REINS: Learning Inertia-Induced Geometry for Physics-Consistent Motion Representation in Clinical Gait Phenotyping

Yiran Ding[1], Yufei Zhang[2], and Zijun Cui[1]

[1] Department of Computer Science and Engineering, Michigan State University, East Lansing, Michigan, USA
[2] Independent Researcher

Problem and Purpose. Computational gait phenotyping for Parkinson's disease (PD) relies on latent motion representations that are learned from data alone. Clinical cohorts are small, heterogeneous, and collected under site-specific protocols, so such representations overfit and transfer poorly [1]. Existing physics-informed motion models apply mechanical structure as a trajectory-level penalty on reconstruction [5], leaving the geometry of the representation space untouched. We ask whether that geometry can itself carry the biomechanics, following a broader program that treats geometric structure as the object of motion representation learning. Human gait is the motion of an articulated rigid-body system, and the generalized inertia matrix $M(q)$ of such a system is a Riemannian metric on configuration space, which is the classical kinetic-energy metric [3], applied to the whole moving human body by Neilson *et al.* [4]. $M(q)$ is fixed by mechanics and body shape, so it is available as an inductive bias no matter how little clinical data exists. We test whether a representation built on this metric buys cross-cohort and small-sample robustness.

Methods. REINS defines the representation space by the inertia metric and reads it out geodesically (Fig. 1). The encoder–decoder and the per-body-part mass and inertia model are taken from PhysPT [5]. Since kinetic energy is the quadratic form $T(q, \dot{q}) = \frac{1}{2}\dot{q}^\top M(q)\dot{q}$, we supervise reconstruction with an inertia-

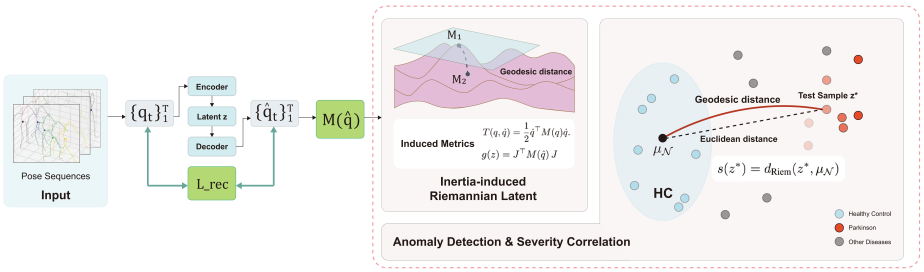

Fig. 1: REINS. Pose sequences are encoded to a latent z; reconstruction is supervised under the inertia metric $M(\hat{q})$ computed at the decoder output. The latent inherits the pullback metric $g(z) = J^\top M(\hat{q})J$, and each subject is scored by geodesic distance to the healthy-control reference.

Table 1: Downstream clinical performance. Left: subject-level anomaly detection on CARE-PD DNE at K=20 HC training subjects (HC/all = HC vs. PD+other movement disease). Right: zero-shot UPDRS-gait severity correlation on four unseen cohorts (n=30, 23, 14, 43). All models share one backbone.

	AUROC ↑		Spearman ρ ↑			
Method	HC/PD	HC/all	PD-GaM	BMCLab	T-SDU-PD	3DGait
Euclidean latent, Euclidean read-out						
Data-driven	0.575	0.609	0.280	0.220	0.257	0.214
Physics-supervised	0.517	0.538	0.251	0.290	0.234	0.170
M-induced Riemannian latent, geodesic read-out						
REINS (ours)	**0.896**	**0.868**	**0.538**	**0.523**	**0.402**	**0.306**
Chance	0.500	0.500	0.000	0.000	0.000	0.000

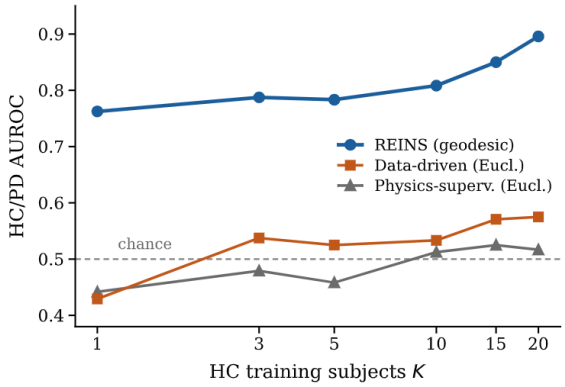

Fig. 2: The physical prior does not need data. HC/PD AUROC as the HC training cohort shrinks (nested subsets).

consistent loss $\mathcal{L}_{\text{rec}} = \sum_t (\dot{q}_t - \hat{\dot{q}}_t)^\top M(q_t)(\dot{q}_t - \hat{\dot{q}}_t)$, i.e. velocity error measured in the inertia metric, in place of a trajectory-level Euler–Lagrange residual. The latent space inherits the pullback metric $g(z) = J^\top M(\hat{q})J$ through the decoder Jacobian J [2], and a test subject is scored by the geodesic distance from its embedding to the healthy-control (HC) reference, approximated on a k-nearest-neighbour graph over embeddings with edges weighted by local Riemannian length and shortest paths by Dijkstra. We fine-tune on CARE-PD [1] DNE healthy controls only; PD and other-movement-disease subjects are never seen in training, and clinical labels are used only for evaluation. Baselines share the identical backbone and differ only in latent geometry and read-out distance: Data-driven Euclidean (reconstruction loss only) and Physics-supervised Euclidean (full PhysPT objective), both read out by Mahalanobis distance.

Results and Conclusions. The inertia geometry separates pathological gait. Both Euclidean read-outs sit near chance on HC vs. PD, including the one

trained with a complete trajectory-level physics objective; re-equipping the same encoder with a geodesic distance under M recovers most of the separability gap (Table 1, left). The advantage does not depend on data volume: the separation survives as the HC training cohort shrinks from 20 subjects to 1 (Fig. 2). The same score transfers zero-shot to four external cohorts spanning different sites, protocols, and severity distributions, with no fine-tuning, label exposure, or per-cohort calibration (Table 1, right). Mechanistically, adding $\mathcal{L}_{\text{rec}}$ cuts predicted-vs.-ground-truth kinetic-energy error to 22.7%, against 82.1% for reconstruction-only and 193.4% for the full trajectory-level physics objective, at matched pose accuracy (9.79 vs. 9.05 mm test MPJPE). We conclude that for clinical motion analysis physics belongs in the metric. Future work broadens the reference cohort across age, body size, and recording protocol.

References

1. Adeli, V., Klabucar, I., Rajabi, J., Filtjens, B., Mehraban, S., Wang, D., Seo, H., Hoang, T.H., Do, M.N., Muller, C., et al.: Care-pd: A multi-site anonymized clinical dataset for parkinson's disease gait assessment. arXiv preprint arXiv:2510.04312 (2025)
2. Arvanitidis, G., Hansen, L.K., Hauberg, S.: Latent space oddity: on the curvature of deep generative models. arXiv preprint arXiv:1710.11379 (2017)
3. Bullo, F., Lewis, A.D.: Geometric control of mechanical systems: modeling, analysis, and design for simple mechanical control systems, vol. 49. Springer (2005)
4. Neilson, P.D., Neilson, M.D., Bye, R.T.: A riemannian geometry theory of human movement: The geodesic synergy hypothesis. Human movement science **44**, 42–72 (2015)
5. Zhang, Y., Kephart, J.O., Cui, Z., Ji, Q.: Physpt: Physics-aware pretrained transformer for estimating human dynamics from monocular videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2305–2317 (2024)