

# **Research Proposal: A Data-Driven Distributional Approach to Gait Assessment: Quantifying Deviation from Normal Gait Patterns**

Stimuloop Research Consortium

November 4, 2025

## **Abstract**

We propose a data-driven gait assessment metric that quantifies patient improvement in neurological rehabilitation by modeling complete gait cycles as probability distributions. The method learns normative patterns from healthy controls and measures patient deviation using distributional distance metrics. We process 4D spatio-temporal motion capture data, segment into gait cycles, and construct individual and population-level distributions. This approach provides objective, comprehensive assessment of gait patterns applicable across rehabilitation scenarios.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Research Objectives</b>	<b>3</b>
<b>3</b>	<b>Literature Review</b>	<b>3</b>
3.1	Current Approaches . . . . .	3
3.2	Probabilistic Approaches . . . . .	3
3.3	Generative Models . . . . .	4
3.4	Physics-Constrained Models . . . . .	4
3.5	Distributional Assessment Framework . . . . .	4
3.6	Distance Measures . . . . .	4
3.7	Clinical Context . . . . .	4
<b>4</b>	<b>Methodology</b>	<b>5</b>
4.1	Framework Overview . . . . .	5
4.2	Data Representation . . . . .	5
4.3	Gait Cycle Segmentation and Normalization . . . . .	5
4.4	Individual Gait Distribution Estimation . . . . .	5
4.5	Normative Distribution Construction . . . . .	5
4.5.1	Distribution Aggregation . . . . .	6
4.5.2	Demographic Conditioning . . . . .	6
4.6	Deviation Quantification . . . . .	6
4.7	Study Design and Participants . . . . .	6
4.7.1	Phase 1: Normative Database Construction . . . . .	6
4.7.2	Phase 2: Clinical Validation . . . . .	6
4.8	Validation . . . . .	7
<b>5</b>	<b>Expected Outcomes</b>	<b>7</b>
5.1	Primary Outcome . . . . .	7

# 1 Introduction

Current gait assessment methods rely on discrete parameters or subjective scales, failing to capture the holistic nature of human locomotion (Baker, 2013). Traditional approaches examine isolated parameters, missing complex multidimensional changes in neurological conditions (Whittle, 2014).

We propose modeling gait as probability distributions over complete cycles to capture natural variability and movement quality (Hausdorff et al., 2001). This distributional approach addresses the gap in comprehensive gait assessment tools that can quantify deviation from healthy patterns and detect rehabilitation improvements (Horst et al., 2019).

## 2 Research Objectives

1. Develop a distributional gait assessment metric quantifying deviation from healthy patterns
2. Establish normative gait distributions with demographic conditioning (age, sex, height)
3. Validate metric sensitivity for detecting improvements in stroke and Parkinson’s patients
4. Compare distributional distance measures for optimal gait quality quantification
5. Create a robust assessment tool applicable across rehabilitation scenarios

## 3 Literature Review

### 3.1 Current Approaches

Traditional gait analysis uses discrete measurements (stride length, cadence, joint angles) that fail to capture coordinated locomotion dynamics (Baker, 2013; Whittle, 2014).

### 3.2 Probabilistic Approaches

Probabilistic models show promise for comprehensive gait representation. Gholami et al. (2019) demonstrated hidden Markov models for temporal dependencies. Stihl et al. (2025) used Hierarchical Variational Sparse Heteroscedastic Gaussian Processes for multiple sclerosis analysis, capturing variability patterns and neurological impairment hallmarks.

### 3.3 Generative Models

Recent AI-driven generative models address data scarcity in clinical populations. Peppes et al. (2023) developed FoGGAN for Parkinson’s episodes. Adeli et al. (2025) introduced GAITGen for conditional 3D gait generation. Yamada et al. (2025) demonstrated physics-based approaches achieving superior performance through synthetic data diversity. Rezvani et al. (2025) applied diffusion models for Parkinson’s severity assessment.

### 3.4 Physics-Constrained Models

Takeishi and Kalousis (2021) developed physics-integrated VAEs embedding biomechanical models for realistic gait generation. Ghosh et al. (2025) validated biomechanical feasibility requirements, ensuring plausible and interpretable synthetic patterns.

### 3.5 Distributional Assessment Framework

Our approach leverages distributional perspectives for assessment rather than generation. We focus on measuring deviation from normative patterns using probabilistic representations (Adeli et al., 2025; Takeishi and Kalousis, 2021). This provides direct quantitative measures of gait abnormality while capturing variability inherent in neurological conditions. Zhang et al. (2025) demonstrated 65% improvement in estimation accuracy through distributional modeling, supporting our hypothesis that comprehensive distributional representations provide more sensitive assessment tools.

### 3.6 Distance Measures

Distance metric selection critically impacts assessment sensitivity and interpretability. We will evaluate information-theoretic measures, geometric distances respecting gait structure, and optimal transport distances. Selection criteria include computational efficiency, clinical sensitivity, noise robustness, and practitioner interpretability.

### 3.7 Clinical Context

Post-stroke gait shows asymmetry and compensatory patterns (Dobkin, 2005). Parkinson’s disease presents reduced step length and increased variability (Keus et al., 2014). Traditional metrics miss complex interactions between movement aspects. Rehabilitation requires sensitive measures detecting gradual improvements (Mehrholtz et al., 2017).

## 4 Methodology

### 4.1 Framework Overview

We model gait as probability distributions over complete cycles with three components: (1) normative distribution learning, (2) patient-specific estimation, and (3) distributional distance computation.

### 4.2 Data Representation

Gait data is represented as 4-dimensional spatio-temporal graphs, where each node corresponds to an anatomical landmark with 3D spatial coordinates  $(x, y, z)$  and temporal information  $t$ . For a single gait cycle, we have:

$$G_i = \{(x_i^j(t), y_i^j(t), z_i^j(t), t) : j \in \mathcal{J}, t \in [0, T_i]\} \quad (1)$$

where  $i$  indexes the gait cycle,  $j$  indexes anatomical landmarks in the set  $\mathcal{J}$ , and  $T_i$  is the duration of cycle  $i$ .

### 4.3 Gait Cycle Segmentation and Normalization

Raw motion capture recordings undergo automatic segmentation into individual gait cycles using heel-strike detection algorithms. Each cycle is temporally normalized to a standard duration to enable comparison across different walking speeds:

$$\tilde{G}_i = \{(x_i^j(\tau), y_i^j(\tau), z_i^j(\tau), \tau) : j \in \mathcal{J}, \tau \in [0, 1]\} \quad (2)$$

where  $\tau = t/T_i$  represents normalized time within the gait cycle.

### 4.4 Individual Gait Distribution Estimation

For each person  $p$ , let  $\mathcal{G}_p = \{\tilde{G}_{p,1}, \tilde{G}_{p,2}, \dots, \tilde{G}_{p,n_p}\}$  represent normalized gait cycles. We estimate  $P_p(\tilde{G})$  using density modeling techniques selected based on dimensionality, sample size, and distributional characteristics. Approaches include parametric, non-parametric, or hybrid methods optimized for variability capture and computational tractability.

### 4.5 Normative Distribution Construction

For healthy controls  $\mathcal{H} = \{P_1, P_2, \dots, P_N\}$ , we construct the normative distribution  $P_{norm}(\tilde{G}|\mathbf{z})$ .

### 4.5.1 Distribution Aggregation

We evaluate:

- Gaussian mixture models:  $P_{norm} = \sum_{k=1}^K \pi_k \mathcal{N}(\mu_k, \Sigma_k)$
- Kernel density estimation with adaptive bandwidth
- Variational autoencoders for latent space modeling
- Weighted ensemble averaging

### 4.5.2 Demographic Conditioning

We will also explore conditioning on demographic variables  $\mathbf{z} = \{age, sex, height, weight\}$ . Methods include similarity-based weighting schemes and parametric conditioning approaches to ensure appropriate normative comparisons across different population characteristics.

## 4.6 Deviation Quantification

We quantify gait deviation through distributional distance measures comparing patient distributions to normative patterns. We evaluate information-theoretic, geometric, optimal transport, and statistical distance measures based on clinical sensitivity, computational efficiency, noise robustness, interpretability, and stability across populations.

## 4.7 Study Design and Participants

### 4.7.1 Phase 1: Normative Database Construction

- **Healthy controls:**  $N = 30$  participants
- **Age range:** 40-70 years, stratified by decade
- **Gender:** Balanced representation
- **Exclusion criteria:** No history of neurological, orthopedic, or cardiovascular conditions affecting gait

### 4.7.2 Phase 2: Clinical Validation

- **Stroke patients:**  $n = 15$ , recruited from rehabilitation centers
- **Parkinson’s patients:**  $n = 15$ , recruited from movement disorder clinics
- **Stimuloop participants:**  $n = 30$ , undergoing specific rehabilitation protocol
- **Assessment timeline:** Baseline, 4 weeks, 8 weeks, 12 weeks post-intervention

## 4.8 Validation

We assess construct validity (correlation with clinical measures), sensitivity (healthy vs. patient differences), responsiveness (intervention detection), and test-retest reliability. Additional validation includes synthetic data testing (Yamada et al., 2025), generative augmentation robustness (Zhang et al., 2025), and cross-modal consistency with generative model representations.

## 5 Expected Outcomes

1. Validated distributional gait assessment metric quantifying deviation from normative patterns
2. Normative database with demographic conditioning capabilities
3. Clinical validation demonstrating sensitivity to rehabilitation improvements
4. Enhanced objective tracking of rehabilitation progress

### 5.1 Primary Outcome

The **Gait Deviation Index (GDI)**:

$$GDI_{patient} = \min\{D(P_{patient}, P_{norm}(\cdot|\mathbf{z}_{patient})), D(P_{patient}, P_{norm})\} \quad (3)$$

where  $D$  is the optimal distance measure, quantifying likelihood of gait patterns under healthy population distribution.

## References

- Adeli, V., Ehsani, K., Rose, C., Kapoor, H., and Adeli, E. (2025). Gaitgen: Disentangled motion-pathology impaired gait generative model – bringing motion generation to the clinical domain. *arXiv preprint arXiv:2503.22397*.
- Baker, R. (2013). *Measuring Walking: A Handbook of Clinical Gait Analysis*. Mac Keith Press, London.
- Dobkin, B. H. (2005). Rehabilitation after stroke. *New England Journal of Medicine*, 352(16):1677–1684.
- Gholami, M., Rezaei, S., and Khodayari, A. (2019). A probabilistic approach to gait analysis using hidden markov models. *Medical Engineering & Physics*, 66:54–63.

- Ghosh, M., Dhara, B. C., and Mukherjee, J. (2025). Gait data generation using lightweight generative deep learning framework. *Journal of Biomechanics*, 192:112951.
- Hausdorff, J. M., Rios, D. A., and Edelberg, H. K. (2001). Gait variability and fall risk in community-living older adults: a 1-year prospective study. *Archives of Physical Medicine and Rehabilitation*, 82(8):1050–1056.
- Horst, F., Lapuschkin, S., Samek, W., Müller, K.-R., and Schöllhorn, W. I. (2019). Explaining the unique nature of individual gait patterns with deep learning. *Scientific Reports*, 9(1):1–13.
- Keus, S. H., Bloem, B. R., Hendriks, E. J., Bredero-Cohen, A. B., and Munneke, M. (2014). European physiotherapy guideline for parkinson’s disease. *KNGF/ParkinsonNet, The Netherlands*.
- Mehrholtz, J., Thomas, S., Kugler, J., Pohl, M., and Elsner, B. (2017). Treadmill training and body weight support for walking after stroke. *Cochrane Database of Systematic Reviews*, (8).
- Peppes, N., Daskalakis, E., Alexakis, T., Adamopoulou, E., and Demestichas, K. (2023). Foggan: Generating realistic parkinson’s disease freezing of gait data using gans. *Sensors*, 23(19):8158.
- Rezvani, A., Xu, R., Little, M. A., Pontil, M., and Ciccarelli, O. (2025). Diffusegait-net: Improving parkinson’s disease gait severity assessment with a diffusion model framework. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 33:2858–2869.
- Stihi, A., Kaski, S., and Solin, A. (2025). Hierarchical gaussian processes for characterizing gait variability in multiple sclerosis. *Data-Centric Engineering*, 6:e36.
- Takeishi, N. and Kalousis, A. (2021). Physics-integrated variational autoencoders for robust and interpretable generative modeling. In *Advances in Neural Information Processing Systems*, volume 34.
- Whittle, M. W. (2014). *Gait Analysis: An Introduction*. Butterworth-Heinemann, Oxford, 5th edition.
- Yamada, Y., Dembia, C. L., Ong, C. F., Uhlrich, S. D., Hicks, J. L., and Delp, S. L. (2025). Utility of synthetic musculoskeletal gaits for generalizable healthcare applications. *Nature Communications*, 16:6188.
- Zhang, W., Jiao, Y., Liu, X., Wang, S., and Li, Q. (2025). Real-time forecasting of pathological gait via imu navigation: a few-shot and generative learning framework for wearable devices. *Discover Electronics*, 2:51.