Article | Received 01 February 2026; Revised 14 April 2026; Accepted 29 April 2026; Published 28 May 2026
https://doi.org/10.55092/aias20260004

Transfer learning from gait cycle percentage prediction to gait phase classification using wearable sensors

Huanghe Zhang[1,2]

[1] School of Control Science and Engineering, Center for Robotics, Shandong University, Jinan 250100, China

[2] International Joint Research Center for Perception and Control of Intelligent Rehabilitation Systems of Sichuan Province, Chengdu 610041, China

E-mail: zhanghuanghe@sdu.edu.cn.

Highlights:

- Regression pre-training improves downstream gait phase classification accuracy over training from scratch.
- A compact Deep Neural Network achieves the best transfer result with an F1-score of 0.9788 and sub-0.07 ms Central Processing Unit latency.
- Cross-population validation confirms robustness on an independent young-adult dataset.

Abstract: Reliable gait phase classification is essential for wearable-based locomotion analysis. Although gait cycle percentage prediction and gait phase classification are biomechanically related, knowledge transfer across these distinct objectives remains underexplored. In this paper, we propose a regression-to-classification transfer learning framework that utilizes temporal representations learned from continuous gait cycle progression to improve discrete phase recognition. We pre-train neural backbones on a regression task and transfer the learned representations to the classification task through model transfer (fine-tuning backbone weights) and feature transfer (using the backbone as a fixed feature extractor). To identify the optimal configuration for resource-constrained environments, we compare a compact Deep Neural Network (DNN) with 0.3 M parameters and a Transformer model across multiple sliding window sizes. Our experimental results demonstrate that model transfer achieves a superior F1-score of 0.9788, outperforming the feature transfer baseline and models trained from scratch. Efficiency tests show that the compact DNN achieves a Central Processing Unit (CPU) latency below 0.07 ms, supporting real-time data processing. Validation on an independent dataset further confirms cross-population robustness, achieving a classification accuracy of 92.3%. These findings suggest that regression pre-training captures effective temporal features, providing a practical framework for wearable-based gait analysis.

Keywords: gait analysis; wearable sensor; transfer learning; gait cycle percentage; gait phase

1. Introduction

With the rapid growth of the aging population, gait analysis has become increasingly important for health monitoring and physical rehabilitation. Older adults often exhibit gait abnormalities such as reduced gait speed, increased variability, and asymmetry due to age-related functional decline [1,2]. These changes not only increase fall risk but may also serve as early indicators of neurological disorders, including Parkinson's and Alzheimer's disease [3,4]. Therefore, continuous monitoring and objective assessment of gait patterns in older adults are essential for timely health management and rehabilitation interventions [5].

Conventional gait assessment systems, such as optical motion-capture platforms and instrumented walkways, are expensive, restricted to laboratory environments, and unsuitable for daily or home-based monitoring [6–8]. In contrast, inertial measurement units (IMUs) provide a practical alternative owing to their portability, low power consumption, and reliable measurement of locomotion dynamics [9,10]. A foot-mounted IMU captures tri-axial acceleration and angular velocity signals, enabling detection of key gait events such as heel-strike (HS) and toe-off (TO) and thereby delineating stance and swing phases. As illustrated in Figure 1, these signals support two complementary gait representations. Gait cycle percentage describes continuous gait progression between consecutive HS events, whereas gait phase partitions the gait cycle into discrete functional categories that facilitate clinical interpretation and functional assessment [11].

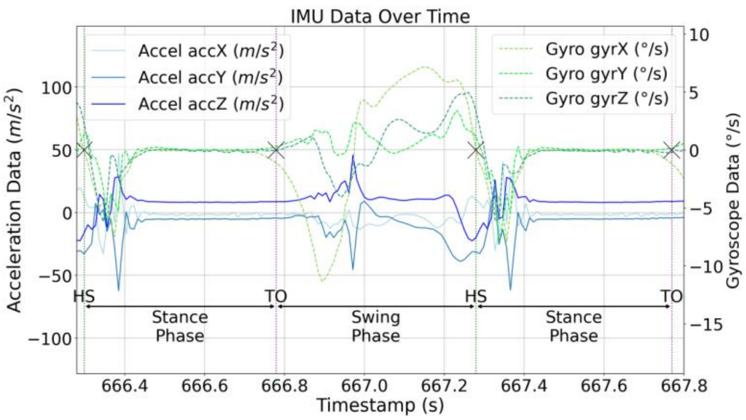


Figure 1. Visualization of IMU signals over time, highlighting stance phase, swing phase, and key events: heel-strike and toe-off.

Extracting robust gait information from raw IMU signals in real-world settings remains challenging due to heterogeneous movement patterns and sensor noise [12]. Deep learning methods have shown strong potential for IMU-based gait modeling by learning task-relevant representations directly from sequential data [13]. Prior studies have developed deep models for gait cycle percentage prediction [14] and for gait phase classification [15,16]. However, these problems are typically treated independently, with separate models trained for regression and classification [17,18]. This practice can introduce redundant training and may miss an opportunity to reuse representations, even though gait cycle percentage and gait phase originate from the same biomechanics and share temporal structure [19].

Transfer learning provides a principled approach for reusing learned representations, especially when labeled data are limited or computational resources are constrained [20]. In gait analysis, prior work has largely focused on transferring models to new subjects or conditions within the same learning objective [21]. Recent studies have further explored domain-level adaptation, such as Sansgiri *et al.* who adapted models from treadmill data to outdoor environments using limited target samples [22]. Despite these advances, knowledge transfer across different objective types remains less explored in wearable gait analysis. Most existing frameworks prioritize domain shifts while overlooking the intrinsic hierarchical relationships between different gait tasks.

A key observation motivates this regression-to-classification transfer setting. Gait phases correspond to specific intervals of the gait cycle percentage, implying a hierarchical relationship between continuous gait progression and discrete phase categories. Learning gait cycle percentage as a regression task encourages models to capture smooth temporal dynamics and fine-grained progression information, which may form a more informative representation for phase classification than training a classifier from scratch. In our previous work, we demonstrated accurate gait cycle percentage prediction in frail older adults using a foot-mounted IMU and deep neural networks [23].

Building on this foundation, we identify a critical research gap in creating efficient, task-adaptive models for wearable edge devices. While heavy architectures like VGG16 and Xception have been repurposed for clinical gait detection through 2D image encoding of kinematic time series [24], their massive computational overhead makes them impractical for real-time execution. Furthermore, personalized clinical transfer learning often relies on complex sensing modalities like surface electromyography (sEMG) [25] or high-precision motion capture systems [26], which limit their utility in daily community-based rehabilitation.

This study fills these gaps by proposing a compact regression-to-classification transfer framework that leverages the biomechanical structure of the gait cycle. We improve discrete gait phase classification by transferring knowledge learned from continuous progression regression. To identify the most effective implementation for resource-constrained environments, we systematically examine two transfer strategies. Our primary approach is model transfer, where the backbone weights are fine-tuned to adapt learned temporal representations to the classification objective. We also implement feature transfer, which utilizes the pre-trained backbone as a fixed feature extractor, as a comparative baseline to quantify the performance gains of weight adaptation. These strategies are compared across a compact Deep Neural Network (DNN) architecture and a Transformer model. To further substantiate the practical utility and robustness of our framework, we perform performance benchmarking of single-sample inference latency and memory footprint. Furthermore, we conduct cross-population validation on the Gold Standard Database to evaluate whether the representations learned from frail older adults can be effectively generalized to a healthy young population. The contributions of this work are as follows:

(1) We study regression-to-classification transfer for wearable gait analysis by leveraging a gait cycle percentage regression model to improve gait phase classification through feature transfer and model transfer strategies.

(2) We compare DNN and Transformer backbones under identical transfer protocols and show that compact DNN models can achieve competitive performance with substantially fewer parameters, supporting deployment in resource-limited wearable systems.

(3) We systematically evaluate sliding window sizes (10, 30, 50 samples) and attention mechanisms, identifying practical design choices that balance recognition performance and computational efficiency.

(4) We provide quantitative experiments of the real-time deployment feasibility of our compact DNN on edge-level CPUs and demonstrate the cross-population robustness of the proposed framework through validation on an independent dataset.

2. Methods

2.1. Experimental dataset

All experiments were conducted on the publicly available GSTRIDE database [27], which provides foot-mounted IMU gait recordings from $N = 158$ older adults (age 70–98 years). The cohort includes both fallers and non-fallers, with basic demographic and clinical characteristics summarized in Table 1.

Table 1. Participant characteristics of the GSTRIDE dataset ($N = 158$).

Characteristic	Value
Women, n (%)	113 (71.5%)
Men, n (%)	45 (28.5%)
Fallers, n (%)	86 (54.4%)
Non-fallers, n (%)	72 (45.6%)
Age (years), mean ± SD	82.6 ± 6.2
Weight (kg), mean ± SD	64.2 ± 13.1
Height (cm), mean ± SD	156.8 ± 10.2
Walking test duration (min), mean ± SD	21.4 ± 7.1
Cognitive deterioration status index, mean ± SD	2.05 ± 1.61
Frailty index, mean ± SD	8.59 ± 2.73

Each subject wore a foot-mounted IMU that recorded tri-axial linear acceleration and angular velocity, $(a_x, a_y, a_z, \omega_x, \omega_y, \omega_z)$, at a sampling rate of either 104 Hz or 128 Hz, as illustrated in Figure 2. As reported in prior work, such differences in device configuration and sampling rate have a negligible impact on gait-related estimation performance in this dataset [28]. In our experiments, window lengths are defined in the number of samples, and model inputs are constructed accordingly.

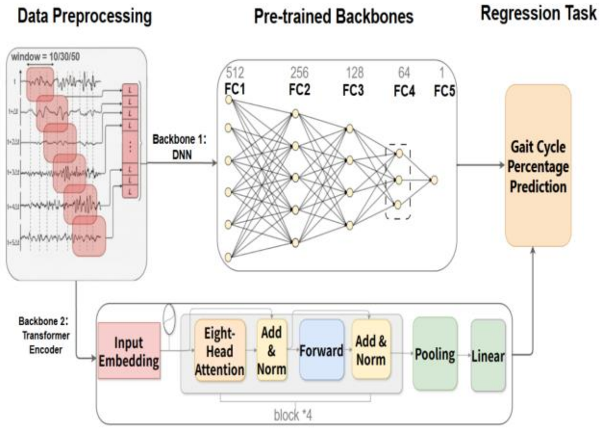
Figure 2. Foot-mounted IMU placement used in the GSTRIDE dataset [27].

To prevent subject leakage, we split the dataset at the subject level into 140 subjects for training, 8 for validation, and 10 for testing. The raw IMU streams were segmented into fixed-length sequences using a sliding-window scheme with window lengths $L \in \{10,30,50\}$. For a window starting at time index t, the input consists of L consecutive samples $\{t, \ldots, t + L - 1\}$, and the corresponding label is assigned at the next time step $t + L$ (one-step-ahead labeling). The resulting input sequence is represented as

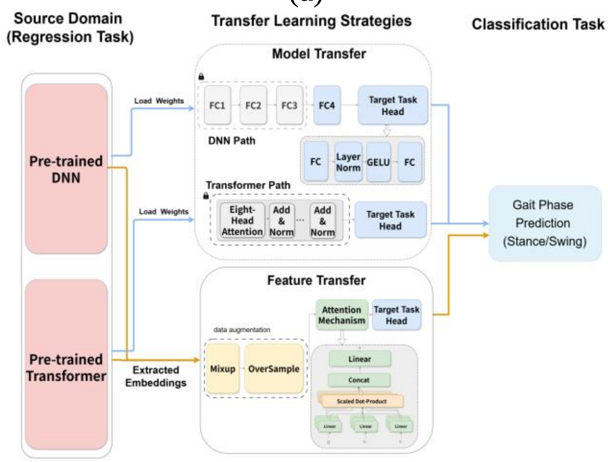
$$X_L(t) = \begin{bmatrix} a_x(t) & a_x(t+1) & \cdots & a_x(t+L-1) \\ a_y(t) & a_y(t+1) & \cdots & a_y(t+L-1) \\ a_z(t) & a_z(t+1) & \cdots & a_z(t+L-1) \\ \omega_x(t) & \omega_x(t+1) & \cdots & \omega_x(t+L-1) \\ \omega_y(t) & \omega_y(t+1) & \cdots & \omega_y(t+L-1) \\ \omega_z(t) & \omega_z(t+1) & \cdots & \omega_z(t+L-1) \end{bmatrix} \in \mathbb{R}^{6\times L} \quad (1)$$

2.2. Transfer learning framework

As shown in Figure 3, we consider two regression-to-classification transfer strategies: model transfer and feature transfer. In both cases, a source model is first trained on gait cycle percentage prediction (regression). Figure 3a illustrates this pre-training stage, where a backbone (DNN or Transformer) processes IMU windows and a fully connected head outputs a single scalar representing the predicted gait cycle percentage. The learned representations are then reused to improve the target task of gait phase classification. Figure 3b shows how the pre-trained backbone is adapted: model transfer replaces the output head and fine-tunes the entire network, while feature transfer freezes the backbone and trains a lightweight classifier on the extracted embeddings, optionally incorporating attention and feature gating modules.



(a)



(b)

Figure 3. Overview of the regression-to-classification transfer framework. FC: fully connected layers. **(a)** Regression pre-training architecture: The backbone is first pre-trained for gait cycle percentage regression; **(b)** Transfer to gait phase classification: The pre-trained representations are then transferred to gait phase classification using either model transfer or feature transfer.

2.2.1. Pre-trained models for the regression task

We pre-train two neural backbones: a compact DNN and a Transformer-based model. The DNN is specifically designed to balance representation power with computational efficiency for wearable deployment. It comprises five fully connected (FC) layers with a progressively narrowing architecture. For a typical input window of $L = 50$ (yielding a 300-dimensional feature vector), the hidden layers are configured with dimensions of 512, 256, 128, and 64, respectively.

Each hidden layer is followed by a Rectified Linear Unit (ReLU) activation function to capture non-linear gait dynamics. The transformation at each layer i is defined as:

$$\mathrm{h}_i = \mathrm{ReLU}(\mathrm{W}_i \mathrm{h}_{i-1} + \mathrm{b}_i), \quad i = 1,2,3,4, \tag{2}$$

where h_0 denotes the flattened input vector from the sliding window. The final regression output is produced by a linear head:

$$\hat{y}_{\mathrm{reg}} = \mathrm{W}_5 \mathrm{h}_4 + \mathrm{b}_5 \tag{3}$$

where $\hat{y}_{\mathrm{reg}}$ represents the predicted gait cycle percentage.

2.2.2. Model transfer

In model transfer, the pre-trained regression model provides initialization for the phase classification network. Let $\mathrm{W}^s = \{\mathrm{W}_1^s, \dots, \mathrm{W}_n^s\}$ denote the source parameters. The target classification parameters W^c are initialized as

$$\mathrm{W}_\mathrm{i}^c = \begin{cases} \mathrm{W}_\mathrm{i}^s, & \mathrm{i} = 1, \dots, \mathrm{n}-1, \\ \text{random initialization}, & \mathrm{i} = \mathrm{n}, \end{cases} \tag{4}$$

where layers 1 to $n-1$ form the shared backbone and the final layer is replaced by a newly initialized classification head.

To preserve transferable representations while enabling adaptation, we adopt selective freezing during fine-tuning [29]. Let $\mathcal{F}$ denote the set of frozen layers. During backpropagation, parameters are updated as

$$\mathrm{W}_i^c \leftarrow \begin{cases} \mathrm{W}_i^c, & i \in \mathcal{F}, \\ \mathrm{W}_i^c - \eta \nabla_{\mathrm{W}_i^c} \mathcal{L}, & i \notin \mathcal{F}, \end{cases} \tag{5}$$

where η is the learning rate and $\mathcal{L}$ is the classification loss.

For the DNN, we freeze the first three fully connected layers and fine-tune the remaining backbone layer and the classification head. For the Transformer, we freeze the input projection and encoder blocks, and fine-tune the pooling/multilayer perceptron (MLP) head for gait phase classification.

2.2.3. Feature transfer

In feature transfer, the regression-pretrained model is used as a fixed feature extractor to produce embeddings for gait phase classification. Let $\mathrm{F} \in \mathbb{R}^{N \times d}$ denote the extracted features for N windows and feature dimension d. For the DNN backbone, we use the activations of the fourth fully connected layer (fc4) as features ($d = 64$). For the Transformer, token outputs are aggregated via temporal average pooling to obtain a d_{model}-dimensional embedding.

We train a lightweight classifier on top of F. To improve robustness, we apply Mixup augmentation

$$\tilde{x} = \lambda x_i + (1-\lambda)x_j, \quad \tilde{y} = \lambda y_i + (1-\lambda)y_j \tag{6}$$

where $\lambda \sim \mathrm{Beta}(\alpha, \alpha)$. Optionally, we incorporate an H-head attention module to reweight discriminative feature components

$$\mathrm{Attn}(Q, K, V) = \mathrm{Softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V. \tag{7}$$

We also consider a feature-gating mechanism

$$F_g = F \odot \sigma\big(W_2\,\mathrm{ReLU}(W_1 F)\big) \tag{8}$$

to suppress irrelevant dimensions adaptively, where $\sigma(\cdot)$ is the sigmoid function.

To address class imbalance, we employ a balanced focal loss

$$\mathcal{L}_{\mathrm{FL}}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{9}$$

where p_t is the predicted probability for the ground-truth class, α_t is a class-balancing factor, and γ controls the focusing strength.

2.3. Implementation details

For regression pre-training, we use Adam with weight decay 1×10^{-5} and learning rate 1×10^{-3} for both backbones. Models are optimized with mean squared error (MSE) using a batch size of 2048. The DNN contains five fully connected layers, and the Transformer uses four encoder blocks. We adopt StepLR with step size 5 epochs and decay factor 0.5.

For phase classification (model transfer and feature transfer), we keep weight decay 1×10^{-5} and optimize a combined objective based on cross-entropy and focal loss (with $\gamma = 2.0$). We use CosineAnnealingLR with initial learning rate 1×10^{-3} and minimum learning rate 1×10^{-6}. Training runs for up to 30 epochs with early stopping (patience 5). Unless otherwise specified, the batch size is 1024.

For baseline comparison, we train a five-layer DNN with the same architecture as the regression DNN directly on raw IMU windows for gait phase classification (*i.e.*, without transfer).

We further apply Mixup augmentation ($\alpha = 0.2$) and label smoothing (0.1) to improve robustness. All experiments are implemented in PyTorch 2.3.0 with Python 3.12 and Compute Unified Device Architecture (CUDA) 12.1 on Ubuntu 22.04, using an NVIDIA RTX 4090 (24 GB), an Intel Xeon Gold 6430 CPU, and 120 GB RAM.

3. Results

3.1. Base model performance on the regression task

Figure 4 summarizes the performance of the two pretrained backbones (DNN and Transformer) on gait cycle percentage prediction. This regression task provides the source representations for subsequent regression-to-classification transfer. Across all window sizes, the DNN achieves consistently lower regression errors than the Transformer. Both backbones attain their best regression performance at the 30-sample window, indicating that this window length offers a favorable balance between temporal context and noise accumulation. The pretrained regression models are then used as source models for feature transfer and model transfer in the gait phase classification task.

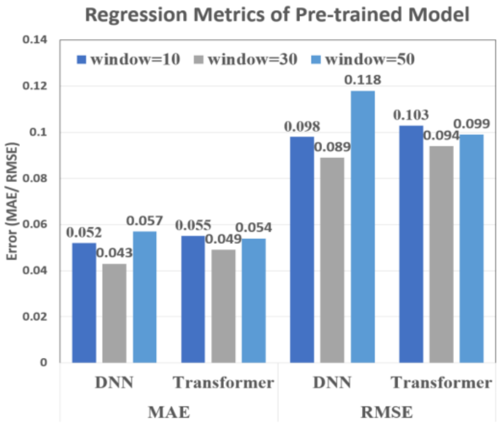


Figure 4. Regression performance of the pretrained backbones for gait cycle percentage prediction under different window sizes.

3.2. Feature transfer performance

Figure 5 and Table 2 report phase-averaged F1-scores and parameter counts for the feature-transfer configurations under window lengths $L \in \{10,30,50\}$. Overall, feature transfer yields stable performance across settings (typically > 0.90 F1), indicating that regression-pretrained representations are informative for gait phase discrimination.

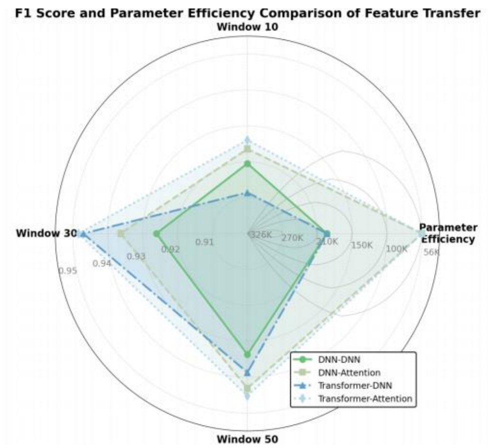


Figure 5. Phase-averaged F1-score *versus* parameter count for feature-transfer configurations. Higher F1 and fewer parameters indicate better accuracy–efficiency trade-offs.

Table 2. Regression and feature transfer performance comparison.

Backbone	Window	Regression Parameters	Transfer Method	Transfer Parameters	F1-score
DNN	10	203.8K	DNN	203.8K	0.9164
			Attention	56.9K	0.9198
	30	265.2K	DNN	265.2K	0.9219
			Attention	87.6K	0.9305
	50	326.7K	DNN	326.7K	0.9283
			Attention	118.3K	0.9362
Transformer	10	2689.2K	DNN	203.8K	0.9096
			Attention	56.9K	0.9219
	30	2720.1K	DNN	265.2K	0.9395
			Attention	87.6K	0.9406
	50	2750.7K	DNN	326.7K	0.9325
			Attention	118.3K	0.9380

Effect of backbone choice. The impact of selecting different regression backbones as feature extractors is moderate and depends on the window length. At $L = 30$, Transformer-based feature extraction achieves an F1-score of 0.9395, compared with 0.9219 using DNN-based extraction, whereas at $L = 10$, the DNN-extracted features slightly outperform Transformer-extracted features (0.9164 *vs.* 0.9096). These results suggest that the choice of feature extractor should be considered jointly with the temporal window configuration.

Effect of attention-based classifier. Incorporating the attention-based classifier provides modest but consistent improvements. For example, at $L = 30$, the F1-score increases from 0.9219 to 0.9305 when using DNN-extracted features, and from 0.9395 to 0.9406 when using Transformer-extracted features.

Parameter efficiency. Transformer backbones involve substantially more parameters for regression pretraining (e.g., 2720.1K *vs.* 265.2K at $L = 30$), yet yield only marginal gains in phase classification under feature transfer. Moreover, the attention-based classifier substantially reduces the number of parameters *in the classification stage* (e.g., from 265.2K to 87.6K at $L = 30$) while maintaining or improving performance, highlighting a favorable accuracy–efficiency trade-off for deployment-oriented scenarios.

3.3. Model transfer approach

Model transfer adapts the regression-pretrained backbone to gait phase classification via head replacement and fine-tuning. Table 3 reports phase-wise precision, recall, and F1-scores under different architectures, fine-tuning strategies, and window sizes. Overall, model transfer substantially improves phase classification compared with the baseline trained from scratch. While the baseline yields phase-wise F1-scores in the range of 0.9124–0.9432, the transferred models consistently achieve F1-scores above 0.96 across all configurations, indicating that regression pretraining provides a strong initialization for the classification objective.

Among the evaluated settings, standard DNN fine-tuning achieves the best overall performance, reaching F1-scores of 0.9788 (stance) and 0.9761 (swing) at the 50-sample window, corresponding to improvements of 5.6% and 3.81% over the baseline for the same window size. Incorporating attention yields only marginal changes (typically $< 0.1\%$) and occasionally slightly degrades performance, suggesting that the DNN backbone already captures the discriminative temporal patterns required for this two-phase classification task.

The Transformer backbone achieves comparable but slightly lower performance, with a maximum stance-phase F1-score of 0.9711 at the 10-sample window. Unlike the DNN, Transformer performance does not show a monotonic improvement with larger windows, and slightly decreases at the 50-sample window (0.9673 for stance and 0.9642 for swing). At $L = 50$, this corresponds to a 1.15% (stance) and 1.19% (swing) decrease relative to DNN fine-tuning under the same window size. Collectively, these results indicate that model transfer with standard DNN fine-tuning provides an effective and computationally efficient solution for gait phase classification.

Table 3. Model performance comparison across different architectures, attention variants, and window sizes. All models except Baseline are initialized from regression pretraining and fine-tuned for classification (model transfer). Baseline is trained from scratch.

Model	Training	Attention	Window	Phase	Precision	Recall	F1-score
DNN	Transfer	✗	10	Stance	0.9622	0.9816	0.9718
			10	Swing	0.9786	0.9561	0.9672
			30	Stance	0.9801	0.9707	0.9754
			30	Swing	0.9669	0.9775	0.9722
			50	Stance	0.9849	0.9727	0.9788
			50	Swing	0.9693	0.9830	0.9761
DNN	Transfer	✓	10	Stance	0.9730	0.9726	0.9728
			10	Swing	0.9688	0.9692	0.9690
			30	Stance	0.9814	0.9697	0.9755
			30	Swing	0.9659	0.9791	0.9724
			50	Stance	0.9836	0.9733	0.9784
			50	Swing	0.9699	0.9815	0.9757
Transformer	Transfer	--	10	Stance	0.9821	0.9603	0.9711
			10	Swing	0.9558	0.9801	0.9678
			30	Stance	0.9788	0.9587	0.9686
			30	Swing	0.9540	0.9763	0.9651
			50	Stance	0.9861	0.9492	0.9673
			50	Swing	0.9445	0.9847	0.9642
Baseline	Scratch	--	10	Stance	0.9406	0.8976	0.9186
			10	Swing	0.8900	0.9360	0.9124
			30	Stance	0.9666	0.9200	0.9427
			30	Swing	0.9143	0.9641	0.9385
			50	Stance	0.9586	0.9283	0.9432
			50	Swing	0.9218	0.9547	0.9380

3.4. Impact of sliding window size

We further examine the effect of window length on performance and computational cost. Across methods, increasing the window size generally improves classification performance, with diminishing returns beyond 30 samples. In particular, the 30-sample window provides a favorable accuracy–efficiency trade-off, capturing sufficient temporal context for phase discrimination while avoiding the additional computation associated with larger windows. Larger windows may offer small gains in some settings but increase training cost, especially for Transformer-based models.

Training-time measurements further highlight the efficiency advantage of transfer learning. For example, with a 50-sample window, training the DNN from scratch requires approximately 30 minutes, whereas the corresponding model-transfer setting completes in under 14 minutes. Feature-transfer configurations are even more efficient, typically requiring less than 8 minutes. In addition, for the same window size, Transformer training is substantially slower than DNN training (e.g., approximately 36 minutes *versus* 15 minutes for the regression pretraining setting at $L = 50$). Overall, these results demonstrate that regression-to-classification transfer not only improves gait phase classification accuracy but also reduces training time compared with training from scratch, supporting its practicality for wearable gait analysis pipelines.

3.5. Real-time inference performance and memory footprint

To further substantiate the practical efficiency of compact DNNs for wearable deployment, we measured single-sample inference latency and model memory footprint for all configurations. In the experimental setup, the batch size was set to 1 to simulate real-time single-sample inference on edge devices. The CPU test represents low-power embedded platforms without Graphics Processing Unit (GPU), while the GPU test evaluates performance limits with hardware acceleration, utilizing CUDA Events for precise timing to eliminate synchronization errors. To ensure reliability and eliminate overhead from JIT compilation and cache initialization, we pre-heated the CPU for 50 trials and the GPU for 100 trials before formal experiments. The mean and 95th percentile latencies were recorded over 500 CPU and 1000 GPU trials, as detailed in Table 4.

The results indicate that compact DNN models (0.2–0.3 M parameters) achieve exceptionally low CPU inference latencies, ranging from 0.052 ms to 0.066 ms. Given that the typical sampling rate in the GSTRIDE dataset is 104–128 Hz (corresponding to a time budget of approximately 7.8–9.6 ms per sample), the DNN models consume less than 1% of the available real-time processing budget. In contrast, while the Transformer models (approximately 2.7 M parameters) remain within the real-time threshold, they exhibit 15–20 × higher CPU latency and a significantly larger memory footprint (approximately 10.4 MB) compared to the DNN (approximately 1.0 MB). Despite the minor differences in GPU memory usage due to framework initialization overhead, the vast gap in CPU latency highlights the absolute advantage of compact DNNs on micro-controller platforms with extremely limited computational resources. These metrics confirm that the proposed regression-to-classification transfer framework not only improves recognition accuracy but also provides an efficient and practically viable solution for resource-constrained wearable gait analysis systems.

Table 4. Inference latency and memory footprint comparison.

Architecture	Window Size	Size (MB)	CPU Latency Mean (ms)	CPU P95 (ms)	GPU Mem (MB)
DNN	10	0.78	0.052	0.054	8.91
	30	1.01	0.058	0.061	9.14
	50	1.25	0.066	0.066	9.38
Transformer	10	10.26	0.903	0.925	11.22
	30	10.37	1.106	1.132	11.30
	50	10.49	1.312	1.348	11.38

3.6. Cross-population validation on an independent dataset

To evaluate the generalizability of our framework across different age groups, we conducted additional experiments on an independent, publicly available IMU dataset collected by Losing and Hasenjäger [30].

This dataset includes gait recordings from 20 healthy young adults, providing a clear contrast to the frail elderly population in the GSTRIDE dataset.

Based on the superior performance of model transfer observed in Section 3.3, we selected it as the representative strategy for this cross-population assessment. We compared the model transfer approach (pre-trained on GSTRIDE regression and fine-tuned on the target classification task) against a baseline model trained from scratch on the same independent dataset. As shown in Table 5, the model transfer approach achieved an accuracy of 92.3%, significantly outperforming the baseline (90.1%). These results demonstrate that the temporal representations learned from the continuous gait cycle percentage regression are highly transferable, capturing fundamental gait dynamics that remain robust despite age-related variations in walking speed and stability.

Table 5. Cross-dataset performance comparison (Window = 50).

Method	Stance F1-score	Swing F1-score	Accuracy
Baseline (from scratch)	0.9199	0.8707	90.1%
Model Transfer (Ours)	0.9382	0.9015	92.3%

4. Discussion

Transfer learning in gait analysis has evolved rapidly, with recent studies focusing on domain adaptation across environments [22] or sensing modalities [25,26]. For instance, Sansgiri *et al.* demonstrated successful treadmill-to-outdoor adaptation for older adults [22], while Zhao *et al.* leveraged healthy-side sEMG signals to predict affected-side gait trajectories [25]. However, most existing frameworks address population or environment shifts within a single objective, often overlooking the intrinsic hierarchical relationships between different gait tasks. This study fills this research gap by investigating a regression-to-classification transfer setting, reusing representations learned from continuous gait cycle percentage prediction to improve discrete gait phase classification.

Across architectures and window sizes, model transfer consistently outperformed feature transfer, with the best model-transfer configuration achieving an F1-score of 0.9788. Feature transfer also provided competitive performance, reaching a maximum F1-score of 0.9406 (Transformer features, 30-sample window), while DNN features achieved 0.9362 (50-sample window). These results indicate that regression-pretrained representations capture informative temporal structure that can be effectively repurposed for discrete phase recognition. Conceptually, gait cycle percentage and gait phases offer continuous and discrete views of the same locomotion process [31]; learning continuous progression via regression can encourage smoother, temporally consistent representations, which align well with the subsequent partitioning required for phase classification. This perspective is also consistent with prior findings that gait-related temporal descriptors support both continuous estimation and categorical decision-making in downstream applications [32–34].

A pivotal finding of this study is the favorable accuracy–efficiency trade-off provided by compact DNNs, a critical requirement for real-time edge deployment. While heavy visual architectures such as VGG16 and Xception have been repurposed for gait-based diagnosis (e.g., knee osteoarthritis) with high accuracy [24], their massive computational footprint limits their utility in wearable sensors. In contrast, our results in Section 3.5 substantiate that a compact 0.3 M parameter DNN can achieve CPU inference latencies as low as 0.052–0.066 ms. Given the typical IMU sampling rates of 104–128 Hz, our framework consumes less than 1% of the available real-time processing budget (approximately 8–10 ms per sample), representing a 15–20× speed advantage over Transformer models without sacrificing classification robustness.

From a practical standpoint, feature transfer offers a clear computational advantage. The reduction in training time (e.g., from approximately 30 minutes to under 8 minutes in our setting) stems from precomputing regression embeddings and training only a lightweight classifier, thereby avoiding repeated updates to the full backbone during classification training. This efficiency, combined with the superior generalization observed on the cross-population dataset (92.3% accuracy for young adults), suggests that regression-to-classification transfer allows models to capture fundamental gait dynamics that transcend age-related domain shifts while minimizing the computational burden of model adaptation.

We further observed that increasing window size generally improves performance, but the effect is bounded, with a maximum gain of about 3.4% in our experiments. Model transfer benefits more consistently from larger windows, suggesting that end-to-end fine-tuning can exploit additional temporal context while learning to suppress irrelevant variation. For feature transfer, the effect depends on the backbone: DNN-derived features improve steadily with longer windows, whereas Transformer-derived features peak at moderate window lengths (e.g., 30 samples). This difference may reflect how each architecture aggregates temporal information: longer windows provide additional context that can compensate for limited receptive fields in DNN representations, while attention-based representations may saturate once the window already covers the most informative temporal relations and may become more sensitive to redundant or noisy correlations beyond that range.

While the validation on an independent dataset confirms the framework's robustness across healthy young and frail elderly populations, several limitations remain. First, this study does not yet include data from individuals with neurological conditions, such as Parkinson's disease or stroke [25]. Such gaits often exhibit high variability and non-periodic episodes, which may pose a greater challenge for representation transfer than age-related variations. Second, our experiments primarily utilized natural walking data under relatively controlled conditions, unlike the complex outdoor level, incline, and decline terrains explored by Sansgiri *et al.* [22]. In real-world environments, abrupt changes in walking speed and diverse terrains could significantly impact IMU signal characteristics, requiring more sophisticated adaptation strategies such as unsupervised domain adaptation.

To address these challenges, several directions are worth exploring. Validating the regression-to-classification transfer strategy on even broader cohorts and additional benchmark datasets would strengthen evidence for generalizability beyond frail older adults. We aim to extend the framework to richer phase definitions, such as multi-class gait sub-phases, and explore unsupervised domain adaptation techniques to further align feature distributions across various pathological states and locomotion conditions. Furthermore, integrating complementary sensing modalities, such as sEMG or pressure/force signals [25], will help capture more comprehensive motor information. Finally, conducting real-time evaluations in unconstrained environments remains a priority to translate this approach into practical continuous monitoring and rehabilitation systems for diverse clinical populations.

5. Conclusion

This study demonstrates the effectiveness of regression-to-classification transfer for IMU-based gait analysis. By transferring knowledge learned from gait cycle percentage prediction to gait phase classification, we achieve higher accuracy and better computational efficiency. Across architectures, compact DNNs with substantially fewer parameters (0.2–0.3 M) achieve performance comparable to, and in some settings slightly better than, Transformer models (2.7 M). Model transfer consistently yields the best results, producing phase-averaged F1-scores above 0.97, whereas feature transfer remains competitive and offers clear efficiency advantages. Attention modules provide modest but consistent gains in several configurations.

From a practical perspective, transfer learning substantially reduces training cost, decreasing training time from roughly 30 minutes to under 8 minutes in our experiments. These results support an efficient workflow in which a single regression-pretrained foundation model can be adapted to downstream gait phase recognition with minimal additional training, reducing redundant model development while maintaining high accuracy. Overall, the proposed regression-to-classification transfer strategy provides a promising pathway toward scalable and resource-efficient wearable gait analysis tools for healthcare and rehabilitation applications.

Data availability statement

All code for data preprocessing, model training, transfer-learning evaluation, cross-population validation, figure generation, and supplementary auditing is available at https://github.com/HuangheZhang/Regression-to-Classification-Gait-Transfer.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used generative AI tools, including GPT-5.4, solely for language polishing and formatting improvement. The AI tools were not involved in data analysis, interpretation of results, or formulation of conclusions. All AI-generated content was carefully reviewed and revised by the authors, who take full responsibility for the content of this manuscript.

Acknowledgments

This work was supported in part by the Taishan Scholars Project (Young Expert Program) under Grant NO.tsqn202408040, in part by the Young Scientists Fund of the National Natural Science Foundation of China under Grant 62403281, in part by the Shandong Excellent Young Scientists Fund Program (Overseas) under Grant 2024HWYQ-019, and in part by the Open Project Fund of International Joint Research Center for Perception and Control of Intelligent Rehabilitation Systems of Sichuan Province under Grant No. 25-H-01.

Conflicts of interest

The author declares no conflicts of interest.

References

[1] Florence CS, Bergen G, Atherly A, Burns E, Stevens J, *et al.* Medical costs of fatal and nonfatal falls in older adults. *J. Am. Geriatrics Soc.* 2018, 66(4):693–698.

[2] Callisaya ML, Blizzard L, Schmidt MD, McGinley JL, Srikanth VK. Ageing and gait variability—a population-based study of older people. *Age Ageing* 2010, 39(2):191–197.
[3] Collyer TA, Murray AM, Woods RL, Storey E, Chong T, *et al.* Association of dual decline in cognition and gait speed with risk of dementia in older adults. *JAMA Netw. Open* 2022, 5(5):2214647–2214647.
[4] Sabo A, Mehdizadeh S, Iaboni A, Taati B. Estimating parkinsonism severity in natural gait videos of older adults with dementia. *IEEE J. Biomed. Health. Inf.* 2022, 26(5):2288–2298.
[5] Momin MS, Sufian A, Barman D, Dutta P, Dong M, *et al.* In-home older adults' activity pattern monitoring using depth sensors: a review. *Sensors* 2022, 22(23):9067.
[6] Balazia M, Sojka P. Gait recognition from motion capture data. *ACM Trans. Multimed. Comput. Commun. Appl.* 2018, 14(1s):1–18.
[7] Parati M, Ambrosini E, De Maria B, Gallotta M, Dalla Vecchia LA, *et al.* The reliability of gait parameters captured via instrumented walkways: a systematic review and meta-analysis. *Eur. J. Phys. Rehabil. Med.* 2022, 58(3):363.
[8] Li J, Liu X, Wang Z, Zhao H, Zhang T, *et al.* Real-time human motion capture based on wearable inertial sensor networks. *IEEE Internet Things J.* 2021, 9(11):8953–8966.
[9] Gao L, Zhang G, Yu B, Qiao Z, Wang J. Wearable human motion posture capture and medical health monitoring based on wireless sensor networks. *Measurement* 2020, 166:108252.
[10] Jarchi D, Pope J, Lee TK, Tamjidi L, Mirzaei A, *et al.* A review on accelerometry-based gait analysis and emerging clinical applications. *IEEE Rev. Biomed. Eng.* 2018, 11:177–194.
[11] Agostini V, Balestra G, Knaflitz M. Segmentation and classification of gait cycles. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2013, 22(5):946–952.
[12] García-de-Villa S, Ruiz LR, Neira G, Álvarez MN, Huertas-Hoyas E, *et al*. Validation of an imu-based gait analysis method for assessment of fall risk against traditional methods. *IEEE J. Biomed. Health. Inf.* 2024, 29(1):107–117.
[13] Alharthi AS, Yunas SU, Ozanyan KB. Deep learning for monitoring of human gait: a review. *IEEE Sens. J.* 2019, 19(21):9575–9591.
[14] Prado A, Cao X, Ding X, Agrawal SK. Prediction of gait cycle percentage using instrumented shoes with artificial neural networks. In *Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, May 31–August 31, 2020, pp. 2834–2840.
[15] Yu S, Yang J, Huang TH, Zhu J, Visco CJ, *et al.* Artificial neural network-based activities classification, gait phase estimation, and prediction. *Ann. Biomed. Eng.* 2023, 51(7):1471–1484.
[16] Han YC, Wong KI, Murray I. Gait phase detection for normal and abnormal gaits using imu. *IEEE Sens. J.* 2019, 19(9):3439–3448.
[17] Cai J, Guan Z, Wang J, Ding Z, Li Y, *et al.* Impact of gait parameters and their variability on fall risk assessment accuracy using wearable sensor. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2025, 33:1996–2003.
[18] Guan Z, Cai J, Wang J, Li Y, Song R, *et al.* Accuracy and precision of wearable-derived gait parameters: how these affect the performance of models for fall prediction in the elderly. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2025, 33:4255–4266.

[19] Wu J, Becsek B, Schaer A, Maurenbrecher H, Chatzipirpiridis G, *et al*. Real-time gait phase detection on wearable devices for real-world free-living gait. *IEEE J. Biomed. Health. Inf.* 2022, 27(3):1295–1306.

[20] Tan C, Sun F, Kong T, Zhang W, Yang C, *et al.* A survey on deep transfer learning. In *Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks*, Rhodes, Greece, October 4–7, 2018, pp. 270–279.

[21] Gu X, Guo Y, Deligianni F, Lo B, Yang G. Cross-subject and cross-modal transfer for generalized abnormal gait pattern recognition. *IEEE Trans. Neural Networks Learn. Syst.* 2020, 32(2):546–560.

[22] Sansgiri S, Matikainen-Tervola E, Rantakokko M, Finni T, Rantalainen T, *et al.* From treadmill to outdoor overground walking: enhancing ground contact timing detection for older adults using transfer learning. *Exp. Gerontol.* 2026, 215:113056.

[23] Wang J, Guan Z, Liang T, Ding Z, Ma X, *et al.* Multi-task learning for gait phase and gait cycle percentage prediction with wearable sensors in frail older adults. *IEEE J. Biomed. Health. Inf.* 2025, 29(12):1–13.

[24] Moura FA, Pelegrinelli AR, Catelli DS, Lamontagne M, Silva Torres R. Knee osteoarthritis prediction from gait kinematics: exploring the potential of deep neural networks and transfer learning methods for time series classification. *J. Biomech.* 2025, 190:112881.

[25] Zhao Y, Li J, Zhang J, Li K, Wang D, *et al.* Personalized gait trajectory prediction of hemiplegic patients based on transfer learning. *IEEE Sens. J.* 2025, 25(16):30970–30983.

[26] Świtoński A, Josiński H, Harężlak K, Augustyn D, Szczęsna A, *et al.* Chaoticity of gait motion capture data determined by transfer learning in human sex recognition challenge. *Eng. Appl. Artif. Intell.* 2026, 167:113782.

[27] García-de-Villa S, Neira G, Álvarez MN, Huertas-Hoyas E, Ruiz LR, *et al.* A database with frailty, functional and inertial gait metrics for the research of fall causes in older adults. *Sci. Data* 2023, 10(1):566.

[28] García-Villamil G, Ruiz L, Jiménez AR, Seco F, Rodríguez-Sánchez M. Influence of imu's measurement noise on the accuracy of stride-length estimation for gait analysis. In *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation*, Lloret de Mar, Spain, November 29–December 2, 2021, pp. 5–16.

[29] Guo Y, Shi H, Kumar A, Grauman K, Rosing T, *et al.* Spottune: transfer learning through adaptive fine-tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, California, USA, June 16–20, pp. 4805–4814.

[30] Losing V, Hasenjäger M. A multi-modal gait database of natural everyday-walk in an urban environment. *Sci. Data* 2022, 9(1):473.

[31] Özateş ME, Yaman A, Salami F, Campos S, Wolf S I, *et al.* Identification and interpretation of gait analysis features and foot conditions by explainable ai. *Sci. Rep.* 2024, 14(1):5998.

[32] De Vos M, Prince J, Buchanan T, FitzGerald JJ, Antoniades CA. Discriminating progressive supranuclear palsy from parkinson's disease using wearable technology and machine learning. *Gait Posture* 2020, 77:257–263.

[33] Lin S, Gao C, Li H, Huang P, Ling Y, *et al.* Wearable sensor-based gait analysis to discriminate early parkinson's disease from essential tremor. *J. Neurol.* 2023, 270(4):2283–2301.

[34] Osorio R, Pastene F, Ortega P, Aqueveque P. Gait subphases classification based on hidden markov models using in-shoes capacitive pressure sensors: preliminary results. In *Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Glasgow, UK, July 11–15, 2022, pp. 756–759.