GAITDIFFUSION: DIFFUSION-BASED GAIT GENERATION CONDITIONED ON PARKINSON'S SEVERITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Generating patient-specific gait patterns is essential for digital twinning and clinical assessment, data augmentation due to the scarcity of patient-specific data, and assistive robotics where gait patterns must adapt to individual pathology and therapeutic goals while preserving biomechanical realism. However, learning to generate full-body gait at clinical scale remains challenging. Here we present GaitDiffusion, a two-stage latent diffusion framework that generates full-body gait sequences conditioned on a clinical severity level scale, i.e., the Unified Parkinson's Disease (PD) Rating Scale (UPDRS). PD is a suitable testbed for pathology-conditioned generation as gait patterns degrade with clinical severity, a progression that hand-crafted healthy templates cannot represent. GaitDiffusion includes a 1D convolutional VAE with surrogate UPDRS classification and class-prototype guidance that compresses sliding-window motion into a compact latent space with embedded clinical structure, and a Diffusion Transformer (DiT) with adaptive layer normalization and classifier-free guidance, which denoises class-conditioned latents. A range-of-motion (ROM) supervision loss enforces kinematic amplitude directly on the decoded output. Trained on the multi-site CARE-PD dataset, GaitDiffusion reproduces the clinical progression of bradykinesia: mean sagittal-leg ROM decreases monotonically with severity, matching real gait with a ROM error of $\pm 5.8\%$ on average, while achieving a mean dynamic time warping of 0.585 across all four classes and an overall latent-space Fréchet inception distance of 6.28. A classifier trained on real gait features assigns the generated gait its intended UPDRS severity in 84.4% of cases, supporting use of the generated trajectories as severity-appropriate reference patterns for exoskeleton control and as synthetic training data for downstream PD gait analysis. Compared with the strongest prior pathology-conditioned generator, GaitDiffusion reduces the Average Variance Error and the Absolute Stooped Posture Mean Difference by 97.9% and 8.3%, respectively, while maintaining a competitive Absolute Arm Swing Mean Difference. Moreover, its continuous latent space enables differentiable biomechanical supervision. Overall, GaitDiffusion shows stable full-body pathology-conditioned gait generation using an ambulatory medical scale.

1 INTRODUCTION

Generating patient-specific gait patterns is essential for many applications, such as digital twinning in precision medicine and clinical assessment (Simonetti et al., 2024), data augmentation for training AI models due to the usual small amount of patient data (Frid-Adar et al., 2018), and personalized control of assistive robotic devices (Zhang et al., 2017). Particularly, in lower-limb exoskeletons, generating gait reference trajectories that match a patient's specific movement is essential for effective therapy and to provide safe and effective assistance (Vallery et al., 2008). Furthermore, pathology-conditioned gait generators allow us to deploy data-driven learning methods, e.g., reinforcement learning, in wearable robotics (Chavarrías et al., 2025).

However, a major bottleneck for current data-driven approaches is the scarcity of clinical gait data. For instance, in Parkinson's disease (PD), gait datasets are small, annotations require expert clinicians, and severity classes are inherently imbalanced, with severely impaired patients underrepresented in research datasets. Collecting multi-site motion capture with expert annotations is costly,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

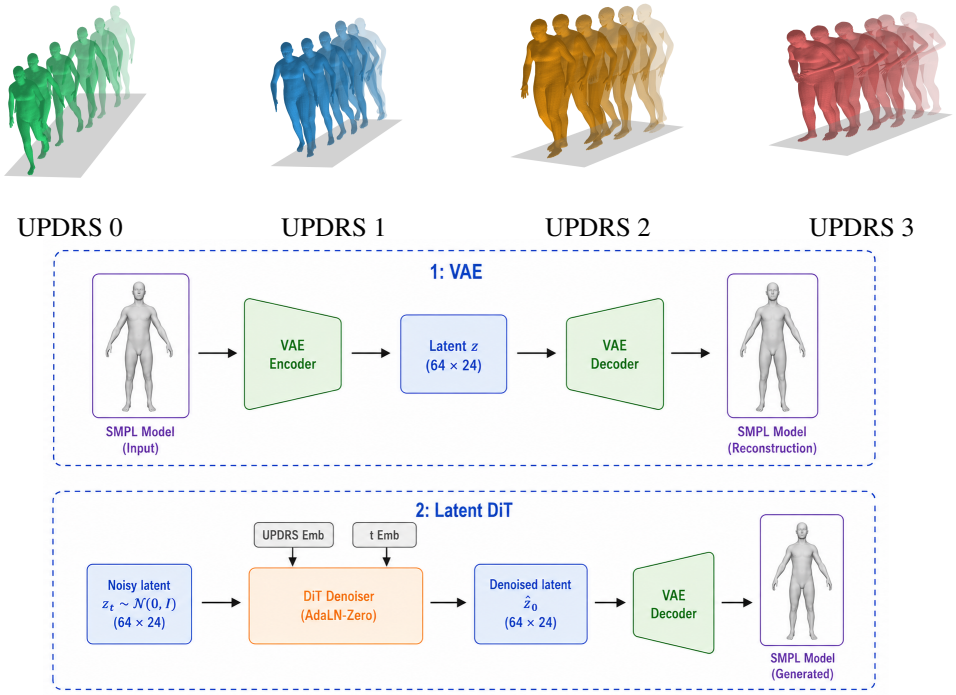

Figure 1: **Overview of GaitDiffusion.** *Top*: Generated gait sequences for each UPDRS severity with GaitDiffusion (0 = Normal, 1 = Slight, 2 = Mild, 3 = Moderate) *Bottom*: the two-stage architecture. Stage 1) a 1D convolutional VAE that compresses a 263-dimensional full-body motion representation into a clinically structured continuous latent. Stage 2) trains a Diffusion Transformer (DiT) with AdaLN-Zero conditioning and classifier-free guidance (CFG) to denoise UPDRS-conditioned latents, with an auxiliary range-of-motion loss on decoded output.

time-intensive, and especially burdensome for clinical populations. This makes it infeasible to collect the large volume of labeled data that data-driven personalization demands.

Generative models can reduce this data overhead. Instead of collecting additional clinical motion capture data, generative modeling could learn the underlying pathology gait true distribution and synthesize plausible gait patterns conditioned, for instance, on clinical severity, augmenting training data for downstream tasks. Prior motion generation work has explored generative adversarial, variational, and transformer-based models for producing realistic and temporally coherent motion (Barsoum et al., 2018; Petrovich et al., 2021; Guo et al., 2020). Recently, diffusion models have emerged as a strong approach for motion synthesis, providing stable training, high sample quality, and flexible conditioning (Ho et al., 2020; Tevet et al., 2022; Zhang et al., 2024).

Parkinson's disease (PD), where gait impairment is a hallmark, presents a suitable testbed for pathology-conditioned full-body gait generation. PD gait shows high variability within subjects and across pathology stages, with complex dynamics, making it very challenging to learn a generalizable representation. PD gait progresses from subtle asymmetry in early stages to severe bradykinesia (slowness and reduced movement amplitude), diminished arm swing, and a stooped posture as the disease advances (Mirelman et al., 2019). The Unified Parkinson's Disease Rating Scale (UPDRS) gait subscore provides a standardized clinical rating of this progression on a 0 to 4 scale (0 = Normal, 1 = Slight, 2 = Mild, 3 = Moderate, 4 = Severe) where level 4 corresponds to patients typically unable to walk without substantial assistance (Goetz et al., 2008)

1.1 Related Work and Contribution

Clinical Gait Analysis and Generation. Traditional machine learning approaches to PD gait severity estimation extract engineered spatiotemporal features from instrumented walkways or mo-

tion capture and train classifiers to predict UPDRS levels (Klucken et al., 2013). While effective in controlled settings, these methods are limited by sensor availability and the difficulty of collecting labeled clinical gait data. PD cohorts are small, UPDRS annotations require expert clinicians, and severity classes are inherently imbalanced; severely impaired patients are underrepresented in research datasets. Generative modeling has emerged as a promising direction to address this scarcity by augmenting training data with synthetic gait samples that preserve clinically meaningful features. Recent datasets such as CARE-PD (Adeli et al., 2025a) have introduced multi-site SMPL-based 3D gait recordings with UPDRS annotations, opening the door to data-driven generative modeling.

Generative approaches to clinical gait have recently emerged. GAITGen (Adeli et al., 2025b) generated SMPL-based gait sequences conditioned on clinical impairment severity with a masked-token generator, demonstrating that synthetic data improves downstream PD severity estimation. Other recent works include DiffuseGaitNet (Rezvani et al., 2025), which used a diffusion model to generate clinical-feature-conditioned PD gait video frames that improve downstream severity prediction; Wan & Zhao (2026), which proposed pathology disentanglement for impaired gait analysis; Dinesh et al. (2026), which conditioned a diffusion model on subject-specific parameters to generate personalized walking patterns across speeds; and Chandrasekaran et al. (2026), who combined LLM-based semantic augmentation with a pathology-preserving VQ tokenizer to synthesize gait sequences for downstream classification. Bamani et al. (2026) extended clinical gait generation beyond Parkinson's disease, proposing a conditional skeleton diffusion model for GMFCS-conditioned gait synthesis in cerebral palsy assessment, while Benasco et al. (2026) conditioned a diffusion model on gait parameters such as step length to generate lower-limb trajectories.

Motion Representation with VAEs. These clinical generators inherit their generative design from the broader human-motion literature, where compressing high-dimensional motion into a compact latent space has become standard practice for efficient generation. ACTOR (Petrovich et al., 2021) introduced a VAE for action-conditioned motion synthesis, while MoMask (Guo et al., 2023) combined a residual VQ-VAE with masked transformers to generate diverse motion sequences from text descriptions. Motion Latent Diffusion (MLD) (Chen et al., 2023) demonstrated that applying diffusion in a learned VAE latent space reduces computational cost compared to raw-motion diffusion while improving generation quality. However, none of these methods incorporate clinical supervision into the latent representation. MAD-VAE (Morlock & Wang, 2020) introduced class-prototype losses for supervised representation learning, but this approach has not been explored for embedding clinical structure into motion latents.

Diffusion Generative Models. Diffusion models have since become a leading approach for motion synthesis, whether applied directly to motion or within such learned latent spaces. Denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020) generate data by learning to reverse a fixed Markovian noising process, progressively denoising Gaussian noise into structured samples through iterative refinement. In human motion, the Motion Diffusion Model (MDM) (Tevet et al., 2022) established state-of-the-art text-to-motion and action-to-motion generation using a transformer backbone with $\mathbf{x}_0$-prediction. Predicting the clean sample rather than the noise at each step enables direct supervision through auxiliary geometric losses on predicted motion, though these losses have yet to be extended to clinically meaningful biomechanical features. The Diffusion Transformer (DiT) (Peebles & Xie, 2023) replaced the conventional U-Net denoiser with transformer blocks incorporating adaptive layer normalization (AdaLN-Zero), where conditioning signals modulate scale, shift, and gate parameters at every layer. Classifier-free guidance (CFG) (Ho & Salimans, 2022) provides a mechanism to control the trade-off between sample diversity and conditioning fidelity by blending conditional and unconditional predictions during sampling. Conditioning on continuous clinical severity scores through this mechanism, however, has not been studied in the clinical gait setting.

Among diffusion-based clinical gait generators, Rezvani et al. (2025) generates video frames to improve severity assessment rather than to produce full-body motion, Dinesh et al. (2026) and Benasco et al. (2026) address able-bodied speed/step-length adaptation with lower-limb sagittal angles only and Tan et al. (2025) predicts biomechanical kinematics and kinetics without severity conditioning. None of these works tackles full-body pathology-conditioned gait generation across an ambulatory medical scale for severity, such as the UPDRS 0–3.

Closest Work and Contribution. The closest PD-relevant generative model to this work is GAITGen (Adeli et al., 2025b), which generates gait sequences conditioned on PD impairment severity, showing that synthetic data improves downstream severity estimation. The key architectural difference from our method is the generative backbone. GAITGen is a two-stage masked-token generator: a Mask Transformer predicts base residual-VQ tokens and a Residual Transformer refines the remaining levels (Guo et al., 2023; Chang et al., 2022), so generation reduces to selecting discrete codebook indices. Our proposed GaitDiffusion instead uses a diffusion transformer that predicts the clean latent $\mathbf{z}_0$ at every denoising step (Ho et al., 2020; Tevet et al., 2022).

This yields two decisive advantages. First, differentiability: a continuous $\mathbf{z}_0$ prediction lets biomechanical losses such as range-of-motion (ROM) be applied directly to the decoded output and back-propagated into the denoiser, whereas GAITGen's nearest-neighbor token selection is non-differentiable and blocks any such gradient. Second, conditioning control: the UPDRS class modulates the AdaLN-Zero blocks and is amplified at inference through the classifier-free guidance scale (Peebles & Xie, 2023; Ho & Salimans, 2022), rather than committing to a discrete pathology token. The continuous-versus-discrete bottleneck is itself a trade-off. A discrete codebook discards the quantization residual and, under hard nearest-neighbor assignment, can leave codes unused (codebook collapse) (van den Oord et al., 2018; Razavi et al., 2019; Esser et al., 2021), which matters when adjacent UPDRS levels differ by small amplitudes, but it provides a token space compatible with motion transformers pretrained on large able-bodied datasets (Guo et al., 2023; 2022), including hierarchical residual codes that capture local detail. A continuous latent avoids these failures and preserves fine amplitude structure, but risks posterior collapse, where the KL term smooths away subtle pathology (Hoffman et al., 2017; Alemi et al., 2018).

This leads to the main gap addressed in this paper: generating clinically plausible gait across an ambulatory medical scale (UPDRS severity levels) with a continuous latent diffusion architecture that embeds clinical structure directly into the latent space. Here, we hypothesize that a VAE with surrogate UPDRS classification and class-prototype guidance, combined with a Diffusion Transformer conditioned on clinical severity, can synthesize biomechanically plausible gait patterns across UPDRS 0–3 levels. Several design questions remain open: whether clinical structure should be embedded directly in the latent space, what biomechanical features should guide generation through auxiliary losses, and how continuous diffusion architectures compare against discrete alternatives for clinical conditioning. We address this gap with a two-stage latent diffusion framework. Specifically, this work presents the following contributions: 1) a 1D convolutional VAE with surrogate UPDRS classification and class-prototype loss that embeds the clinical severity ordering into a continuous latent space; 2) a Diffusion Transformer with adaptive layer normalization and classifier-free guidance that provides class separability and amplitude accuracy; and 3) a lightweight range-of-motion supervision loss on the decoded output that corrects ROM magnitude while preserving the VAE's learned ordering. Whether continuous, biomechanically-supervised latents outperform discrete pathology tokens for clinical gait generation is investigated in this work.

2 Materials and Methods

Given a gait sequence $\mathbf{x} \in \mathbb{R}^{T \times D}$ of length T and dimensionality D with clinical severity label c_p (e.g., UPDRS-gait levels $c_p \in \{0, 1, 2, 3\}$), we aim to learn a generative model that produces synthetic gait sequences conditioned on the clinical severity. The proposed framework consists of two stages: (1) a 1D convolutional variational autoencoder (VAE) that compresses the high-dimensional gait sequences into a low-dimensional latent representation $\mathbf{z} \in \mathbb{R}^{64 \times 24}$ (64 channels, 24 timesteps, achieved via $4\times$ temporal downsampling from the 96-frame input), and (2) a latent diffusion model that generates new latent codes conditioned on the UPDRS class, which are then decoded back into gait sequences using the frozen weights of the VAE decoder.

2.1 Data & Preprocessing

The proposed model was trained using the CARE-PD dataset (Adeli et al., 2025a), a multi-site collection of Skinned Multi-Person Linear (SMPL) (Loper et al., 2015) 3D body meshes aggregated from nine clinical cohorts, of which the four cohorts with UPDRS-gait annotations (PD-GaM, BMCLab, 3DGait, and T-SDU-PD) were used, totaling 110 subjects. The SMPL meshes were converted to the HumanML3D 263-dimensional representation (Guo et al., 2022), which encodes 22

joints with 6-degree-of-freedom rotations, local joint velocities, foot contact labels, plus global root translation and rotation. The data was segmented into overlapping windows of 96 frames (3.2 s at 30 fps) with a stride of 48 frames (50% overlap), capturing 2–4 gait cycles per window to balance stride-to-stride temporal coverage with sliding-window density. This yielded approximately 9,500 training windows. Sliding windows were used without cycle segmentation, as gait event detection is unreliable on Parkinson's patients and temporal irregularity is itself a clinically informative signal. Each window was normalized per-channel to zero mean and unit variance using training-set statistics.

The dataset provides UPDRS-gait scores on a 0 to 3 scale. All four classes, UPDRS 0 (normal), UPDRS 1 (slight), UPDRS 2 (mild), and UPDRS 3 (moderate), are used as conditioning targets for both the surrogate classifier and the diffusion model.

Subject-level train/eval splits were used to prevent data leakage: 86 subjects for training and 24 for evaluation (BMCLab: 19/4, PD-GaM: 22/8, 3DGait: 34/9, T-SDU-PD: 11/3). Per-channel normalization (zero mean, unit variance) was computed from the training split only, with a standard deviation floor to prevent division by near-zero values in saturated channels.

2.2 Stage 1: Variational Autoencoder

The VAE compresses gait windows $\mathbf{x} \in \mathbb{R}^{263\times 96}$ into a low-dimensional latent space $\mathbb{R}^{64\times 24}$ via a 1D convolutional encoder $q_\phi(\mathbf{z}|\mathbf{x})$ and reconstructs the input through a decoder $p_\theta(\mathbf{x}|\mathbf{z})$. The latent space is shaped by three complementary objectives: fidelity to the original motion, separability by clinical severity, and compactness of per-class distributions.

2.2.1 Standard VAE Loss

The encoder parametrizes a Gaussian posterior $\mathcal{N}(\boldsymbol{\mu}_\phi(\mathbf{x}), \boldsymbol{\sigma}^2_\phi(\mathbf{x}))$, from which $\mathbf{z}$ is sampled via the reparametrization trick (Kingma & Welling, 2022). The standard VAE objective $\mathcal{L}_{\text{VAE}}$ in equation 1 combines a reconstruction term with a Kullback-Leibler (KL) divergence regularizer:

$$\mathcal{L}_{\text{VAE}} = \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\|\mathbf{x} - p_\theta(\mathbf{x}|\mathbf{z})\|_2^2\right]}_{\mathcal{L}_{\text{recon}}} + \beta \underbrace{D_{\text{KL}}\big(q_\phi(\mathbf{z}|\mathbf{x}) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I})\big)}_{\mathcal{L}_{\text{KL}}} \tag{1}$$

where β is kept small to preserve pathological gait features that would otherwise be smoothed away by aggressive KL regularization.

2.2.2 Surrogate Severity-level Classifier

To enforce that the latent space carries clinically relevant structure, a lightweight classifier head f_ψ is trained on the mean-pooled latent vector $\bar{\mathbf{z}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{z}_t \in \mathbb{R}^d$ to predict the UPDRS-gait score c_p of each window:

$$\mathcal{L}_{\text{surr}} = \text{CrossEntropy}\big(f_\psi(\bar{\mathbf{z}}),\ c_p\big). \tag{2}$$

2.2.3 Class-Prototype Loss

To further sharpen per-class structure, each severity class k is assigned a learnable prototype vector $\mathbf{c}_k \in \mathbb{R}^d$ (Morlock & Wang, 2020), randomly initialized and optimized via gradient descent alongside the encoder. The prototype operates on the mean-pooled latent $\bar{\mathbf{z}} \in \mathbb{R}^d$ rather than the full spatio-temporal code, leaving the KL divergence free to regularize per-timestep dynamics independently. A two-term loss pulls each sample toward its class prototype while repelling prototypes apart:

$$\mathcal{L}_{\text{proto}} = \gamma_{\text{prox}} \underbrace{\sum_k \sum_{\mathbf{x}\in\mathcal{C}_k} \|\bar{\mathbf{z}}(\mathbf{x}) - \mathbf{c}_k\|_2^2}_{\mathcal{L}_{\text{proximity}}} - \gamma_{\text{dist}} \underbrace{\sum_{k\neq\ell} \|\mathbf{c}_k - \mathbf{c}_\ell\|_2^2}_{\mathcal{L}_{\text{distance}}} \tag{3}$$

where $\mathcal{C}_k$ denotes training windows of class k.

2.2.4 TOTAL VAE OBJECTIVE

The complete VAE training loss (equation 4) is the sum of all four terms:

$$\mathcal{L}_{\text{VAE}}^{\text{total}} = \mathcal{L}_{\text{recon}} + \beta\,\mathcal{L}_{\text{KL}} + \lambda\,\mathcal{L}_{\text{surr}} + \mathcal{L}_{\text{proto}}. \tag{4}$$

We fixed $\beta = 0.001$, $\lambda = 0.3$, $\gamma_{\text{prox}} = 0.3$, $\gamma_{\text{dist}} = 0.001$ for all experiments. The KL weight was kept small ($\beta = 0.001$) to avoid smoothing away subtle pathological features: low-β VAEs trade latent compression for reconstruction fidelity (Hoffman et al., 2017; Alemi et al., 2018; Chérief-Abdellatif et al., 2022), and small KL weights ($\lambda_{\text{KL}} = 0.01$) are likewise used for the HumanML3D learned motion latent space (Guo et al., 2022). The surrogate weight was selected from a sweep over $\lambda \in \{0.1, 0.3, 0.5\}$ by latent class separation, and the class-prototype proximity and distance weights from a log-scale grid, $\gamma_{\text{prox}} \in [10^{-3}, 1]$ and $\gamma_{\text{dist}} \in [10^{-3}, 10^{-1}]$, ranked by downstream DTW and ROM error.

2.3 STAGE 2: LATENT DIFFUSION TRANSFORMER

The second stage operates on the frozen VAE latent space $\mathbf{z} \in \mathbb{R}^{64\times 24}$, learning to generate class-conditioned samples through a denoising diffusion process (Ho et al., 2020). The forward process gradually adds Gaussian noise to a clean latent $\mathbf{z}_0$ over $T = 500$ timesteps, producing a sequence of increasingly noisy samples $\{\mathbf{z}_t\}_{t=0}^{T}$. The reverse process learns to denoise step-by-step, recovering $\mathbf{z}_0$ from $\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. At each timestep, the learned denoising distribution is:

$$p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t) = \mathcal{N}(\mathbf{z}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{z}_t, t, \mathbf{c}), \boldsymbol{\sigma}_\theta(\mathbf{z}_t, t, \mathbf{c})), \tag{5}$$

where $\mathbf{c}$ is the conditioning vector and $\boldsymbol{\mu}_\theta$, $\boldsymbol{\sigma}_\theta$ are parameterized by the network.

Following MDM (Tevet et al., 2022), we adopt an $\mathbf{x}_0$-prediction objective rather than noise prediction. At each timestep t, the model directly predicts the clean latent $\hat{\mathbf{z}}_0 = G_\theta(\mathbf{z}_t, t, \mathbf{c})$, where $\mathbf{c}$ is the conditioning vector. This formulation enables range-of-motion (ROM) supervision to be computed on the decoded output $\hat{\mathbf{x}}_0 = \text{VAE}_{\text{dec}}(\hat{\mathbf{z}}_0)$, directly supervising the kinematic amplitude of each generated gait window. The training objective combines the diffusion loss with a velocity smoothness term and ROM guidance:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t,\mathbf{z}_0,\boldsymbol{\epsilon}}\left[\|\mathbf{z}_0 - G_\theta(\mathbf{z}_t, t, \mathbf{c})\|_2^2\right] + \lambda_v \|\Delta\hat{\mathbf{z}}_0 - \Delta\mathbf{z}_0\|_2^2 + w_{\text{ROM}}\,\mathcal{L}_{\text{ROM}}, \tag{6}$$

where $\Delta\mathbf{z}$ denotes the first-order temporal difference in latent space (weight $\lambda_v = 0.1$), and $\mathcal{L}_{\text{ROM}}$ is the L1 error between the per-class batch mean of the six sagittal leg joint ranges of motion (hip, knee, ankle, left/right) of the decoded output and the clinical target $\mathbf{t}_{\text{ROM}}^{(c)}$ computed from real training data (weight $w_{\text{ROM}} = 0.1$).

The denoising network G_θ is a Diffusion Transformer (DiT) with adaptive layer normalization (AdaLN-Zero) blocks (Peebles & Xie, 2023). The architecture consists of 4 transformer layers with 4 attention heads and 256-dimensional embeddings. Each block modulates its internal parameters (scale, shift, gate) based on a timestep embedding and a class conditioning vector, enabling the model to adapt its denoising behavior to both the noise level and the target UPDRS severity.

Conditioning is provided by a learned embedding of the UPDRS class $c_p \in \{0, 1, 2, 3\}$, projected to 256 dimensions, forming the conditioning vector $\mathbf{c} = \text{MLP}(\text{Emb}(c_p))$. During training, classifier-free guidance (CFG) (Ho & Salimans, 2022) is applied by randomly dropping the conditioning with probability 25%, replacing it with a learned null token. At inference, the guidance scale s controls the trade-off between sample diversity and condition adherence:

$$\hat{\mathbf{z}}_0 = G_\theta(\mathbf{z}_t, t, \emptyset) + s \cdot \big(G_\theta(\mathbf{z}_t, t, \mathbf{c}) - G_\theta(\mathbf{z}_t, t, \emptyset)\big), \tag{7}$$

where $s > 1$ amplifies the conditioning signal. We use $s = 1.5$ as the default guidance scale.

2.4 SAMPLING & GENERATION

At inference, latent codes are generated by reversing the diffusion process over $T = 500$ timesteps starting from Gaussian noise $\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. At each reverse step t, the DiT predicts a clean latent $\hat{\mathbf{z}}_0$ via classifier-free guidance as defined in Eq. equation 7, and the previous-step latent is sampled from the DDPM posterior:

$$\mathbf{z}_{t-1} \sim \mathcal{N}\big(\boldsymbol{\mu}_\theta(\mathbf{z}_t, \hat{\mathbf{z}}_0),\ \tilde{\beta}_t\mathbf{I}\big), \tag{8}$$

where $\boldsymbol{\mu}_\theta$ is computed analytically from $\hat{\mathbf{z}}_0$ and $\mathbf{z}_t$, and $\tilde{\beta}_t$ follows the standard DDPM variance schedule (Ho et al., 2020). For a target UPDRS class c, the corresponding learned embedding is injected through the AdaLN-Zero blocks at every denoising step. The resulting clean latent $\hat{\mathbf{z}}_0$ is decoded through the frozen VAE decoder to obtain the final gait sequence $\hat{\mathbf{x}}_0 = \text{VAE}_{\text{dec}}(\hat{\mathbf{z}}_0)$; the VAE encoder is not used during generation.

2.5 METRICS

We evaluate generated gait across four complementary dimensions:

Waveform fidelity. Dynamic Time Warping (DTW) with Sakoe-Chiba band (10% of sequence length) provides per-sample temporal alignment without requiring paired samples.

Biomechanical plausibility. Per-joint Range of Motion (ROM) error, arm swing amplitude, and trunk inclination angle. Cross-class ordering of these biomarkers is the primary signal of UPDRS conditioning (Mirelman et al., 2019).

Distributional quality. Latent-space Fréchet Inception Distance (FID) computed on 64-dim VAE latent vectors, Maximum Mean Discrepancy (MMD^2) on 132-dim biomechanical features, and Precision/Recall with k=3 nearest-neighbor. Diversity ratio is synth/real mean pairwise L2 distance in VAE latent space. As FID is computed on the VAE's own latent representations, it reflects within-architecture distributional fit rather than an independent perceptual metric.

Generalization. We train a Random Forest classifier on VAE latent features to predict UPDRS severity. Train-on-Real Test-on-Real (TRTR) via 10-fold cross-validation sets the data ceiling; Train-on-Real Test-on-Synth (TRTS) measures conditioning correctness.

Pathology-specific metrics. Following GAITGen (Adeli et al., 2025b), AVE (Average Variance Error), AAMD (Absolute Arm Swing Mean Difference), and ASMD (Absolute Stooped Posture Mean Difference) measure joint-variance, arm-swing, and stooped-posture differences between synthetic and real gait (AAMD/ASMD leg-length normalized).

3 RESULTS

3.1 GENERATION QUALITY AND COMPARISON TO BASELINES

Table 1 reports ROM and DTW for the sagittal lower-limb joints that define gait (left and right legs pooled). The model generates gait trajectories that closely match the biomechanical characteristics of each UPDRS class. Synthetic mean ROM deviates from real gait by $\pm 5.8\%$ on average across real classes 0–3, and DTW remains low (0.292 to 0.505) for classes 0–2. DTW is highest for UPDRS 3 (1.180), which suffers from a data bottleneck (158 evaluation windows, vs. 742–1546 for U0–U2) and heterogeneous gait patterns present in the most advanced PD levels. ROM decreases monotonically with severity ($44.84 > 39.06 > 27.43 > 16.26°$), confirming that UPDRS conditioning propagates correctly through the latent space.

As shown in Table 2, distributional quality is strong for classes 0–2 where per-class latent-space FID remains below 2.0, indicating the VAE latent distribution is well-matched. Precision is high across classes 0–2 (0.65–0.83; U3: 0.09), confirming generated samples lie within the real-data manifold for the well-represented classes; recall is uniformly low (0.13–0.22 for U0–U3), indicating limited mode coverage because the ROM supervision loss anchors generation near class-mean amplitudes. U3 is unreliable due to the small real sample size (n=158), and even a class-weighted loss cannot overcome this limit for distribution estimation. The pooled latent-space FID of 6.28 and diversity ratio of 1.227 show that the generated samples are not fully collapsed.

Table 2 reports per-class TRTR (real-data ceiling) and TRTS. TRTS reaches an overall accuracy of 84.4%, and classes 1–3 are recovered well (90.4%–97.7%), whereas U0 remains ambiguous (55.8%), mostly confused with slight gait. Notably, TRTS exceeds TRTR for U1–U3, indicating the synthetic set is less diverse than real gait rather than a gain in conditioning.

Figure 2 confirms this at the joint level. Overlaying the real per-class mean sagittal waveforms (gray, ± 1 std) with their severity-conditioned synthetic counterparts (class-colored, phase-aligned on the left knee) shows that the hip, knee, and ankle cycles recover the real gait morphology and

Table 1: Per-joint lower-limb range of motion (ROM) and dynamic time warping (DTW) per UPDRS class. ROM and DTW are reported as mean ± std over windows; *Err* is the signed relative ROM error, negative when the synthetic value is below real (under-generation) and positive when above. Left and right legs are pooled per joint (right leg is phase-aligned to the left one).

	Hip	Knee	Ankle	**Mean**
UPDRS 0 (Normal)				
DTW	0.411 ± 0.026	0.494 ± 0.033	0.181 ± 0.012	0.362 ± 0.024
Real ROM (°)	51.58 ± 10.45	63.70 ± 18.72	28.19 ± 10.20	47.82 ± 13.12
Synth ROM (°)	52.26 ± 1.95	58.11 ± 2.60	24.15 ± 1.34	44.84 ± 1.96
Err (%)	$+1.3$	-8.8	-14.3	-7.3
UPDRS 1 (Slight)				
DTW	0.524 ± 0.027	0.720 ± 0.030	0.271 ± 0.009	0.505 ± 0.022
Real ROM (°)	42.96 ± 9.49	58.92 ± 16.40	24.95 ± 10.18	42.27 ± 12.02
Synth ROM (°)	41.88 ± 1.75	53.20 ± 2.00	22.09 ± 1.35	39.06 ± 1.70
Err (%)	-2.5	-9.7	-11.4	-7.9
UPDRS 2 (Mild)				
DTW	0.203 ± 0.018	0.504 ± 0.029	0.170 ± 0.017	0.292 ± 0.021
Real ROM (°)	29.10 ± 11.17	42.07 ± 18.59	16.47 ± 8.05	29.22 ± 12.60
Synth ROM (°)	27.67 ± 1.67	40.07 ± 2.22	14.56 ± 1.42	27.43 ± 1.77
Err (%)	-4.9	-4.8	-11.6	-7.1
UPDRS 3 (Moderate)				
DTW	1.425 ± 0.231	1.210 ± 0.102	0.907 ± 0.091	1.180 ± 0.141
Real ROM (°)	15.96 ± 7.68	23.59 ± 11.39	8.71 ± 3.66	16.09 ± 7.58
Synth ROM (°)	14.37 ± 2.61	25.27 ± 5.11	9.13 ± 1.90	16.26 ± 3.20
Err (%)	-9.9	$+7.1$	$+4.8$	$+0.7$

Table 2: Distributional and generalization metrics per UPDRS class.

	LS-FID ↓	MMD^2 ↓	Precision ↑	Recall ↑	Div. ratio	TRTR ↑	TRTS ↑
UPDRS 0	1.24	0.065	0.837	0.189	0.719	86.9	55.8
UPDRS 1	0.89	0.040	0.788	0.144	0.739	80.6	93.5
UPDRS 2	1.75	0.057	0.649	0.131	0.678	80.1	90.4
UPDRS 3	29.14	0.123	0.091	0.222	0.649	54.8	97.7
Overall	**6.28**	0.036	0.556	0.199	1.227	81.3	84.4

reproduce the monotonic amplitude contraction with severity ($44.84 > 39.06 > 27.43 > 16.26°$). Synthetic means track the real waveforms closely for UPDRS 0–2, consistent with their low DTW, while UPDRS 3 retains the correct phase but deviates more, matching its elevated DTW (1.180).

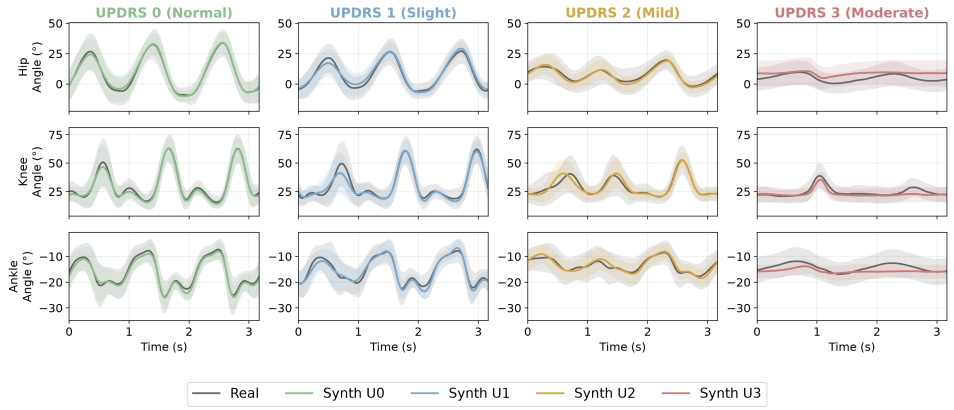

Figure 2: Per-class sagittal joint angle waveforms (Rows: hip, knee, ankle; columns: UPDRS 0–3), real gait (gray) overlaid with severity-conditioned synthetic gait (class-colored), phase-aligned on the left knee.

Compared to prior work (Table 3), our model achieves the best AVE (0.004), outperforming GAIT-Gen, and falls closely to GAITGen's reported ASMD 0.048 and AAMD 0.096, 0.044 and 0.109 for our method respectively. Baseline values are copied from (Adeli et al., 2025b) and were not retrained or re-evaluated by us. Note that GAITGen trains on the PD-GaM cohort and evaluates cross-cohort on unseen sites, while our model trains on all four CARE-PD cohorts with a subject-level split.

Table 3: Comparison with SOTA motion generation baselines. Baseline metric values are copied directly from (Adeli et al., 2025b) and were not re-trained or re-evaluated by us. Values are normalized by leg length. Relative to the strongest baseline (Adeli et al., 2025b), GaitDiffusion reduces AVE by 97.9% and ASMD by 8.3%, while its AAMD is 13.5% higher.

Method	AVE ↓	AAMD ↓	ASMD ↓
StableMoFusion (Huang et al., 2024)	1.437	1.568	0.201
MMM (Pinyoanuntapong et al., 2024)	1.037	0.791	0.118
SnapMoGen (Guo et al., 2025)	0.745	0.526	0.166
MoMask (Guo et al., 2023)	0.898	0.440	0.106
GAITGen Adeli et al. (2025b)	0.194	**0.096**	0.048
GaitDiffusion (Ours)	**0.004**	0.109	**0.044**

Table 4: Ablation: VAE-only, DiT without auxiliary supervision, and full GaitDiffusion.

	DTW per class				Mean	ROM		TRTS	LS-FID
Variant	U0	U1	U2	U3	DTW	order	err	Overall	Overall
VAE-only (no DiT)	0.883	0.642	0.378	0.858	0.690	✓	24.8%	37.5%	—
DiT (no aux)	**0.281**	**0.163**	**0.277**	**1.067**	**0.447**	✓	20.6%	83.9%	29.52
GaitDiffusion	0.362	0.505	0.292	1.180	0.585	✓	**5.8%**	**84.4%**	6.28

3.2 Ablation: Component Contributions

Table 4 reports a three-way ablation. We compare three configurations: (1) VAE-only, sampling latents directly from the VAE's per-class distribution $\mathcal{N}(\boldsymbol{\mu}_c, 0.5\,\boldsymbol{\sigma}_c)$ and decoding, bypassing the DiT entirely; (2) DiT without auxiliary supervision (no ROM), isolating the contribution of class-conditioned denoising alone; and (3) full GaitDiffusion with ROM supervision.

Each component plays a distinct role. The VAE embeds the clinical ROM ordering, which decreases monotonically with severity in all variants; the DiT drives class separability, raising TRTS from 37.5% (VAE-only) to 83.9% (no auxiliary supervision) and 84.4% (full), and finally the ROM supervision governs amplitude accuracy (24.8% → 20.6% → 5.8%). Table 4 also reflects a precision-diversity tradeoff: without amplitude constraints, the DiT produces waveforms that match real gait well (DTW 0.447) but the aggregate distribution drifts from the real (FID 29.52). ROM supervision anchors the distribution (FID 6.28) at the cost of per-sample waveform fidelity (DTW 0.585).

4 Conclusion

We presented GaitDiffusion, a two-stage latent diffusion framework for generating full-body PD severity-conditioned gait sequences. A class-prototype VAE embeds clinical severity structure into a continuous latent space, from which a Diffusion Transformer with classifier-free guidance generates condition-specific samples. A three-way ablation (VAE-only, DiT without auxiliary supervision, and full GaitDiffusion) confirms that the VAE's class-prototype loss and surrogate classifier embed the clinical ROM ordering independently, while the DiT's AdaLN-Zero conditioning provides the class separability (TRTS: 37.5% → 83.9% → 84.4%), and ROM supervision corrects the amplitude (ROM error: 24.8% → 5.8%) and distributional fit (latent-space FID: 29.52 → 6.28) at a cost of per-sample waveform fidelity (DTW: 0.447 → 0.585). Our model achieves a mean DTW of 0.585 across all four classes (0.386 across 0–2), a pooled latent-space FID of 6.28, and an overall TRTS conditioning accuracy of 84.4%. Against the strongest prior pathology-conditioned baseline, GaitDiffusion lowers the Average Variance Error by 97.9% and the Absolute Stooped Posture Mean Difference by 8.3%, while its Absolute Arm Swing Mean Difference is 13.5% higher, demonstrating that latent diffusion is competitive for clinically-conditioned gait generation. ROM decreases monotonically with severity ($44.84 > 39.06 > 27.43 > 16.26$), demonstrating that UPDRS conditioning propagates faithfully through the continuous latent diffusion pipeline.

The generated sequences have potential utility as reference trajectories for exoskeleton control, synthetic training data for downstream PD gait analysis, and personalized monitoring in medical treatment. UPDRS 3 (moderate) generation remains limited by data sparsity ($n = 158$ evaluation windows), and per-class conditioning is weakest for U0 (TRTS: 55.8%), where over one-third of generated normal gait is confounded with mild, reflecting class proximity in the learned latent space. Future work includes extending conditioning beyond discrete UPDRS classes to additional control variables such as gait speed, subject height, medication state, and disease duration, enabling fine-grained personalization of generated motion to individual patient profiles, and validating the generated patterns in a control loop for patient-adaptive assistive therapy.

Reproducibility statement and AI disclosure. We provide *GaitDiffusion* code and the scripts to run the model in inference and to train with the databases.

In this work, we used generative AI tools for providing feedback on experiments, helping implement the code, supporting qualitative data analysis, assisting with translation and language for writing, modifying figures for clarity, and summarizing and analysing existing literature. We have NOT used generative AI tools for generating synthetic data sets, clean and reformat dataset or helping develop theoretical models or conceptual frameworks, and formulate mathematical claims, provide critical ingredients for proving mathematical claims or assist in the writing of proofs are not applicable to this work. We have reviewed all AI-assisted input. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Vida Adeli, Ivan Klabučar, Javad Rajabi, Benjamin Filtjens, Soroush Mehraban, Diwei Wang, Trung Hieu Hoang, Minh Do, Hyewon Seo, Candice MULLER, Daniel Coelho, Claudia de Oliveira, Pieter Ginis, Moran Gilat, Alice Nieuwboer, Joke Spildooren, J. Mckay, Hyeokhyen Kwon, Gari Clifford, Christine Esper, Stewart Factor, Imari Genias, Amirhossein Dadashzadeh, Leia Shum, Alan Whone, Majid Mirmehdi, Andrea Iaboni, and Babak Taati. Care-pd: A multi-site anonymized clinical dataset for parkinson's disease gait assessment. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38. Curran Associates, Inc., 2025a.

Vida Adeli, Soroush Mehraban, Majid Mirmehdi, Alan Whone, Benjamin Filtjens, Amirhossein Dadashzadeh, Alfonso Fasano, Andrea Iaboni, and Babak Taati. Gaitgen: Disentangled motion-pathology impaired gait generative model – bringing motion generation to the clinical domain, 2025b. URL https://arxiv.org/abs/2503.22397.

Alexander A. Alemi, Ben Poole, Ian Fischer, Joshua V. Dillon, Rif A. Saurous, and Kevin Murphy. Fixing a broken elbo, 2018. URL https://arxiv.org/abs/1711.00464.

Eran Beeri Bamani, Joao Buzzatto, and Hermano Igo Krebs. Diffusion-augmented spatio-temporal graph convolution for clinical gait and motor function assessment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2026.

Emad Barsoum, John Kender, and Zicheng Liu. Hp-gan: Probabilistic 3d human motion prediction via gan. In *CVPR Workshops*, 2018.

Damian Benasco, Juan Carballeira-Lopez, Jaime Ramos-Rojas, Julio S. Lora-Millan, Antonio J. Del-Ama, David Rodriguez-Cianca, and Pablo Lanillos. Diffusion-based generation of gait trajectories, 2026. URL https://arxiv.org/abs/2609.14642.

Mritula Chandrasekaran, Sanket Kachole, Jarik Francik, and Dimitrios Makris. Llm-conditioned synthesis of pathological gaits via structured gait-language representations. *arXiv preprint arXiv:2606.06048*, 2026.

Huiwen Chang, Han Zhang, Lu Jiang, Ce Liu, and William T. Freeman. Maskgit: Masked generative image transformer, 2022. URL https://arxiv.org/abs/2202.04200.

Andrés Chavarrías, David Rodriguez-Cianca, and Pablo Lanillos. Adaptive torque control of exoskeletons under spasticity conditions via reinforcement learning. In *2025 International Conference On Rehabilitation Robotics (ICORR)*, pp. 705–711. IEEE, 2025.

Xin Chen, Biao Jiang, Wen Liu, Zilong Huang, Bin Fu, Tao Chen, Jingyi Yu, and Gang Yu. Executing your Commands via Motion Diffusion in Latent Space, May 2023. URL http://arxiv.org/abs/2212.04048. arXiv:2212.04048 [cs.CV].

Badr-Eddine Chérief-Abdellatif, Yuyang Shi, Arnaud Doucet, and Benjamin Guedj. On pac-bayesian reconstruction guarantees for vaes. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera (eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 3066–3079. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/cherief-abdellatif22a.html.

Diya Dinesh, Adrian Krieger, Changseob Song, Dongho Park, Aaron J Young, and Inseung Kang. Generating personalized lower-limb kinematics across walking speeds using subject-conditioned diffusion. *arXiv preprint arXiv:2607.07533*, 2026.

Patrick Esser, Robin Rombach, and Björn Ommer. Taming transformers for high-resolution image synthesis, 2021. URL https://arxiv.org/abs/2012.09841.

Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, December 2018. ISSN 0925-2312. doi: 10.1016/j.neucom.2018.09.013. URL http://dx.doi.org/10.1016/j.neucom.2018.09.013.

Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard Dodel, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, 23(15):2129–2170, 2008.

Chuan Guo, Xinxin Zuo, Sen Wang, Shihao Zou, Qingyao Sun, Annan Deng, Minglun Gong, and Li Cheng. Action2motion: Conditioned generation of 3d human motions. In *ACM Multimedia*, 2020.

Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. Generating diverse and natural 3d human motions from text. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5152–5161, 2022.

Chuan Guo, Yuxuan Mu, Muhammad Gohar Javed, Sen Wang, and Li Cheng. Momask: Generative masked modeling of 3d human motions, 2023. URL https://arxiv.org/abs/2312.00063.

Chuan Guo, Inwoo Hwang, Jian Wang, and Bing Zhou. Snapmogen: Human motion generation from expressive texts, 2025. URL https://arxiv.org/abs/2507.09122.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *arXiv preprint arxiv:2006.11239*, 2020.

Matthew D Hoffman, Carlos Riquelme, and Matthew J Johnson. The β-vae's implicit prior. In *Workshop on Bayesian Deep Learning, NIPS*, pp. 1–5, 2017.

Yiheng Huang, Hui Yang, Chuanchen Luo, Yuxi Wang, Shibiao Xu, Zhaoxiang Zhang, Man Zhang, and Junran Peng. Stablemofusion: Towards robust and efficient diffusion-based motion generation framework, 2024. URL https://arxiv.org/abs/2405.05691.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022. URL https://arxiv.org/abs/1312.6114.

Jochen Klucken, Jens Barth, Patrick Kugler, Johannes Schlachetzki, Thore Henze, Franz Marxreiter, Zacharias Kohl, Ralph Steidl, Joachim Hornegger, Bjoern Eskofier, et al. Unbiased and mobile gait analysis detects motor impairment in parkinson's disease. *PloS one*, 8(2):e56956, 2013.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Smpl: a skinned multi-person linear model. *ACM Trans. Graph.*, 34(6), November 2015. ISSN 0730-0301. doi: 10.1145/2816795.2818013. URL https://doi.org/10.1145/2816795.2818013.

Anat Mirelman, Paolo Bonato, Richard Camicioli, Terry D Ellis, Nir Giladi, Jamie L Hamilton, Chris J Hass, Jeffrey M Hausdorff, Elisa Pelosin, and Quincy J Almeida. Gait impairments in parkinson's disease. *The Lancet Neurology*, 18(7):697–708, 2019.

Frederick Morlock and Dingsu Wang. MAD-VAE: Manifold Awareness Defense Variational Autoencoder, October 2020. URL http://arxiv.org/abs/2011.01755. arXiv:2011.01755 [cs.CR].

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023.

Mathis Petrovich, Michael J Black, and Gül Varol. Action-conditioned 3d human motion synthesis with transformer vae. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10985–10995, 2021.

Ekkasit Pinyoanuntapong, Pu Wang, Minwoo Lee, and Chen Chen. Mmm: Generative masked motion model, 2024. URL https://arxiv.org/abs/2312.03596.

Ali Razavi, Aaron van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with vq-vae-2, 2019. URL https://arxiv.org/abs/1906.00446.

Arshak Rezvani, Nasrin Ravansalar, Mohammad Ali Akhaee, Andrew J. Greenshaw, Russell Greiner, Maryam S. Mirian, Muhammad Yousefnezhad, and Martin J. McKeown. Diffusegaitnet: Improving parkinson's disease gait severity assessment with a diffusion model framework. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 33:2858–2869, 2025. doi: 10.1109/TNSRE.2025.3589074.

Donatella Simonetti, Maartje Hendriks, Bart Koopman, Noel Keijsers, and Massimo Sartori. A wearable gait lab powered by sensor-driven digital twins for quantitative biomechanical analysis post-stroke. *Wearable Technologies*, 5:e13, 2024.

Tian Tan, Tom Van Wouwe, Keenon F. Werling, C. Karen Liu, Scott L. Delp, Jennifer L. Hicks, and Akshay S. Chaudhari. GaitDynamics: A Generative Foundation Model for Analyzing Human Walking and Running. *Research Square*, pp. rs.3.rs–6206222, March 2025. ISSN 2693-5015. doi: 10.21203/rs.3.rs-6206222/v1. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC11957236/.

Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H. Bermano. Human Motion Diffusion Model, October 2022. URL http://arxiv.org/abs/2209.14916. arXiv:2209.14916 [cs.CV].

Heike Vallery, Edwin HF Van Asseldonk, Martin Buss, and Herman Van Der Kooij. Reference trajectory generation for rehabilitation robots: complementary limb motion estimation. *IEEE transactions on neural systems and rehabilitation engineering*, 17(1):23–30, 2008.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning, 2018. URL https://arxiv.org/abs/1711.00937.

Xiaoyue Wan and Xu Zhao. Diffusion-based personalized pathology disentanglement for impaired gait analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 9648–9656, 2026.

Juanjuan Zhang, Pieter Fiers, Kirby A Witte, Rachel W Jackson, Katherine L Poggensee, Christopher G Atkeson, and Steven H Collins. Human-in-the-loop optimization of exoskeleton assistance during walking. *Science*, 356(6344):1280–1284, 2017.

Mingyuan Zhang et al. Motiondiffuse: Text-driven human motion generation with diffusion model. *IEEE TPAMI*, 46(6):4115–4128, 2024.