

Running speed modulates joint kinetics and ground reaction forces during outdoor running: a wearable sensor study

Lucas Höschler^{*1}, Christina Halmich^{2 3}, Saša Cigoja⁴, Martin Ullrich⁴, Anne D. Koelewijn⁵, Hermann Schwameder¹

¹ Department of Sport and Exercise Science, University of Salzburg, Hallein, Austria

² Department of Artificial Intelligence and Human Interfaces, University of Salzburg, Salzburg, Austria

³ Human Motion Analytics Research Group, Salzburg Research Forschungsgesellschaft mbH, Salzburg, Austria

⁴ Innovation, adidas AG, Herzogenaurach, Germany

⁵ Department of Electrical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

* lucas.hoeschler@stud.plus.ac.at

ORIGINAL ARTICLE

Submitted: July 25, 2025

Accepted: January 20, 2026

Published: March 30, 2026

Editor-in-Chief

Claudio R. Nigg, University of Bern, Switzerland

Guest Editors

Thorsten Stein, Karlsruhe Institute of Technology, Germany

Bernd Stetter, Karlsruhe Institute of Technology, Germany

ABSTRACT

Wearable sensors, such as inertial measuring units (IMUs), enable biomechanical analyses of running in real-world conditions. While the effects of running speed on joint kinetics and ground reaction forces (GRFs) have been extensively studied in laboratories, field-based evidence remains scarce. This study used IMU data and machine learning to analyze the influence of different endurance running speeds on lower-limb joint moments and GRFs during outdoor running. First, a convolutional neural network (CNN) was trained and validated to estimate ipsilateral GRFs and lower-limb joint moments from three IMUs (foot, shank, pelvis) using an independent dataset. The model was then applied to predict these metrics for twenty-nine recreational runners performing an incremental speed protocol on a 400 m outdoor track, while wearing IMUs at the same locations. Speed differences were analyzed using statistical parametric mapping (SPM) methods. Running speed affected all kinetic parameters during most of the stance phase. While GRFs, ankle and hip moments increased significantly across all speed increments, knee moments were mostly unaffected by speed increases beyond 10 km/h. These findings suggest that reducing running speed may help mitigate ankle- and hip-related loading, which could be relevant for injury prevention. The observed speed-dependent effects on running kinetics were consistent with laboratory-based findings, reinforcing the validity of our approach and supporting the feasibility of IMU-based methods for biomechanical analysis in

ecological conditions. This study demonstrated the potential of wearable sensors and machine learning to complement traditional lab-based methods and enhance our understanding of running biomechanics in real-world settings.

Keywords

field study, machine learning, inertial sensor, joint loading, SPM, running biomechanics

Citation:

Höschler, L., Halmich, C., Cigoja, S., Ullrich, M., Koelewijn, A. D., & Schwameder, H. (2026). Running speed modulates joint kinetics and ground reaction forces during outdoor running: a wearable sensor study. *Current Issues in Sport Science*, 11(3), Article 004. <https://doi.org/10.36950/2026.3ciss004>

Introduction

Wearable sensor technology, including inertial measuring units (IMUs) and pressure insoles, has the potential to transform running biomechanics research by providing a cost-effective and portable alternative to traditional motion-capture systems and force plates (Benson et al., 2022; Blazey et al., 2021; Lee & Lee, 2022; Mundt, 2023; Weygers et al., 2020). These sensors allow for biomechanical assessments in real-world conditions, enabling studies on parameters such as ground reaction forces (GRFs) and joint moments outside the laboratory. Their ability to continuously capture 3D movement data over extended periods can help provide valuable insights for athletes, coaches, and medical professionals. Additionally, they can enhance lab-based research by enabling data collection in unsupervised, ecological settings (Davis et al., 2024; Plesek et al., 2024). However, despite these advantages, many IMU-based studies remain confined to controlled lab environments, limiting their full potential for real-world running analysis (Benson et al., 2022).

To fully leverage the potential of wearable sensors outside the laboratory, accurate methods are needed to estimate kinetic parameters such as joint moments and GRF time-series directly from sensor data. Such

kinetic data are critical for monitoring mechanical load and identifying injury risks (Stefanyshyn et al., 2006), making them highly relevant for both performance optimization and preventive strategies. In recent years, significant progress has been made in developing methods that combine IMU data with advanced modeling techniques, such as physics-based methods (e.g., optimal-control simulations) (Dorschky et al., 2019, 2025) and machine learning (ML) (Carter et al., 2024a, 2024b; Höschler et al., 2024; Höschler, Halmich, Schranz, Fritz, et al., 2025; Hossain et al., 2023; Liew et al., 2021; Mundt et al., 2020; Stetter et al., 2020). Validation results against lab-based reference systems have demonstrated excellent accuracy for sagittal plane joint moments as well as vertical and antero-posterior GRFs (Carter et al., 2024a; Hossain et al., 2023), with slightly lower accuracy for kinetics in the minor planes (Höschler et al., 2024; Höschler, Halmich, Schranz, Fritz, et al., 2025; Mundt et al., 2020; Stetter et al., 2020). Despite these promising results, their application to real-world outdoor running conditions remains limited.

So far, field studies in running with wearables have primarily focused on kinematics (Debertin et al., 2024; Fraeulin et al., 2021; Genitrini et al., 2023, 2024; Zrenner et al., 2020), spatio-temporal parameters (Davis et al., 2024; DeLong Lempke et al., 2024, 2025; Genitrini

et al., 2023, 2024; Hollis et al., 2021; Kozinc et al., 2024; Long, Outerleys, et al., 2023), or indirect kinetic proxies (DeJong Lempke et al., 2024, 2025; Gregory et al., 2019; Kozinc et al., 2024; Long, Outerleys, et al., 2023) rather than kinetic time-series data. While these approaches provide valuable insights, their ability to assess mechanical load, injury mechanisms, or detailed kinetic profiles is limited. Existing studies typically rely on commercially available IMU-based systems, commonly reporting a combination of discrete kinetic metrics, spatio-temporal parameters, and IMU-derived measures. For instance, some wearables estimate running power or peak impact forces, providing a simplified representation of mechanical load (Cerezuola-Espejo et al., 2020; Jaén-Carrillo et al., 2020). Others focus on spatio-temporal parameters such as step frequency, ground contact time, and stride length, which are widely used to assess running performance and efficiency (van Hooren et al., 2024). Additionally, some studies extract IMU-derived metrics, such as sample entropy or acceleration variance, as indicators of movement variability and neuromuscular control (Harrison et al., 2024). Although these metrics offer valuable insights, they have fundamental limitations. Discrete kinetic values, like estimated running power, do not capture the full complexity of joint loading across the stride cycle. Spatio-temporal parameters, though useful for monitoring performance trends, do not provide direct information on joint moments or GRFs. Similarly, IMU-derived variability measures describe movement stability but cannot replace traditional force-based assessments. Critically, none of these approaches provide kinetic time-series data, meaning they possibly lack the resolution to analyze joint synergies, mechanical load distribution, or injury-related biomechanical patterns (Pataky et al., 2015).

One of the most direct modulators of mechanical load during running is running speed. Since repeated high joint loads accelerate tissue microdamage via mechanical fatigue and reduce load capacity over time, understanding how speed affects joint kinetics is crucial for preventing overuse injuries and guiding training

adaptations (Hamstra-Wright et al., 2021). Laboratory studies have consistently shown speed-related effects on kinetic parameters. At endurance running speeds (below 5 m/s), GRFs increase linearly (Dorn et al., 2012; Nilsson & Thorstensson, 1989; Pataky et al., 2013). Similarly, ankle moments increase, but with greater magnitude than knee moments, where the effect diminishes at higher speeds (Arampatzis et al., 1999; Dorn et al., 2012; Schache et al., 2011). Hip kinetics findings are less consistent—some studies report increased moments, particularly during swing rather than stance (Dorn et al., 2012; Fukuchi et al., 2017), while others have found no differences between speed conditions (Orendurff et al., 2018). While these studies show speed related effects on running kinetics in controlled conditions, evidence from real-world running environments remains limited. Two field studies using commercially available IMU systems have reported speed-dependent changes in spatio-temporal metrics (Hollis et al., 2021; Kozinc et al., 2024) and discrete IMU-derived load proxies (Hollis et al., 2021), but comprehensive analyses of continuous joint moment and GRF time-series in ecological running conditions are still lacking.

Therefore, the aim of this study is to bridge the gap between laboratory-based kinetic analysis and real-world running by applying wearable sensor-based ML methods to estimate joint moment and GRF time-series during outdoor running at different running speeds. First, we train and validate a convolutional neural network (CNN) to estimate GRFs and lower-body joint moments from IMU data. We then apply this model in a field study to examine how increasing running speed influences these kinetic variables under real-world conditions. Focusing on running speed as a primary modulator of mechanical load, we specifically investigate how increases in running speed affect lower-body joint kinetics and GRFs.

Methods

Machine Learning

We developed and trained a CNN with eight 1D convolutional layers to estimate 12 kinetic time-series metrics (3D joint moments of ankle, knee and hip, as well as 3D GRFs) from IMU data. The training dataset consisted of synchronized IMU (3D accelerometer and gyroscope signals) and lab-based reference kinetic data from 24 recreational runners across various treadmill running conditions, including different footwear, slopes ($0 \pm 5\%$ incline), and speeds (7–14 km/h), as described in detail in Höschler, Halmich, Schranz, Fritz, et al. (2025).

Input data came from five lower-body IMUs attached at following locations: bilateral dorsal foot (on top of the shoelaces), bilateral shank (approx. midpoint between ankle and knee on the tibia), and pelvis (approx. midpoint between the posterior superior iliac spines). Previous results have shown that these locations provide the highest accuracy for knee moment estimation (Höschler, Halmich, Schranz, Koelewijn, et al., 2025).

Only the data from three sensors (foot, shank, pelvis) were used for the prediction of ipsilateral running kinetics (i.e., left foot, left shank and pelvis IMU were used for predicting left side joint moments and GRFs).

To improve numerical stability, all input signals and reference metrics were normalized using z-score normalization with fixed mean and standard deviation values computed across the dataset. The loss function L (1) computed the grand mean of the root mean square error (RMSE) across all 12 outputs j between predicted (\hat{y}) and reference (y) time series with i samples (2).

$$L = \frac{1}{12} \sum_{j=1}^{12} \text{RMSE}(\hat{y}_j, y_j) \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{i} \sum_{t=1}^i (\hat{y}_t - y_t)^2} \quad (2)$$

Table 1
Optimized Hyperparameters of the Convolutional Neural Network

Architecture									
Parameter	Search Space	Layers							
		Input	Layer 2	Aux 1	Aux 2	Layer 5	Layer 6	Layer 7	Output
In Channels	{8, 16, 32, 64, 128, 256}	18 ^a	256 ^b	32 + 2 ^c	32 ^b	32 ^b	256 ^b	256 ^b	128 ^b
Out Channels	{8, 16, 32, 64, 128, 256}	256	32	32	32	256	256	128	12 ^e
Kernel Size	{7, 15, 27, 51}	7	15	1 ^d	1 ^d	51	51	15	51
Training Parameters									
Parameter	Search Space	Value							
lr_{min}	{0.0001, 0.001}	1.1 × 10 ⁻⁴							
lr_{max}	{1:10} × lr _{min}	3.4 × 10 ⁻⁴							
Epochs	{25:50}	50							
Batch Size	{8, 16, 32, 64, 128, 256}	8							
Noise ratio	{0, 0.05, 0.1, 0.2}	0							
Dropout ratio	{0, 0.1, 0.2, 0.3}	0.2							

Architectural parameters and search spaces are displayed for the eight 1D convolutional layers (Input, Layer 2, Aux 1, Aux 2, Layer 5, Layer 6, Layer 7, Output). For each layer, the number of input channels, output channels and the kernel size were optimized. Additionally, the values and search spaces for training parameters are displayed.

^a The number of input channels of the Input layer were fixed to 18 (3 sensor locations x 6 input signals – 3D accelerations and angular velocities).

^b The number of input channels had to match the number of output channels of the previous layer.

^c Two dimensions were added to the first auxiliary layer (Aux 1) to incorporate additional information of the running slope.

^d The kernel size of the auxiliary layers was fixed to 1.

^e The number of output channels of the output layer was fixed to 12 (to output 3D ankle, knee, and hip moments plus 3D GRFs).

The model was trained on continuous data including stance and flight phases using a sliding window approach. A window length of 1 s with 30 % overlap resulted in the best performance during pilot testing. Optimal model architecture and training parameters of the CNN model (Table 1) were found through Bayesian hyperparameter optimization using a Tree-structured Parzen Estimator (TPE) sampler implemented in the *Optuna* package for Python (Akiba et al., 2019).

Model Validation

To ensure an independent performance evaluation of the CNN, the dataset was randomly split into a training set (20 participants) and a test set (4 participants) as recommended (Halilaj et al., 2018). During hyperparameter optimization, a three-fold cross-validation was applied within the training set. After optimization, the model was trained on the training set and evaluated on the separate test set. Model performance was assessed by root mean square error (RMSE), normalized RMSE (nRMSE, relative to the value range in the reference data) and intra-class correlation (ICC, two-way mixed effects, absolute agreement, single measurement (Koo & Li, 2016; Pini et al., 2022)). ICC values above 0.9, above 0.75, above 0.5, and below 0.5 were interpreted as “excellent”, “good”, “moderate”, and “poor” agreement (Koo & Li, 2016). Performance met-

rics were computed for each running condition (defined by the combination of subject, shoe, slope, speed, and side) over the entire condition (CONT) and specifically during stance phases (PHSS) (Höschler, Halmich, Schranz, Fritz, et al., 2025). For each of the four test participants, mean performance was obtained by averaging across all their respective condition-level metrics.

Participants

Twenty-nine recreational runners were recruited for the field study (participant characteristics are summarized in Table 2), none of whom were part of the dataset to train the CNN. Inclusion criteria matched those of the training dataset (Höschler, Halmich, Schranz, Fritz, et al., 2025): age 18–40 years, running experience ≥ 2 years, ≥ 1 training session per week, no injuries in the past three months, and no acute or chronic musculoskeletal or neurological conditions. An additional criterion was the absence of hearing impairment because auditory cues were used to regulate running speed during trials. All participants provided written informed consent, and the experimental procedures were approved by the University of Salzburg’s ethics committee (EK-GZ 32/2024).

Table 2

Participants' Anthropometric and Running-Related Characteristics

Sex	Age [years]	Height [cm]	Mass [kg]	Weekly distance [km]	Running Experience [years]
Male (14)	27 (5) 21-37	179 (6) 170-188	72 (8) 61-88	30 (38) 6-160	13 (8) 2-25
Female (15)	26 (5) 18-38	166 (7) 153-181	60 (8) 45-70	15 (5) 5-25	13 (7) 3-30

Displayed are mean (bold), standard deviation (in brackets), and the range from minimum to maximum.

Wearable Sensors and Data Collection

Wearable instrumentation included five wireless light-weight IMU sensors (Wave Track, Cometa, Milano, Italy) with measuring ranges of ± 16 g (3D accelerometer) and $\pm 2000^\circ/\text{s}$ (3D gyroscope) at 2000 Hz sampling rate. The sensors were calibrated prior to testing to remove any systematic offset according to the manufacturer's instructions. The sensors were attached to the participants' bilateral dorsal foot, bilateral shank, and pelvis using tape and elastic bandages.

The data collection was conducted at an outdoor track-and-field stadium with a 400 m tartan track. After attaching the sensors, the participants completed a warm-up of 800 m running at self-selected speed. The study protocol consisted of six consecutive laps, each at a progressively increasing speed (Table 3). Participants started running at 8 km/h in the first lap and

increased their speed by 1 km/h per lap, reaching 13 km/h in the final lap. This speed range was well represented in the training set of the ML model (7–14 km/h), (Höschler, Halmich, Schranz, Fritz, et al., 2025). Running speed was controlled using cones placed on the track every 50 meters, which the participants had to reach in a certain time interval. The time intervals were indicated by auditory cues from a wrist-worn watch (Forerunner 245, Garmin, Schaffhausen, Switzerland). A continuous countdown displayed the remaining time to reach the next cone, with the final five seconds of each interval indicated by additional beeps. Participants were instructed to maintain a constant pace within each lap. Except for one participant who was unable to maintain the required 13 km/h on the last lap, all participants completed the protocol and reported that they had no difficulty following the speed targets.

Table 3

Overview of the Study Protocol

Lap	Accumulated Distance [m]	Target Speed [km/h]	Target Speed [m/s]	Time Interval à 50 m [s]	Elapsed Time [s]
1	400	8	2.2	23	184
2	800	9	2.5	20	344
3	1200	10	2.8	18	488
4	1600	11	3.1	16	616
5	2000	12	3.3	15	736
6	2400	13	3.6	14	848

The study protocol consisted of 6 laps à 400 m, with speed increases by 1 km/h every lap. The first lap was run at 8 and the last lap at 13 km/h.

In addition to this one participant dropping out, six participants experienced sensor loss or malfunction affecting at least one sensor at some point of the protocol, resulting in partial data loss. Four participants lost data from a single foot or shank sensor, but as the sensors of unaffected side remained intact, their unilateral data were retained for analysis. Two participants (1 M, 1 F), however, experienced failure of the pelvis sensor, which is essential for estimating kinetics on both sides. As a result, these participants were excluded from the analysis. The remaining dataset contained 22 participants with bilateral data and four with unilateral data only.

Data Processing

After exporting the 3D acceleration and gyroscope data from the IMUs' onboard memory, the recordings were sliced to a total duration of 848 seconds (8 x 50 m time intervals of 23, 20, 18, 16, 15, and 14 s, Table 2) removing any data before the start and after the completion of the protocol. Visual inspection of raw signals confirmed consistent signal quality and no substantial drift. Data processing followed Höschler, Halmich, Schranz, Fritz, et al. (2025) and was implemented in Python using *PyTorch*, *pandas*, *NumPy*, and *SciPy*. IMU data were down-sampled to 400 Hz and low-pass filtered with a bi-directional second-order Butterworth filter using signal-specific cut-off frequencies (mean \pm standard deviation: 24 ± 3.7 Hz; Yu et al., 1999). The data were z-score normalized using the same mean and standard deviation values as for the training dataset.

For kinetic predictions, a pre-trained instance of the validated CNN was loaded into a Python environment and initialized using the state dictionary containing the pretrained weights and bias terms for all layers. The preprocessed IMU data were segmented into 1-second windows and fed into the model to generate time-series predictions of joint moments and GRFs normalized to body mass. The predicted outputs were then concatenated to reconstruct the full-length time series and annotated with time indices, lap segments (Turn 1, Straight 1, Turn 2, Straight 2), and running

speed. To ensure steady state running and to avoid asymmetries due to curved running (Alt et al., 2015), only data from the second straight segment of each lap (final 100 m) were included in the analysis.

Gait events—initial contact (IC) and terminal contact (TC)—were detected using a 20 N threshold of the predicted vertical GRF. The accuracy of this approach, validated against force data from an instrumented treadmill, was -24 ± 22 milliseconds (ms) (mean \pm standard deviation) for IC detection and 3 ± 14 ms for TC detection. Further details on the validation are provided in the supplementary material (Appendix A.1). For each participant and speed condition, individual steps, representing stance phases from IC to TC, were extracted from the continuous data.

The field dataset contained 13,516 individual steps. A single step with an unrealistic stance time shorter than 100 ms was removed. All retained steps were time-normalized to 101 data points. Outlier detection was performed using functional boxplots (Gunning et al., 2024), implemented in the *scikit-fda* Python package (Ramos-Carreño et al., 2024), detecting steps where any kinetic time-series metric (e.g., sagittal ankle moment) exceeded the 1.5 times interquartile range (Figure 1). In total, 2121 outlier steps (15.7 %) were removed, evenly distributed across participants (range: 11–23%, SD: 3.7%). The remaining steps were averaged with left and right-side data pooled together. The resulting field dataset contained 156 observations (26 participants x 6 speeds) à 101 data points for each of the 12 kinetic metrics.

Statistics

To analyze the influence of running speed on kinetics, statistical parametric mapping (SPM) methods were applied using the *spm1d* package in Python (Pataky, 2012). We used a one-way repeated-measures ANOVA (RM-ANOVA) to identify phases of significant differences between running speeds and paired t-test for post-hoc comparisons of all 15 speed pairs. The thresholds of statistical significance were set to $p \leq 0.05$ and 0.0033 for RM-ANOVA and paired t-test (Bonferroni corrected), respectively.

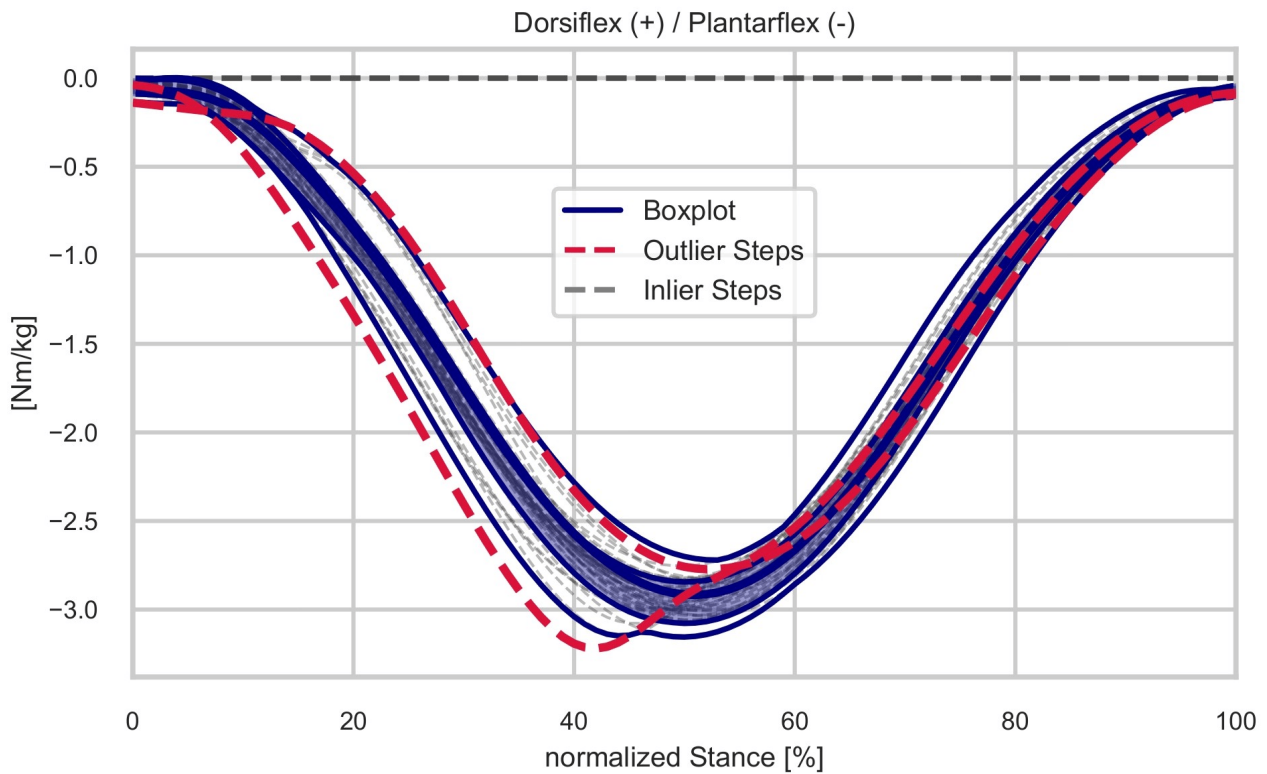


Figure 1 Detection of Outlier Steps Using Functional Boxplots

Representative example of sagittal ankle moment of a participant running at 10 km/h. Analogue to a classical boxplot, the functional boxplot displays the median, upper and lower quartiles, as well as the 1.5 times interquartile range (blue lines). The interquartile range (“box” in classical boxplot) is displayed as shaded blue. Individual steps are displayed as dashed lines: inliers thin grey; outliers thick red.

The assumption of normality of residuals for RM-ANOVA cannot be directly tested with only one mean step per participant per condition. Therefore, we conducted both parametric and non-parametric RM-ANOVAs (10,000 iterations), which yielded highly consistent results. For paired t-tests, normality was confirmed for all speed-pairs using the Shapiro-Wilk test.

Usage of Artificial Intelligence (AI) Tools

The following AI tools were used in the preparation of the manuscript: OpenAI’s ChatGPT (model GPT-4-turbo, version June 2025, <https://openai.com/>) was used for language refinement and clarity. DeepL Write (version June 2025, DeepL SE, <https://www.deepl.com/write>)

was used to improve grammar and style. All outputs were critically reviewed and edited by the authors to ensure accuracy, originality, and alignment with the research objectives.

Results

Model Validation

Predictions of vertical and antero-posterior GRFs showed the highest agreement with reference data (mean PHSS ICC = 0.98 and 0.97) and the lowest relative errors (nRMSE = 0.05 and 0.06). Medio-lateral GRFs demonstrated lower agreement (ICC = 0.69) and higher relative error (nRMSE = 0.19), as well as the

largest variation between the four test participants (PHSS ICC range: 0.47–0.86). RMSE was highest in the vertical GRF component (mean PHSS = 1.09 N/kg).

For ankle moments, sagittal-plane predictions yielded the highest agreement and lowest relative errors (mean CONT ICC = 0.98; nRMSE = 0.03) and a mean RMSE of 0.08 Nm/kg. Agreement was lower and relative errors were higher in the frontal and transverse planes (ICC = 0.90 and 0.93; nRMSE = 0.05 and 0.04). Agreement and accuracy decreased during PHSS, most notably in the frontal plane (ICC = 0.81; nRMSE = 0.17). The sagittal plane showed the highest RMSE (mean PHSS = 0.20 Nm/kg).

For knee moments, sagittal-plane predictions showed highest agreement (mean CONT ICC = 0.97) and lowest relative error (nRMSE = 0.03). Frontal and transverse plane predictions showed lower agreement (ICC = 0.77 and 0.90) and higher relative errors (nRMSE = 0.07 and 0.04). During PHSS, transverse-plane agreement remained high (ICC 0.86), whereas the frontal plane showed lower agreement and higher relative errors (ICC 0.63; nRMSE 0.24). The mean RMSE values during PHSS were 0.18, 0.15 and 0.06 Nm/kg for sagittal, frontal, and transverse plane, respectively. Frontal plane predictions showed a large variation (PHSS ICC range: 0.33–0.85)

For hip moments, agreement was higher and relative error was lower in the sagittal and frontal planes (mean CONT ICC = 0.85 and 0.89; nRMSE = 0.06 and 0.06) than in the transverse plane (ICC 0.58; nRMSE = 0.07). During PHSS, agreement decreased and relative error increased across all planes, especially in the sagittal and transverse plane (mean PHSS ICC = 0.62 and 0.39; nRMSE = 0.17 and 0.20). Sagittal plane predictions showed the highest RMSE (mean PHSS = 0.34 Nm/kg) and large variation between participants (PHSS ICC range: 0.52–0.81)

The complete validation results with the individual values for the four test participants can be found in Appendix A.2.

Influence of Running Speed

The results of the SPM analyses with RM-ANOVA and post-hoc comparisons for GRFs and joint moments are presented in Figure 2, Figure 3, Figure 4 and Figure 5 with continuous SPM F- and t-value plots provided in Appendix A.3.

Three-dimensional GRFs showed a significant speed effect during most of the stance phase. Vertical GRF showed significant speed effects, including increases in the peak region ($p < 0.001$), though the effect diminished at higher speeds. Anterior and posterior forces exhibited similar pattern in their peak regions ($p < 0.001$). Almost all speed increments resulted in significantly higher posterior forces, while anterior forces increased with large effects across all increments. In the medio-lateral direction, lateral forces during late stance differed significantly between all speed increments. Medial forces during midstance decreased at higher speeds, with small but significant post-hoc differences observed only between 13 km/h and speeds below 12 km/h.

For ankle moments, speed had a significant effect on all three moment components throughout most of the stance phase. Plantarflexion moment increased with running speed ($p < 0.001$), with significant differences between all speed increments. Higher speeds also resulted in a delayed onset and steeper slopes of the plantarflexion moment ($p = 0.002$). A similar pattern was observed for inversion and external rotation ($p = 0.007$ and $p = 0.003$), accompanied by an earlier drop in the inversion moment during late stance ($p < 0.001$). Regions near the peak moments (i.e., 40–60 % of stance) in all three planes were significantly affected by speed, with significant differences between almost all speed increments.

Running speed also significantly influenced knee moments throughout most of the stance phase ($p < 0.001$), with increased extension, abduction, and internal rotation moments at higher running speeds. However, the effects plateaued beyond 10 km/h, with less pronounced differences between speed increments at faster speeds.

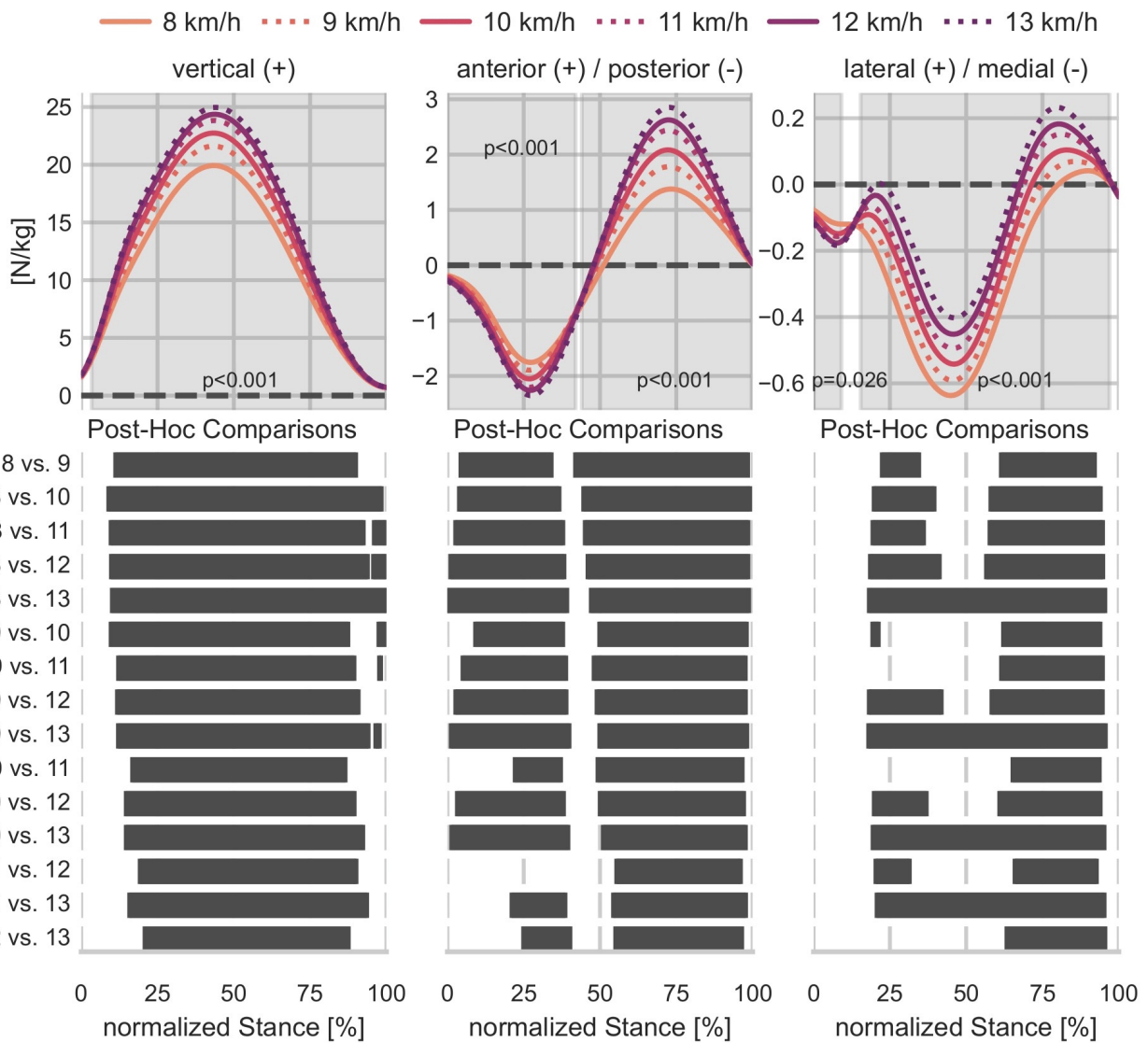


Figure 2 Speed Influence on 3D Ground Reaction Forces

Mean ground reaction force time-series (normalized to body mass) for running speeds between 8 and 13 km/h. Grey shaded bands indicate areas of statistical significance ($\alpha < 0.05$) of SPM repeated measures ANOVA between the six running speeds. The horizontal dark bars below indicate areas of statistical significance ($\alpha < 0.0033$) for pairwise post-hoc comparisons between speed pairs using SPM paired t-test.

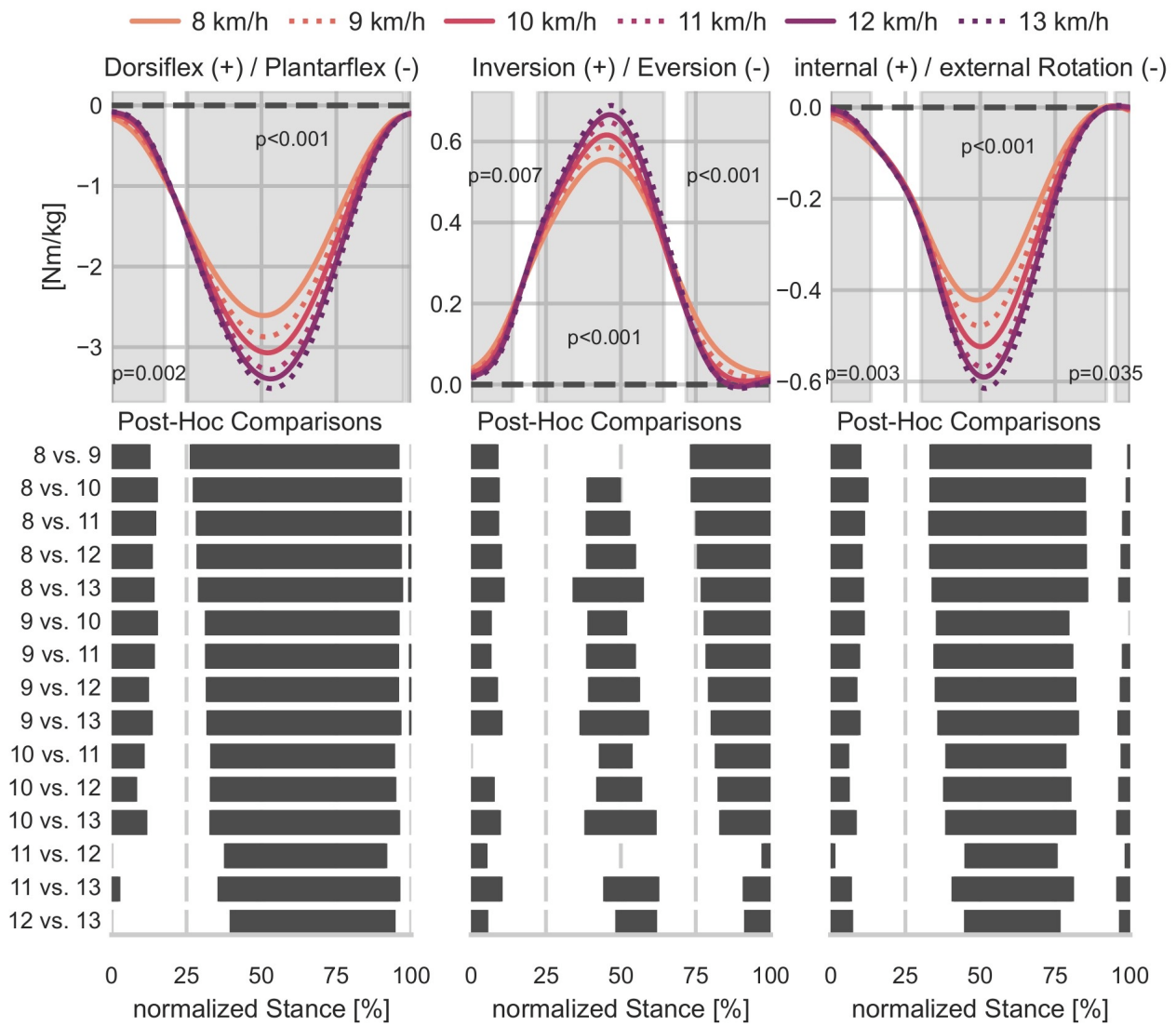


Figure 3 Speed Influence on 3D Ankle Moments

Mean ankle moment time-series (normalized to body mass) for running speeds between 8 and 13 km/h. Grey shaded bands indicate areas of statistical significance ($\alpha < 0.05$) of SPM repeated measures ANOVA between the six running speeds. The horizontal dark bars below indicate areas of statistical significance ($\alpha < 0.0033$) for pairwise post-hoc comparisons between speed pairs using SPM paired t-test.

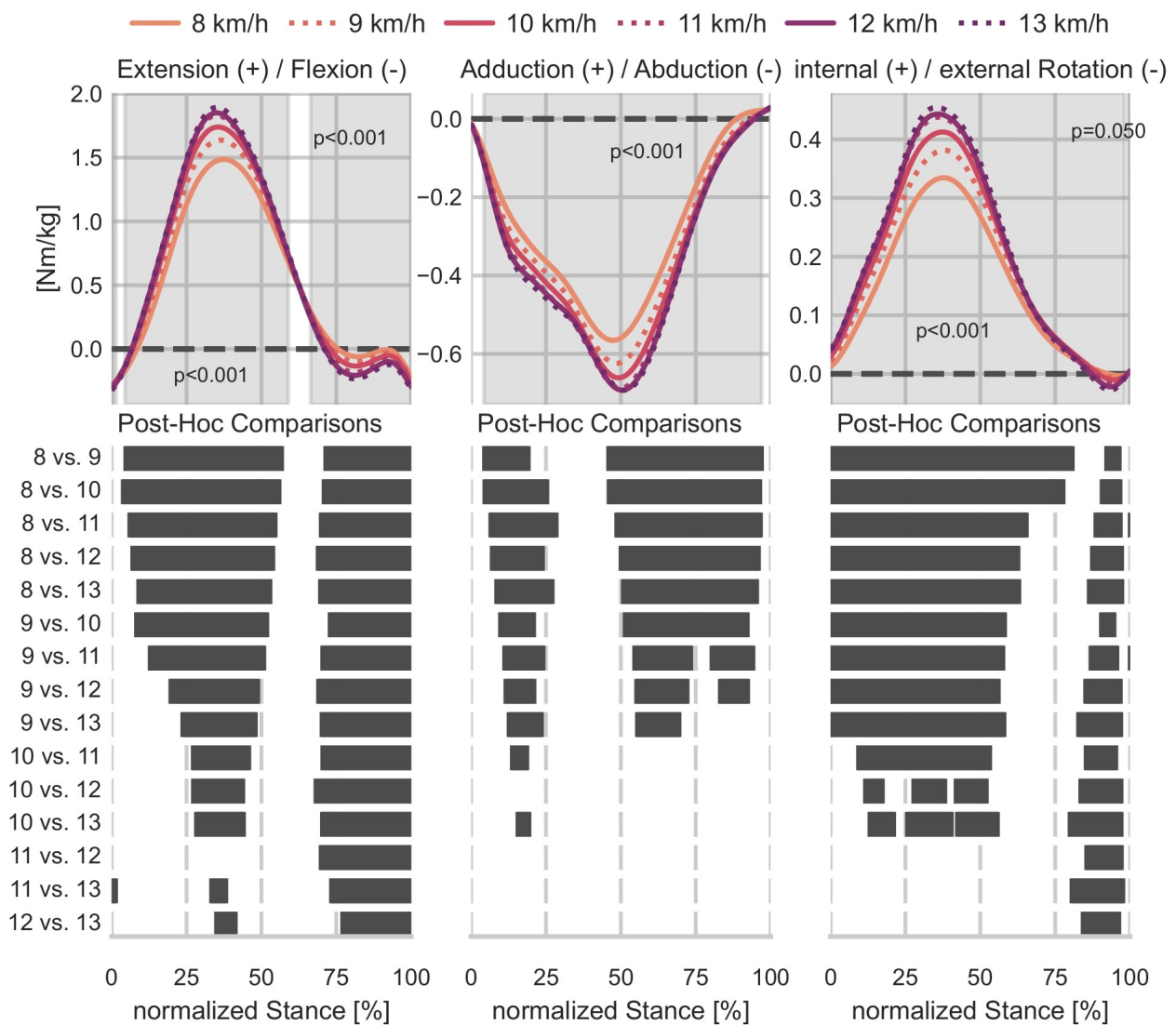


Figure 4 Speed Influence on 3D Knee Moments

Mean knee moment time-series (normalized to body mass) for running speeds between 8 and 13 km/h. Grey shaded bands indicate areas of statistical significance ($\alpha < 0.05$) of SPM repeated measures ANOVA between the six running speeds. The horizontal dark bars below indicate areas of statistical significance ($\alpha < 0.0033$) for pairwise post-hoc comparisons between speed pairs using SPM paired t-test.

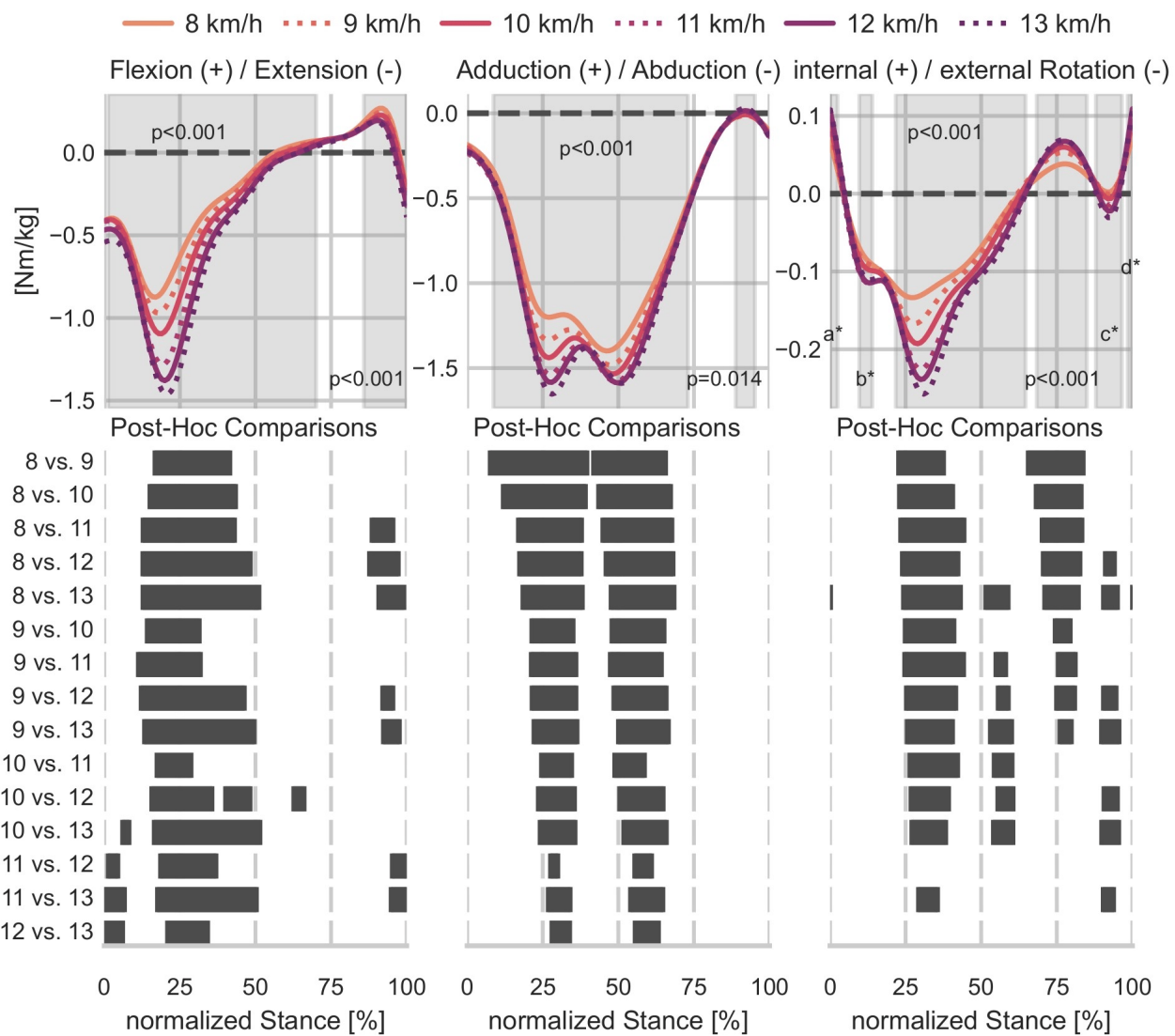


Figure 5 Speed Influence on 3D Hip Moments

Mean hip moment time-series (normalized to body mass) for running speeds between 8 and 13 km/h. Grey shaded bands indicate areas of statistical significance ($\alpha < 0.05$) of SPM repeated measures ANOVA between the six running speeds. The horizontal dark bars below indicate areas of statistical significance ($\alpha < 0.0033$) for pairwise post-hoc comparisons between speed pairs using SPM paired t-test. a*: $p = 0.041$, b*: $p = 0.029$, c*: $p = 0.005$, d*: $p = 0.046$.

For hip moments, increasing running speed resulted in greater extension moments during early stance ($p < 0.001$) and small but significant increases in flexion moments during late stance ($p < 0.001$). Post-hoc comparisons revealed significant increases in extension moments at every speed increment. In the frontal

plane, higher abduction moments were observed during early and midstance at higher running speeds ($p < 0.001$), with significant differences between all speed increments. The differences were more pronounced at slower speeds (8-10 km/h) than at faster speeds. Rotational moment changes were more local-

ized, with speed increases leading to significantly higher external rotation moments in early stance and greater internal rotation moments in late stance. The effects were more pronounced below 10 km/h, diminishing at higher speeds.

Discussion

This study demonstrates the feasibility of using wearable sensors and machine learning to estimate continuous kinetic time-series during outdoor running, enabling the analysis of speed effects on joint loading under real-world conditions. To support this application, we have first developed and validated a CNN model for estimating GRFs and lower-limb joint moments from IMU data, using a previously collected treadmill running dataset for model training. The model achieves high accuracy and agreement across most joints and movement planes. GRFs are predicted with excellent agreement in the vertical and antero-posterior directions and moderate agreement in the medio-lateral direction, consistent with previous studies using wearable sensors and ML models (Carter et al., 2024a; Honert et al., 2022; Wouda et al., 2018). Ankle moments are estimated with good to excellent agreement in all dimensions, aligning with prior findings (Carter et al., 2024b; Liew et al., 2021; Long, Pavicic, et al., 2023; Mundt et al., 2020; Zhang et al., 2022). Knee moment predictions show high accuracy and good to excellent agreement in the sagittal and transverse plane, while frontal plane predictions are less accurate and vary considerably between participants, confirming results from previous work (Carter et al., 2024b; Höschler et al., 2024; Höschler, Halmich, Schranz, Fritz, et al., 2025; Stetter et al., 2020). Hip moment predictions, particularly in the sagittal and transverse plane, are less accurate with only poor to moderate agreement and considerable variation between participants. These findings align with previous reports highlighting the difficulty of estimating hip kinetics from IMU data (Carter et al., 2024b; Liew et al., 2021). Overall, the model generalizes to multiple joints and GRFs. However, the low performance from some joints and dimensions (i.e., medio-lateral GRFs,

frontal plane knee moments, sagittal and transverse hip moments), as well as the variability between participants for some metrics, should be considered when interpreting the model's estimates. Overall, our results demonstrate that ML models trained on wearable sensor data can generalize well to estimate running kinetics across multiple joints.

Building on this validated framework, we have applied the trained CNN in a field study to examine how running speed modulates joint kinetics and GRFs during outdoor running. Running speed significantly affects all kinetic time-series, with higher speeds resulting in greater joint moments and GRFs. This aligns with lab-based findings (Dorn et al., 2012; Fukuchi et al., 2017). Faster running results in increased GRFs, particularly in the vertical and anterior directions, due to the greater push-off forces and longer aerial phases required to increase step length at submaximal speeds (Dorn et al., 2012). The increase in posterior force likely results from a greater foot-to-center-of-mass distance at ground contact, which increases braking forces because the GRF vector is directed more posteriorly relative to the body's center of mass (Nilsson & Thorstensson, 1989). These force patterns, in turn, explain the increases in joint moments, following fundamental inverse dynamics principles where greater external forces lead to greater internal joint forces and moments (Selbie et al., 2004).

At the ankle, running speed increases lead to greater net joint moments in all three planes. Previous studies have demonstrated the important role of plantar flexor muscles in generating the high forces required for propulsion (Dorn et al., 2012; Fukuchi et al., 2017; Orendurff et al., 2018). Consequently, the ankle joint experiences the largest increase in loading compared to hip and knee joint with speed. This relationship has been linked to higher injury risks in the posterior lower leg and plantar foot at faster running speeds (Petersen et al., 2014). Additionally, our results also show increased inversion and internal rotation moments, as well as higher rates of moment change. Since excessive eversion (pronation) movements have been associated with injury development (Papagian-

naki et al., 2020), these results may suggest that higher speeds place greater demands on the ankle joint and elevate injury risks. Runners with ankle-related issues may benefit from running at slower speeds to lower their joint loading.

Knee moments increase with increasing speeds up to approximately 10 km/h and then plateau at higher speeds. The additional force production required at higher speeds appears to be increasingly handled by the ankle plantar flexors rather than the knee extensors (Petersen et al., 2014). This shift likely represents an efficient strategy as shorter ground contact times enhance elastic energy recycling in the Achilles tendon (Abdelsattar et al., 2018), while the plantar flexors work at favorable positions on their force-length and force-velocity profiles (Dorn et al., 2012), allowing for effective force production with minimal metabolic cost (Fletcher & MacIntosh, 2015). Our findings align with previous studies that also have reported no further increase in knee moments at fast running speeds (Fukuchi et al., 2017; Schache et al., 2011). This suggests that running at faster speeds (> 10 km/h) places similar mechanical demands on the knee joint.

Hip moments increase with speed, though the effects in the non-sagittal planes diminish at higher speeds. Our results align with Fukuchi et al. (2017), who have found larger increases in 3D hip moments between 8.75 to 12.6 km/h than between 12.6 to 16.2 km/h. However, other studies with similar speed range have reported no significant speed effects on sagittal hip moments (Orendurff et al., 2018). The increases in frontal and transverse plane hip moments in our study suggest greater stabilization demands at higher speeds, though these effects level off beyond 10 km/h.

Overall, we have observed consistent increases in GRFs, ankle moments, and sagittal hip moments with speed, whereas knee and non-sagittal hip moments increase at slower speeds but plateau beyond 10 km/h. This suggests that transitioning from slow (< 10 km/h) to moderate (10 - 13 km/h) speeds results mainly in greater ankle and sagittal hip loading, while knee and non-sagittal hip moments remain unchanged. Slower running may provide an effective

strategy to reduce ankle and hip loading, potentially lowering injury risks for recreational runners. The observed effects of running speed on kinetics during outdoor running align with previous lab-based findings, building confidence in the wearable-based methods for real-world biomechanical assessments.

When applying trained ML models to new scenarios, it is essential to consider whether model predictions reflect true biomechanical patterns or are biased toward the conditions under which the model was trained (Halilaj et al., 2018). In our case, the model has been trained on treadmill-based laboratory data, which raises the question of whether the observed speed effects in outdoor running genuinely represent in-field biomechanics or are merely replications of lab-based trends encoded in the training data. We acknowledge this as a key limitation. However, several factors support the validity of our approach. Prior research has demonstrated strong biomechanical similarities between treadmill and overground running (van Hooren et al., 2020), and a comparison of IMU signals at 10 km/h reveals high agreement between treadmill and outdoor conditions (Appendix A.4). Furthermore, the model reproduces well-known speed-dependent trends in GRFs and joint moments and was able to predict these effects from IMU data collected under different (ecological) conditions, without providing speed data as input. This supports the model's generalizability and suggests that it captures meaningful biomechanical changes rather than reproducing training-set patterns.

To promote valid and generalizable predictions, several aspects have been carefully controlled. We have used consistent inclusion criteria, ensured identical sensor placement, and covered a running speed range that was well represented in the training data. Nonetheless, caution is warranted when applying ML models beyond their original domain. If sufficient similarity between training and application data cannot be ensured, transfer learning can offer a solution by fine-tuning a pre-trained model to new conditions (e.g., different populations, environments, or movement tasks) using a limited amount of additional data (Goodfellow

et al., 2016). Alternatively, individualized models, particularly beneficial in high-performance settings, have shown advantages in capturing subject-specific biomechanics and improving predictive accuracy (Honert et al., 2022; Moghadam et al., 2023). Future work could explore these approaches to further enhance model accuracy and generalizability.

Furthermore, some predicted metrics show limited reliability and substantial between-participant variability, specifically the medio-lateral GRFs, frontal plane knee moments, sagittal and transverse plane hip moments. Increasing training dataset size, either through additional data collection, combining compatible datasets, or applying data-augmentation strategies (Halmich et al., 2025), may help improve and stabilize generalizability for these metrics.

In addition, this study was conducted on a 400 m track and analysis was restricted to straight-line running. Although these conditions ensure consistency between training and application conditions, which is a key requirement for valid ML predictions, they do not fully represent the variability encountered in unconstrained outdoor running (e.g., uneven terrain, turns, variable surfaces). Thus, the present findings should be viewed as an intermediate step toward “real-world” running analysis. Future work should extend model validation and inference to more heterogeneous outdoor environments to evaluate robustness and ecological validity.

Beyond the considerations around model generalizability, further limitations should be noted. We have not performed an SPM regression, which could have provided a more detailed analysis regarding the magnitude of speed effects. This decision has been guided by our expectation of non-linear effects, particularly for knee moments. Additionally, we have focused on mean values across participants, which creates a “non-existing” time-series that may obscure individual variations in biomechanical responses to speed changes. To address this, we repeated the RM-ANOVA analysis using 20 randomly selected steps per condition and participant. The resulting significant regions were highly consistent with the averaged-step analysis and all biomechanical interpretations remained

unchanged. These results are provided in Appendix A.5. Future research could examine inter-individual differences and explore subgroup analyses based on factors such as sex, running experience, or physiological responses (e.g., heart rate).

Conclusion

This study demonstrates that increasing running speed significantly affects lower-body joint kinetics and GRFs during outdoor running. Our results align with previous laboratory-based findings, confirming that higher speeds lead to increased joint loading, particularly at the ankle and hip, while knee loading plateaus beyond 10 km/h. These findings provide insights into the mechanical demands of running at different speeds and suggest that runners aiming to reduce joint stress—especially at the hip and ankle—may benefit from running at slower speeds. This is specifically important in the case of local pain and rehabilitation phases.

Beyond its biomechanical contributions, this study demonstrates the potential of wearable sensors and ML-based methods as a complement to lab-based measurements for analyzing running during controlled outdoor conditions. This approach is cost-effective, scalable, and enables extended data collection during regular training sessions. Such long-duration measurements open opportunities for future research on fatigue, injury risk, individual responses to training, and potentially real-time monitoring of joint loading. These methods also provide a foundation for longitudinal studies aimed at understanding how load distribution changes over time and how these changes relate to injury development and prevention. As ML models continue to improve and training datasets become more diverse, they may help shift the study of running mechanics from isolated lab-based snapshots toward more continuous and ecologically valid evaluations in real-world environments.

References

- Abdelsattar, M., Konrad, A., & Tilp, M. (2018). Relationship between achilles tendon stiffness and ground contact time during drop jumps. *Journal of Sports Science & Medicine, 17*(2), 223–228.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In A. Teredesai, V. Kumar, Y. Li, R. Rosales, E. Terzi, & G. Karypis (Eds.), *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2623–2631). <https://doi.org/10.1145/3292500.3330701>
- Alt, T., Heinrich, K., Funken, J., & Potthast, W. (2015). Lower extremity kinematics of athletics curve sprinting. *Journal of Sports Sciences, 33*(6), 552–560. <https://doi.org/10.1080/02640414.2014.960881>
- Arampatzis, A., Brüggemann, G. P., & Metzler, V. (1999). The effect of speed on leg stiffness and joint kinetics in human running. *Journal of Biomechanics, 32*(12), 1349–1353. [https://doi.org/10.1016/S0021-9290\(99\)00133-5](https://doi.org/10.1016/S0021-9290(99)00133-5)
- Benson, L. C., Clermont, C. A., Watari, R., Exley, T., & Ferber, R. (2019). Automated accelerometer-based gait event detection during multiple running conditions. *Sensors, 19*(7), Article 1483. <https://doi.org/10.3390/s19071483>
- 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), Article 1722. <https://doi.org/10.3390/s22051722>
- Bernhart, S., Kranzinger, S., Berger, A., & Peternell, G. (2022). Ground contact time estimating wearable sensor to measure spatio-temporal aspects of gait. *Sensors, 22*(9), Article 3132. <https://doi.org/10.3390/s22093132>
- Blazey, P., Michie, T. V., & Napier, C. (2021). A narrative review of running wearable measurement system accuracy and reliability: Can we make running shoe prescription objective? *Footwear Science, 13*(2), 117–131. <https://doi.org/10.1080/19424280.2021.1878287>
- Carter, J., Chen, X., Cazzola, D., Trewartha, G., & Preatoni, E. (2024a). Consumer-priced wearable sensors combined with deep learning can be used to accurately predict ground reaction forces during various treadmill running conditions. *PeerJ, 12*, e17896. <https://doi.org/10.7717/peerj.17896>
- Carter, J., Chen, X., Cazzola, D., Trewartha, G., & Preatoni, E. (2024b). Estimating joint moments during treadmill running using various consumer based wearable sensor locations. *ISBS Proceedings Archive, 42*(1), 152.
- Cerezuela-Espejo, V., Hernández-Belmonte, A., Courel-Ibáñez, J., Conesa-Ros, E., Martínez-Cava, A., & Pallarés, J. G. (2020). Running power meters and theoretical models based on laws of physics: Effects of environments and running conditions. *Physiology & Behavior, 223*, 112972. <https://doi.org/10.1016/j.physbeh.2020.112972>
- Davis, J. J., Meardon, S. A., Brown, A. W., Raglin, J. S., Harezlak, J., & Gruber, A. H. (2024). Are gait patterns during in-lab running representative of gait patterns during real-world training? An experimental study. *Sensors, 24*(9), Article 2892. <https://doi.org/10.3390/s24092892>
- Debertin, D., Wargel, A., & Mohr, M. (2024). Reliability of xsens IMU-based lower extremity joint angles during in-field running. *Sensors, 24*(3), Article 871. <https://doi.org/10.3390/s24030871>
- DeJong Lempke, A. F., Audet, A. P., Wasserman, M. G., Melvin, A. C., Soldes, K., Heithoff, E., Shah, S., Kozloff, K. M., & Lepley, A. S. (2025). Biomechanical differences and variability during sustained motorized treadmill running versus outdoor overground running using wearable sensors. *Journal of Biomechanics, 178*, 112443. <https://doi.org/10.1016/j.jbiomech.2024.112443>

- DeJong Lempke, A. F., Hunt, D. L., Willwerth, S. B., d'Hemecourt, P. A., Meehan, W. P., & Whitney, K. E. (2024). Biomechanical changes identified during a marathon race among high-school aged runners. *Gait & Posture*, *108*, 44–49. <https://doi.org/10.1016/j.gaitpost.2023.11.009>
- Dorn, T. W., Schache, A. G., & Pandy, M. G. (2012). Muscular strategy shift in human running: Dependence of running speed on hip and ankle muscle performance. *The Journal of Experimental Biology*, *215*(Pt 11), 1944–1956. <https://doi.org/10.1242/jeb.064527>
- Dorschky, E., Nitschke, M., Mayer, M., Weygers, I., Gassner, H., Seel, T., Eskofier, B. M., & Koelewijn, A. D. (2025). Comparing sparse inertial sensor setups for sagittal-plane walking and running reconstructions. *Frontiers in Bioengineering and Biotechnology*, *13*. <https://doi.org/10.3389/fbioe.2025.1507162>
- Dorschky, E., Nitschke, M., Seifer, A. -K., van de Bogert, A. J., & Eskofier, B. M. (2019). Estimation of gait kinematics and kinetics from inertial sensor data using optimal control of musculoskeletal models. *Journal of Biomechanics*, *95*, 109278. <https://doi.org/10.1016/j.jbiomech.2019.07.022>
- Fletcher, J. R., & MacIntosh, B. R. (2015). Achilles tendon strain energy in distance running: Consider the muscle energy cost. *Journal of Applied Physiology*, *118*(2), 193–199. <https://doi.org/10.1152/jappphysiol.00732.2014>
- Fraeulin, L., Maurer-Grubinger, C., Holzgreve, F., Groneberg, D. A., & Ohlendorf, D. (2021). Comparison of joint kinematics in transition running and isolated running in elite triathletes in overground conditions. *Sensors*, *21*(14), Article 4869. <https://doi.org/10.3390/s21144869>
- Fukuchi, R. K., Fukuchi, C. A., & Duarte, M. (2017). A public dataset of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics. *PeerJ*, *5*, e3298. <https://doi.org/10.7717/peerj.3298>
- Genitrini, M., Fritz, J., Stöggel, T., & Schwameder, H. (2023). Performance level affects full body kinematics and spatiotemporal parameters in trail running—a field study. *Sports*, *11*(10), Article 188. <https://doi.org/10.3390/sports11100188>
- Genitrini, M., Fritz, J., Stöggel, T., & Schwameder, H. (2024). Spatiotemporal parameters and kinematics differ between race stages in trail running – a field study. *Frontiers in Sports and Active Living*, *6*, Article 1406824. <https://doi.org/10.3389/fspor.2024.1406824>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <http://www.deeplearning-book.org>
- Gregory, C., Koldenhoven, R. M., Higgins, M., & Hertel, J. (2019). External ankle supports alter running biomechanics: A field-based study using wearable sensors. *Physiological Measurement*, *40*(4), 44003. <https://doi.org/10.1088/1361-6579/ab15ad>
- Gunning, E., Warmenhoven, J., Harrison, A. J., & Bargary, N. (2024). Exploring variation in biomechanical data. In E. Gunning, J. Warmenhoven, A. J. Harrison, & N. Bargary (Eds.), *Functional data analysis in biomechanics: A concise review of core techniques, applications and emerging areas* (pp. 25–37). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-68862-1_3
- Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J. L., Hastie, T. J., & Delp, S. L. (2018). Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *Journal of Biomechanics*, *81*, 1–11. <https://doi.org/10.1016/j.jbiomech.2018.09.009>
- Halmich, C., Höschler, L., Schranz, C., & Borgelt, C. (2025). Data augmentation of time-series data in human movement biomechanics: A scoping review. *PLOS ONE*, *20*(7), e0327038. <https://doi.org/10.1371/journal.pone.0327038>

- Hamstra-Wright, K. L., Huxel Bliven, K. C., & Napier, C. (2021). Training load capacity, cumulative risk, and bone stress injuries: A narrative review of a holistic approach. *Frontiers in Sports and Active Living*, 3, Article 665683. <https://doi.org/10.3389/fspor.2021.665683>
- Harrison, K., Honert, E. C., & Feeney, D. (2024). The effect of footwear fit on movement complexity in trail. *ISBS Proceedings Archive*, 42(1), 366–369.
- Hollis, C. R., Koldenhoven, R. M., Resch, J. E., & Hertel, J. (2021). Running biomechanics as measured by wearable sensors: Effects of speed and surface. *Sports Biomechanics*, 20(5), 521–531. <https://doi.org/10.1080/14763141.2019.1579366>
- Honert, E. C., Hoitz, F., Blades, S., Nigg, S. R., & Nigg, B. M. (2022). Estimating running ground reaction forces from plantar pressure during graded running. *Sensors*, 22(9), Article 3338. <https://doi.org/10.3390/s22093338>
- Höschler, L., Halmich, C., Schranz, C., Fritz, J., Čigoja, S., Ullrich, M., Koelewijn, A. D., & Schwameder, H. (2025). Wearable-based estimation of continuous 3D knee moments during running using a convolutional neural network. *Sports Biomechanics*, 1–19. <https://doi.org/10.1080/14763141.2025.2481164>
- Höschler, L., Halmich, C., Schranz, C., Fritz, J., Koelewijn, A. D., & Schwameder, H. (2024). Towards real-time assessment: Wearable-based estimation of 3D knee kinetics in running and the influence of preprocessing workflows. *ISBS Proceedings Archive*, 42(1), 406–409.
- Höschler, L., Halmich, C., Schranz, C., Koelewijn, A. D., & Schwameder, H. (2025). Evaluating the influence of sensor configuration and hyperparameter optimization on wearable-based knee moment estimation during running. *International Journal of Computer Science in Sport*, 24(2), 80–106. <https://doi.org/10.2478/ijcss-2025-0014>
- 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>
- Jaén-Carrillo, D., Roche-Seruendo, L. E., Cartón-Llorente, A., Ramírez-Campillo, R., & García-Pinillos, F. (2020). Mechanical power in endurance running: A scoping review on sensors for power output estimation during running. *Sensors*, 20(22), Article 6482. <https://doi.org/10.3390/s20226482>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kozinc, Ž., Smajla, D., & Šarabon, N. (2024). The reliability of wearable commercial sensors for outdoor assessment of running biomechanics: The effect of surface and running speed. *Sports Biomechanics*, 23(11), 2330–2343. <https://doi.org/10.1080/14763141.2021.2022746>
- Lee, C. J., & Lee, J. K. (2022). Inertial motion capture-based wearable systems for estimation of joint kinetics: A systematic review. *Sensors*, 22(7), Article 2507. <https://doi.org/10.3390/s22072507>
- Liew, B. X. W., Rügamer, D., Zhai, X., Wang, Y., Morris, S., & Netto, K. (2021). Comparing shallow, deep, and transfer learning in predicting joint moments in running. *Journal of Biomechanics*, 129, 110820. <https://doi.org/10.1016/j.jbiomech.2021.110820>
- Long, T., Outerleys, J., Yeung, J. and B., T. and Fernandez, & Besier, T. F. (2023). Predicting ankle and knee sagittal kinematics and kinetics using an ankle-mounted inertial sensor. *Computer Methods in Biomechanics and Biomedical Engineering*, 1–14. <https://doi.org/10.1080/10255842.2023.2224912>

- Long, T., Pavicic, P., & Stapleton, D. (2023). Kinetic and spatiotemporal characteristics of running during regular training sessions for collegiate male distance runners using shoe-based wearable sensors. *Journal of Athletic Training, 58*(4), 338–344. <https://doi.org/10.4085/1062-6050-0703.21>
- 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. *Scientific Reports, 13*(1), 5046. <https://doi.org/10.1038/s41598-023-31906-z>
- Mundt, M. (2023). Bridging the lab-to-field gap using machine learning: A narrative review. *Sports Biomechanics, 1*–20. <https://doi.org/10.1080/14763141.2023.2200749>
- Mundt, M., Thomsen, W., Witter, T., Koeppe, A., David, S., Bamer, F., Potthast, W., & Markert, B. (2020). Prediction of lower limb joint angles and moments during gait using artificial neural networks. *Medical & Biological Engineering & Computing, 58*(1), 211–225. <https://doi.org/10.1007/s11517-019-02061-3>
- Nilsson, J., & Thorstensson, A. (1989). Ground reaction forces at different speeds of human walking and running. *Acta Physiologica Scandinavica, 136*(2), 217–227. <https://doi.org/10.1111/j.1748-1716.1989.tb08655.x>
- Orendurff, M. S., Kobayashi, T., Tulchin-Francis, K., Tullock, A. M. H., Villarosa, C., Chan, C., Kraus, E., & Strike, S. (2018). A little bit faster: Lower extremity joint kinematics and kinetics as recreational runners achieve faster speeds. *Journal of Biomechanics, 71*, 167–175. <https://doi.org/10.1016/j.jbiomech.2018.02.010>
- Papagiannaki, M., Samoladas, E., Maropoulos, S., & Arabatzi, F. (2020). Running-related injury from an engineering, medical and sport science perspective. *Frontiers in Bioengineering and Biotechnology, 8*, Article 533391. <https://doi.org/10.3389/fbioe.2020.533391>
- Pataky, T. C. (2012). One-dimensional statistical parametric mapping in python. *Computer Methods in Biomechanics and Biomedical Engineering, 15*(3), 295–301. <https://doi.org/10.1080/10255842.2010.527837>
- Pataky, T. C., Robinson, M. A., & Vanrenterghem, J. (2013). Vector field statistical analysis of kinematic and force trajectories. *Journal of Biomechanics, 46*(14), 2394–2401. <https://doi.org/10.1016/j.jbiomech.2013.07.031>
- Pataky, T. C., Vanrenterghem, J., & Robinson, M. A. (2015). Zero- vs. one-dimensional, parametric vs. non-parametric, and confidence interval vs. hypothesis testing procedures in one-dimensional biomechanical trajectory analysis. *Journal of Biomechanics, 48*(7), 1277–1285. <https://doi.org/10.1016/j.jbiomech.2015.02.051>
- Petersen, J., Nielsen, R. O., Rasmussen, S., & Sørensen, H. (2014). Comparisons of increases in knee and ankle joint moments following an increase in running speed from 8 to 12 to 16 km·h⁻¹. *Clinical Biomechanics, 29*(9), 959–964. <https://doi.org/10.1016/j.clinbiomech.2014.09.003>
- Pini, A., Markström, J. L., & Schelin, L. (2022). Test-retest reliability measures for curve data: An overview with recommendations and supplementary code. *Sports Biomechanics, 21*(2), 179–200. <https://doi.org/10.1080/14763141.2019.1655089>
- Plesek, J., Hamill, J., Burda, M., Elavsky, S., Skypala, J., Urbaczka, J., Silvernail, J. F., Zahradnik, D., Uchytil, J., & Daniel, J. (2024). Running distance and biomechanical risk factors for plantar fasciitis: A one-year prospective 4HAIE cohort study. *Medicine and Science in Sports and Exercise*, Advance online publication. <https://doi.org/10.1249/MSS.0000000000003617>
- Ramos-Carreño, C., Torrecilla, J. L., Carbajo-Berrocal, M., Marcos, P., & Suárez, A. (2024). scikit-fda: A python package for functional data analysis. *Journal of Statistical Software, 109*(2), Article 2. <https://doi.org/10.18637/jss.v109.i02>

- Schache, A. G., Blanch, P. D., Dorn, T. W., Brown, N. A. T., Rosemond, D., & Pandy, M. G. (2011). Effect of running speed on lower limb joint kinetics. *Medicine and Science in Sports and Exercise*, *43*(7), 1260–1271. <https://doi.org/10.1249/MSS.0b013e3182084929>
- Selbie, W. S., Hamill, J., & Kepple, T. M. (2004). Three-dimensional kinetics. In D. G. E. Robertson, G. E. Caldwell, J. Hamill, G. Kamen, & S. N. Whittlesey (Eds.), *Research methods in biomechanics*. Human Kinetics.
- Stefanyshyn, D. J., Stergiou, P., Lun, V. M., Meeuwisse, W. H., & Worobets, J. T. (2006). Knee angular impulse as a predictor of patellofemoral pain in runners. *The American Journal of Sports Medicine*, *34*(11), 1844–1851. <https://doi.org/10.1177/0363546506288753>
- 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*, Article 9. <https://doi.org/10.3389/fbioe.2020.00009>
- van Hooren, B., Fuller, J. T., Buckley, J. D., Miller, J. R., Sewell, K., Rao, G., Barton, C., Bishop, C., & Willy, R. W. (2020). Is motorized treadmill running biomechanically comparable to overground running? A systematic review and meta-analysis of cross-over studies. *Sports Medicine*, *50*(4), 785–813. <https://doi.org/10.1007/s40279-019-01237-z>
- van Hooren, B., Lennartz, R., Cox, M., Hoitz, F., Plasqui, G., & Meijer, K. (2024). Differences in running technique between runners with better and poorer running economy and lower and higher mileage: An artificial neural network approach. *Scandinavian Journal of Medicine & Science in Sports*, *34*(3), e14605. <https://doi.org/10.1111/sms.14605>
- Weygers, I., Kok, M., Konings, M., Hallez, H., de Vroey, H., & Claeys, K. (2020). Inertial sensor-based lower limb joint kinematics: A methodological systematic review. *Sensors*, *20*(3), Article 673. <https://doi.org/10.3390/s20030673>
- Wouda, F. J., Giuberti, M., Bellusci, G., Maartens, E., Reenalda, J., 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*, Article 218. <https://doi.org/10.3389/fphys.2018.00218>
- Yu, B., Gabriel, D., Noble, L., & An, K. -N. (1999). Estimate of the optimum cutoff frequency for the butterworth low-pass digital filter. *Journal of Applied Biomechanics*, *15*(3), 318–329. <https://doi.org/10.1123/jab.15.3.318>
- Zhang, L., Zhu, X., Gutierrez-Farewik, E. M., & Wang, R. (2022). Ankle joint torque prediction using an NMS solver informed-ANN model and transfer learning. *IEEE Journal of Biomedical and Health Informatics*, *26*(12), 5895–5906. <https://doi.org/10.1109/JBHI.2022.3207313>
- Zrenner, M., Feldner, C., Jensen, U., Roth, N., Richer, R., & Eskofier, B. M. (2020). Evaluation of foot kinematics during endurance running on different surfaces in real-world environments. In M. Lames, A. Danilov, E. Timme, & Y. Vassilevski (Eds.), *Proceedings of the 12th International Symposium on Computer Science in Sport (IACSS 2019) (Advances in intelligent systems and computing, Vol. 1028)* (pp. 106–113). Springer. https://doi.org/10.1007/978-3-030-35048-2_13

Acknowledgements

We would like to thank all individuals who participated in this study.

Funding

The project TexSense is funded within the context of WISS 2025, der Wissenschafts- und Innovationsstrategie 2025, by the federal state of Salzburg (grant 20102-445 F2101127-FPR).

Competing interests

The authors have declared that no competing interests exist.

Data availability statement

The dataset acquired for this study and example code for reproducing the statistical analysis are available at: https://github.com/luceskywalker/TexSense_OutdoorRunning

A Appendix

A.1 Gait Event Detection

The detection of gait events –initial contact (IC) and terminal contact (TC)– is essential to separate continuous time-series data into individual steps. In laboratory studies, typically a 20 N threshold of the vertical GRF is used. In the absence of measured force data during outdoor running, we applied the 20 N threshold to the estimated vertical GRF data using IMU data and the trained ML model.

To validate this approach, we compared IC and TC events detected from the estimated GRF to ground-truth events from an instrumented treadmill. An ML model was trained on the training dataset containing data from 20 participants to estimate 3D GRFs and lower-limb joint moments. Performance was tested using an independent test set of four participants, covering 27 running conditions per participant with variations in footwear ($n=3$), speed (self-selected ± 1 km/h), and treadmill slope (level ± 5 % inclination). Steps with stance times shorter than 100 or longer than 500 ms were discarded. In total, 14,358 bilateral steps were analyzed.

The accuracy of event detection was -24 ± 22 milliseconds (ms) for IC and 3 ± 14 ms for TC (mean \pm standard deviation). Estimated stance times were 27 ± 26 ms longer than the ground truth, as IC was systematically detected earlier, while TC detection was nearly perfect (Supplementary Figure 6).

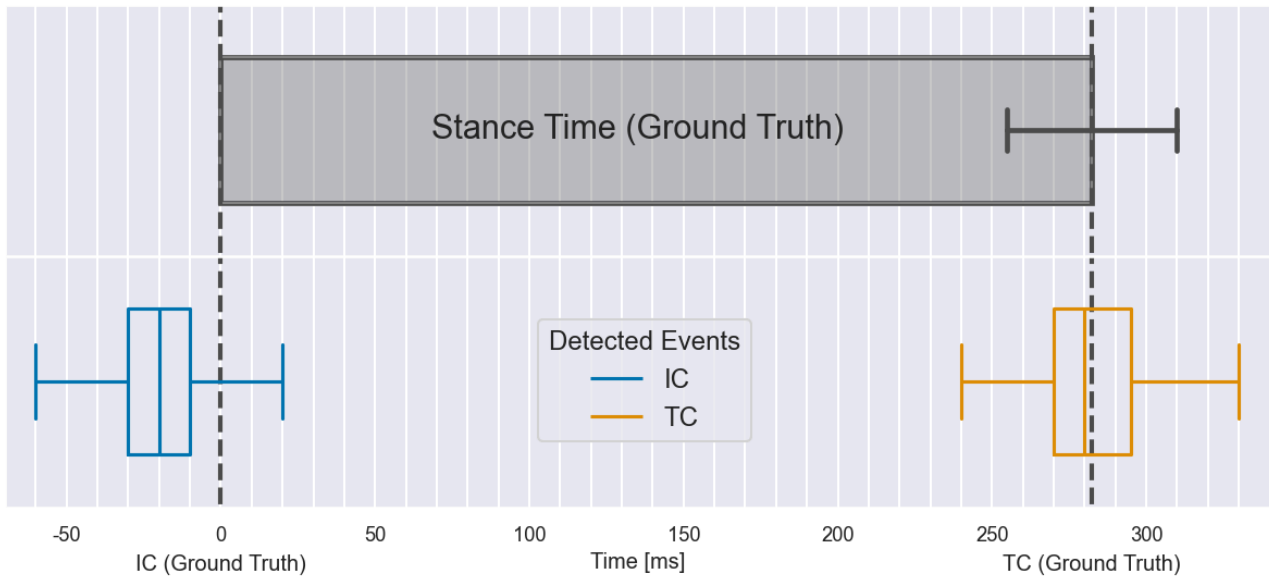


Figure 6 Gait Event Detection Accuracy

Detection accuracy of initial contact (IC, blue) and terminal contact (TC, orange) using a 20 N threshold of the estimated vertical GRF compared against the ground truth reference from the instrumented treadmill (grey). Boxes show the upper and lower quartiles as well as the median, with whiskers indicating the 1.5 times interquartile range. Smaller values than the ground truth represent a premature event detection.

These results are consistent with previous IMU-based gait detection studies. Bernhart et al. (2022) reported a -51 ms median error in ground contact time using a pelvis-mounted IMU during walking. For running, Benson et al. (2019) evaluated multiple algorithms with foot- and pelvis-mounted IMUs, reporting mean errors of -15 to 50 ms for IC and 20 to 30 ms for TC compared to force plate data. Our results fall within this range, supporting the validity of the proposed approach.

A.2 Model Validation Results

The full results of the performance evaluation of the extended model by intra-class correlation (ICC), root mean squared error (RMSE) and normalized RMSE (nRMSE) are provided in Supplementary Tables 4, 5 and 6.

Table 4
Intra-Class Correlation (ICC) Between Estimated and Ground-Truth Kinetics

Metric	Dimension	CONT				mean	PHSS				mean
		TP1	TP2	TP3	TP4		TP1	TP2	TP3	TP4	
GRF	vertical	0.99	0.99	1.00	0.99	0.99	0.98	0.99	0.99	0.97	0.98 (0.92-0.99)
		(0.93-1.00)	(0.99-1.00)	(0.98-1.00)	(0.98-1.00)	(0.98-1.00)	(0.80-0.99)	(0.97-1.00)	(0.95-1.00)	(0.94-0.99)	
		0.96	0.97	0.98	0.97	0.97 (0.89 -0.99)	0.96	0.97	0.98	0.97	
AP	ML	0.91	0.68	0.76	0.72	0.78	0.86	0.47	0.73	0.59	0.69 (0.30-0.89)
		(0.78-0.97)	(0.33-0.91)	(0.40-0.90)	(0.35-0.93)	(0.45-0.93)	(0.65-0.94)	(0.06-0.83)	(0.32-0.90)	(0.17-0.88)	
		0.98	0.98	0.99	0.97	0.98	0.97	0.96	0.98	0.94	0.96 (0.90-0.99)
Ankle Moment	sagittal	0.92	0.97	1.00	0.95	0.95-1.00	0.83	0.99	0.98	0.94	0.96 (0.90-0.99)
		(0.92-1.00)	(0.97-0.99)	(0.97-1.00)	(0.95-0.99)	(0.95-1.00)	(0.83-0.99)	(0.94-0.99)	(0.93-1.00)	(0.88-0.98)	
		0.81	0.91	0.97	0.92	0.90	0.67	0.84	0.94	0.80	0.81 (0.55-0.95)
Knee Moment	frontal	(0.43-0.97)	(0.81-0.96)	(0.91-0.99)	(0.82-0.98)	(0.74-0.98)	(0.12-0.94)	(0.66-0.91)	(0.81-0.98)	(0.61-0.95)	
		0.95	0.97	0.92	0.90	0.93	0.91	0.93	0.90	0.84	0.89 (0.78-0.97)
		(0.88-0.99)	(0.93-0.98)	(0.85-0.99)	(0.79-0.98)	(0.89-0.99)	(0.79-0.98)	(0.86-0.96)	(0.80-0.98)	(0.67-0.97)	
Ankle Moment	sagittal	0.98	0.96	0.98	0.96	0.97	0.97	0.94	0.98	0.95	0.96 (0.85-0.99)
		(0.86-0.99)	(0.92-0.98)	(0.92-0.99)	(0.92-0.99)	(0.91-0.99)	(0.78-0.99)	(0.86-0.98)	(0.89-0.99)	(0.87-0.98)	
		0.91	0.70	0.91	0.53	0.77	0.83	0.43	0.85	0.33	0.63 (0.41-0.87)
Knee Moment	frontal	(0.82-0.97)	(0.50-0.91)	(0.84-0.98)	(0.06-0.96)	(0.56-0.96)	(0.64-0.95)	(0.14-0.79)	(0.67-0.96)	(0.18-0.79)	
		0.92	0.77	0.94	0.90	0.90	0.89	0.66	0.95	0.85	0.86 (0.67-0.95)
		(0.78-0.97)	(0.62-0.91)	(0.89-0.98)	(0.82-0.96)	(0.78-0.96)	(0.68-0.98)	(0.45-0.88)	(0.86-0.99)	(0.70-0.95)	
Hip Moment	sagittal	0.85	0.90	0.83	0.85	0.85	0.68	0.81	0.54	0.52	0.62 (0.46-0.78)
		(0.79-0.90)	(0.84-0.93)	(0.73-0.88)	(0.76-0.91)	(0.78-0.91)	(0.55-0.78)	(0.72-0.89)	(0.30-0.70)	(0.29-0.73)	
		0.91	0.88	0.87	0.91	0.89	0.87	0.82	0.82	0.87	0.85 (0.71-0.92)
Hip Moment	frontal	(0.79-0.96)	(0.84-0.92)	(0.78-0.95)	(0.87-0.95)	(0.82-0.95)	(0.68-0.94)	(0.74-0.89)	(0.66-0.91)	(0.77-0.93)	
		0.61	0.62	0.53	0.59	0.58	0.41	0.49	0.32	0.39	0.39 (0.13-0.72)
		(0.44-0.76)	(0.44-0.72)	(0.36-0.70)	(0.34-0.77)	(0.40-0.74)	(0.18-0.69)	(0.25-0.61)	(0.07-0.56)	(0.02-0.70)	

Intra-Class Correlation (ICC) between the ML model's estimation and ground-truth reference for 3D lower-limb joint moments and ground reaction forces (GRFs). Displayed are the mean and 95 % confidence intervals for individual test-set participants (TP1-4) and the mean across them (bold). CONT: continuous predictions during stance and swing phases, PHSS: predictions during stance phases only. AP: antero-posterior, ML: medio-lateral.

Table 5
Root Mean Square Error (RMSE) Between Estimated and Ground-Truth Kinetics

Metric	Dimension	CONT				mean	PHSS				mean
		TP1	TP2	TP3	TP4		TP1	TP2	TP3	TP4	
GRF [N/kg]	vertical	0.47 (0.28-1.04)	0.36 (0.22-0.64)	0.35 (0.21-0.65)	0.58 (0.38-0.85)	0.45 (0.35-0.58)	1.18 (0.64-2.70)	0.85 (0.50-1.60)	0.81 (0.46-1.63)	1.41 (0.85-2.22)	1.09 (0.81-1.41)
	AP	0.10 (0.06-0.20)	0.09 (0.07-0.11)	0.08 (0.06-0.14)	0.10 (0.07-0.05)	0.09 (0.08-0.10)	0.25 (0.15-0.52)	0.23 (0.16-0.28)	0.19 (0.13-0.35)	0.24 (0.17-0.35)	0.22 (0.19-0.25)
Ankle Moment [Nm/kg]	ML	0.07 (0.05-0.12)	0.11 (0.06-0.17)	0.07 (0.05-0.10)	0.09 (0.05-0.14)	0.08 (0.07-0.11)	0.19 (0.12-0.31)	0.28 (0.15-0.43)	0.17 (0.12-0.25)	0.23 (0.13-0.37)	0.21 (0.17-0.28)
	sagittal	0.08 (0.05-0.14)	0.08 (0.06-0.11)	0.06 (0.03-0.12)	0.11 (0.07-0.17)	0.08 (0.06-0.11)	0.20 (0.12-0.36)	0.20 (0.13-0.28)	0.14 (0.07-0.28)	0.26 (0.16-0.41)	0.20 (0.14-0.26)
Knee Moment [Nm/kg]	frontal	0.04 (0.02-0.08)	0.04 (0.03-0.07)	0.02 (0.02-0.04)	0.04 (0.02-0.06)	0.03 (0.02-0.04)	0.10 (0.05-0.20)	0.11 (0.07-0.18)	0.06 (0.03-0.09)	0.09 (0.05-0.15)	0.09 (0.06-0.11)
	transverse	0.02 (0.01-0.04)	0.02 (0.01-0.02)	0.02 (0.01-0.03)	0.03 (0.01-0.05)	0.02 (0.02-0.03)	0.06 (0.03-0.10)	0.04 (0.03-0.06)	0.06 (0.03-0.08)	0.08 (0.03-0.14)	0.06 (0.04-0.08)
Hip Moment [Nm/kg]	sagittal	0.08 (0.06-0.14)	0.09 (0.06-0.13)	0.07 (0.05-0.15)	0.10 (0.07-0.14)	0.09 (0.07-0.10)	0.17 (0.11-0.32)	0.19 (0.11-0.28)	0.12 (0.09-0.31)	0.22 (0.12-0.31)	0.17 (0.12-0.19)
	frontal	0.06 (0.04-0.09)	0.07 (0.05-0.10)	0.07 (0.04-0.09)	0.09 (0.05-0.13)	0.07 (0.06-0.09)	0.12 (0.07-0.19)	0.15 (0.08-0.22)	0.13 (0.07-0.18)	0.19 (0.10-0.31)	0.15 (0.13-0.19)
Ankle Moment [Nm/kg]	transverse	0.04 (0.03-0.06)	0.05 (0.04-0.07)	0.04 (0.02-0.05)	0.04 (0.03-0.05)	0.04 (0.04-0.05)	0.06 (0.04-0.11)	0.09 (0.06-0.15)	0.04 (0.02-0.07)	0.06 (0.04-0.09)	0.06 (0.04-0.09)
	sagittal	0.23 (0.19-0.30)	0.19 (0.15-0.24)	0.26 (0.21-0.33)	0.19 (0.16-0.22)	0.22 (0.19-0.26)	0.39 (0.30-0.51)	0.27 (0.20-0.35)	0.42 (0.35-0.52)	0.24 (0.17-0.30)	0.34 (0.24-0.42)
Hip Moment [Nm/kg]	frontal	0.17 (0.13-0.23)	0.18 (0.15-0.22)	0.22 (0.16-0.28)	0.15 (0.12-0.18)	0.18 (0.15-0.22)	0.23 (0.17-0.37)	0.24 (0.19-0.34)	0.32 (0.21-0.47)	0.21 (0.16-0.25)	0.25 (0.21-0.32)
	transverse	0.07 (0.06-0.09)	0.07 (0.05-0.09)	0.11 (0.08-0.14)	0.06 (0.05-0.07)	0.08 (0.06-0.11)	0.11 (0.09-0.13)	0.10 (0.07-0.16)	0.17 (0.10-0.24)	0.07 (0.05-0.11)	0.11 (0.07-0.17)

Root mean square error (RMSE) between the ML models estimation and ground-truth reference for 3D lower-limb joint moments and ground reaction forces (GRFs). Displayed are the mean and 95 % confidence intervals for individual test-set participants (TP1-4) and the mean across them. CONT: continuous predictions during stance and swing phases, PHSS: predictions during stance phases only. AP: antero-posterior, ML: medio-lateral.

Table 6
Normalized Root Mean Square Error (nRMSE) Between Estimated and Ground-Truth Kinetics

Metric	Dimension	CONT				mean	PHSS				mean
		TP1	TP2	TP3	TP4		TP1	TP2	TP3	TP4	
GRF	vertical	0.02	0.01	0.01	0.02	0.02	0.05	0.04	0.04	0.06	0.05
		(0.01-0.06)	(0.01-0.02)	(0.01-0.03)	(0.01-0.04)	(0.01-0.04)	(0.04-0.17)	(0.04-0.07)	(0.03-0.09)	(0.04-0.10)	(0.04-0.11)
	AP	0.02	0.02	0.01	0.02	0.02	0.07	0.06	0.04	0.06	0.06
	(0.01-0.07)	(0.01-0.02)	(0.01-0.04)	(0.01-0.03)	(0.01-0.04)	(0.03-0.21)	(0.02-0.08)	(0.02-0.10)	(0.04-0.09)	(0.03-0.12)	
ML	sagittal	0.03	0.05	0.03	0.04	0.04	0.13	0.26	0.16	0.23	0.19
		(0.02-0.05)	(0.02-0.08)	(0.02-0.05)	(0.02-0.07)	(0.02-0.06)	(0.08-0.21)	(0.13-0.41)	(0.11-0.26)	(0.12-0.35)	(0.11-0.31)
	0.02	0.03	0.02	0.04	0.03	0.07	0.08	0.05	0.11	0.08	
	(0.01-0.06)	(0.02-0.04)	(0.01-0.04)	(0.02-0.06)	(0.02-0.05)	(0.04-0.17)	(0.05-0.11)	(0.03-0.11)	(0.06-0.18)	(0.02-0.14)	
Ankle Moment	frontal	0.06	0.04	0.03	0.05	0.05	0.22	0.14	0.10	0.19	0.17
		(0.03-0.14)	(0.03-0.06)	(0.02-0.06)	(0.02-0.09)	(0.02-0.09)	(0.09-0.51)	(0.11-0.20)	(0.05-0.20)	(0.08-0.35)	(0.02-0.31)
	transverse	0.04	0.03	0.03	0.05	0.04	0.12	0.09	0.12	0.17	0.13
	(0.02-0.06)	(0.02-0.04)	(0.02-0.05)	(0.02-0.10)	(0.02-0.06)	(0.05-0.21)	(0.07-0.13)	(0.05-0.19)	(0.06-0.32)	(0.02-0.21)	
Knee Moment	sagittal	0.03	0.04	0.02	0.03	0.03	0.07	0.10	0.05	0.08	0.07
		(0.02-0.07)	(0.02-0.06)	(0.02-0.05)	(0.02-0.05)	(0.02-0.06)	(0.05-0.20)	(0.06-0.17)	(0.04-0.12)	(0.05-0.13)	(0.05-0.15)
	frontal	0.05	0.09	0.04	0.10	0.07	0.15	0.39	0.12	0.36	0.24
	(0.03-0.08)	(0.04-0.16)	(0.02-0.07)	(0.05-0.16)	(0.04-0.12)	(0.05-0.27)	(0.17-0.68)	(0.07-0.22)	(0.14-0.70)	(0.11-0.47)	
Hip Moment	transverse	0.06	0.09	0.04	0.06	0.06	0.12	0.30	0.07	0.14	0.14
		(0.03-0.11)	(0.04-0.13)	(0.02-0.06)	(0.03-0.10)	(0.03-0.10)	(0.06-0.27)	(0.12-0.47)	(0.04-0.13)	(0.08-0.24)	(0.09-0.27)
	sagittal	0.06	0.05	0.08	0.07	0.06	0.14	0.10	0.24	0.18	0.17
	(0.05-0.08)	(0.04-0.06)	(0.05-0.10)	(0.04-0.08)	(0.05-0.08)	(0.11-0.17)	(0.08-0.13)	(0.17-0.32)	(0.14-0.23)	(0.13-0.23)	
frontal	0.05	0.06	0.05	0.06	0.06	0.10	0.12	0.12	0.13	0.12	
	(0.04-0.07)	(0.04-0.07)	(0.03-0.07)	(0.05-0.08)	(0.04-0.07)	(0.07-0.17)	(0.10-0.14)	(0.09-0.16)	(0.09-0.19)	(0.09-0.17)	
transverse	0.07	0.07	0.08	0.07	0.07	0.20	0.18	0.20	0.20	0.20	
(0.04-0.09)	(0.05-0.09)	(0.04-0.11)	(0.05-0.10)	(0.05-0.10)	(0.05-0.10)	(0.14-0.28)	(0.15-0.24)	(0.14-0.29)	(0.12-0.27)	(0.14-0.27)	

Normalized root mean square error (nRMSE) between the ML model's estimation and ground-truth reference for 3D lower-limb joint moments and ground reaction forces (GRFs). Displayed are the mean and 95 % confidence intervals for individual test-set participants (TP1-4) and the mean across them. CONT: continuous predictions during stance and swing phases, PHSS: predictions during stance phases only. AP: antero-posterior, ML: medio-lateral.

A.3 SPM Results

Continuous F and t-values of the SPM analysis are displayed in Supplementary Figure 7, 8, 9 and 10. Post-hoc comparisons are shown for speed increments of 1 km/h, only.

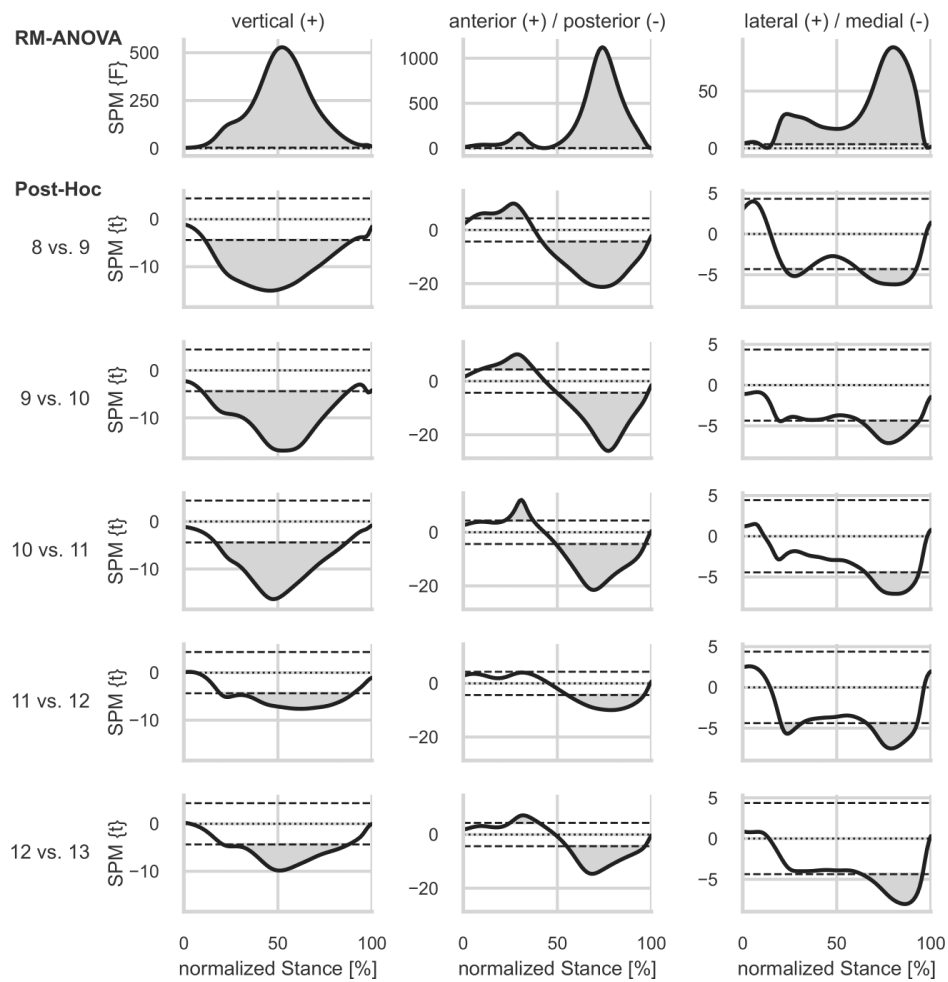


Figure 7 SPM Results Ground Reaction Forces

Continuous F-values of SPM repeated measures ANOVA (RM-ANOVA) and t-values of SPM paired t-tests for pairwise post-hoc comparisons of ground reaction forces. Areas where the thresholds of significance (dashed horizontal lines) are exceeded are filled in grey.

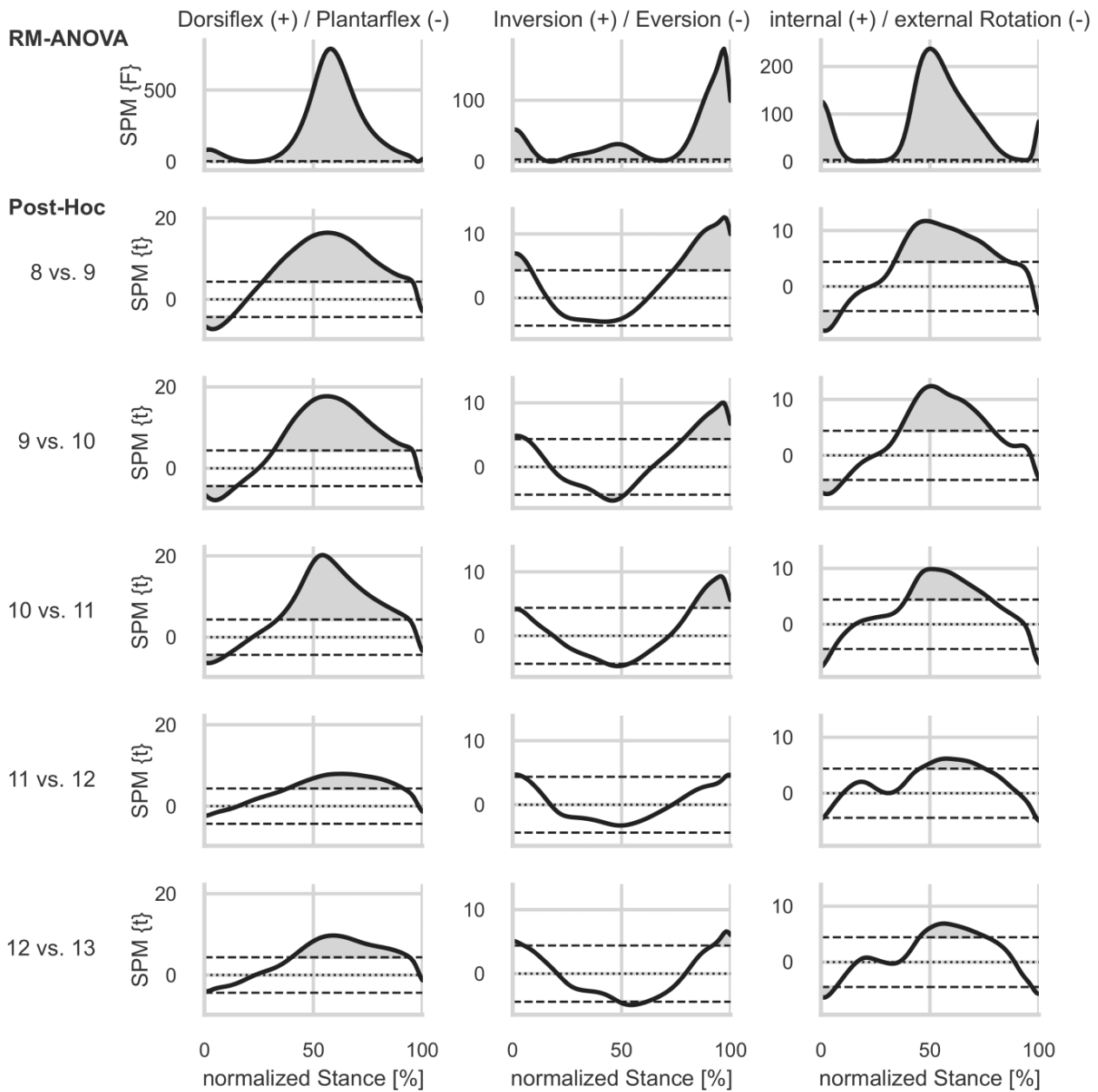


Figure 8 SPM Results Ankle Moments

Continuous F-values of SPM repeated measures ANOVA (RM-ANOVA) and t-values of SPM paired t-tests for pairwise post-hoc comparisons of ankle moments. Areas where the thresholds of significance (dashed horizontal lines) are exceeded are filled in grey.

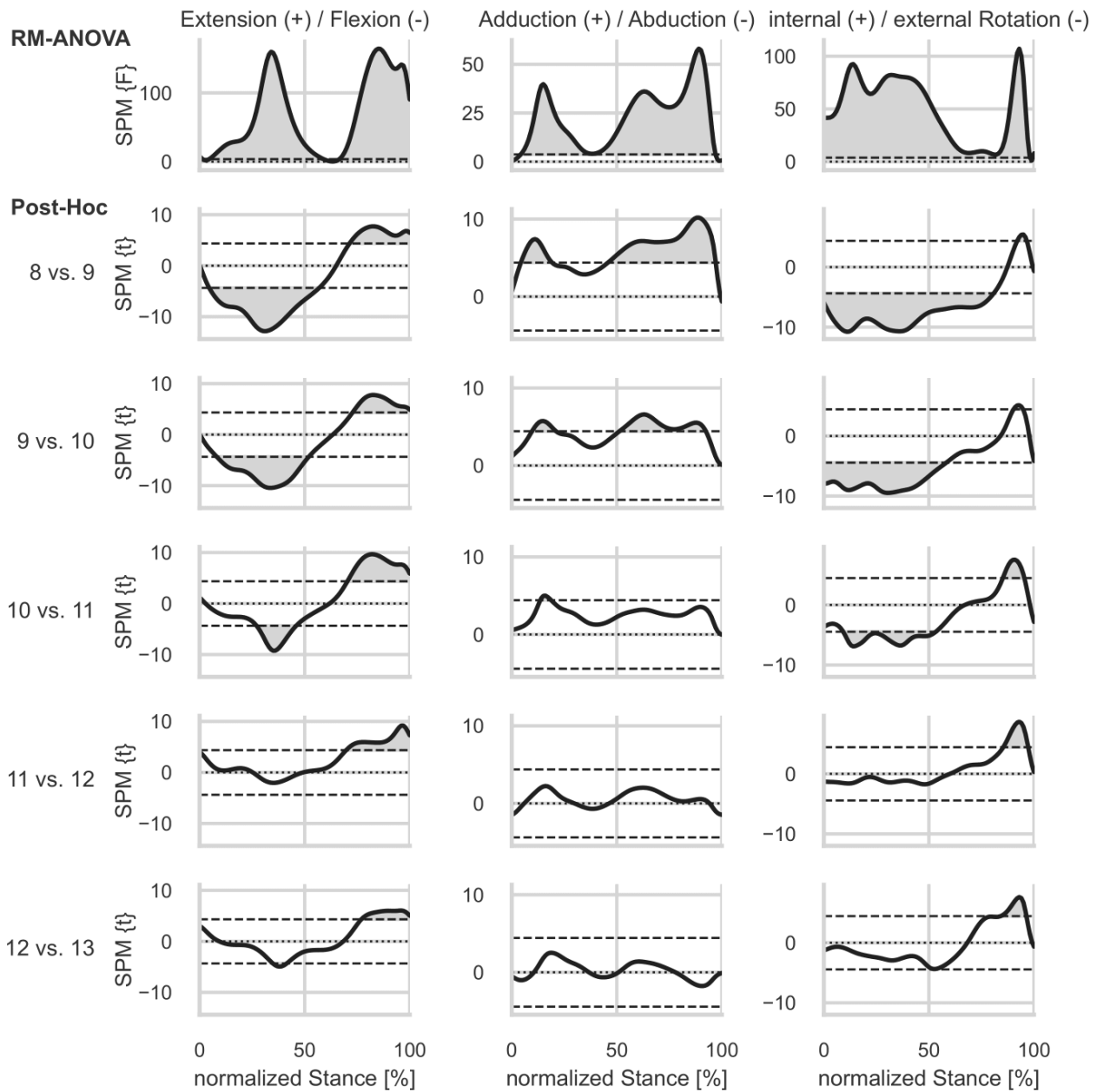


Figure 9 SPM Results Knee Moments

continuous F-values of SPM repeated measures ANOVA (RM-ANOVA) and t-values of SPM paired t-tests for pairwise post-hoc comparisons of knee moments. Areas where the thresholds of significance (dashed horizontal lines) are exceeded are filled in grey.

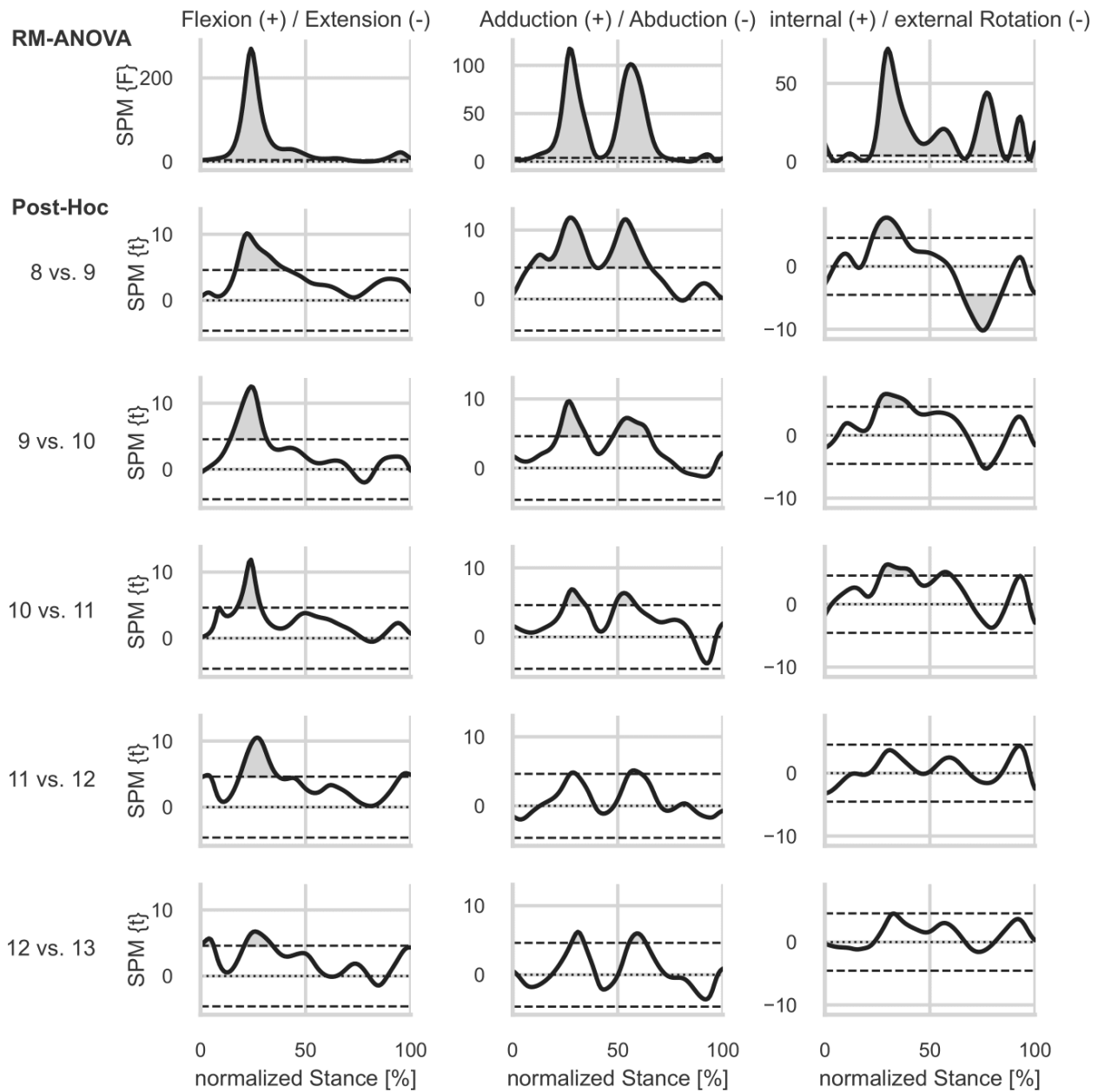


Figure 10 SPM Results Hip Moments

Continuous F-values of SPM repeated measures ANOVA (RM-ANOVA) and t-values of SPM paired t-tests for pairwise post-hoc comparisons of hip moments. Areas where the thresholds of significance (dashed horizontal lines) are exceeded are filled in grey.

A.4 IMU Signal Comparison

To ensure comparability between IMU data from lab and outdoor runs, we analyzed sensor signals at a running speed of 10 km/h from both datasets. From the lab dataset, we selected all trials where participants ran at 10 km/h (level), yielding 48 trials from 16 participants. For the outdoor dataset, we extracted the second straight segment of the 10 km/h lap for all 26 participants.

We segmented all data into stride cycles (initial contact – initial contact) and time-normalized them to 201 data points. Within each trial, we identified and removed outliers using functional boxplots, discarding strides exceeding 1.5 times the interquartile range. The remaining strides were then averaged within each trial. For the lab dataset, we further averaged the three mean strides per participant of the three different footwear conditions to obtain a single representative stride per subject. Finally, we computed the grand mean and standard deviation across subjects for each dataset and performed a correlation analysis between the dataset means to assess similarity.

The results of the correlation analysis (Supplementary Table 7) show predominantly very strong correlations (> 0.8) between the sensor signals of lab and field running. Only the gyroscope signals at the pelvis and foot display slightly lower correlations in some dimensions (> 0.64). The mean and standard deviation plots (Supplementary Figure 11) show almost perfect signal overlap, with the aforementioned exceptions. These results confirm the signal comparability between the two datasets and justify the application of the ML model to the field data.

Table 7

IMU Signal Correlations Between Lab and Field Running

Location	Side	Sensor					
		ACC			GYR		
		Dimension					
		x	y	z	x	y	z
Pelvis	left	0.98	0.89	0.83	0.95	0.66	0.65
	right	0.99	0.93	0.83	0.95	0.70	0.74
Shank	left	0.97	0.96	0.95	0.94	0.99	0.99
	right	0.86	0.91	0.93	0.87	0.99	0.99
Foot	left	0.99	0.97	0.98	0.92	1.00	0.99
	right	0.96	0.89	0.95	0.77	1.00	0.98

Correlations between the mean IMU signals of lab and field running at 10 km/h. Accelerometer (ACC) and gyroscope (GYR) signals at the three sensor dimensions (x, y, z). The sensor dimensions do not correspond to anatomical segment coordinates as no sensor-to-segment calibration was performed. The general IMU axis orientation was x: pointing down (forward in case of the foot), y: pointing left (right in case of the pelvis), z: pointing up (foot), forward (shank), backward (pelvis).

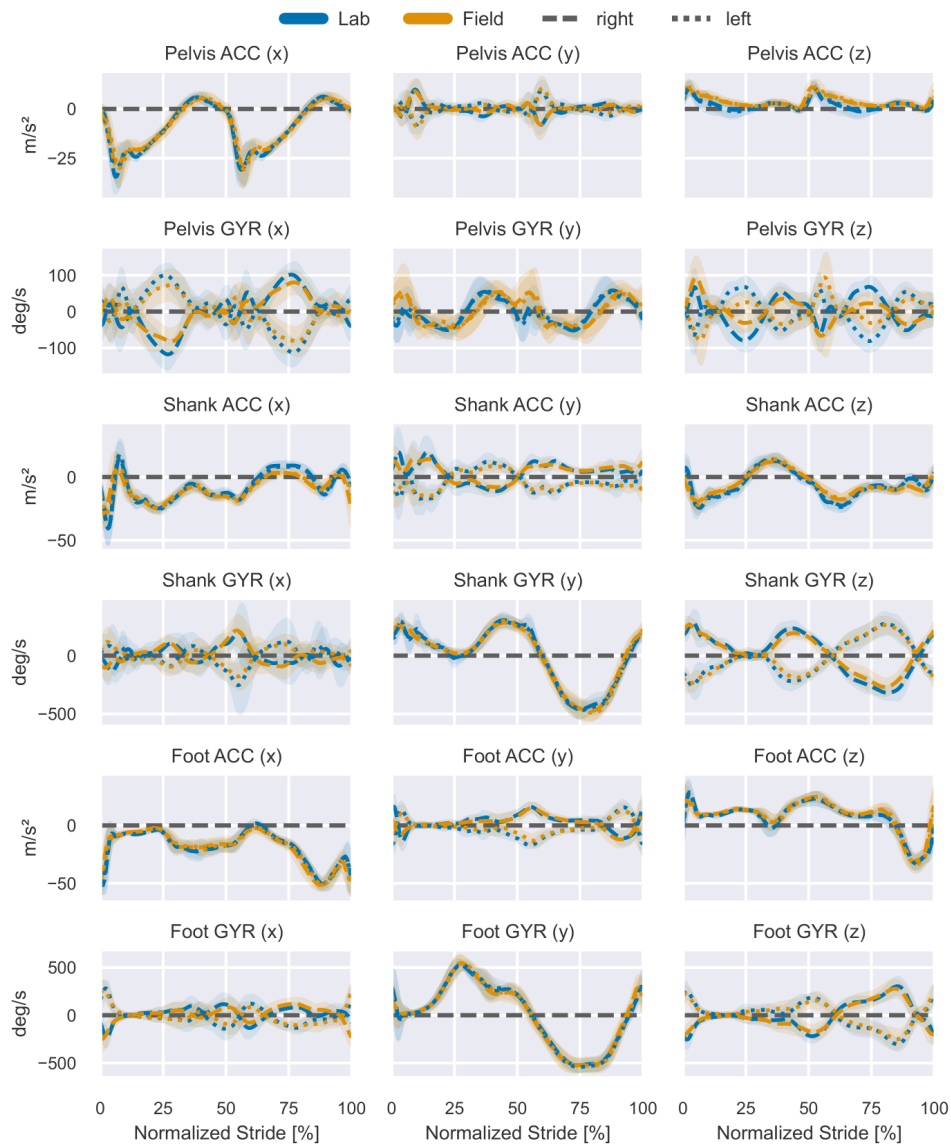


Figure 11 Comparison of IMU Signals During Lab and Field Running

IMU signal comparison between stride-normalized lab (blue) and field data (orange) at 10 km/h level running. Displayed are the mean \pm standard deviations across each dataset. Right side data are indicated as dashed lines, left side data are indicated as dotted lines. ACC: accelerometer signals (acceleration in m/s^2), GYR: gyroscope signals (angular velocity in deg/s).

A.5 Multi-Step Analysis

To account for within-participant variability and avoid relying solely on averaged waveforms, we additionally performed a repeated-measures (multi-step) analysis. For each participant and speed condition, 20 randomly

selected steps were used, and the statistical analysis was repeated following the same procedures described in the main text. The results for GRFs and joint moments are presented in Supplementary Figure 12, 13, 14 and 15.

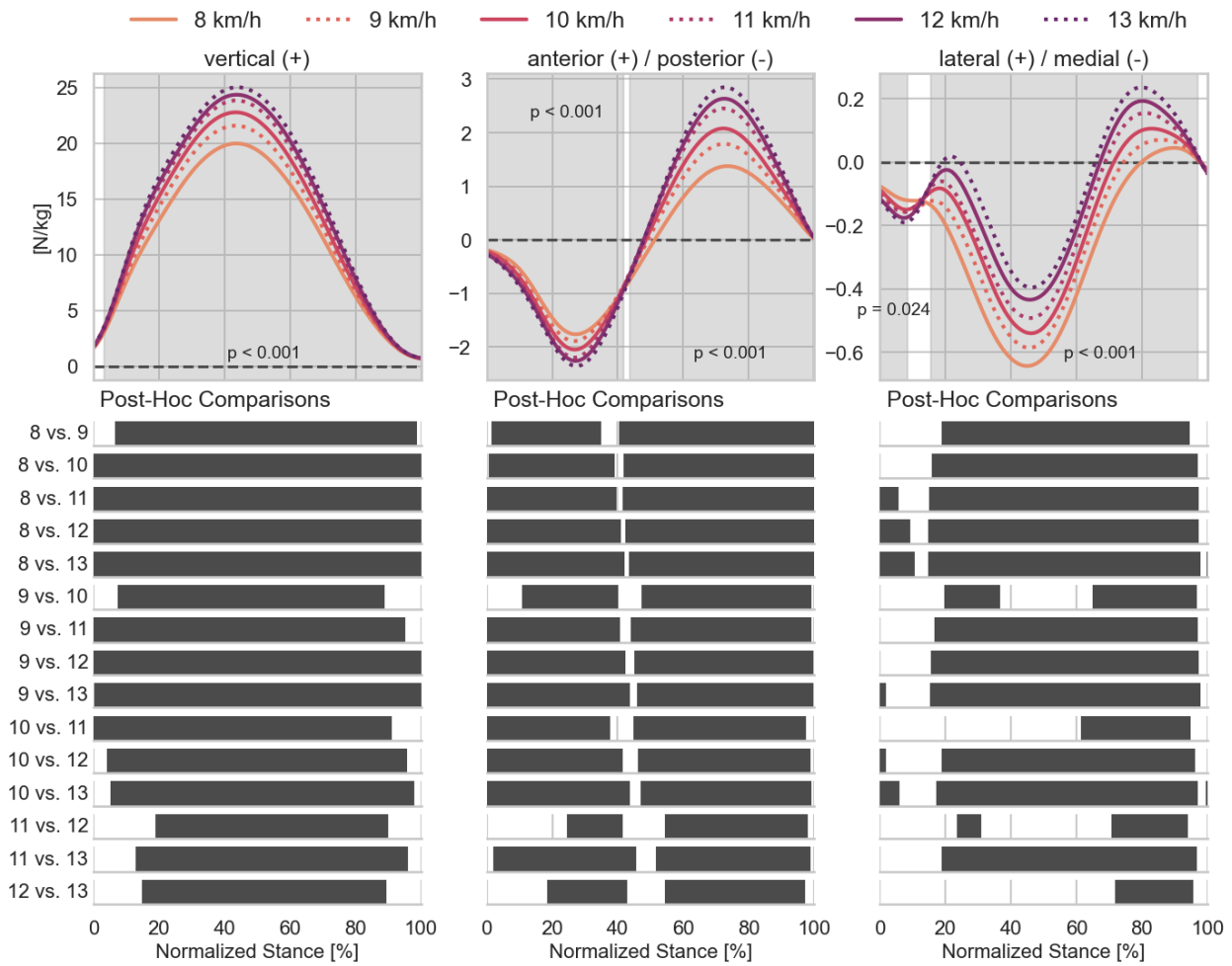


Figure 12 Speed Influence on 3D Ground Reaction Forces (Multi-Step Analysis)

Mean ground reaction force time-series (normalized to body mass) for running speeds between 8 and 13 km/h. Grey shaded bands indicate areas of statistical significance ($\alpha < 0.05$) of SPM repeated measures ANOVA between the six running speeds. The horizontal dark bars below indicate areas of statistical significance ($\alpha < 0.0033$) for pairwise post-hoc comparisons between speed pairs using SPM paired t-test.

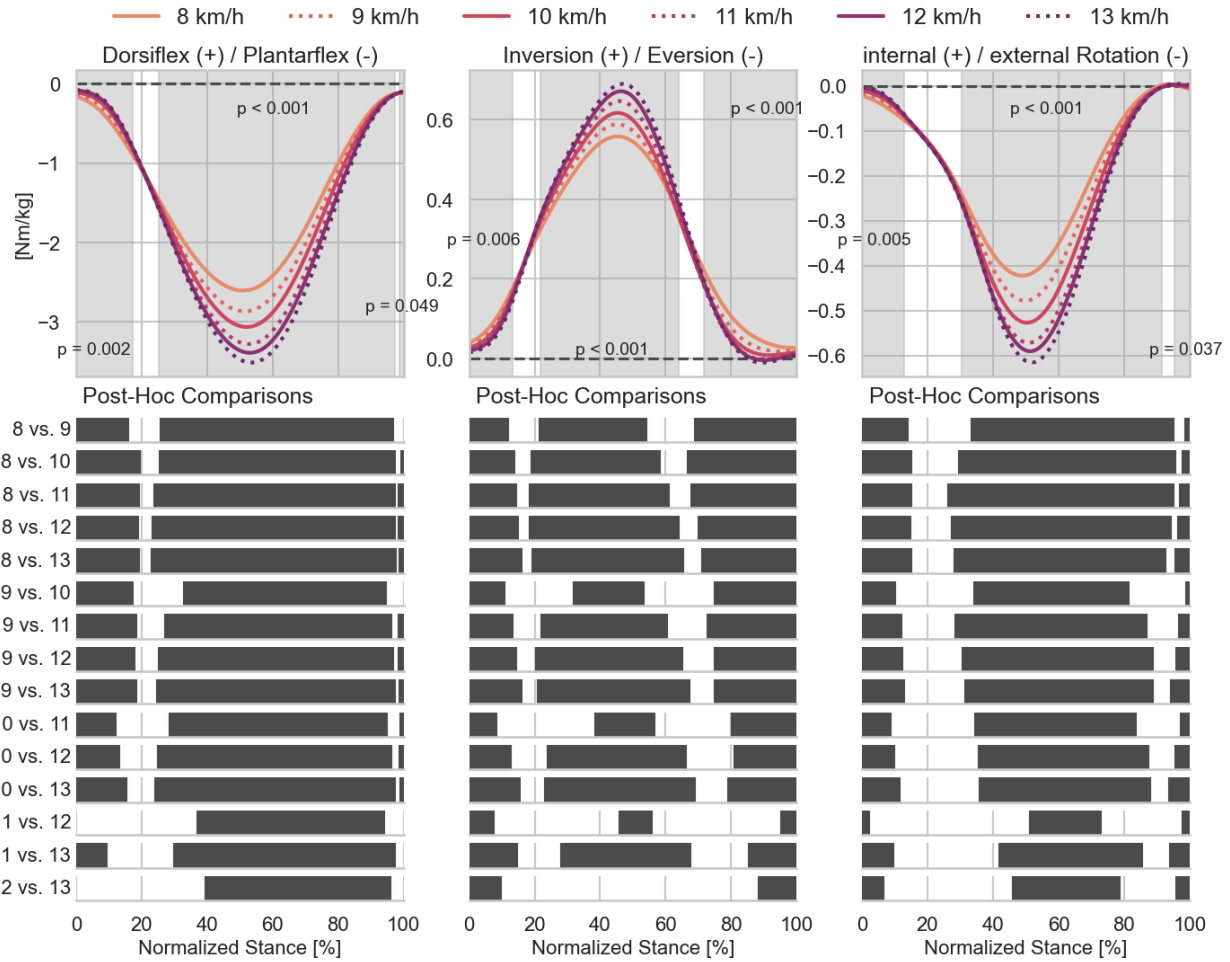


Figure 13 Speed Influence on 3D Ankle Moments (Multi-Step Analysis)

Mean ankle moment time-series (normalized to body mass) for running speeds between 8 and 13 km/h. Grey shaded bands indicate areas of statistical significance ($\alpha < 0.05$) of SPM repeated measures ANOVA between the six running speeds. The horizontal dark bars below indicate areas of statistical significance ($\alpha < 0.0033$) for pairwise post-hoc comparisons between speed pairs using SPM paired t-test.

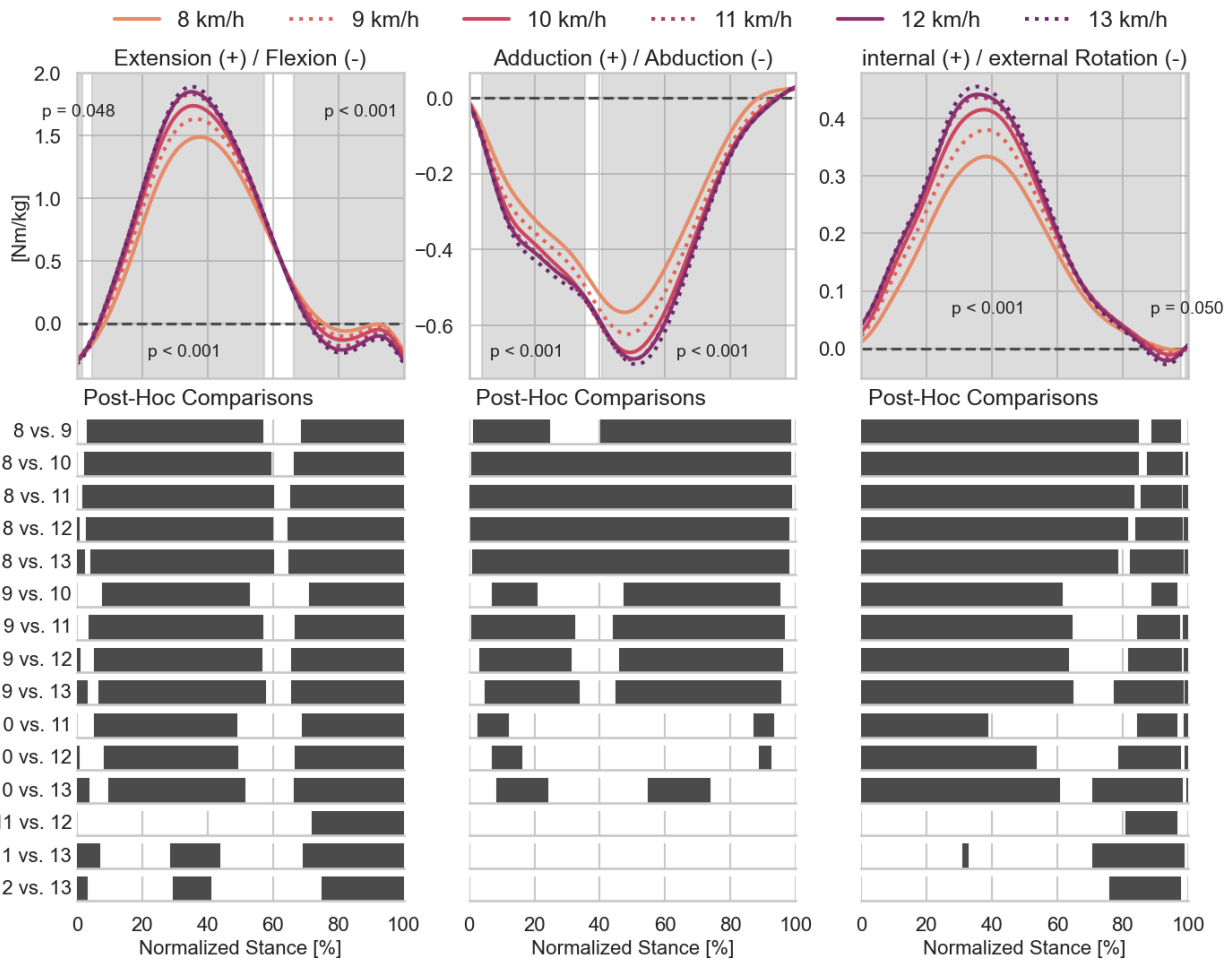


Figure 14 Speed Influence on 3D Knee Moments (Multi-Step Analysis)

Mean knee moment time-series (normalized to body mass) for running speeds between 8 and 13 km/h. Grey shaded bands indicate areas of statistical significance ($\alpha < 0.05$) of SPM repeated measures ANOVA between the six running speeds. The horizontal dark bars below indicate areas of statistical significance ($\alpha < 0.0033$) for pairwise post-hoc comparisons between speed pairs using SPM paired t-test.

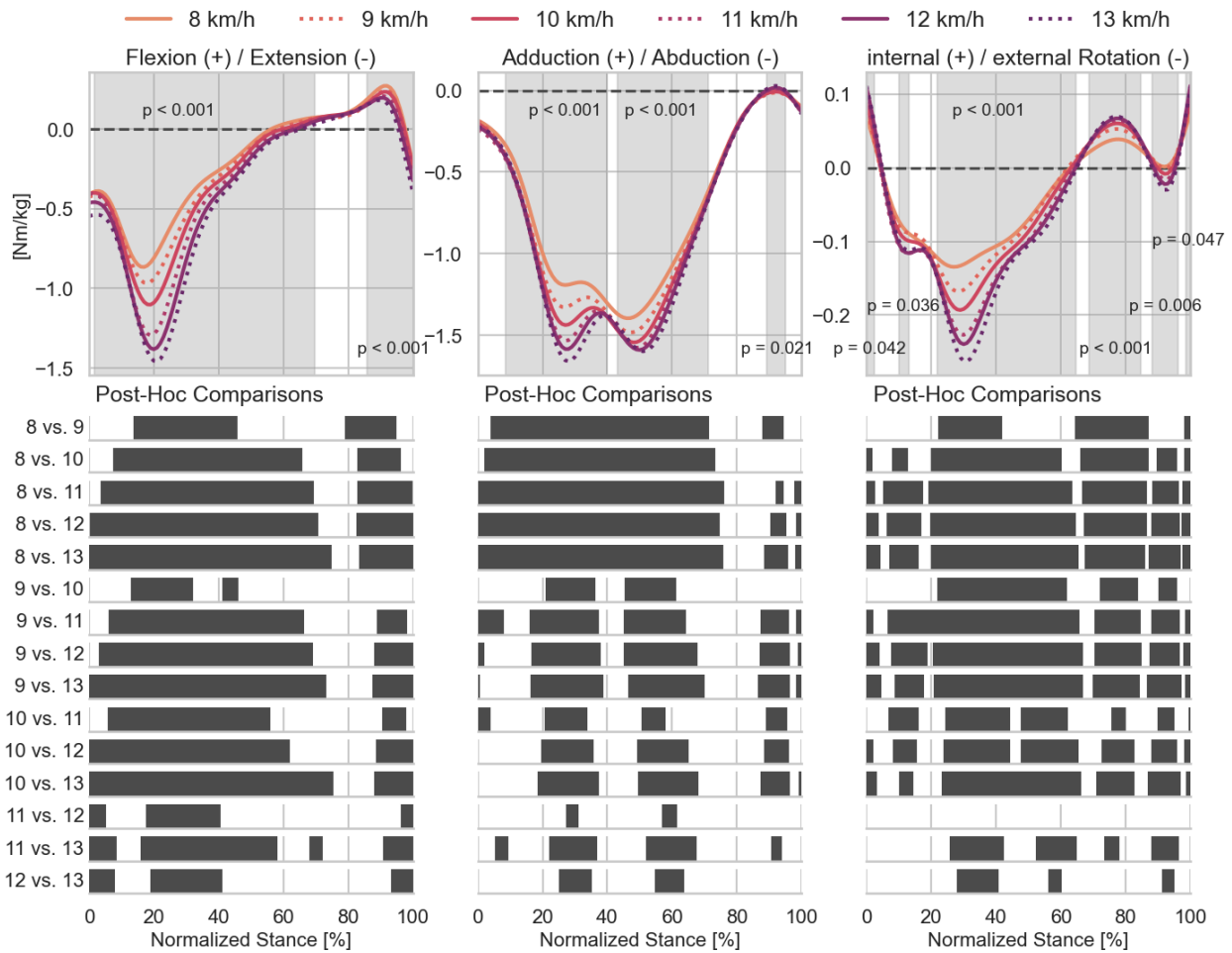


Figure 15 Speed Influence on 3D Hip Moments (Multi-Step Analysis)

Mean hip moment time-series (normalized to body mass) for running speeds between 8 and 13 km/h. Grey shaded bands indicate areas of statistical significance ($\alpha < 0.05$) of SPM repeated measures ANOVA between the six running speeds. The horizontal dark bars below indicate areas of statistical significance ($\alpha < 0.0033$) for pairwise post-hoc comparisons between speed pairs using SPM paired t-test.