6.59 ± 2.11 | 6.98 ± 2.18 | 7.24 ± 2.47 |
| Stair descent | apGRF | 11.60 ± 1.83 | 10.93 ± 2.33 | 11.03 ± 2.22 | 13.43 ± 3.73 | 11.37 ± 2.11 |
| | mlGRF | 22.45 ± 3.76 | 22.58 ± 6.52 | 22.18 ± 5.16 | 26.00 ± 7.23 | 23.05 ± 5.34 |
| | vGRF | 6.54 ± 1.92 | 5.99 ± 1.78 | 6.24 ± 1.85 | 6.89 ± 2.21 | 6.36 ± 1.85 |
| Running | apGRF | 10.72 ± 3.62 | 10.38 ± 3.21 | 10.91 ± 3.66 | 14.40 ± 5.44 | 10.32 ± 3.92 |
| | mlGRF | 32.24 ± 13.04 | 39.70 ± 17.95 | 36.19 ± 17.25 | 49.27 ± 20.33 | 32.15 ± 12.61 |
| | vGRF | 4.55 ± 1.82 | 5.24 ± 2.62 | 4.67 ± 2.19 | 6.48 ± 2.69 | 4.90 ± 2.12 |
| Running spin turn | apGRF | 13.48 ± 4.52 | 12.45 ± 3.99 | 14.15 ± 5.70 | 16.56 ± 6.79 | 15.03 ± 5.95 |
| | mlGRF | 18.65 ± 11.14 | 22.04 ± 17.76 | 24.60 ± 17.95 | 25.76 ± 19.63 | 23.88 ± 16.09 |
| | vGRF | 6.07 ± 2.55 | 6.02 ± 2.50 | 6.37 ± 2.88 | 8.82 ± 5.15 | 6.31 ± 2.76 |
| Running step turn | apGRF | 13.11 ± 6.31 | 13.59 ± 5.29 | 14.31 ± 7.21 | 16.27 ± 5.84 | 14.72 ± 6.62 |
| | mlGRF | 14.07 ± 7.02 | 16.14 ± 10.76 | 21.56 ± 19.92 | 26.53 ± 19.45 | 18.64 ± 13.80 |
| | vGRF | 6.46 ± 1.92 | 6.39 ± 2.25 | 6.86 ± 2.71 | 8.40 ± 3.13 | 6.99 ± 2.39 |
| **Overall** | **apGRF** | **11.88 ± 4.60** | **11.43 ± 4.23** | **12.01 ± 5.20** | **14.24 ± 5.43** | **12.22 ± 5.31** |
| | **mlGRF** | **21.89 ± 10.00** | **23.79 ± 14.08** | **25.04 ± 14.45** | **29.46 ± 17.21** | **23.66 ± 11.45** |
| | **vGRF** | **6.04 ± 2.24** | **6.02 ± 2.28** | **6.12 ± 2.37** | **7.44 ± 3.16** | **6.31 ± 2.36** |

Table 3: Mean and standard deviation of the coefficient of variation (CV) [%] for the reference ground reaction force and movement speed [m/s] for the individual movement tasks.

| | **Walking** | **Stair ascent** | **Stair decent** | **Running** | **Running spin turn** | **Running step turn** |
|---|---|---|---|---|---|---|
| **CV apGRF** [%] | 36.93 ± 28.09 | 66.65 ± 34.73 | 65.06 ± 33.03 | 44.33 ± 28.70 | 56.76 ± 23.92 | 42.78 ± 23.94 |
| **CV mlGRF** [%] | 56.01 ± 30.81 | 56.27 ± 32.88 | 42.36 ± 17.76 | 91.82 ± 29.01 | 47.86 ± 26.17 | 39.82 ± 18.59 |
| **CV vGRF** [%] | 9.86 ± 5.59 | 12.68 ± 7.11 | 14.85 ± 6.20 | 14.54 ± 3.81 | 25.92 ± 13.99 | 19.12 ± 7.78 |
| **Speed** [m/s] | 1.38 ± 0.11 | 0.53 ± 0.04 | 0.59 ± 0.07 | 2.67 ± 0.32 | 2.58 ± 0.27 | 2.64 ± 0.30 |