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

Stimuloop Research Consortium

November 4, 2025

Abstract

This research proposal presents a novel data-driven gait assessment metric designed to quantify patient improvement following neurological rehabilitation, particularly for stroke and Parkinson's disease patients participating in the Stimuloop program. Our approach captures the holistic nature of human gait by learning probability distributions over complete gait cycles from healthy control subjects, then measuring deviation from these normative patterns using appropriate distributional distance metrics. The method processes 4-dimensional spatio-temporal graph data from motion capture systems, segments recordings into gait cycles, and constructs person-specific and population-level distributions. This purely data-driven approach provides a comprehensive assessment tool that captures the full complexity of gait patterns while remaining broadly applicable across different rehabilitation scenarios.

1 Introduction

Gait assessment is fundamental to evaluating rehabilitation outcomes in neurological patients, particularly those recovering from stroke or living with Parkinson's disease (**??**). Current clinical gait assessment methods typically rely on discrete parameters or subjective rating scales, which fail to capture the holistic, dynamic nature of human locomotion (**?**). This research proposes a paradigm shift toward a purely data-driven approach that models gait as probability distributions over complete movement cycles.

1.1 Motivation: The Need for Holistic Gait Assessment

Traditional gait analysis focuses on isolated kinematic and kinetic parameters such as stride length, cadence, or joint angles at specific time points (**?**). However, neurological conditions affect gait in complex, multidimensional ways that cannot be adequately captured by examining individual parameters in isolation. Rehabilitation programs like Stimuloop require sensitive metrics that can detect subtle improvements in overall movement quality and coordination.

1.2 The Distributional Perspective

Human gait exhibits natural variability even within healthy populations, and this variability contains important information about movement quality and control (**?**). By modeling gait as probability distributions over complete cycles, we can capture not only the central tendencies of movement patterns but also their inherent variability and the relationships between different phases of the gait cycle.

1.3 Research Gap

While machine learning approaches have been applied to gait analysis (**??**), most methods focus on classification or parameter extraction rather than comprehensive distributional modeling. There is a critical need for assessment tools that can quantify how much a patient's gait deviates from healthy patterns while being sensitive to improvements following rehabilitation interventions.

2 Research Objectives

The primary objectives of this research are:

1. To develop a data-driven gait assessment metric based on probability distributions over complete gait cycles that quantifies deviation from healthy normative patterns

2. To establish comprehensive normative gait distributions from a large dataset of healthy control subjects, with considerations for demographic conditioning (age, sex, height)

3. To validate the metric's sensitivity in detecting gait improvements in stroke and Parkinson's disease patients undergoing rehabilitation

4. To evaluate and compare various distributional distance measures for quantifying gait quality and deviation from normative patterns

5. To create a robust, broadly applicable assessment tool suitable for various rehabilitation scenarios beyond the Stimuloop program

3 Literature Review

3.1 Current Approaches to Gait Assessment

Traditional gait analysis relies heavily on discrete measurements such as temporal-spatial parameters (stride length, step width, cadence) and joint kinematics at specific gait events (**??**). While these methods provide valuable insights, they often fail to capture the coordinated, dynamic nature of human locomotion and may miss subtle changes that occur during rehabilitation.

3.2 Distributional and Probabilistic Approaches

Recent advances in probabilistic and generative modeling have helped to move gait analysis beyond traditional parameter-based methods toward comprehensive distributional representations. **?** applied hidden Markov models to gait data, demonstrating the value of probabilistic frameworks for capturing temporal dependencies in movement patterns.

The emergence of hierarchical probabilistic models has shown particular promise. **?** introduced Hierarchical Variational Sparse Heteroscedastic Gaussian Processes (HVSHGP) for multiple sclerosis gait analysis, treating entire gait cycle waveforms as functions to be learned rather than extracting discrete parameters. This approach successfully captured both group-level differences and individual variability patterns, revealing subtle deviations in swing phase dynamics and quantifying stride-to-stride heteroscedasticity—hallmarks of neurological impairment.

3.3 Generative Models for Gait Synthesis and Augmentation

The past five years have witnessed constant growth in AI-driven generative models for gait analysis, addressing critical data scarcity issues in clinical populations. **?** developed FoGGAN, a GAN-based framework specifically for generating realistic Parkinson's freezing-of-gait episodes, demonstrating that synthetic data augmentation can significantly improve downstream classification accuracy.

More sophisticated approaches have emerged for comprehensive gait generation. **?** introduced GAITGen, a conditional residual vector-quantized VAE combined with masked transformers, capable of generating full 3D gait cycles conditioned on Parkinson's disease severity levels. This work demonstrated the feasibility of disentangling normal motion dynamics from pathology-specific factors, enabling controlled generation of clinically meaningful gait patterns.

Recent work by **?** represents a paradigm shift toward physics-based generative approaches. By leveraging musculoskeletal simulations with varied anatomical

and locomotion parameters, they created vast synthetic gait datasets that achieved comparable performance to real data across multiple clinical tasks. Remarkably, models trained exclusively on synthetic data showed superior generalization when fine-tuned with limited real-world data, demonstrating the power of diverse synthetic training distributions.

Diffusion models have also entered the gait analysis domain. **?** applied denoising diffusion models to Parkinson's gait severity assessment, learning probabilistic generative representations that could synthesize novel gait variations conditioned on clinical features. This approach effectively enriched limited clinical datasets by generating synthetic samples that captured previously unseen movement variations.

3.4 Physics-Constrained Generative Models

A notable advancement has been the integration of physical constraints into generative models. **?** developed physics-integrated VAEs that embed simplified biomechanical models (e.g., inverted pendulum dynamics) into the latent space, ensuring physically realistic gait generation. **?** further validated this concept by demonstrating that synthetic gait cycles must satisfy Newtonian mechanics and biomechanical feasibility through musculoskeletal simulation testing.

These physics-constrained approaches address a fundamental limitation of purely data-driven generative models: ensuring that synthetic gait patterns remain biomechanically plausible and clinically interpretable.

3.5 From Generation to Assessment: The Distributional Paradigm

While recent generative models excel at creating synthetic gait data, our proposed approach leverages the distributional perspective for a complementary purpose: comprehensive assessment rather than generation. The success of generative models like GAITGen and physics-integrated VAEs demonstrates that probability distributions can effectively capture the complex, high-dimensional nature of gait patterns (**??**).

Our distributional assessment framework builds upon these advances by using similar probabilistic representations—but focusing on measuring deviation from normative patterns rather than generating new samples. This approach addresses several limitations of current generative models for clinical assessment:

- **Assessment Focus**: Rather than generating new gait patterns, we quantify how much a patient's gait distribution deviates from healthy norms

- **Interpretability**: Distance metrics provide direct, quantitative measures of gait abnormality rather than requiring interpretation of latent space representations

4

- **Robustness**: By modeling complete distributions rather than point estimates, our approach captures gait variability and uncertainty inherent in neurological conditions

- **Clinical Relevance**: The framework directly addresses rehabilitation monitoring needs, providing sensitive measures of improvement over time

The work by **?** particularly supports our approach, demonstrating that generative augmentation of pathological gait patterns leads to more robust assessment algorithms. Their TimeGAN-based augmentation achieved 65% improvement in stride length estimation accuracy, suggesting that distributional modeling of gait variability enhances clinical measurement precision.

Furthermore, the success of **?** in achieving superior performance through synthetic data diversity aligns with our hypothesis that comprehensive distributional representations—capturing the full range of healthy gait variability—will provide more sensitive assessment tools than traditional parameter-based methods.

3.6 Potential Integration of Generative Models

While our primary focus is on distributional assessment, the recent advances in generative gait modeling present opportunities for future enhancement of our framework. Generative models could potentially address several challenges in our approach:

- **Data Augmentation**: Following **?** and **?**, we could use physics-constrained generative models to augment sparse regions of our normative distribution, particularly for underrepresented demographic groups

- **Counterfactual Analysis**: Leveraging approaches like GAITGen (**?**), we could generate "what-if" scenarios to better understand how specific pathological features contribute to distributional distances

- **Personalized Baselines**: Generative models could help create individualized normative baselines by conditioning on patient-specific characteristics beyond basic demographics

However, our core distributional assessment approach remains independent of generative modeling, ensuring robust clinical applicability without dependence on complex generative architectures.

3.7 Distance Measures for Distributional Comparison

The selection of appropriate distance metrics for comparing probability distributions is a critical methodological decision that will significantly impact the sensitivity and interpretability of our gait assessment approach. Various families of

distance measures offer different theoretical properties and practical advantages for capturing deviations from normative patterns.

Information-theoretic measures, such as divergences based on entropy and mutual information, provide principled frameworks for quantifying distributional differences. Geometric measures that respect the underlying structure of the data space may be particularly suitable for gait analysis, where temporal and spatial relationships are meaningful. Additionally, optimal transport-based distances can capture the effort required to transform one distribution into another, potentially offering intuitive interpretations of rehabilitation progress.

The choice of distance measure will be determined through empirical evaluation, considering factors such as computational efficiency, sensitivity to clinically relevant changes, robustness to measurement noise, and interpretability for clinical practitioners. This data-driven selection process will ensure that our framework utilizes the most appropriate metric for the specific characteristics of gait data and clinical assessment requirements.

3.8   Neurological Gait Impairments

Stroke and Parkinson's disease affect gait in characteristic ways. Post-stroke gait is often characterized by asymmetry, reduced walking speed, and compensatory movement patterns (**?**). Parkinson's disease typically presents with reduced step length, increased step-to-step variability, and freezing episodes (**?**). Traditional metrics often capture these changes as isolated parameter deficits, missing the complex interactions between different aspects of the movement pattern.

3.9   Rehabilitation Assessment Needs

Effective rehabilitation requires sensitive outcome measures that can detect clinically meaningful changes (**?**). Current assessment tools often lack the sensitivity to capture gradual improvements or may be influenced by compensatory strategies that mask underlying recovery. A distributional approach could provide more comprehensive assessment by capturing improvements in overall movement coordination and quality.

4   Methodology

4.1   Overview of the Distributional Gait Assessment Framework

Our approach models gait as probability distributions over complete gait cycles, enabling holistic assessment of movement patterns. The framework consists of three

main components: (1) normative distribution learning from healthy controls, (2) patient-specific distribution estimation, and (3) distributional distance computation for deviation quantification.

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:

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

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:

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

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

4.4 Individual Gait Distribution Estimation

For each person p, we estimate their characteristic gait distribution from multiple recorded cycles. Let $\mathcal{G}_p = \{\tilde{G}_{p,1}, \tilde{G}_{p,2}, \ldots, \tilde{G}_{p,n_p}\}$ represent the set of normalized gait cycles for person p.

The individual gait distribution $P_p(\tilde{G})$ will be estimated using appropriate density modeling techniques. The choice of density estimation method will be determined based on the dimensionality of the data, sample size, and distributional characteristics observed in the gait cycles. Potential approaches include parametric models (if suitable distributional assumptions can be validated), non-parametric methods, or hybrid approaches that combine multiple modeling strategies.

The density estimation approach will be selected to optimally capture the underlying variability structure in individual gait patterns while maintaining computational tractability and statistical robustness for the available sample sizes.

4.5 Normative Distribution Construction

The normative distribution is constructed from healthy control subjects. We consider two approaches:

4.5.1 Unconditional Normative Distribution

$$P_{norm}(\tilde{G}) = \frac{1}{N} \sum_{p=1}^{N} P_p(\tilde{G}) \tag{3}$$

where N is the number of healthy control subjects.

4.5.2 Conditional Normative Distribution

For demographic conditioning on variables $\mathbf{z} = (age, sex, height)^T$:

$$P_{norm}(\tilde{G}|\mathbf{z}) = \sum_{p=1}^{N} w_p(\mathbf{z}) P_p(\tilde{G}) \tag{4}$$

where weights $w_p(\mathbf{z})$ will be determined using appropriate statistical methods for incorporating demographic covariates. The specific weighting approach will be selected based on the distributional characteristics of the demographic variables and their relationship to gait patterns, ensuring robust conditioning while maintaining statistical validity.

4.6 Deviation Quantification Using Distributional Distances

The core innovation of our approach lies in quantifying gait deviation through distributional distance measures. Rather than comparing discrete parameters, we will assess how much a patient's complete gait distribution differs from the normative distribution.

We will systematically evaluate multiple classes of distance measures to determine the most clinically relevant and mathematically appropriate metric for gait assessment:

- **Information-theoretic measures**: Divergences that quantify information loss or gain when comparing distributions

- **Geometric distances**: Metrics that respect the underlying geometry and structure of the gait space

- **Optimal transport distances**: Measures that quantify the minimal effort required to transform one distribution into another

- **Statistical measures**: Classical statistical distance measures adapted for high-dimensional distributional comparison

The distance measure $D(P_{patient}, P_{norm})$ will be selected based on empirical evaluation of:

1. Sensitivity to clinically meaningful changes

2. Computational efficiency for real-time assessment

3. Robustness to measurement noise and sample size variations

4. Interpretability for clinical practitioners

5. Statistical stability across different patient populations

This comprehensive evaluation approach ensures that our final metric optimally captures the likelihood of observed gait patterns under the healthy population distribution, providing a principled foundation for quantifying deviation from normative gait.

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 Statistical Analysis and Validation

4.8.1 Metric Validation

- **Construct validity**: Correlation with established clinical measures

- **Sensitivity**: Ability to detect known group differences (healthy vs. patient)

- **Responsiveness**: Ability to detect change following intervention

- **Test-retest reliability**: Stability across repeated measurements

4.8.2 Enhanced Validation Using Generative Models

Drawing inspiration from recent advances in gait synthesis, we will incorporate additional validation approaches:

- **Synthetic Validation**: Following **?**, we will validate our distributional distance measures using physics-based synthetic gait data with known deviations from normal patterns

- **Generative Augmentation Testing**: Using approaches similar to **?**, we will test whether our framework remains robust when normative distributions are augmented with synthetic healthy gait cycles

- **Cross-Modal Validation**: We will compare our distributional assessments with state-of-the-art generative model latent representations (e.g., GAITGen embeddings) to ensure consistency across different mathematical frameworks

5 Expected Outcomes

This research is expected to produce:

1. **Novel Assessment Metric**: A validated data-driven gait assessment tool that quantifies deviation from normative patterns using distributional distances

2. **Normative Database**: Comprehensive probability distributions representing healthy gait patterns, with demographic conditioning capabilities

3. **Clinical Validation**: Demonstrated sensitivity to gait improvements in stroke and Parkinson's patients, with established minimal

4. **Clinical Impact**: Enhanced ability to track rehabilitation progress objectively with sensitivity improvements demonstrated through comparison with traditional methods and recent AI-driven approaches

5.1 Primary Outcome Measure

The primary outcome will be the **Gait Deviation Index (GDI)**, defined as:

$$GDI_{patient} = \min\{D(P_{patient}, P_{norm}(\cdot|\mathbf{z}_{patient})), D(P_{patient}, P_{norm})\} \qquad (5)$$

where D represents the optimal distributional distance measure selected through our empirical evaluation process, providing a single numeric score representing overall gait quality. This score quantifies the likelihood of the assessed gait patterns under the healthy population distribution, with higher values indicating greater deviation from normative patterns.

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*.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. pages 214–223.

Baker, R. (2013). *Measuring Walking: A Handbook of Clinical Gait Analysis*. Mac Keith Press, London.

Begg, R. K. and Kamruzzaman, J. (2005). A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics*, 38(3):401–408.

Cover, T. M. and Thomas, J. A. (2006). Elements of information theory.

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*.

Mehrholz, 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). Diffusegaitnet: 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.

Villani, C. (2003). Topics in optimal transportation. *Graduate Studies in Mathematics*, 58.

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.