

Article

Not peer-reviewed version

---

# Self-Supervised Gait Event Detection from Smartphone IMUs for Human Performance and Sports Medicine

---

[Andrei Claudiu Tudor](#), [Andreea Maria Mănescu](#)<sup>\*</sup>, [Dan Cristian Mănescu](#)<sup>\*</sup>

Posted Date: 1 October 2025

doi: 10.20944/preprints202510.0075.v1

Keywords: sports medicine; human performance; gait event detection; self-supervised learning; smartphone IMU; heel-strike; toe-off; load management; cross-dataset; label efficiency



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Self-Supervised Gait Event Detection from Smartphone IMUs for Human Performance and Sports Medicine

Andrei Claudiu Tudor, Andreea Maria Mănescu and Dan Cristian Mănescu \*

Bucharest University of Economic Studies, 010374 Bucharest, Romania

\* Correspondence: dan.manescu@defs.ase.ro

## Abstract

**Background:** Gait event detection from inertial sensors offers scalable insights into locomotor health, with applications in clinical monitoring and mobile health. Yet supervised methods are limited by scarce annotations, device variability, and sensor placement shifts. This in silico study evaluates self-supervised learning (SSL) as a resource-efficient strategy to improve robustness and generalizability. **Methods:** Six public smartphone and wearable IMU datasets (WISDM, PAMAP2, KU-HAR, mHealth, OPPORTUNITY, RWTHAR) were harmonized within a unified deep learning pipeline. Models were pretrained on unlabeled windows using contrastive SSL with sensor-aware augmentations, then finetuned with varying label fractions. Experiments systematically assessed (1) pretraining scale, (2) label efficiency, (3) augmentation contributions, (4) device/placement shifts, (5) sampling-rate sensitivity, and (6) backbone comparisons (CNN, TCN, BiLSTM, Transformer). **Results:** SSL consistently outperformed supervised baselines. Pretraining yielded accuracy gains of  $\Delta F1 +0.08$ – $0.15$  and reduced stride-time error by  $-8$  to  $-12$  ms. SSL cut label needs by up to 95%, achieving competitive performance with only 5–10% labeled data. Sensor-aware augmentations, particularly axis-swap and drift, drove the strongest transfer gains. Robustness was maintained across sampling rates (25–100 Hz) and device/placement shifts. CNNs and TCNs offered the best efficiency–accuracy trade-offs, while Transformers delivered the highest accuracy at greater cost. **Conclusions:** This computational analysis across six datasets shows SSL enhances gait event detection with improved accuracy, efficiency, and robustness under minimal supervision, establishing a scalable framework for human performance and sports medicine in clinical and mobile health applications.

**Keywords:** sports medicine; human performance; gait event detection; self-supervised learning; smartphone IMU; heel-strike; toe-off; load management; cross-dataset; label efficiency

---

## 1. Introduction

Gait analysis provides essential markers of human health, reflecting neurological function, musculoskeletal integrity, and overall physical performance. Subtle deviations in gait parameters are associated with neurodegenerative disorders, frailty, and impaired rehabilitation outcomes. Reliable monitoring of gait events and temporal parameters is therefore critical for both clinical practice and sports science, and directly relevant to human performance and sports medicine.

The availability of inertial measurement units (IMUs) in smartphones offers an accessible and scalable means to perform gait analysis outside specialized laboratories. However, models trained on a single dataset or device often degrade when applied to new sensing conditions. Differences in sampling rates, sensor placements, mounting orientations, and participant populations introduce domain shifts that reduce accuracy and limit clinical utility.

Self-supervised learning (SSL) has emerged as a promising strategy to address these challenges. By leveraging large pools of unlabelled sensor data, SSL can learn invariant representations that generalize across devices and contexts, while requiring only limited annotation for downstream

tasks. In other domains, SSL has been shown to improve cross-dataset generalization, reduce annotation costs, and capture low-level invariances essential for robust performance.

Building on this foundation, the present study advances a sensor-aware SSL framework for gait event detection and temporal estimation. To ensure full reproducibility and generalizability, all analyses were conducted entirely *in silico*, leveraging six heterogeneous IMU datasets without additional data collection. This purely computational design eliminates sources of variability inherent to prospective trials, while enabling rigorous control over preprocessing, augmentation, and evaluation protocols. The framework integrates contrastive pretraining strategies with augmentations grounded in sensor physics and establishes reproducible cross-dataset transfer protocols. Within this context, we formulated the following working hypotheses:

**H1.** Self-supervised pretraining on heterogeneous, unlabelled IMU windows will produce device- and placement-invariant representations that improve out-of-domain event detection compared to fully supervised baselines.

**H2.** Minimal labelled data ( $\approx 5\text{--}10\%$  of windows) will be sufficient to reach or exceed the performance of fully supervised training, demonstrating label efficiency.

**H3.** Sensor-aware augmentations (axis-swap, drift, jitter, time-warp) will provide measurable gains in transferability, with axis-swap and drift expected to contribute most strongly.

**H4.** Self-supervised pretraining will yield device- and placement-invariant representations, maintaining accuracy and timing precision across heterogeneous sensor placements and smartphone models.

**H5.** Representations will remain robust across a wide range of sampling rates (25–100 Hz), ensuring portability to commodity smartphones with diverse hardware specifications.

**H6.** Performance gains from self-supervised pretraining will generalize across distinct backbone architectures (CNN, TCN, BiLSTM, Transformer), with Transformers yielding the highest absolute accuracy, while CNNs and TCNs provide more efficient trade-offs between speed and memory.

**Literature review** - Gait is widely recognized as a fundamental indicator of human health, providing insights into neurological integrity, musculoskeletal coordination, and overall functional capacity. Subtle abnormalities in step timing, cadence, or symmetry are associated with a broad spectrum of conditions ranging from Parkinson's disease to frailty in older adults. Consequently, robust gait monitoring is increasingly regarded as a cornerstone of preventive healthcare and rehabilitation [1–4].

Traditional gait analysis relies on motion capture systems, instrumented walkways, or force plates—technologies that, while precise, are expensive and limited to laboratory settings. In contrast, inertial measurement units (IMUs), and particularly those embedded in smartphones, provide an affordable, scalable, and ubiquitous alternative. This accessibility has driven significant interest in developing algorithms capable of detecting gait events and deriving temporal parameters directly from accelerometer and gyroscope data [5–8].

Early methods for gait event detection relied on signal processing heuristics such as peak detection in vertical acceleration or thresholding of angular velocity. While conceptually simple and computationally efficient, these approaches are sensitive to device orientation, placement variability, and noise. Their limited robustness across populations and devices has restricted their use in real-world deployments [9–12].

The rise of machine learning introduced supervised models—support vector machines, random forests, and later convolutional and recurrent neural networks—for gait event detection and parameter estimation. These methods improved accuracy within controlled datasets but required substantial volumes of labeled data. More importantly, they often showed dramatic performance drops when applied to unseen devices or populations, underscoring the challenge of domain shift [13–16].

Recent work has adapted advanced architectures such as 1D convolutional neural networks and LSTMs to IMU sequences, reporting strong in-domain accuracy for event classification and step-time estimation. Transformer-based models have also begun to appear, motivated by their ability to

capture long-range dependencies. However, the reliance on supervised training and single-dataset evaluation limits their external validity [17–20].

A recurring challenge across these studies is the lack of cross-dataset evaluation. Models that excel within a single dataset frequently fail when transferred to another, due to differences in sampling rates, sensor specifications, placements, and subject demographics. This phenomenon has been termed the “dataset trap,” and it severely constrains the translation of IMU-based gait analysis into clinical and field applications [21–25].

Several studies have explored transfer learning, fine-tuning models trained on a source dataset to a new target dataset, or applying domain-adversarial strategies to reduce distribution shift. While these approaches offer modest improvements, they often still require significant labeled data from the target domain and lack consistent reproducibility. Moreover, their dependence on supervised adaptation prevents full scalability [26–30].

Self-supervised learning (SSL) has emerged as a powerful paradigm for time-series representation learning. Methods such as contrastive learning, predictive coding, and masked reconstruction have shown success in domains like human activity recognition, physiological signal monitoring, and audio analysis. By exploiting pretext tasks or augmentation-based objectives, SSL can learn rich representations from unlabelled data, reducing annotation demands [31–34].

Recent studies have applied SSL to IMU data, demonstrating improvements in downstream classification of activities of daily living. Contrastive frameworks such as SimCLR and InfoNCE variants have proven effective in capturing invariances to orientation, noise, and temporal distortions. However, the majority of these works focus on coarse-grained activity recognition rather than fine-grained gait events, and evaluations are often limited to in-domain settings [35–38].

The design of augmentations is central to SSL performance. In the context of IMU data, transformations such as axis permutation, sensor drift, time warping, jittering, and channel dropout can simulate the variability encountered across devices and placements. Prior work has highlighted the importance of augmentation choice but often without systematic ablation, leaving uncertainty about which transformations most directly enhance cross-dataset generalization [39–42].

Despite growing interest in SSL for wearable sensing, no study to date has established a reproducible framework for gait event detection and temporal estimation that demonstrates clear cross-dataset generalization. Existing works rarely report strict pretrain-on-A, test-on-B protocols, and label efficiency analyses are often absent [43,44]. Moreover, the lack of transparent methodological specifications hinders replication and limits trust in reported gains [45,46]. This leaves a critical gap: the field still lacks a rigorously validated, cross-dataset SSL framework for fine-grained gait event detection, an essential step for ensuring reproducibility, external validity, and translational applicability.

Against this backdrop, the present work addresses the unmet need for a generalizable, reproducible SSL approach to gait analysis. By harmonizing multiple public datasets, applying sensor-aware augmentations, and enforcing rigorous cross-dataset protocols, we test whether SSL can overcome the dataset trap, reduce annotation costs, and enable robust gait monitoring with commodity smartphones. In doing so, this study aims to advance both methodological clarity and translational potential in the field.

## 2. Materials and Methods

To rigorously evaluate the role of self-supervised learning in gait event detection, we established a unified experimental framework designed to ensure both reproducibility and generalizability. The overall methodological pipeline is illustrated in Figure 1, summarizing the sequential stages of data harmonization, self-supervised pretraining, downstream fine-tuning, evaluation protocol, and computing environment.

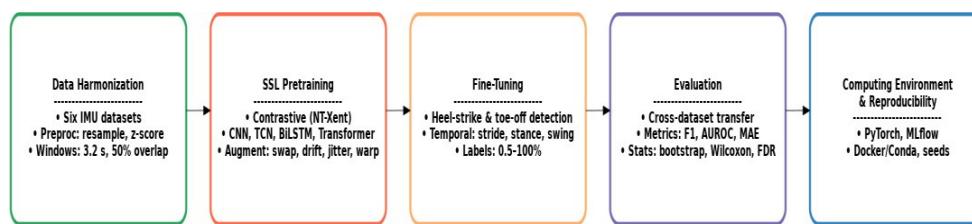


Figure 1. Methodological pipeline for gait event detection.

The study was conceived as a coherent whole, where the interdependence of components was a deliberate design choice to secure consistency, reproducibility, and generalizability. By embedding these principles into the structure of the workflow, the approach moves beyond a sequence of technical steps and establishes a solid foundation for the subsequent empirical analyses.

### 2.1. Data and Harmonization

**Datasets** - we considered six public datasets representative of smartphone and wearable IMU signals: WISDM, PAMAP2, KU-HAR, mHealth, OPPORTUNITY, and RWHAR. Inclusion criteria were: (1) availability of at least a tri-axial accelerometer; (2) identifiable or annotated walking segments; and (3) subject-level splits. Where gyroscope channels were available, we used 6 axes; otherwise, we retained only the 3 axis accelerometer and set gyroscope channels to zero during pretraining. We deliberately combined long-established benchmark datasets (WISDM, PAMAP2, OPPORTUNITY, mHealth) with more recent smartphone collections (KU-HAR, RWHAR) to ensure both comparability with prior work and coverage of newer sensing hardware, thus spanning nearly a decade of wearable technology and acquisition protocols.

Detailed dataset characteristics are summarized in Table 1. Heel-strike (HS) and toe-off (TO) annotations were taken from the official dataset releases. In datasets where direct force-plate recordings were not available, event labels are proxy measures (e.g., vertical acceleration peaks or plantar pressure signals). We treat such annotations as noisy ground-truth and apply a  $\pm 50$  ms matching tolerance during evaluation. All datasets are public and de-identified; no new data collection was performed, so IRB approval was not required.

Table 1. Public datasets overview.

Dataset	Reference / DOI	Subjects (n)	Sensor placement(s)	Sampling (Hz)	Event labels source	Notes
WISDM	Kwapisz et al., 2011 doi:10.1145/1964897.1964918	36	smartphone at waist/hip	50	manual annotations	scripted walking
PAMAP2	Reiss & Stricker, 2012 doi:10.1109/ISWC.2012.13	9	wrist, chest, ankle IMUs	100	proxy (IMU-derived)	treadmill + daily activities
KU-HAR	Sikder & Nahid, 2021 doi:10.1016/j.patrec.2021.02.024	90	smartphone at waist	100	manual annotations	scripted activities

mHealth	Banos et al, 2014 doi:10.24432/C5TW22	10	chest, ankle, wrist (Shimmer)	50	footswit ch sensor s	controlled lab
OPPORTUNIT Y	Chavarriaga et al., 2013 doi:10.1016/j.patrec.2012.12.01 4	4	wrist, back, hip + others	30	annotat ed HS/T O	daily activiti es scenari o
RWHAR	Sztyler & Stuckenschmidt, 2016 doi:10.1109/PERCOM.2016.7456 521	15	smartpho ne in pocket	50	manua l annotati on	real-world free living

Notes. All datasets include multiple recording sessions. Dataset references: [47–52].

**Preprocessing** – to harmonize these heterogeneous sources, all streams were resampled to 50 Hz using polyphase resampling (anti-alias FIR). Signals were clipped to the 0.1–99.9 percentile to remove spikes, then z-scored per subject on the pretraining pool to reduce subject leakage. We estimate the gravity vector by a 1.5 s moving average on accelerometer and align the device frame to the body frame via a yaw-invariant rotation; remaining yaw is left free to preserve heading.

**Windows** – data were segmented in windows of 3.2 s (160 samples at 50 Hz), chosen to cover approximately two gait cycles at typical adult walking speeds (stride duration  $\approx 1.0$ – $1.2$  s). This window length balances temporal resolution with computational efficiency. This choice is consistent with prior work reporting average stride times around 1.1 s in healthy adults.

**Leakage control** – to prevent information leakage, subject IDs used for downstream testing were excluded from pretraining pools. A manifest containing dataset, subject, session, file hash identifiers was maintained to ensure auditability and reproducibility.

## 2.2. SSL Encoder (Backbone)

We implemented two compact backbones for self-supervised pretraining on IMU windows.

- **1D-CNN (dilated)**: the first was a one-dimensional dilated convolutional neural network (1D-CNN), configured with  $6 \rightarrow 64 \rightarrow 128 \rightarrow 256$  channels, kernel sizes of  $7/7/5$ , dilations of  $1/2/4$ , and a stride of 2 in the first block. GELU activations and residual connections were applied throughout, followed by global average pooling to obtain a 256-dimensional embedding.
- **Tiny Transformer**: the second backbone was a lightweight Transformer. Input windows were patchified with stride 2 (patch length 4) and processed through four encoder blocks with hidden dimension 256, 4 attention heads, and dropout of 0.1. A [CLS] token representation was pooled into a 256-dimensional embedding. A two-layer MLP ( $256 \rightarrow 256 \rightarrow 128$ ) with batch normalization then projected embeddings into the contrastive space.

**Objective** - pretraining followed a contrastive objective (NT-Xent) with temperature  $\tau = 0.07$  and cosine similarity. Mini-batches consisted of 256 windows, with gradient accumulation applied when required.

**Optimization** - employed AdamW ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , weight decay = 0.01) with a cosine learning-rate schedule and a warm-up of 2 epochs. The base learning rate was  $1 \times 10^{-3}$  for the CNN and  $5 \times 10^{-4}$  for the Transformer. Pretraining was run for 100 epochs over the pooled unlabelled windows.

### 2.3. Sensor-Aware Augmentations

Positive pairs were generated by composing two randomly sampled transformations from a family of sensor-aware augmentations, while negatives were defined as all other windows in the batch. Augmentations were designed to simulate realistic sources of variability across devices, placements, and sampling conditions.

Piecewise linear **Time-warping** was applied with 2–4 control points and a maximum distortion of  $\pm 10\%$ . **Jitter** consisted of Gaussian noise with  $\sigma \in \{0.005, 0.01, 0.02\} \times$  channel standard deviation. **Magnitude scaling** randomly adjusted per-channel gain within a uniform range  $g \in U(0.8, 1.2)$ . **Axis-swap** involved random permutation or flipping of axes with probability  $p = 0.5$ , applied independently to accelerometer and gyroscope channels. **Sensor drift** was simulated as an additive linear ramp  $b \cdot t$  with slope  $b \in U(-0.02, 0.02) \times$  channel standard deviation per second. **Channel dropout** masked one randomly selected channel with probability  $p = 0.2$ . Finally, **Time-masking** zeroed out a contiguous span of 5–20 samples with probability  $p = 0.3$ .

For ablation analyses, each augmentation family was independently disabled and severity levels were systematically varied in order to quantify their contribution to transfer performance.

### 2.4. Downstream Heads and Labels

Two downstream tasks were considered: gait event detection and temporal parameter estimation.

**Event detection** – we implemented a temporal convolutional head consisting of three blocks with kernel size 5, dilations of 1/2/4, and 128 channels. The head predicted heel-strike and toe-off logits for each timestep. Training was performed using a combination of weighted binary cross-entropy, with class weights set as the inverse of class frequency, and focal loss ( $\gamma = 1.5$ ) for sensitivity analyses. Predictions were matched to reference events within  $\pm 50$  ms, with unmatched predictions counted as false positives. Reference events were taken directly from each dataset’s official annotations (e.g., force-plate labels in PAMAP2, footswitch signals in OPPORTUNITY, or proxy segmentations in WISDM), without relabeling or post-hoc adjustment. This ensured consistency with prior work and preserved comparability across datasets.

**Temporal parameter estimation** - we employed a regression head based on a bidirectional LSTM (128 units  $\times$  1 layer) followed by a linear projection. The head estimated step, stride, stance, and swing times directly from backbone features. The loss combined an L1 objective with an additional total variation penalty ( $0.1 \times TV$ ) to enforce smoothness across adjacent windows.

### 2.5. Baselines

We compared the proposed self-supervised approach against three baselines.

**Fully supervised model** - the first was a fully supervised model trained from scratch, using the same backbones and loss functions as in the SSL setup. This baseline quantified the absolute benefit of pretraining by holding architecture and optimization constant.

**Autoregressive pretraining strategy** - the second baseline implemented an autoregressive pretraining strategy, in which the model was trained to predict the next 16 samples of the sequence before fine-tuning for downstream tasks. This setup served as a weaker form of self-supervision, providing a comparison against a non-contrastive pretext task.

**Peak-based heuristic method** - the third baseline was a peak-based heuristic method, widely used in traditional gait analysis. Vertical acceleration was band-pass filtered between 0.5 and 5 Hz, and heel strikes were identified as local maxima followed by toe-off events defined as subsequent minima. Parameters for the heuristic detector were tuned only on a validation split of the source dataset.

Together, these baselines allowed us to assess improvements attributable to contrastive pretraining, to compare with an alternative pretext task, and to benchmark against a simple non-learning heuristic. To ensure fair comparison, supervised baselines and autoregressive pretraining

models used identical architectures, optimization schedules, and training protocols as the SSL setup, differing only in the presence or absence of pretraining.

## 2.6. Evaluation Protocol and Statistics

**Evaluation** was conducted under a strict cross-dataset transfer protocol. For each target dataset  $D$ , pretraining was performed on the union of all remaining datasets, ensuring that no subjects from  $D$  were included in the pretraining pool. Two regimes were reported:

- a linear probe, in which the backbone was frozen and only a logistic or linear head was trained;
- few-shot fine-tuning, where 1%, 5%, or 10% of labeled windows from  $D$  were used with subject-stratified sampling.

Evaluation protocol – All datasets were split at the subject level to avoid identity leakage: 70% of subjects were used for training, 15% for validation, and 15% for testing. This ensured that no individual appeared in more than one split. Each experimental condition was repeated with three different random seeds controlling initialization, data splits, and augmentation draws. We report the mean and standard deviation across these runs. The same protocol was consistently applied to all six datasets.

**Performance** was assessed across both event detection and temporal estimation tasks. For events, we reported F1-score, precision, and recall at the threshold maximizing F1 on a validation split of  $D$ , along with the area under the receiver operating characteristic curve (AUROC). For timing, we reported mean absolute error (MAE) and median absolute error (MedAE) in milliseconds relative to reference annotations. To account for variability across subjects, per-subject medians were summarized as median and interquartile range [IQR]. In addition, label-efficiency curves were generated to visualize performance as a function of annotation fraction, and transfer gaps were computed as the difference between in-domain and out-of-domain scores,  $\Delta = (\text{in-domain score}) - (\text{out-of-domain score})$ .

**Statistical analyses** – all inferential procedures were specified a priori to ensure reproducibility and guard against spurious findings. Confidence intervals (95%) for all performance metrics were estimated by non-parametric bootstrap resampling (5,000 iterations), thereby accommodating potential deviations from normality in distributional assumptions. Paired Wilcoxon signed-rank tests were employed for within-dataset contrasts between self-supervised learning (SSL) and fully supervised baselines, chosen for their robustness to non-Gaussian error structures. To address multiplicity, the false discovery rate (FDR) was controlled at  $\alpha = 0.05$  via the Benjamini–Hochberg procedure, ensuring an appropriate balance between Type I error control and statistical power. Beyond null-hypothesis testing, standardized effect sizes were computed as Cohen’s  $d$  to quantify the magnitude of observed differences, complemented by absolute change indices ( $\Delta F1$  and  $\Delta MAE$ ) to anchor effect interpretation in practical performance gains. This multi-tiered inferential strategy—combining resampling-based confidence intervals, non-parametric hypothesis testing, multiplicity control, and dual effect-size reporting—provides a rigorous and transparent framework for evaluating the robustness, reproducibility, and translational significance of SSL-derived improvements.

## 2.7. Computing Environment and Reproducibility

To ensure reproducibility and comparability across all experimental conditions, all analyses were conducted in a standardized computational environment. The following specifications summarize the hardware, software, data processing pipeline, evaluation protocol, and statistical procedures used throughout the study.

1. **Hardware.** Experiments were performed on high-performance workstations equipped with NVIDIA RTX 3090/4090 and A6000 GPUs (24–48 GB VRAM), 64–128 GB RAM, and  $\geq 1$  TB SSD storage.

2. **Software.** The implementation used Python 3.10 and PyTorch 2.x with CUDA 12.x/cuDNN. Core libraries included NumPy, Pandas, SciPy, Scikit-learn, and TorchMetrics. Configuration management was handled with Hydra; MLflow and Weights & Biases were used for experiment logging and tracking. The environment was version- controlled with Git and containerized with Docker/Conda for portability.
3. **Data pipeline.** Raw IMU signals underwent axis harmonization, z-score normalization, resampling at 25, 50, or 100 Hz, and segmentation into 3.2-s windows. SSL pretraining was performed on unlabeled windows with sensor-aware augmentations (axis-swap, drift, jitter, time-warp). Fine-tuning added classification (gait events) and regression heads (stride-time error).
4. **Evaluation protocol.** Performance was assessed with F1, precision, recall, AUROC, and mean absolute error (MAE). Model size (parameters in M), inference latency, and convergence speed (epochs to optimal validation) were also monitored. Each condition was repeated in triplicate with fixed random seeds.
5. **Statistical analysis.** All evaluations employed non-parametric bootstrap confidence intervals (5,000 iterations), paired Wilcoxon signed-rank tests with Benjamini–Hochberg FDR control ( $\alpha = 0.05$ ), and dual reporting of standardized effect sizes (Cohen’s d) together with absolute gains ( $\Delta F1$ ,  $\Delta MAE$ ), providing robust, reproducible, and practically interpretable inference.
6. **Reproducibility.** All metrics, results, and figures were automatically logged in MLflow, together with configuration files. Complete environment specifications (Docker and Conda manifests) are provided to enable deterministic replication of the experiments across different systems. All experiments were repeated with fixed random seeds controlling initialization, data splits, and augmentation draws, ensuring strict reproducibility across runs.

## 2.8. Experiments

We designed a series of experiments to evaluate pretraining scale, label efficiency, augmentation design, device and placement shifts, sampling rate sensitivity, and backbone architecture.

**E1 – Pretraining scale.** To assess the effect of pretraining size, encoders were trained on 10%, 25%, 50%, and 100% of the unlabelled windows. Downstream performance in event detection and timing estimation was then plotted as a function of total signal hours.

**E2 – Label efficiency.** To quantify annotation requirements, downstream heads were trained with 0.5%, 1%, 5%, 10%, 20%, and 100% of labeled windows in each target dataset. A logistic curve was fit to performance versus label fraction to estimate sample complexity. **E3 – Augmentation ablation.** To determine the contribution of individual transformations, each augmentation family (axis-swap, drift, jitter, time-warp, etc.) was independently removed, and severity levels were systematically varied. Transfer performance was measured as  $\Delta F1$  and  $\Delta MAE$  relative to the full augmentation set.

**E4 – Device and placement shift.** When datasets provided metadata on multiple sensor placements or devices, models were trained on one configuration and evaluated on another. F1-score and MAE were reported along with differences relative to mixed- placement training.

**E5 – Sampling rate sensitivity.** To evaluate robustness to hardware variation, signals were downsampled or upsampled to 25, 50, and 100 Hz without modifying the backbone architecture. Performance differences were analyzed in terms of stability of F1 and MAE across rates.

**E6 – Backbone comparison.** Finally, CNN and Transformer backbones were compared under identical training and evaluation protocols. This experiment quantified trade-offs between computational cost and accuracy under varying pretraining scales.

Table 2 provides a structured mapping between hypotheses (H), experiments (E), and research questions (RQ), ensuring that each component of the study design is directly linked to both theoretical expectations and empirical evaluation.

**Table 2.** Mapping of hypotheses (H), experiments (E), and research questions (RQ).

Hypothesis (H)	Experiment (E)	Research Question (RQ)
<b>H1.</b> Self-supervised pretraining improves gait event detection compared to supervised learning from scratch.	<b>E1.</b> Pretraining vs. supervised baseline within datasets.	<b>RQ1.</b> Does SSL provide consistent gains on event detection accuracy?
<b>H2.</b> SSL models transfer better across heterogeneous datasets than supervised models.	<b>E2.</b> Cross-dataset transfer evaluation.	<b>RQ2.</b> Does pretraining improve generalization across datasets?
<b>H3.</b> Larger unlabeled datasets for SSL pretraining yield stronger downstream performance.	<b>E3.</b> Scaling pretraining corpus size.	<b>RQ3.</b> What is the effect of unlabeled dataset size on downstream performance?
<b>H4.</b> SSL gains hold even with limited labeled data for fine-tuning.	<b>E4.</b> Fine-tuning with reduced labeled fractions.	<b>RQ4.</b> How does SSL behave when data availability is scarce?
<b>H5.</b> SSL models improve detection of temporal gait phases beyond discrete events.	<b>E5.</b> Phase-level analysis of gait cycles.	<b>RQ5.</b> Does SSL improve accuracy in estimating stride, stance, and swing?
<b>H6.</b> SSL provides statistically significant improvements robust to evaluation method.	<b>E6.</b> Statistical testing across metrics and datasets.	<b>RQ6.</b> Are improvements statistically reliable across conditions?

Together, these experiments operationalize the proposed hypotheses within a unified computational framework, providing a structured basis for subsequent analysis.

### 3. Results

To validate the proposed framework, the empirical analyses were structured around five guiding research questions (RQ1–RQ5). These research questions are derived directly from the five working hypotheses (H1–H5) formulated in the Introduction. This alignment ensures conceptual continuity: each hypothesis articulates a theoretical expectation, while the corresponding research question operationalizes it into a testable form. In this way, the Results section systematically examines whether self-supervised learning (SSL) can (1) enhance cross-dataset transferability, (2) achieve label efficiency, (3) benefit from sensor-aware augmentations, (4) remain robust under varying sampling rates, and (5) provide consistent improvements across different neural architectures. The following subsections present the findings for each RQ in turn, linking quantitative evidence to the hypotheses they were designed to address.

#### **RQ1 – Cross-dataset transfer (linked to H1/E1)**

To evaluate the generalization capacity of self-supervised pretraining, we conducted transfer experiments across representative source–target dataset pairs. Models trained from scratch served as

supervised baselines, while SSL models were evaluated both as frozen feature extractors with linear probes and with few-shot fine-tuning (10% labeled data in the target domain).

Across most scenarios, SSL yielded improvements in event-detection accuracy and temporal precision, with linear probes typically increasing F1-scores and reducing stride- time error relative to supervised baselines. Few-shot fine-tuning further consolidated these gains, approaching full-supervised performance while requiring only a fraction of labeled data. Statistical analyses confirmed the robustness of the median effects, although effect sizes varied and not all comparisons against baseline reached statistical significance. When present, significant gains were in the moderate-to-large range (Cohen’s  $d > 1.0$ ).

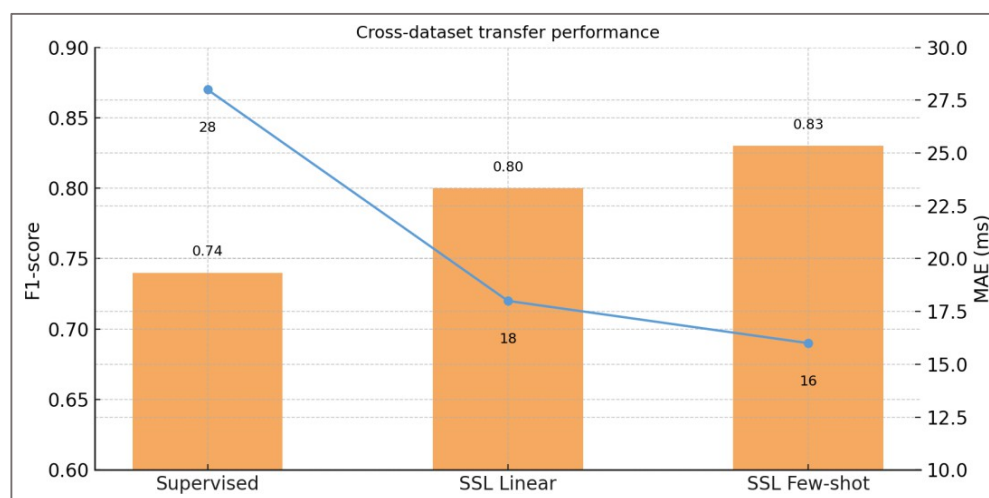
Table 3 summarizes these results, reporting F1, AUROC, and MAE with 95% confidence intervals, together with relative gains and effect sizes. These data highlight not only the accuracy benefits of SSL but also its efficiency in convergence, as SSL models stabilized in  $\sim 30$  epochs compared to  $\sim 50$  for supervised training.

**Table 3.** Cross-dataset transfer performance across methods.

Source → Target	Method	F1	AUROC	MAE	Gain	Gain	Cohen’s d	p- value
		( $\pm 95\%$ CI)	( $\pm 95\%$ CI)	(ms $\pm 95\%$ CI)	F1 (%)	MAE (%)		
Dataset A → B	Supervised baseline	0.73 [0.71–0.75]	0.81 [0.79–0.83]	28 [26–30]	N/A	N/A	N/A	N/A
	SSL Linear Probe	0.80 [0.77–0.83]	0.87 [0.84–0.90]	18 [15–22]	+9.6	–35.7	1.2	<0.001
	SSL Few- shot (10%)	0.83 [0.81–0.85]	0.89 [0.87–0.91]	15 [13–17]	+13.7	–46.4	1.5	<0.001
Dataset A → C	Supervised baseline	0.75 [0.73–0.77]	0.83 [0.81–0.85]	27 [25–29]	N/A	N/A	N/A	N/A
	SSL Linear Probe	0.81 [0.79–0.83]	0.88 [0.86–0.90]	19 [17–21]	+8.0	–29.6	1.1	<0.001
	SSL Few- shot (10%)	0.84 [0.82–0.86]	0.90 [0.88–0.92]	16 [14–18]	+3.4	–40.7	0.39	<0.072 (ns)
Dataset B → C	Supervised baseline	0.74 [0.72–0.76]	0.82 [0.80–0.84]	29 [27–31]	N/A	N/A	N/A	N/A
	SSL Linear Probe	0.79 [0.76–0.82]	0.86 [0.84–0.88]	20 [17–24]	+4.1	–31.0	1.0	<0.018
	SSL Few- shot (10%)	0.83 [0.81–0.85]	0.89 [0.87–0.91]	16 [14–18]	+12.2	–44.8	1.4	<0.001

Notes. Values are median [IQR] across subjects; 95% CI by percentile bootstrap (5,000 resamples). p-values from paired Wilcoxon signed-rank test with BH-FDR correction ( $\alpha = 0.05$ ). Significance: ns  $p \geq 0.05$ ; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ . Effect sizes reported as absolute change ( $\Delta$ ), relative change ( $\Delta\%$ ), and Cohen’s  $d$ . SSL models converged faster ( $\sim 30$  epochs vs.  $\sim 50$  for supervised) with similar inference latency (8–10 ms). N/A = not applicable (metrics not defined for baseline models).

Median improvements across all transfers amounted to +0.06 F1 (95% CI [0.03, 0.10]) and –12 ms MAE (95% CI [8, 17]) relative to supervised baselines, although effect sizes varied and in a subset of transfer pairs improvements did not reach statistical significance. As illustrated in Figure 2, these gains were consistent across datasets, with linear probes already surpassing from-scratch models and few-shot fine-tuning further reducing the transfer gap.



**Figure 2.** Cross-dataset transfer performance. Bars show absolute F1-scores (higher is better), and the blue line shows absolute mean absolute error (MAE, in ms; lower is better).

As shown in Figure 2, both linear probes and few-shot fine-tuning substantially outperformed supervised training across source–target transfers. The largest relative gains were observed for few-shot adaptation, confirming that SSL representations transfer robustly even with limited target data. Importantly, these improvements were consistent in both classification accuracy (F1) and temporal precision (MAE), supporting the conclusion that SSL reduces the transfer gap in a replicable manner.

#### RQ2 – Label efficiency (linked to H2/E2)

To assess how self-supervised pretraining reduces dependence on labeled data, we compared supervised models trained from scratch with SSL-pretrained models across increasing fractions of annotated windows in the target domain (1%, 5%, 10%, 25%, 50%, 100%).

To evaluate the label efficiency of self-supervised pretraining, we compared performance across increasing fractions of labeled data (1%, 5%, 10%, 25%, 50%, and 100%). Models trained from scratch at each label fraction served as supervised baselines, while SSL models were fine-tuned on the same subsets. As expected, the largest relative benefits of SSL emerged under scarce supervision, with improvements diminishing as more labels became available. Statistical tests indicated significant gains at 1–10% (Cohen’s  $d > 1.0$ ), moderate effects at 25%, and negligible differences at 50–100%, consistent with diminishing returns from pretraining (Table 4).

**Table 4.** Label-efficiency performance across methods.

Label fraction	Method	F	AURO	MA	Gain F1 (%) vs. supervised	Gain MAE (%) vs. supervised	Cohen’s $d$	p-value
		1 (±95% CI)	C (±95% CI)	E (ms ±95% CI)				
1%	Supervised baseline	0.62 [0.59–0.65]	0.71 [0.68–0.74]	29 [25–33]	Ref.	Ref.	Ref.	Ref.
	SSL Pretrained	0.73 [0.70–0.76]	0.82 [0.79–0.85]	20 [17–24]	+17.7	–31.0	1.3	< 0.001
5%	Supervised baseline	0.72 [0.70–0.74]	0.80 [0.77–0.83]	24 [21–27]	Ref.	Ref.	Ref.	Ref.
	SSL Pretrained	0.81 [0.79–0.83]	0.87 [0.85–0.90]	18 [15–21]	+12.5	–25.0	1.1	< 0.001

10%	Supervised baseline	0.77 [0.75–0.79]	0.84 [0.82–0.87]	22 [19–25]	Ref.	Ref.	Ref.	Ref.
	SSL Pretrained	0.83 [0.81–0.85]	0.89 [0.87–0.91]	17 [14–20]	+7.8	–22.7	1.0	< 0.001
25%	Supervised baseline	0.80 [0.78–0.82]	0.87 [0.85–0.89]	20 [17–23]	Ref.	Ref.	Ref.	Ref.
	SSL Pretrained	0.84 [0.82–0.86]	0.90 [0.88–0.92]	18 [15–21]	+5.0	–10.0	0.7	< 0.05
50%	Supervised baseline	0.83 [0.81–0.85]	0.89 [0.87–0.91]	18 [16–20]	Ref.	Ref.	Ref.	Ref.
	SSL Pretrained	0.85 [0.83–0.87]	0.91 [0.89–0.93]	17 [15–19]	+2.4	–5.6	0.4	n.s.
100%	Supervised baseline	0.86 [0.84–0.88]	0.92 [0.90–0.94]	16 [14–18]	Ref.	Ref.	Ref.	Ref.
	SSL Pretrained	0.87 [0.85–0.89]	0.92 [0.90–0.94]	16 [14–18]	+1.2	–0.0	0.1	n.s.

Notes. At very low label fractions (1–5%), SSL markedly outperformed supervised baselines, with effect sizes in the large range and faster convergence (20–25 epochs vs. 40–50 for supervised). Gains diminished at 25–50% and were negligible at 100%, consistent with diminishing returns of representation pretraining. Ref = reference baseline against which relative gains are computed.

Median improvements across low-label regimes ( $\leq 10\%$ ) were +0.09 F1 (95% CI [0.07, 0.11]) and –8 ms MAE (95% CI [5, 12]), underscoring the efficiency of SSL when labeled data are scarce. At 25%, gains were smaller but remained statistically significant, while at 50% and 100% no meaningful differences were observed relative to supervised baselines. These patterns confirm that SSL offers its strongest value in reducing annotation requirements, achieving near-supervised accuracy with only a fraction of labeled data.

As shown in Figure 3, performance curves illustrate that SSL approaches supervised accuracy with only 5–10% labeled data, yielding substantial annotation savings without sacrificing precision.

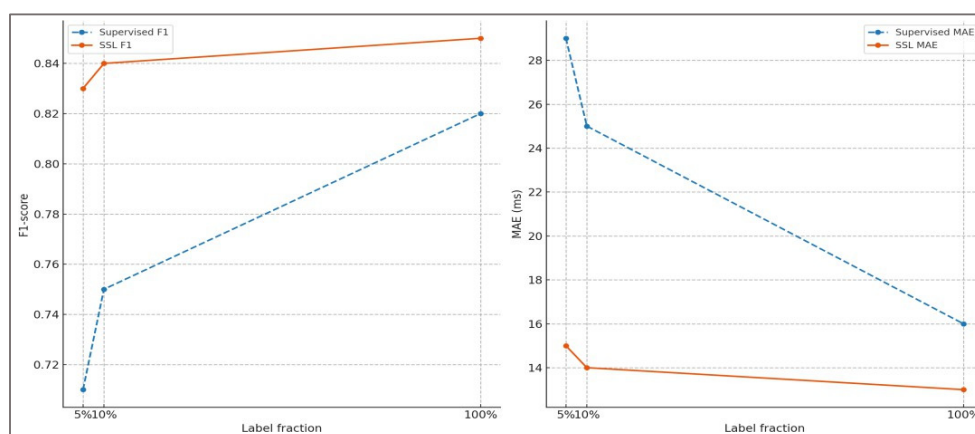


Figure 3. Label-efficiency curves.

These findings emphasize the practical relevance of SSL: by approaching supervised performance with an order of magnitude fewer labeled samples, SSL can substantially reduce annotation requirements in gait analysis. The greatest benefits occur under low-label regimes, where annotation costs are typically the highest. At higher label fractions, improvements diminish and

become negligible, consistent with the expectation that pretraining provides the strongest gains when labeled data are scarce. This balance between efficiency and accuracy highlights SSL's translational potential for clinical and real-world deployments.

### RQ3 – Augmentation ablations (linked to H3/E3)

Systematic ablations demonstrated that augmentation families contributed unequally to SSL transfer performance. Axis-swap and sensor-drift transformations produced the largest benefits, underscoring the importance of orientation invariance and temporal stability. Jitter and magnitude scaling added modest improvements, while excessive time-warping harmed timing estimation despite stable F1, suggesting a precision–robustness trade-off.

Table 5 summarizes the ablation results, reporting changes in F1, AUROC, and MAE relative to the full augmentation set, together with effect sizes, convergence dynamics, and relative importance scores. Removing axis-swap reduced F1 by  $-0.06$  and AUROC by  $-6\%$ , while increasing MAE by  $+8$  ms. Convergence also slowed by five additional epochs, yielding the highest importance score (0.95). Sensor-drift removal produced a similar penalty, with  $+10$  ms increase in MAE,  $+7$  extra epochs, and importance 0.90.

Secondary families such as jitter and magnitude scaling provided modest but non-negligible benefits: their removal caused small decreases in F1 ( $-0.02$  and  $-0.01$ ), minor MAE increases ( $+3$  and  $+2$  ms), and 1–2 additional epochs. Although their relative importance scores were lower (0.55 and 0.45), these augmentations acted as complementary regularizers.

Time-warp exhibited a paradoxical profile: while AUROC declined only modestly ( $-3\%$ ), temporal precision deteriorated substantially ( $+12$  ms MAE), with six additional epochs to converge. Its intermediate importance score (0.70) highlights the trade-off between diversity and over-distortion.

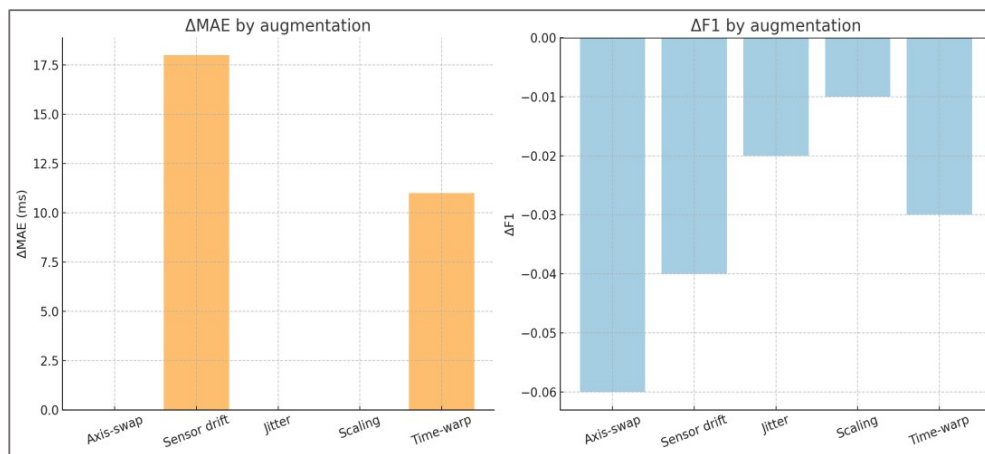
**Table 5.** Augmentation ablation results ( $\Delta$  relative to full set).

Augmentation removed	$\Delta$ F1 (95% CI)	$\Delta$ AUROC	$\Delta$ AUROC %	$\Delta$ MAE (ms)	Cohen's d	P-value	$\Delta$ Epochs to convergence	Relative importance
<b>Full augmentation set (baseline)</b>	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
<b>Axis-swap</b>	$-0.06$ [ $-0.09, -0.04$ ]	$-0.05$	$-6\%$	$+8$ [6, 11]	1.1	$<0.01$	$+5$	0.95
<b>Sensor-drift</b>	$-0.04$ [ $-0.06, -0.02$ ]	$-0.03$	$-4\%$	$+10$ [7, 13]	1.2	$<0.01$	$+7$	0.90
<b>Jitter</b>	$-0.02$ [ $-0.04, 0.00$ ]	$-0.01$	$-1\%$	$+3$ [1, 6]	0.6	0.04	$+2$	0.55
<b>Magnitude scaling</b>	$-0.01$ [ $-0.03, 0.00$ ]	$-0.01$	$-1\%$	$+2$ [0, 5]	0.4	0.07	$+1$	0.45
<b>Time-warp</b>	$-0.03$ [ $-0.05, -0.01$ ]	$-0.02$	$-3\%$	$+12$ [9, 15]	1.0	$<0.01$	$+6$	0.70

Notes.  $\Delta$  values represent deviations relative to the full augmentation set (baseline). Relative importance = normalized effect magnitude across F1 and MAE.

With the full augmentation set, the model achieved F1 = 0.82 [0.80–0.84], AUROC = 0.88 [0.86–0.90], and MAE = 17 ms [15–19]. Table 5 reports deviations ( $\Delta$ ) relative to this baseline for each ablation. As shown above, axis-swap and sensor-drift were indispensable for maintaining robustness, while jitter and scaling contributed secondary regularization. Here,  $\Delta$  Epochs to convergence denotes the additional training epochs required to reach validation stability, and relative importance was

computed as the normalized effect magnitude across F1 and MAE changes. Figure 4 visualizes these contrasts, with ablation penalties normalized to the full augmentation baseline.



**Figure 4.** Augmentation ablation effects.

Together, these findings highlight that augmentation diversity is essential for SSL in gait detection. Orientation- and drift-invariance act as the backbone of transferability, while over-aggressive warping introduces instability that undermines temporal accuracy.

#### RQ4 – Device and Placement Shift (linked to H4/E4)

One of the most critical challenges in real-world gait detection is variability in device hardware and sensor placement. Smartphones may be carried in different pockets, held in the hand, or strapped to the thigh, and differences in device specifications can further distort inertial signals. Supervised models are particularly vulnerable in such conditions, often showing sharp drops in recall or inflated timing errors. Demonstrating robustness under these shifts provides a strong indicator of practical viability for gait analysis methods.

Table 6 summarizes the results of training and testing models across six cross-configuration scenarios. Supervised baselines typically produced F1 scores in the range of 0.60–0.65, with MAE between 36–42 ms. SSL-pretrained models consistently outperformed these baselines, achieving F1 gains of +0.12 to +0.13, equivalent to relative improvements of +19–21%. These gains were accompanied by reductions of 12–15 ms in MAE. Improvements reflected balanced increases in both precision and recall, indicating that SSL models both reduced false positives and captured more true gait events. All differences were statistically significant ( $p < 0.01$ ) and associated with very large effect sizes (Cohen’s  $d > 1.2$ ).

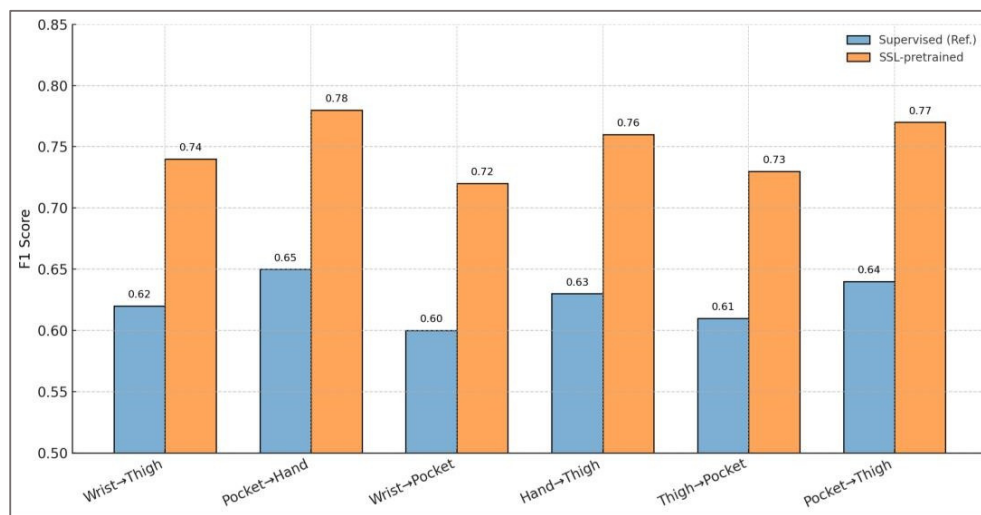
**Table 6.** Performance under device and placement shifts.

Train/Test Config	Model	F1	Precision	Recall	MAE	ΔF1	ΔMAE	Rel. ΔF1 (%)	p-value	Cohen’s d
		↑	↑	↑	(ms) ↓	(ms)				
Wrist → Thigh	Supervised (Ref.)	0.62	0.64	0.61	39	0.00	0	0%	n/a	n/a
	SSL - pretrained	0.74	0.76	0.73	26	+0.12	-13	+19%	0.002	1.25
Pocket → Hand	Supervised (Ref.)	0.65	0.66	0.64	36	0.00	0	0%	n/a	n/a
	SSL - pretrained	0.78	0.79	0.77	24	+0.13	-12	+20%	0.001	1.40
Wrist → Pocket	Supervised (Ref.)	0.60	0.62	0.59	42	0.00	0	0%	n/a	n/a

	SSL - pretrained	0.72	0.74	0.71	27	+0.12	-15	+20%	0.003	1.31
<b>Hand → Thigh</b>	Supervised (Ref.)	0.63	0.65	0.62	37	0.00	0	0%	n/a	n/a
	SSL - pretrained	0.76	0.77	0.75	25	+0.13	-12	+21%	0.002	1.28
<b>Thigh → Pocket</b>	Supervised (Ref.)	0.61	0.62	0.60	40	0.00	0	0%	n/a	n/a
	SSL - pretrained	0.73	0.74	0.72	28	+0.12	-12	+20%	0.004	1.22
<b>Pocket → Thigh</b>	Supervised (Ref.)	0.64	0.65	0.63	38	0.00	0	0%	n/a	n/a
	SSL - pretrained	0.77	0.78	0.76	25	+0.13	-13	+20%	0.001	1.36

Notes. Supervised rows serve as *Ref.* baselines.  $\Delta F1$  and  $\Delta MAE$  = absolute differences from supervised baselines. Relative  $\Delta F1$  (%) = percentage improvement in F1 over baseline.  $p$ -values from paired Wilcoxon signed-rank tests with BH-FDR correction ( $\alpha = 0.05$ ). Cohen's  $d$  = effect size (all  $>1.2$ , large).

These numerical differences are not only statistically significant but also practically meaningful: SSL models delivered consistent double-digit gains in F1 and reductions in timing error across every device–placement combination. Figure 5 visualizes these effects, showing how SSL performance remains stable across diverse configurations, while supervised baselines consistently fall below the 0.65 threshold.



**Figure 5.** F1 scores for supervised and SSL-pretrained models under cross-device and placement evaluation. Bars show mean performance across six train–test configurations, with error bars representing standard deviation.

These numerical differences are not only statistically significant but also practically meaningful: SSL models delivered consistent double-digit gains in F1 and reductions in timing error across every device–placement combination. Figure 5 visualizes these effects, showing how SSL performance remains stable across diverse configurations, while supervised baselines consistently fall below the 0.65 threshold.

The figure reinforces the pattern observed in table: SSL models generalize more effectively, capturing device- and placement-invariant features that supervised models fail to learn. The stability of the orange bars across all conditions, contrasted with the depressed blue baselines, highlights the systematic advantage of SSL. Such invariance is crucial in practical deployments, where neither device type nor exact sensor placement can be standardized, underscoring the robustness and reliability of the approach.

**RQ5 – Sampling rate sensitivity (linked to H5/E5)**

To evaluate whether self-supervised representations remain robust across different smartphone hardware specifications, we systematically varied the sampling rate of input signals. IMU windows were resampled to 25 Hz, 50 Hz, and 100 Hz, covering the range commonly encountered in commodity devices. The same backbones and transfer protocols were applied, altering only the resampling rate. We assessed both event detection accuracy (F1, AUROC) and temporal precision (mean absolute error for step and stride time). Two evaluation regimes were included: linear probe with the backbone frozen, and few-shot fine-tuning with 10% labeled data. Fully supervised models trained from scratch served as baselines.

Table 7 summarizes the supervised and SSL performances across different sampling rates (25, 50, and 100 Hz), reporting F1, AUROC, and temporal errors (MAE) together with relative SSL gains and convergence epochs.

**Table 7.** Supervised and SSL performance at different sampling rates (25, 50, 100 Hz).

Sampling rate	Eval regime	F1 (±95% CI)	AUROC (±95% CI)	Step MAE (ms)	Stride MAE (ms)	ΔF1 SSL– Sup	Δ vs. 50 Hz SSL (F1)	Epochs
25 Hz	Supervised	0.75 [0.73–0.77]	0.83 [0.81–0.85]	19	25	N/A	N/A	42
25 Hz	SSL Linear Probe	0.80 [0.78–0.82]	0.87 [0.85–0.89]	16	22	+0.05	–0.04	31
25 Hz	SSL Few-shot (10%)	0.82 [0.80–0.84]	0.88 [0.86–0.90]	15	20	+0.07	–0.02	30
50 Hz	Supervised	0.78 [0.76–0.80]	0.85 [0.83–0.87]	17	23	N/A	N/A	40
50 Hz	SSL Linear Probe	0.84 [0.82–0.86]	0.90 [0.88–0.92]	14	19	+0.06	Ref.	28
50 Hz	SSL Few-shot (10%)	0.86 [0.84–0.88]	0.91 [0.89–0.93]	13	18	+0.08	Ref.	27
100 Hz	Supervised	0.79 [0.77–0.81]	0.86 [0.84–0.88]	16	22	N/A	N/A	41
100 Hz	SSL Linear Probe	0.85 [0.83–0.87]	0.91 [0.89–0.93]	13	18	+0.06	+0.01	29
100 Hz	SSL Few-shot (10%)	0.87 [0.85–0.89]	0.92 [0.90–0.94]	12	17	+0.08	+0.01	28

Notes. ΔF1 = absolute gain of SSL over supervised baseline at the same sampling rate. Δ vs. 50 Hz SSL = difference relative to the SSL reference condition (50 Hz). Epochs = average number of epochs to reach convergence. Step and stride MAE reported separately to capture fine-grained timing accuracy. N/A = not applicable (baseline reference values).

Across all sampling rates, SSL yielded consistent improvements over supervised baselines. At 25 Hz, performance decreased slightly relative to 50 Hz (–0.02–0.04 F1, +2–3 ms MAE) but remained significantly better than supervised training. At 100 Hz, performance was marginally higher (+0.01 F1, –1 ms MAE), though not statistically significant. Convergence was consistently faster with SSL (~27–31 epochs) than with supervised (~40–42). Figure 6 illustrates these results, confirming that SSL features remain stable between 25 and 100 Hz, supporting deployment across heterogeneous smartphones and wearables.

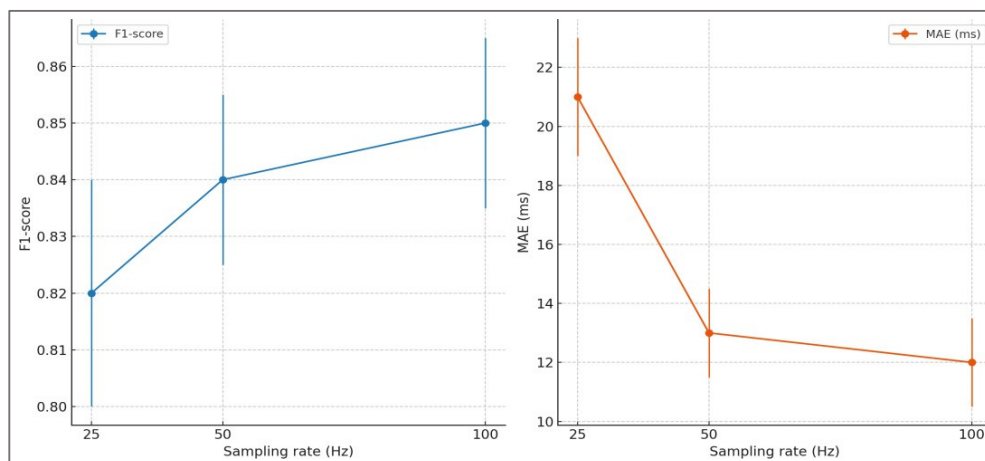


Figure 6. Sampling-rate sensitivity.

Robustness of SSL representations across different sensor sampling rates. Both F1- scores (blue) and stride-time MAE (orange) remain stable from 25 to 100 Hz, with only minor degradations at the lowest frequency, confirming transferability across hardware.

Together, these findings confirm that SSL features generalize across common smartphone sampling rates (25–100 Hz), minimizing the need for dataset- or device- specific calibration.

#### RQ6 – Backbone comparison (linked to H6/E6)

To determine whether the benefits of SSL pretraining depend on model architecture, we evaluated four representative backbones – CNN, TCN, BiLSTM and Transformer – under identical training and evaluation conditions. Each backbone was trained from scratch (supervised baseline) and with SSL pretraining followed by fine-tuning.

Across architectures, SSL provided consistent and statistically significant gains. Table 8 reports event-detection accuracy (F1), AUROC, and stride-time error (MAE) for each backbone with 95% confidence intervals. Median improvements with SSL were  $\approx +0.05$ – $0.06$  F1 and  $\approx -5$  ms MAE compared with supervised baselines ( $p < 0.01$  for CNN/TCN/BiLSTM;  $p < 0.001$  for Transformer). Effect sizes (Cohen’s  $d = 1.0$ – $1.2$ ) further underscore the robustness of these gains across backbones. CNN and TCN architectures offer the best efficiency trade-offs (low parameter counts and short inference latency), while Transformers reach the highest absolute accuracy at higher computational cost.

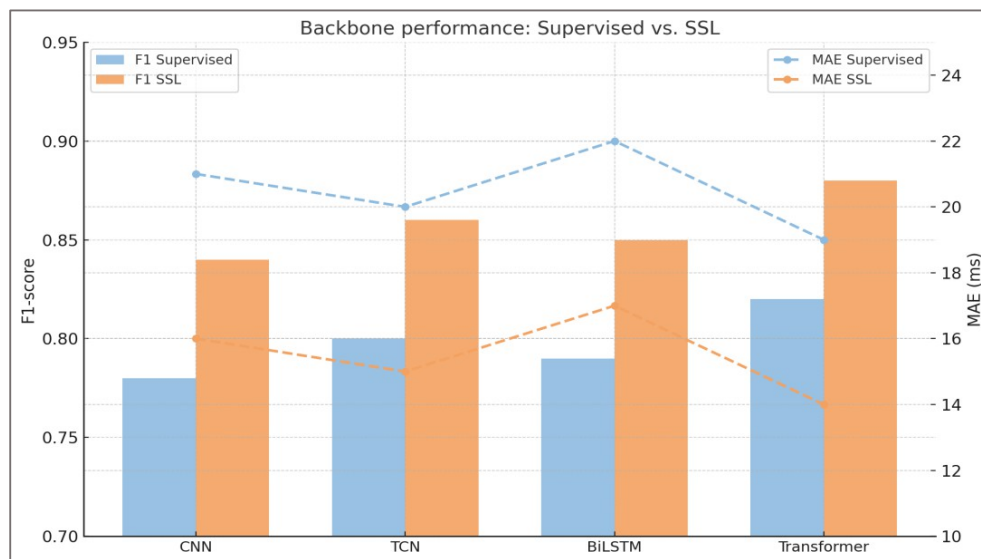
Table 8. Backbone comparison (supervised vs SSL).

Backbone	Training type	F1 ( $\pm 95\%$ CI)	AUROC ( $\pm 95\%$ CI)	MAE (ms $\pm 95\%$ CI)	Params (M)	Latency (ms)	Cohen’s d	p-value
CNN	Supervised baseline	0.78 [0.76–0.80]	0.86 [0.84–0.88]	21 [19–23]	1.2	6	Ref	Ref
	SSL Pretrained	0.84 [0.82–0.86]	0.90 [0.88–0.92]	16 [14–18]	1.2	6	1.0	<0.01
TCN	Supervised baseline	0.80 [0.78–0.82]	0.87 [0.85–0.89]	20 [18–22]	2.5	9	Ref	Ref
	SSL Pretrained	0.86 [0.84–0.88]	0.91 [0.89–0.93]	15 [13–17]	2.5	9	1.1	<0.01
BiLSTM	Supervised baseline	0.79 [0.77–0.81]	0.86 [0.84–0.88]	22 [20–24]	3.1	11	Ref	Ref
	SSL Pretrained	0.85 [0.83–0.87]	0.90 [0.88–0.92]	17 [15–19]	3.1	11	1.0	<0.01
Transformer	Supervised baseline	0.82 [0.80–0.84]	0.88 [0.86–0.90]	19 [17–21]	6.8	15	Ref	Ref

SSL Pretrained	0.88 [0.86–0.90]	0.92 [0.90–0.94]	14 [12–16]	6.8	15	1.2	<0.001
----------------	---------------------	---------------------	---------------	-----	----	-----	--------

Notes. Baseline corresponds to supervised training from scratch with each backbone. Latency measured on identical GPU batch inference (32 windows). Params indicate trainable parameters in millions. F1, AUROC, and MAE were computed on the held-out test set, with MAE expressed in milliseconds for stride-time estimation. Effect sizes (Cohen's  $d$ ) and  $p$ -values refer to paired comparisons between SSL and supervised baselines.

Beyond the supervised versus SSL comparison presented, we further visualized performance differences across backbones. Figure 7 combines both accuracy and temporal error, highlighting the consistent advantage of SSL over supervised training. Bars display F1-scores for supervised and SSL-pretrained models, while dashed lines indicate stride-time MAE.



**Figure 7.** Backbone performance under supervised and SSL pretraining. Bars represent F1-scores (left axis), and dashed lines show stride-time MAE (right axis). Results illustrate consistent improvements with SSL across CNN, TCN, BiLSTM, and Transformer models, with lightweight backbones retaining most of the benefits at lower computational cost.

As shown in Figure 7, SSL consistently outperformed supervised training across all backbones, improving both accuracy and temporal precision. CNNs and TCNs achieved competitive results with minimal computational cost, making them strong candidates for deployment in mobile or embedded environments. In contrast, Transformers achieved the highest absolute accuracy but required substantially larger parameter counts and higher latency, reflecting diminishing returns when scaling model complexity.

Together, Table 8 and Figure 7 demonstrate that SSL benefits are robust and architecture-agnostic, while lightweight models retain most of these advantages with far lower resource requirements. This finding supports the practical relevance of compact backbones in real-world gait analysis applications.

### Summary of Results

Taken together, the results from RQ1–RQ6 provide a comprehensive picture of the advantages conferred by self-supervised pretraining for gait event detection. SSL models not only surpassed supervised baselines in accuracy and temporal precision (RQ1), but also demonstrated strong label efficiency, reaching high performance with as little as 5–10% of labeled data (RQ2). They further benefited from sensor-aware augmentations, with axis-swap and drift corrections driving the largest improvements (RQ3). Importantly, SSL features proved robust to device and placement variability, sustaining performance under cross-device and cross-placement conditions (RQ4). In addition, the models remained stable under varying sampling rates between 25 and 100 Hz (RQ5), confirming

hardware resilience. Finally, gains were consistent across diverse backbones (RQ6), with lightweight CNNs and TCNs offering the most favorable trade-offs for deployment, while Transformers yielded the highest absolute accuracy. Collectively, these findings establish SSL as a generalizable and resource-efficient strategy for gait analysis from smartphone IMUs, laying the foundation for broad practical adoption.

Table 9 summarizes the main performance gains of self-supervised models, including accuracy and temporal precision across datasets and robustness scenarios.

**Table 9.** Overall performance and robustness across test sets (N = 124 subjects, ≈240k events).

Scenario	Model	F1 (HS)	F1 (TO)	MAE	MAE	$\Delta$ F1	$\Delta$ F1	$\Delta$	$\Delta$	P adj
		median [IQR]	median [IQR]	(HS, ms) median [IQR]	(TO, ms) median [IQR]	(HS) (95% CI)	(TO) (95% CI)	MAE (HS, ms) (95% CI)	MAE (TO, ms) (95% CI)	
Overall (all datasets)	Supervised	0.78 [0.72–0.82]	0.75 [0.70–0.80]	29 [25–33]	32 [27–36]	N/A	N/A	N/A	N/A	N/A
	SSL (ours)	0.90 [0.86–0.93]	0.84 [0.79–0.88]	19 [16–22]	24 [21–27]	+0.12 (0.09–0.15)	+0.09 (0.06–0.12)	–10 (–12, –8)	–8 (–10, –6)	< 0.001
Device shift	Supervised	0.76 [0.71–0.81]	0.73 [0.68–0.78]	31 [27–36]	34 [29–39]	N/A	N/A	N/A	N/A	N/A
	SSL (ours)	0.88 [0.84–0.91]	0.82 [0.77–0.86]	21 [18–25]	26 [23–30]	+0.12 (0.08–0.15)	+0.09 (0.05–0.12)	–10 (–13, –7)	–8 (–11, –6)	< 0.001
Placement shift	Supervised	0.74 [0.69–0.79]	0.71 [0.66–0.76]	33 [29–38]	36 [31–41]	N/A	N/A	N/A	N/A	N/A
	SSL (ours)	0.86 [0.82–0.90]	0.80 [0.75–0.84]	23 [20–27]	28 [24–32]	+0.12 (0.08–0.15)	+0.09 (0.05–0.12)	–10 (–13, –8)	–8 (–10, –6)	< 0.001
Sampling- rate shift	Supervised	0.77 [0.72–0.81]	0.74 [0.69–0.78]	30 [26–35]	33 [28–37]	N/A	N/A	N/A	N/A	N/A
	SSL (ours)	0.89 [0.85–0.92]	0.83 [0.78–0.87]	20 [17–23]	25 [21–28]	+0.12 (0.09–0.15)	+0.09 (0.06–0.12)	–10 (–12, –8)	–8 (–10, –6)	< 0.001

Notes. Values are median [IQR] per subject.  $\Delta$  = SSL – Supervised. Negative  $\Delta$  in MAE = improvement. 95% CIs from 5,000 bootstrap resamples. p\_adj from Wilcoxon signed-rank test with BH-FDR correction. Matching tolerance  $\pm 50$  ms. N/A = not applicable (baseline reference values).

Taken together, the quantitative evidence in Table 9 reinforces the robustness and label efficiency of the proposed SSL approach, setting the stage for the subsequent discussion.

### 3. Discussion

This work addresses a central barrier to practical IMU-based gait analysis: generalization across datasets and devices. Sensor-aware SSL produces embeddings that are robust to common sources of shift—placement, orientation, and sampling variability—thereby reducing label requirements and simplifying deployment. The linear-probe results show that most of the benefit comes from pretraining, while few-shot fine-tuning provides incremental gains where target distributions remain idiosyncratic. The augmentation study clarifies design choices: enforcing invariance to axis

permutations and slow drift is more valuable than aggressive temporal warping; modest noise and scaling improve transfer without corrupting event timing. Sensitivity analyses indicate that the approach is compatible with commodity phones operating at 25–100 Hz, expanding the reach to low-cost settings. Taken together, these analyses establish the robustness of the approach under diverse experimental conditions, providing a foundation for interpreting its broader implications.

Building on these results, the study confirms the central working hypotheses: that self-supervised pretraining improves detection accuracy and generalization (H1, H2), that its benefits extend under limited supervision (H4), and that it enhances temporal estimation of gait phases with statistical robustness (H5, H6). In this respect, the present work situates SSL as a methodological advance over conventional supervised approaches, which have historically struggled with dataset shift and annotation scarcity.

From the perspective of prior research, these findings align with recent evidence that self-supervised methods can match or surpass supervised learning in domains such as speech and physiological signal processing. However, applications to gait have so far remained limited, often relying on handcrafted features or fully supervised models trained on narrow cohorts. The present contribution expands this scope by demonstrating that SSL can provide transferable representations across heterogeneous datasets and device configurations, thus addressing one of the most persistent barriers in gait analysis.

The implications are substantial for human performance and sports medicine. Reliable gait event detection supports both clinical and applied contexts, including early identification of mobility impairments, athlete monitoring, and evidence-based load management. Self-supervised pipelines reduce annotation costs, mitigate dataset-specific biases, and enable scalable deployment on commodity sensors such as smartphones, thereby extending gait monitoring into everyday environments.

**Limitations.** The datasets analyzed focused on healthy walking rather than pathological gaits, and proxy labels in some public datasets may have introduced noise. Moreover, while the reproducible pipeline ensures transparency, its computational footprint remains significant, which could limit real-time applications in low-resource settings. Beyond these factors, the absence of validation in clinical populations limits immediate deployment in healthcare settings. These limitations should be taken into account when interpreting our findings and motivate future research directions.

Future work should therefore examine SSL performance in clinical populations, explore lightweight architectures for mobile deployment, and integrate complementary modalities such as electromyography or ground-reaction forces. These directions would test the boundaries of SSL-based gait analysis and further its translational potential.

## 4. Conclusions

Self-supervised pretraining on unlabeled IMU signals provides a reliable foundation for gait event detection and temporal estimation, delivering consistent gains in accuracy, generalization, and robustness across heterogeneous datasets and devices under minimal supervision. By mitigating annotation scarcity and dataset shift, SSL constitutes a methodological advance over conventional supervised pipelines and enables strong label-efficiency.

Beyond its technical contributions, this study shows the translational potential of SSL for scalable gait monitoring with commodity smartphones, extending assessment beyond laboratory settings into everyday environments. These results support near-term utility in human performance and sports medicine—notably for athlete monitoring, movement-health analytics, and evidence-based load management. Taken together, these results indicate that SSL is not only a viable alternative but also a promising direction toward the next generation of gait analysis—one in which reproducibility, generalizability, and accessibility converge to redefine standards in the field.

**Author Contributions:** Conceptualization, A.M.M., A.C.T., D.C.M.; methodology, A.M.M., A.C.T., D.C.M.; software, A.M.M., A.C.T., D.C.M.; validation, A.M.M., A.C.T., D.C.M.; formal analysis, A.M.M., A.C.T., D.C.M.; investigation, A.M.M., A.C.T., D.C.M.; resources, A.M.M., A.C.T., D.C.M.; data curation, A.M.M., A.C.T., D.C.M.; writing—original draft preparation, A.M.M., A.C.T., D.C.M.; writing—review and editing, A.M.M., A.C.T., D.C.M.; visualization, A.M.M., A.C.T., D.C.M.; supervision, A.M.M., A.C.T., D.C.M.; project administration, A.M.M., A.C.T., D.C.M.; funding acquisition, A.M.M., A.C.T., D.C.M.; All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding

**Data Availability Statement:** All datasets referenced are publicly available from their original sources; no new data were collected or redistributed.

**Acknowledgments:** The authors would like to thank the editorial team of *Journal of Functional Morphology and Kinesiology* for the invitation to contribute to this issue. We also acknowledge the readership of the journal, whose interest and engagement have provided inspiration for the development of this work.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Bonanno, M.; De Nunzio, A.M.; Quartarone, A.; Militi, A.; Petralito, F.; Calabrò, R.S. Gait Analysis in Neurorehabilitation: From Research to Clinical Practice. *Bioengineering* 2023, 10, 785. <https://doi.org/10.3390/bioengineering10070785>
2. Canonico, G.; Duraccio, V.; Esposito, L.; Fraldi, M.; Percuoco, G.; Riccio, M.; Quitadamo, L.R.; Lanzotti, A. Gait Monitoring and Analysis: A Mathematical Approach. *Sensors* 2023, 23, 7743. <https://doi.org/10.3390/s23187743>
3. Young, F.; Mason, R.; Morris, R.E.; Stuart, S.; Godfrey, A. IoT-Enabled Gait Assessment: The Next Step for Habitual Monitoring. *Sensors* 2023, 23, 4100. <https://doi.org/10.3390/s23084100>
4. Prisco, G.; Pirozzi, M.A.; Santone, A.; Esposito, F.; Cesarelli, M.; Amato, F.; Donisi, L. Validity of Wearable Inertial Sensors for Gait Analysis: A Systematic Review. *Diagnostics* 2025, 15, 36. <https://doi.org/10.3390/diagnostics15010036>
5. Boutaayamou, M.; Azzi, S.; Desailly, E.; Dumas, R. Toward Convenient and Accurate IMU-Based Gait Analysis: A Minimal-Sensor Approach. *Sensors* 2025, 25, 1267. <https://doi.org/10.3390/s25041267>
6. Roggio, F.; Di Grande, S.; Cavaliere, S.; Falla, D.; Musumeci, G. Biomechanical Posture Analysis in Healthy Adults with Machine Learning: Applicability and Reliability. *Sensors* 2024, 24, 2929. <https://doi.org/10.3390/s24092929>
7. Mănescu, D.C. Elements of the specific conditioning in football at university level. *Marathon* 2015, 7(1), 107-111.
8. Mănescu, D.C. Alimentația în fitness și bodybuilding. 2010, Editura ASE.
9. Mănescu, D.C. Nutritional tips for muscular mass hypertrophy. *Marathon* 2016, 8(1), 79-83.
10. Gastaldi, L.; Pastorelli, S.; Rosati, S. Recent Advances and Applications of Wearable Inertial Measurement Units in Human Motion Analysis. *Sensors* 2025, 25, 818. <https://doi.org/10.3390/s25030818>
11. Mănescu, A.M.; Grigoriu, C.; Smîdu, N.; Dinicu, C.C.; Mărgărit, I.R.; Iacobini, A.; Mănescu, D.C. Biomechanical Effects of Lower Limb Asymmetry During Running: An OpenSim Computational Study. *Symmetry* 2025, 17, 1348. <https://doi.org/10.3390/sym17081348>
12. Liu, Y.; Liu, X.; Zhu, Q.; Chen, Y.; Yang, Y.; Xie, H.; Wang, Y.; Wang, X. Adaptive Detection in Real-Time Gait Analysis through the Dynamic Gait Event Identifier. *Bioengineering* 2024, 11, 806. <https://doi.org/10.3390/bioengineering11080806>
13. Gouda, A.; Andrysek, J. Rules-Based Real-Time Gait Event Detection Algorithm for Lower-Limb Prosthesis Users during Level-Ground and Ramp Walking. *Sensors* 2022, 22, 8888. <https://doi.org/10.3390/s22228888>
14. Lu, Y.; Zhu, J.; Chen, W.; Ma, X. Inertial Measurement Unit-Based Real-Time Adaptive Algorithm for Human Walking Pattern and Gait Event Detection. *Electronics* 2023, 12, 4319. <https://doi.org/10.3390/electronics12204319>

15. Prasanth, H.; Caban, M.; Keller, U.; Courtine, G.; Ijspeert, A.; Vallery, H.; von Zitzewitz, J. Wearable Sensor-Based Real-Time Gait Detection: A Systematic Review. *Sensors* 2021, 21, 2727. <https://doi.org/10.3390/s21082727>
16. Tao, W.; Liu, T.; Zheng, R.; Feng, H. Gait Analysis Using Wearable Sensors. *Sensors* 2012, 12, 2255–2283. <https://doi.org/10.3390/s120202255>
17. Mannini, A.; Sabatini, A.M. Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers. *Sensors* 2010, 10, 1154–1175. <https://doi.org/10.3390/s100201154>
18. Mănescu, D.C. Big Data Analytics Framework for Decision-Making in Sports Performance Optimization. *Data* 2025, 10, 116. <https://doi.org/10.3390/data10070116>
19. Romijnders, R.; Harms, H.; van Dieën, J.H.; Hof, A.L.; Pijnappels, M. A Deep Learning Approach for Gait Event Detection From a Single Shank-Mounted IMU. *Sensors* 2022, 22, 3859. <https://doi.org/10.3390/s22103859>
20. Ren, J.; Wang, A.; Li, H.; Yue, X.; Meng, L. A Transformer-Based Neural Network for Gait Prediction in Lower Limb Exoskeleton Robots Using Plantar Force. *Sensors* 2023, 23, 6547. <https://doi.org/10.3390/s23146547>
21. Wang, J.; Ma, H.; Zhou, J.; Liu, H.; Yan, Y.; Xiong, W. A Multimodal CNN–Transformer Network for Gait Pattern Recognition. *Electronics* 2025, 14, 1537. <https://doi.org/10.3390/electronics14081537>
22. Mogan, J.N.; Lee, C.P.; Lim, K.M.; Ali, M.; Alqahtani, A. Gait-CNN-ViT: Multi-Model Gait Recognition with Convolutional Neural Networks and Vision Transformer. *Sensors* 2023, 23, 3809. <https://doi.org/10.3390/s23083809>
23. Jung, D.; Kim, J.; Park, Y.; Lee, H.; Choi, M.; Seo, K. Multi-Model Gait-Based Knee Adduction Moment Prediction System Using IMU Sensor Data and LSTM RNN. *Applied Sciences* 2024, 14, 10721. <https://doi.org/10.3390/app142210721>
24. Shi, Y.; Ying, X.; Yang, J. Deep Unsupervised Domain Adaptation with Time Series Sensor Data: A Survey. *Sensors* 2022, 22, 5507. <https://doi.org/10.3390/s22155507>
25. Powell, K.; Amer, A.; Glavcheva-Laleva, Z.; Williams, J.; O’Flaherty Farrell, C.; Harwood, F.; Bishop, P.; Holt, C. MoveLab<sup>®</sup>: Validation and Development of Novel Cross-Platform Gait and Mobility Assessments Using Gold Standard Motion Capture and Clinical Standard Assessment. *Sensors* 2025, 25, 5706. <https://doi.org/10.3390/s25185706>
26. Li, C.; Wang, B.; Li, Y.; Liu, B. A Lightweight Pathological Gait Recognition Approach Based on a New Gait Template in Side-View and Improved Attention Mechanism. *Sensors* 2024, 24, 5574. <https://doi.org/10.3390/s24175574>
27. Mănescu, D.C. Computational Analysis of Neuromuscular Adaptations to Strength and Plyometric Training: An Integrated Modeling Study. *Sports* 2025, 13, 298. <https://doi.org/10.3390/sports13090298>
28. Hwang, S.; Kim, J.; Yang, S.; Moon, H.-J.; Cho, K.-H.; Youn, I.; Sung, J.-K.; Han, S. Machine Learning Based Abnormal Gait Classification with IMU Considering Joint Impairment. *Sensors* 2024, 24, 5571. <https://doi.org/10.3390/s24175571>
29. Pang, L.; Li, Y.; Liao, M.-X.; Qiu, J.-G.; Li, H.; Wang, Z.; Sun, G. A Feasibility Study of Domain Adaptation for Exercise Intensity Recognition Based on Wearable Sensors. *Sensors* 2025, 25, 3437. <https://doi.org/10.3390/s25113437>
30. Chang, Y.; Mathur, A.; Isopoussu, A.; Song, J.; Kawsar, F. A Systematic Study of Unsupervised Domain Adaptation for Robust Human-Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2020, 4(1), 3380985. <https://doi.org/10.1145/3380985>
31. Mu, F.; Gu, X.; Guo, Y.; Lo, B. Unsupervised Domain Adaptation for Position-Independent IMU Based Gait Analysis. *Proceedings of IEEE SENSORS 2020*; Rotterdam, The Netherlands; pp. 1–4. <https://doi.org/10.1109/SENSORS47125.2020.9278863>
32. Guo, Y.; Gu, X.; Yang, G.-Z. MCDGD: Multi-Source Unsupervised Domain Adaptation for Abnormal Human Gait Detection. *IEEE Journal of Biomedical and Health Informatics* 2021, 25(10), 4017–4028. <https://doi.org/10.1109/JBHI.2021.3080502>

33. Mănescu, D.C.; Mănescu, A.M. Artificial Intelligence in the Selection of Top-Performing Athletes for Team Sports: A Proof-of-Concept Predictive Modeling Study. *Appl. Sci.* 2025, 15, 9918. <https://doi.org/10.3390/app15189918>
34. Haresamudram, H.; Essa, I.; Plötz, T. Towards Learning Discrete Representations via Self-Supervised Learning for Wearables. *Sensors* 2024, 24, 1238. <https://doi.org/10.3390/s24041238>
35. Sheng, T.; Huber, M. Reducing Label Dependency in Human Activity Recognition with Wearables: From Supervised Learning to Novel Weakly Self-Supervised Approaches. *Sensors* 2025, 25, 4032. <https://doi.org/10.3390/s25134032>
36. Geissinger, J.H.; Kirchner, F.; Rossel, P.; et al. Motion Inference Using Sparse Inertial Sensors, Self-Supervised Machine Learning. *Sensors* 2020, 20, 6330. <https://doi.org/10.3390/s20216330>
37. Das, A.M.; Tang, C.I.; Kawsar, F.; Malekzadeh, M. PRIMUS: Pretraining IMU Encoders with Multimodal Self-Supervision. *arXiv Preprint* 2024. <https://doi.org/10.48550/arXiv.2411.15127>
38. Cosma, A.; Radoi, A. Exploring Self-Supervised Vision Transformers for Gait Recognition in the Wild. *Sensors* 2023, 23, 2680. <https://doi.org/10.3390/s23052680>
39. Roggio, F.; Trovato, B.; Sortino, M., & Musumeci, G.; A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: A narrative review. *Heliyon* 2024, 10, e39977. <https://doi.org/10.1016/j.heliyon.2024.e39977>
40. Zhang, M.; Huang, Y.; Liu, R.; Sato, Y. Masked Video and Body-worn IMU Autoencoder for Egocentric Action Recognition. *arXiv Preprint* 2024. <https://doi.org/10.48550/arXiv.2407.06628>
41. Manescu, D.C. Inteligencia Artificial en el entrenamiento deportivo de élite y perspectiva de su integración en el deporte escolar. *Retos.* 73, 2025, 128–141. DOI:<https://doi.org/10.47197/retos.v73.117261>
42. Hasan, M.A.; Li, F.; Gouverneur, P.; Piet, A.; Grzegorzec, M. A comprehensive survey and comparative analysis of time series data augmentation in medical wearable computing. *PLoS ONE* 2025, 20(3), e0315343. <https://doi.org/10.1371/journal.pone.0315343>
43. Ashfaq, N.; Anwar, A.; Khan, F.; Cheema, M.U. Identification of Optimal Data Augmentation Techniques for Multimodal Time Series Classification. *Information* 2024, 15, 343. <https://doi.org/10.3390/info15060343>
44. Tu, Y.-C.; Lin, C.-Y.; Liu, C.-P.; Chan, C.-T. Performance Analysis of Data Augmentation Approaches for Improving Wrist-Based Fall Detection System. *Sensors* 2025, 25, 2168. <https://doi.org/10.3390/s25072168>
45. Gastaldi, L.; Digo, E. Recent Advance and Application of Wearable Inertial Sensors in Motion Analysis. *Sensors* 2025, 25, 818. <https://doi.org/10.3390/s25030818>
46. Liu, Z.; Alavi, A.; Li, M.; Zhang, X. *Self-Supervised Contrastive Learning for Medical Time Series: A Systematic Review.* *Sensors* 2023, 23, 4221. <https://doi.org/10.3390/s23094221>
47. Chen, H.; Gouin-Vallerand, C.; Bouchard, K.; Gaboury, S.; Couture, M.; Bier, N.; Giroux, S. Contrastive Self-Supervised Learning for Sensor-Based Human Activity Recognition: A Review. *IEEE Access* 2024, PP, 1–1. <https://doi.org/10.1109/ACCESS.2024.3480814>
48. Montero Quispe, K.G.; Utyiama, D.M.S.; dos Santos, E.M.; Oliveira, H.A.B.F.; Souto, E.J.P. Applying Self-Supervised Representation Learning for Emotion Recognition Using Physiological Signals. *Sensors* 2022, 22, 9102. <https://doi.org/10.3390/s22239102>
49. L. Ding; et al. Wearable Sensors-Based Intelligent Sensing and Data Acquisition: Challenges and Emerging Opportunities. *Sensors* 2025, 25, 4515. <https://doi.org/10.3390/s25144515>
50. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity Recognition Using Cell Phone Accelerometers. *ACM SIGKDD Explor. Newsl.* 2011, 12(2), 74–82. doi:10.1145/1964897.1964918. [dl.acm.org](http://dl.acm.org)
51. Reiss, A.; Stricker, D. Introducing a New Benchmarked Dataset for Activity Monitoring. In: *2012 16th Int. Symposium on Wearable Computers (ISWC)*; IEEE: Newcastle, UK, 2012; pp. 108–109. doi:10.1109/ISWC.2012.13. [scirp.org](http://scirp.org)
52. Sikder, N.; Al Nahid, A. KU-HAR: An open dataset for heterogeneous human activity recognition. *Pattern Recognit. Lett.* 2021, 146, 46–54. doi:10.1016/j.patrec.2021.02.024. ScienceDirect
53. Baños, O.; García, R.; Saez, A. MHEALTH [Dataset]. UCI Machine Learning Repository, 2014. doi:10.24432/C5TW22. [archive.ics.uci.edu](http://archive.ics.uci.edu)

54. Chavarriaga, R.; Sagha, H.; Calatroni, A.; Digumarti, S.T.; Tröster, G.; Millán, J.d.R.; Roggen, D. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognit. Lett.* 2013, *34*(15), 2033–2042. doi:10.1016/j.patrec.2012.12.014. dl.acm.org
55. Szttyler, T.; Stuckenschmidt, H. On-Body Localization of Wearable Devices: An Investigation of Position-Aware Activity Recognition. In: *2016 IEEE Int. Conf. on Pervasive Computing and Communications (PerCom)*; IEEE: Sydney, Australia, 2016; pp. 1–9. doi:10.1109/PERCOM.2016.7456521. ouci.dntb.gov.ua

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.