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 distributional distance metrics such as KL-divergence or Wasserstein distance. 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.

Contents

1 Introduction

Gait assessment is fundamental to evaluating rehabilitation outcomes in neurological patients, particularly those recovering from stroke or living with Parkinson's disease (Dobkin, 2005; Keus et al., 2014). 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 (Baker, 2013). 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 (Whittle, 2014). 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 (Hausdorff et al., 2001). 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 (Horst et al., 2019; Begg and Kamruzzaman, 2005), 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 demonstrate the clinical utility of distributional distance measures (KL-divergence, Wasserstein distance) for quantifying gait quality

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 (Baker, 2013; Whittle, 2014). 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 revolutionized gait analysis, moving beyond traditional parameter-based methods toward comprehensive distributional representations. Gholami et al. (2019) 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. Stihi et al. (2025) 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 explosive growth in AI-driven generative models for gait analysis, addressing critical data scarcity issues in clinical populations. Peppes et al. (2023) 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. Adeli et al. (2025) 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 Yamada et al. (2025) 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. Rezvani et al. (2025) 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 critical advancement has been the integration of physical constraints into generative models. Takeishi and Kalousis (2021) developed physics-integrated VAEs that embed simplified biomechanical models (e.g., inverted pendulum dynamics) into the latent space, ensuring physically realistic gait generation. Ghosh et al. (2025) 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 (Adeli et al., 2025; Takeishi and Kalousis, 2021).

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
- **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 Zhang et al. (2025) 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 Yamada et al. (2025) 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 Ghosh et al. (2025) and Zhang et al. (2025), 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 (Adeli et al., 2025), 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 Biological Data

The choice of distance metric for comparing probability distributions is crucial for our approach. Kullback-Leibler (KL) divergence has been widely used in information theory and machine learning to measure the difference between probability distributions (Cover and Thomas, 2006). However, KL-divergence is asymmetric and can be undefined when distributions have non-overlapping support.

The Wasserstein distance, also known as the Earth Mover's Distance, provides a more robust alternative (Villani, 2003). It measures the minimum cost of transforming one distribution into another and has gained popularity in machine learning applications (Arjovsky et al., 2017). The Wasserstein distance is particularly well-suited for comparing distributions over structured spaces like gait cycles, as it naturally accounts for the geometric structure of the underlying space.

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 (Dobkin, 2005). Parkinson's disease typically presents with reduced step length, increased step-to-step variability, and freezing episodes (Keus et al., 2014). 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 (Mehrholz et al., 2017). 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. We model the individual gait distribution as:

$$P_p(\tilde{G}) = \frac{1}{n_p} \sum_{i=1}^{n_p} K_h(\tilde{G} - \tilde{G}_{p,i}) \tag{3}$$

where K_h is a kernel density estimator with bandwidth h, chosen through cross-validation.

4.5 Normative Distribution Construction

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

4.5.1 Unconditional Normative Distribution

$$P_{norm}(\tilde{G}) = \frac{1}{N} \sum_{p=1}^{N} P_p(\tilde{G}) \tag{4}$$

where N is the number of healthy control subjects.

4.5.2 Conditional Normative Distribution

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

$$P_{norm}(\tilde{G}|\mathbf{z}) = \sum_{p=1}^{N} w_p(\mathbf{z}) P_p(\tilde{G}) \tag{5}$$

where weights $w_p(\mathbf{z})$ are determined by kernel regression:

$$w_p(\mathbf{z}) = \frac{K_\sigma(\mathbf{z} - \mathbf{z}_p)}{\sum_{q=1}^{N} K_\sigma(\mathbf{z} - \mathbf{z}_q)} \tag{6}$$

4.6 Deviation Quantification Using Distributional Distances

We evaluate two primary distance measures for quantifying gait deviation:

4.6.1 Kullback-Leibler Divergence

$$D_{KL}(P_{patient} \| P_{norm}) = \int P_{patient}(\tilde{G}) \log \frac{P_{patient}(\tilde{G})}{P_{norm}(\tilde{G})} d\tilde{G} \tag{7}$$

4.6.2 Wasserstein Distance

$$W_2(P_{patient}, P_{norm}) = \left(\inf_{\gamma \in \Gamma(P_{patient}, P_{norm})} \int \|\tilde{G}_1 - \tilde{G}_2\|^2 d\gamma(\tilde{G}_1, \tilde{G}_2) \right)^{1/2} \tag{8}$$

where $\Gamma(P_{patient}, P_{norm})$ is the set of all joint distributions with marginals $P_{patient}$ and P_{norm}.

4.7 Study Design and Participants

4.7.1 Phase 1: Normative Database Construction

- **Healthy controls**: $N = 200$ participants
- **Age range**: 20-80 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**: $n = 100$, recruited from rehabilitation centers
- **Parkinson's patients**: $n = 100$, recruited from movement disorder clinics
- **Stimuloop participants**: $n = 50$, undergoing specific rehabilitation protocol
- **Assessment timeline**: Baseline, 4 weeks, 8 weeks, 12 weeks post-intervention

4.8 Data Collection Protocol

All participants undergo standardized motion capture assessment:

- **Equipment**: Optical motion capture system with minimum 16 cameras
- **Marker set**: Full-body model with 39 anatomical landmarks
- **Walking trials**: 6 trials of comfortable-speed walking over 10-meter walkway
- **Sampling rate**: 120 Hz minimum
- **Additional measures**: Clinical gait assessments (10-meter walk test, Timed Up and Go) for validation

4.9 Statistical Analysis and Validation

4.9.1 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

4.9.2 Enhanced Validation Using Generative Models

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

- **Synthetic Validation**: Following Yamada et al. (2025), 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 Zhang et al. (2025), 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

4.9.3 Computational Implementation

The framework will be implemented using:

- **Python/NumPy**: Core numerical computations
- **Scikit-learn**: Kernel density estimation and machine learning utilities
- **POT (Python Optimal Transport)**: Wasserstein distance computation
- **GPU acceleration**: For large-scale distance computations

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 detectable change thresholds

4. **Methodological Advancement**: Comprehensive comparison of KL-divergence vs. Wasserstein distance for gait assessment, informed by recent advances in generative modeling, with evidence-based recommendations for optimal distance metrics

5. **Integration Framework**: Guidelines for incorporating generative model insights (data augmentation, physics constraints) into distributional assessment, bridging the gap between generative AI and clinical evaluation

6. **Software Tools**: Open-source implementation with modular design enabling integration with existing generative gait models and broad adoption across rehabilitation centers

7. **Benchmark Dataset**: Validated distributional assessment results on established datasets, enabling comparison with generative modeling approaches from recent literature

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

9. **Scientific Publications**: High-impact publications in rehabilitation medicine, biomechanics, medical informatics, and machine learning journals, contributing to both clinical and methodological literature

5.1 Primary Outcome Measure

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

$$GDI_{patient} = \min\{D(P_{patient}, P_{norm}(\cdot|\mathbf{z}_{patient})), D(P_{patient}, P_{norm})\} \tag{9}$$

where D represents either KL-divergence or Wasserstein distance, providing a single numeric score representing overall gait quality.

5.2 Competitive Advantages Over Generative Approaches

While recent generative models like GAITGen and DiffuseGaitNet excel at synthesis, our distributional assessment approach offers distinct advantages for clinical evaluation:

- **Direct Interpretability**: Distance metrics provide immediately interpretable measures of abnormality, unlike latent space representations that require additional interpretation
- **Computational Efficiency**: No need for complex generative architectures during assessment—once normative distributions are learned, patient evaluation is computationally lightweight
- **Robustness**: Less susceptible to mode collapse or generation artifacts that can affect generative models
- **Clinical Validation**: Direct correspondence between distance metrics and clinical severity measures, facilitating validation and adoption
- **Regulatory Compliance**: Simpler mathematical framework may facilitate regulatory approval for clinical use compared to complex neural architectures

6 Timeline

7 Budget

8 Ethical Considerations

This research involves human subjects and will require institutional review board (IRB) approval. Key ethical considerations include:

- **Informed Consent**: All participants will provide written informed consent, with particular attention to explaining the research purpose and data usage to patients with potential cognitive impairments
- **Data Privacy**: Motion capture data will be de-identified and stored securely according to HIPAA requirements. Demographic conditioning variables will be anonymized
- **Patient Burden**: Data collection sessions will be designed to minimize fatigue and discomfort for neurological patients

Phase	Activity	Duration
1	Algorithm Development & Implementation - Gait cycle segmentation algorithms - Distribution estimation methods - Distance metric computations	Months 1-4
2	Normative Database Construction - Healthy control recruitment (n=200) - Motion capture data collection - Distribution learning and validation	Months 5-12
3	Clinical Validation Study - Patient recruitment and baseline assessment - Longitudinal data collection - Stimuloop program integration	Months 13-24
4	Analysis & Validation - Statistical analysis and metric validation - Clinical correlation studies - Sensitivity and responsiveness testing	Months 25-30
5	Dissemination & Implementation - Software tool development - Publication preparation - Clinical implementation guidelines	Months 31-36

Table 1: Project Timeline

- **Data Sharing**: Anonymized normative datasets will be made available to the research community to promote broader scientific advancement
- **Clinical Translation**: Results will be shared with participants and their clinical teams when appropriate and beneficial for their care

9 Significance and Impact

9.1 Scientific Significance

This research represents a paradigm shift from traditional parameter-based gait analysis to distributional modeling of complete movement patterns. The approach addresses fundamental limitations in current assessment methods by:

- Capturing the holistic nature of gait rather than isolated parameters
- Providing a principled mathematical framework for quantifying deviation from normal patterns
- Enabling personalized assessment through demographic conditioning
- Establishing a foundation for data-driven rehabilitation monitoring

Budget Category	**Amount (USD)**
Personnel (3 years)	
- Principal Investigator (25% effort)	\$150,000
- Postdoctoral Researcher (100% effort)	\$180,000
- Graduate Student (50% effort)	\$90,000
- Research Coordinator (50% effort)	\$120,000
Equipment	
- Motion capture system upgrades	\$50,000
- Computing hardware (GPU cluster)	\$30,000
Data Collection	
- Participant compensation	\$25,000
- Clinical site fees	\$40,000
Other Direct Costs	
- Software licenses	\$15,000
- Travel and dissemination	\$20,000
- Publication fees	\$10,000
Total Direct Costs	\$730,000
Indirect Costs (25%)	\$182,500
Total Project Cost	\$912,500

Table 2: Budget Breakdown

9.2 Clinical Impact

The developed metric will provide clinicians with:

- **Objective Assessment**: Quantitative measures of gait quality that complement subjective clinical evaluations
- **Sensitive Monitoring**: Ability to detect subtle improvements that may be missed by traditional assessments
- **Treatment Planning**: Data-driven insights to guide rehabilitation interventions
- **Outcome Prediction**: Potential for predicting rehabilitation success based on baseline gait patterns

9.3 Broader Applications

While developed for stroke and Parkinson's disease rehabilitation, the framework's data-driven nature makes it broadly applicable to:

- Other neurological conditions (multiple sclerosis, cerebral palsy)
- Orthopedic rehabilitation (joint replacement, injury recovery)
- Aging and fall prevention studies

- Sports performance and injury prevention
- Pharmaceutical clinical trials requiring gait outcome measures

10 Conclusion

This research proposal presents a novel data-driven approach to gait assessment that captures the holistic nature of human locomotion through distributional modeling. By learning normative patterns from healthy controls and quantifying patient deviations using principled distance metrics, our approach addresses fundamental limitations in current gait assessment methods.

The proposed Gait Deviation Index provides a comprehensive, sensitive measure of movement quality that is particularly suited for monitoring rehabilitation progress in neurological patients. The framework's mathematical foundation ensures reproducibility and interpretability, while its data-driven nature enables broad applicability across different clinical populations and rehabilitation programs.

This work represents a significant advancement in personalized rehabilitation medicine, providing clinicians with powerful tools for objective assessment and treatment monitoring. The expected outcomes will not only benefit patients in the Stimuloop program but will establish a new paradigm for gait analysis that can be adopted across the broader rehabilitation community.

The integration of advanced mathematical concepts from optimal transport theory and information theory with clinical rehabilitation needs exemplifies the potential of interdisciplinary research to address real-world healthcare challenges. Upon completion, this project will deliver both theoretical contributions to the field of biomechanics and practical tools that directly improve patient care.

11 Appendix: Algorithm Summary

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.

Algorithm 1 Data-Driven Gait Assessment Algorithm

Phase 1: Normative Database Construction
for each healthy control subject $p = 1, \ldots, N$ **do**
  Collect gait cycles $\mathcal{G}_p = \{\tilde{G}_{p,1}, \ldots, \tilde{G}_{p,n_p}\}$
  Estimate individual distribution $P_p(\tilde{G})$ using kernel density estimation
end for
Learn normative distribution $P_{norm}(\tilde{G}|\mathbf{z})$ with demographic conditioning
Phase 2: Patient Assessment
for each patient **do**
  Collect gait cycles $\mathcal{G}_{patient}$
  Estimate patient distribution $P_{patient}(\tilde{G})$
  Compute deviation measure:
  $GDI = \min\{D(P_{patient}, P_{norm}(\cdot|\mathbf{z}_{patient})), D(P_{patient}, P_{norm})\}$
  where $D \in \{D_{KL}, W_2\}$ (KL-divergence or Wasserstein distance)
end for

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). 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.

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.