

Topological Gait Analysis: A New Framework and its Application to the Study of Human Gait - Additional Information

Shreyam Mishra , Debasish Chatterjee  Senior Member, IEEE, and Neeta Kanekar 

I. INTRODUCTION

This material accompanies the main manuscript titled “*Topological Gait Analysis: A New Framework and its Application to the Study of Human Gait*” by Shreyam Mishra, Debasish Chatterjee, and Neeta Kanekar. This document provides a comprehensive theoretical overview of cubical homology, clinical information and analysis of gait parameters for each subject from [1], [2], a detailed explanation of the method used for tuning the hyperparameter σ and an exploration of an additional dataset [3] comprising subjects with neurodegenerative conditions as well as aging to underscore the generalizability of the proposed topological gait analysis (TGA) framework. We also discuss the minimum amount of gait data (in terms of number of strides) required for the proposed TGA framework to produce accurate results and how its applicability can be extended to subjects with severe conditions who face significant challenges in walking for extended periods.

II. TOPOLOGICAL DATA ANALYSIS USING CUBICAL HOMOLOGY

Homology is a tool used to study the structure and connectivity of multi-dimensional data. ‘Persistence’ homology offers a method to inspect the data across various scales and identify ‘significant’ features that persist over a vast range. We review some basic concepts of cubical homology; for more formal treatment, readers are guided to [4, pp 57–59].

Definition II.1 (Abstract grid cubes). Consider $p \in \mathbb{Z}^d$ and $\ell \in \{0, 1\}^d$. The set $c(p, \ell) := [p_1, p_1 + \ell_1] \times \cdots \times [p_d, p_d + \ell_d]$ is called the *abstract grid cube* in \mathbb{R}^d associated with vectors p, ℓ with p being the base point of $c(p, \ell)$. For the case of $d = 2$, we call the elementary cube $c(p, \ell)$ as a *pixel* in X .

Definition II.2 (Cubical complex). A collection of grid cubes X is called a *grid cubical complex* if whenever $c \in X$, all the grid cubes contained as subsets in c also belong to X .

Definition II.3 (Boundary operator). To determine homologies of the grid complex, we define a linear map called as the

DC acknowledges partial support of the SERB MATRICS grant MTR/2022/000656, Govt. of India, and the IoE Cell, IIT Bombay.

Corresponding author: NK.

SM and DC are with Systems and Control Engineering, and NK is with the Department of Biosciences and Bioengineering,  Indian Institute of Technology Bombay, Powai, Mumbai 400076, India.

Emails: (mishra.shreyam, dchatter, nkanekar)@iitb.ac.in

n th boundary operator on every elementary cube of X as

$$\begin{aligned} \partial_n(c(p, \ell)) &:= \sum_{m=1}^d (-1)^{\ell_1+\cdots+\ell_m} ([p_1, p_1 + \ell_1] \times \cdots \\ &\quad \times \{p_m\} \times \cdots \times [p_d, p_d + \ell_d]) \\ &\quad - [p_1, p_1 + \ell_1] \times \cdots \\ &\quad \times \{p_m + \ell_m\} \times \cdots \times [p_d, p_d + \ell_d], \end{aligned} \quad (1)$$

where $n = \dim(c(p, \ell)) = \ell_1 + \cdots + \ell_d$.

Definition II.4 (Homology group). For every $n \in \mathbb{N}$, let $Q_n(X)$ denote the *free abelian group* generated by all the n dimensional grid cubes in X . The homology groups $QH_n(X)$ are then defined as

$$QH_n(X) = \text{Ker}(\partial_n) / \text{Im}(\partial_{n+1}). \quad (2)$$

Intuitively the 0th homology group H_0 captures the number of connected components, the 1st homology group H_1 counts cycles, H_2 identifies voids, and so forth.

Consider the 5×5 gray-scale image in Figure 1(a). The collection of pixels $(q_{ij})_{i,j=1}^5$ constitute the grid complex X , while the gray-scale values of each pixel shown in the matrix of Figure 1(b) define the function $X \ni q_{ij} \mapsto f(q_{ij})$. The sub-level sets of f are then defined by

$$L_t(f) := \{q_{ij} \in X \mid f(q_{ij}) \leq t\}.$$

Definition II.5 (Filtration). The sequence

$$L_{t_1}(f) \subset L_{t_2}(f) \subset \cdots \subset L_{t_n}(f),$$

for an increasing sequence of values $t_1 < t_2 < \cdots < t_n$ is called a *filtration* of f . Figure 1(c) displays the sub-level sets of the image for $t \in \{40, 75, 100, 150, 190\}$.

The sub-level sets in Figure 1(c) capture the image’s topological features effectively. The two ‘holes’ (relatively whiter pixels) appear to be born together at the filtration value $t = 90$. Formally, this is evaluated using (1) and (2), and is visually represented by the onset of the blue bar, denoted as H_1 , at $t = 90$ in Figure 1(d). The ‘gray-er’ pixel (corresponding to pixel value 130 as shown in Figure 1(b)) disappears at $t = 130$, which is seen by the termination of one of the blue lines at $t = 130$ while the blue line corresponding to the second hole continues to grow until $t = 180$. The second ‘hole’ dies and the blue line tracking its evolution is also terminated.

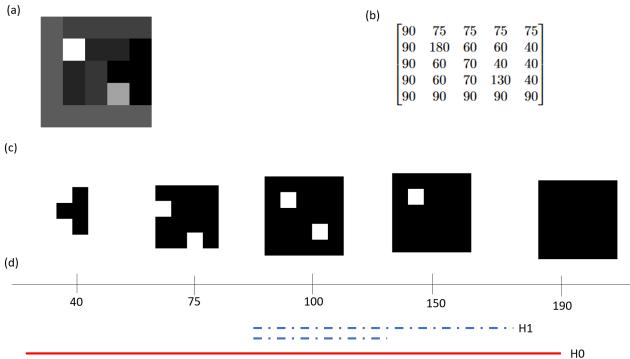


Fig. 1: Application of cubical homology: (a) Shows a gray scale image (b) A matrix representing the pixel values of the image (c) The sequence of filtered cubical complexes (d) Persistence barcode associated with the filtration.

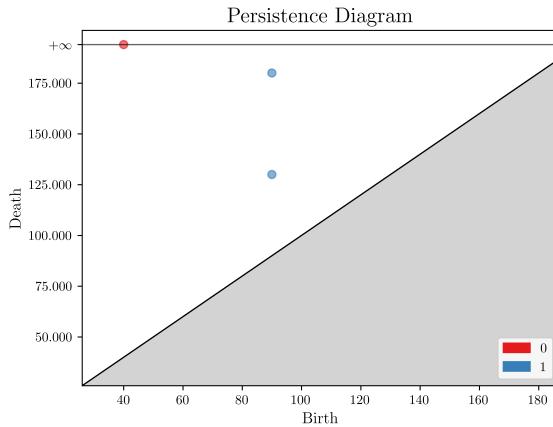


Fig. 2: Persistence diagram for the gray-scale image in Figure 1(a). The two blue dots in the persistence diagram indicate the presence of two local local maximas.

Since the image exists as a single connected component for all values of t , the red H_0 line never terminates. Such a pictorial representation of recording the appearance (birth) and disappearance (death) of topological features as the filtration parameter t is varied is called a *persistence barcode*. This information can also be captured using a single graph known as the *persistence diagram* $D := \{(b_i, d_i), \dots, (b_m, d_m)\}$, which is a collection of points in \mathbb{R}^2 with (b_i, d_i) representing the birth and death values of the i th feature. The persistence diagram for the given example is shown in Figure 2.

III. CLINICAL INFORMATION

The main manuscript covered the analysis of the Gait Dynamics in Neuro-Degenerative Disease Dataset from [1], [2]. Details regarding the data collection protocol and experimental conditions have been mentioned in §II. of the main manuscript. Clinical aspects such as subject age, disease severity along with essential statistics related to the left and right foot stride, stance and swing intervals (henceforth referred as LStr, RStr, LStn, RStn, LSw and RSw respectively) have been summarized in Table I–Table IV. These statistics include the average

(μ) , standard deviation ($\hat{\sigma}$) and coefficient of variation (CV). The coefficient of variation was calculated as $CV = 100(\frac{\hat{\sigma}}{\mu})$. Any other information can be found in the Physionet database [2].

IV. HYPER-PARAMETER TUNING OF σ

To determine an optimal value for σ , the dataset was systematically divided into two sets: 1) training, and 2) validation, each containing a random sample of 50% of the subjects. The purpose of this splitting is to ensure that the hyper-parameter tuning occurs systematically and is not prone to overfitting.

The training set serves the purpose of tuning the hyper-parameter σ through visual inspection of the CDF plots of persistence entropy. The σ value that provided the clearest separation between healthy and neurodegenerative gait was selected. This method is akin to the training-validation-test split commonly used in machine learning experiments.

The selected σ was then tested on the validation set to verify that the trend identified in the training set generalizes to the validation set. This step ensured that the chosen σ is not a consequence of overfitting to the dataset, but rather identifies the underlying patterns in the dataset effectively.

As illustrated in Figures 3–4, $\sigma = 0.0015$ was selected for stride analysis and $\sigma = 0.0020$ for stance analysis. Once an appropriate σ is fixed for each gait parameter, the CDF plots of persistence entropy are plotted to check if the trend identified based on training samples generalizes to a disjoint set of subjects, i.e., the validation set. Indeed, based on Figures 5a, 5b, 5c and 5d, the CDF plots continue to exhibit a clear trend with disease severity, indicating that these trends are not a consequence of overfitting. Although this validation was not explicitly performed for the swing interval analysis reported in the manuscript, it is confident that similar results would be obtained. The systematic tuning and validation process enhances the credibility of the findings and demonstrates that the identified trends are meaningful and reliable.

V. GENERALIZABILITY OF TGA FRAMEWORK

To verify the generalizability of the proposed TGA framework, we applied it to an additional, distinct dataset from PhysioBank [3].

A. Procedure

This database comprises of walking stride interval time series of 15 subjects with 5 subjects in the groups:

- 1) Healthy young adults (Y-HC) (23–29 years old)
- 2) Healthy old adults (O-HC) (71–77 years old)
- 3) Older adults (60–77 years old) with PD

The experimental protocol involved subjects walking continuously on level ground around an obstacle free, path and the stride interval was measured using insole force sensitive resistors. The force signal was sampled at 300 Hz, converted

TABLE I: Clinical information and gait parameters of healthy controls.

Sub.	Age	Sev.	LStr (s)			RStr (s)			LStn (s)			RStn (s)			LSw (% GC)			RSw (% GC)		
			Mean	Std-Dev	CV	Mean	Std-Dev	CV	Mean	Std-Dev	CV	Mean	Std-Dev	CV	Mean	Std-Dev	CV	Mean	Std-Dev	CV
HC-1	57	0	0.73	0.04	5.57	0.69	0.03	4.63	1.07	0.04	3.81	1.07	0.04	3.52	32.39	2.07	6.39	35.55	1.60	4.51
HC-2	22	0	0.71	0.10	14.08	0.71	0.06	7.99	1.16	0.11	9.48	1.15	0.05	4.60	38.88	1.84	4.73	38.06	2.36	6.19
HC-3	23	0	0.70	0.02	3.56	0.75	0.03	3.78	1.09	0.03	3.02	1.09	0.04	3.36	35.71	1.21	3.40	31.32	1.14	3.63
HC-4	52	0	0.65	0.02	2.33	0.65	0.02	2.63	1.04	0.02	1.91	1.04	0.02	1.90	37.28	0.87	2.34	37.07	0.90	2.43
HC-5	47	0	0.70	0.05	6.86	0.72	0.05	6.80	1.11	0.05	4.88	1.11	0.05	4.83	37.13	1.99	5.36	34.84	2.29	6.57
HC-6	30	0	0.65	0.02	3.46	0.71	0.03	3.74	1.03	0.03	2.91	1.03	0.03	2.92	36.75	1.07	2.92	31.34	1.35	4.31
HC-7	22	0	0.67	0.02	3.63	0.67	0.02	2.73	1.07	0.03	2.85	1.07	0.03	2.88	37.45	0.74	1.99	36.81	0.89	2.42
HC-8	22	0	0.68	0.03	4.30	0.68	0.03	3.96	1.06	0.04	3.78	1.06	0.04	3.68	35.72	1.01	2.83	36.36	0.95	2.61
HC-9	32	0	0.64	0.03	4.75	0.65	0.03	4.08	1.01	0.04	3.69	1.01	0.04	3.81	36.67	1.33	3.62	36.02	0.98	2.72
HC-10	38	0	0.63	0.04	6.38	0.64	0.03	4.57	1.00	0.04	4.18	1.00	0.04	3.94	37.25	1.69	4.52	36.24	1.00	2.76
HC-11	69	0	0.69	0.03	4.33	0.67	0.03	5.15	1.04	0.04	3.56	1.04	0.04	3.56	33.16	1.07	3.23	35.34	1.45	4.09
HC-12	74	0	0.73	0.07	9.59	0.72	0.05	6.66	1.14	0.07	6.50	1.14	0.06	5.10	36.65	1.89	5.16	36.73	1.88	5.12
HC-13	61	0	0.70	0.03	4.56	0.70	0.04	5.35	1.11	0.04	3.52	1.11	0.04	3.78	37.08	1.28	3.45	36.52	2.18	5.96
HC-14	20	0	0.71	0.04	5.50	0.71	0.04	5.19	1.12	0.05	4.47	1.12	0.05	4.79	36.65	1.55	4.22	36.05	1.36	3.76
HC-15	20	0	0.89	0.04	4.81	0.89	0.05	5.91	1.40	0.07	5.00	1.40	0.07	5.07	36.17	1.24	3.44	36.14	1.07	2.95
HC-16	40	0	0.74	0.07	9.87	0.72	0.06	8.29	1.11	0.08	7.45	1.11	0.08	6.91	33.82	1.88	5.56	35.30	1.66	4.71
Group Mean			0.70	0.04	5.85	0.71	0.04	5.09	1.10	0.05	4.44	1.10	0.04	4.04	36.17	1.42	3.95	35.61	1.44	4.05

TABLE II: Clinical information and gait parameters of people with PD.

Sub.	Age	Sev.	LStr (s)			RStr (s)			LStn (s)			RStn (s)			LSw (% GC)			RSw (% GC)		
			Mean	Std-Dev	CV	Mean	Std-Dev	CV	Mean	Std-Dev	CV									
PD-1	77	4	0.74	0.05	6.32	0.78	0.05	7.01	1.13	0.04	3.69	1.13	0.05	4.26	34.98	3.06	8.74	31.56	3.59	11.38
PD-2	44	1.5	0.63	0.04	6.55	0.64	0.04	6.97	1.01	0.06	5.54	1.01	0.06	5.56	37.81	1.07	2.83	36.07	2.08	5.76
PD-3	80	2	0.81	0.10	11.94	0.86	0.10	11.74	1.21	0.11	9.36	1.21	0.11	9.30	32.97	2.08	6.31	28.64	2.57	8.99
PD-4	74	3.5	0.86	0.05	6.34	0.77	0.06	7.60	1.25	0.08	6.38	1.25	0.08	6.64	30.85	3.71	12.02	38.29	2.23	5.82
PD-5	75	2	0.71	0.04	6.15	0.72	0.04	5.92	1.06	0.05	4.77	1.06	0.05	4.94	32.92	1.95	5.92	31.85	2.30	7.24
PD-6	53	2	0.66	0.05	7.39	0.67	0.04	6.03	1.04	0.05	4.59	1.04	0.05	4.90	35.91	3.63	10.11	35.56	2.39	6.73
PD-7	64	4	0.84	0.60	70.99	0.82	0.23	28.57	1.23	0.59	47.64	1.22	0.34	28.11	32.98	3.76	11.39	32.75	3.21	9.80
PD-8	64	4	0.97	0.10	10.54	0.95	0.11	11.48	1.37	0.14	10.26	1.36	0.15	10.67	28.80	4.30	14.94	30.42	3.51	11.52
PD-9	68	1.5	0.80	0.09	11.28	0.81	0.10	12.79	1.25	0.11	8.81	1.25	0.11	8.72	36.27	1.45	4.00	35.35	2.15	6.08
PD-10	60	3	0.64	0.06	9.97	0.67	0.05	7.71	0.97	0.08	8.21	0.97	0.07	7.57	33.51	2.39	7.14	30.52	3.11	10.19
PD-11	74	3	0.90	1.71	189.80	0.90	1.42	158.72	1.21	1.73	142.98	1.15	1.42	123.23	29.98	5.13	17.12	24.97	4.69	18.79
PD-12	57	3	0.73	0.04	6.05	0.77	0.05	6.87	1.13	0.05	4.64	1.13	0.05	4.70	35.33	1.74	4.93	32.24	2.25	6.97
PD-13	79	3	0.74	0.12	15.97	0.75	0.10	13.28	1.11	0.11	9.99	1.10	0.11	9.63	33.17	4.00	12.07	31.88	2.54	7.96
PD-14	57	3	0.69	0.09	12.46	0.60	0.06	10.35	1.00	0.13	12.78	1.01	0.20	19.46	30.98	2.58	8.33	39.63	6.12	15.43
PD-15	76	2.5	0.79	0.05	5.93	0.81	0.05	6.45	1.17	0.05	4.46	1.17	0.05	3.88	32.79	3.21	9.78	30.63	3.19	10.41
Group Mean			0.77	0.21	25.18	0.77	0.17	20.10	1.14	0.23	18.94	1.14	0.19	16.77	33.28	2.94	9.04	32.69	3.06	9.54

TABLE III: Clinical information and gait parameters of people with HD.

Sub.	Age	Sev.	LStr (s)			RStr (s)			LStn (s)			RStn (s)			LSw (% GC)			RSw (% GC)		
			Mean	Std-Dev	CV	Mean	Std-Dev	CV	Mean	Std-Dev	CV									
HD-1	42	8	0.56	0.04	6.67	0.54	0.03	6.28	0.90	0.05	5.74	0.90	0.05	5.69	38.19	2.87	7.52	39.49	2.32	5.89
HD-2	41	11	0.81	0.08	9.50	0.81	0.07	8.55	1.23	0.10	8.43	1.23	0.10	8.19	34.17	2.17	6.35	34.58	2.96	8.56
HD-3	66	4	0.79	0.09	11.27	0.77	0.07	9.64	1.20	0.12	9.80	1.20	0.13	10.55	34.16	3.92	11.48	35.40	3.95	11.15
HD-4	47	2	0.67	0.08	11.56	0.78	0.07	9.41	1.04	0.09	8.74	1.04	0.08	7.22	35.91	4.47	12.43	25.37	5.45	21.46
HD-5	36	10	0.65	0.09	13.10	0.66	0.08	12.41	1.06	0.12	11.42	1.05	0.09	8.50	38.48	2.49	6.48	37.08	2.70	7.28
HD-6	41	8	0.67	0.07	10.64	0.66	0.06	9.25	1.06	0.08	7.20	1.06	0.09	8.39	36.16	3.43	9.48	37.20	2.53	6.81
HD-7	71	2	0.81	0.07	8.15	0.79	0.06	7.26	1.20	0.10	7.98	1.20	0.09	7.82	32.27	3.85	11.94	34.33	3.14	9.14
HD-8	53	9	0.68	0.07	9.60	0.69	0.05	6.60	1.09	0.07	6.62	1.08	0.05	4.93	37.72	1.62	4.30	36.44	1.35	3.71
HD-9	54	12	0.65	0.07	10.25	0.70	0.07	10.36	1.03	0.07	6.86	1.03	0.08	7.41	36.86	1.82	4.93	31.98	2.23	6.99
HD-10	47	4	0.83	0.06	6.89	0.85	0.06	7.32	1.26	0.08	6.40	1.26	0.08	6.23	34.12	2.45	7.19	32.73	2.36	7.19
HD-11	33	11	0.74	0.06	8.13	0.87	0.09	10.29	1.16	0.07	6.21	1.16	0.07	6.23	36.04	2.78	7.73	25.46	6.09	23.91
HD-12	47	8	0.70	0.07	10.05	0.72	0.08	11.22	1.08	0.10	8.88	1.08	0.10	9.55	35.31	3.59	10.16	33.40	4.38	13.12
HD-1																				

TABLE IV: Clinical information and gait parameters of people with ALS.

Sub.	Age	Sev.	LStr (s)			RStr (s)			LStn (s)			RStn (s)			LSw (% GC)			RSw (% GC)		
			Mean	Std-Dev	CV	Mean	Std-Dev	CV	Mean	Std-Dev	CV									
ALS-1	68	1	0.87	0.33	38.10	0.91	0.32	34.91	1.30	0.33	25.74	1.30	0.34	25.92	33.32	3.81	11.42	30.59	3.01	9.83
ALS-2	63	14	0.76	0.02	2.30	0.75	0.02	3.11	1.15	0.02	1.84	1.15	0.02	1.77	33.66	0.95	2.81	34.78	1.65	4.74
ALS-3	70	13	0.85	0.47	55.57	0.91	0.52	57.44	1.29	0.48	37.13	1.30	0.52	40.24	35.14	4.50	12.80	30.81	4.20	13.64
ALS-4	70	54	1.64	3.25	198.42	1.58	3.13	197.38	2.06	3.23	156.62	2.05	3.10	151.06	25.98	5.42	20.84	28.66	5.94	20.72
ALS-5	36	5.5	0.86	0.55	64.34	0.85	0.09	10.11	1.31	0.56	42.92	1.27	0.09	7.02	35.14	3.43	9.76	33.43	2.03	6.08
ALS-6	43	17	1.12	0.10	8.87	1.15	0.11	9.40	1.58	0.10	6.49	1.57	0.10	6.65	28.88	2.57	8.92	27.07	2.77	10.24
ALS-7	65	9	1.28	0.11	8.87	1.22	0.11	9.31	1.75	0.13	7.42	1.75	0.12	6.93	27.07	2.00	7.38	30.46	2.58	8.46
ALS-8	51	3	0.77	0.05	6.80	0.79	0.06	7.65	1.20	0.07	6.14	1.20	0.07	6.03	35.19	2.36	6.71	34.06	1.97	5.77
ALS-9	50	54	0.84	0.06	7.36	0.85	0.15	17.46	1.31	0.08	6.49	1.32	0.16	12.41	35.67	1.24	3.47	35.78	2.28	6.38
ALS-10	40	14.5	0.77	0.03	4.16	0.76	0.02	3.14	1.13	0.03	3.03	1.13	0.03	2.90	32.50	1.19	3.66	32.56	1.19	3.65
ALS-11	39	7	0.82	0.05	6.27	0.80	0.05	6.55	1.22	0.06	5.10	1.22	0.06	4.66	32.68	1.71	5.22	33.88	1.63	4.82
ALS-12	62	12	1.62	5.39	333.68	1.65	5.33	322.61	2.28	5.82	254.89	2.33	5.95	255.51	33.34	8.06	24.17	31.47	7.35	23.36
ALS-13	66	34	1.04	0.09	8.85	1.02	0.08	7.69	1.52	0.09	6.10	1.52	0.10	6.29	31.47	2.79	8.88	32.70	2.59	7.92
Group Mean			1.02	0.81	57.20	1.02	0.77	52.83	1.47	0.85	43.07	1.47	0.82	40.57	32.31	3.08	9.70	32.02	3.01	9.66

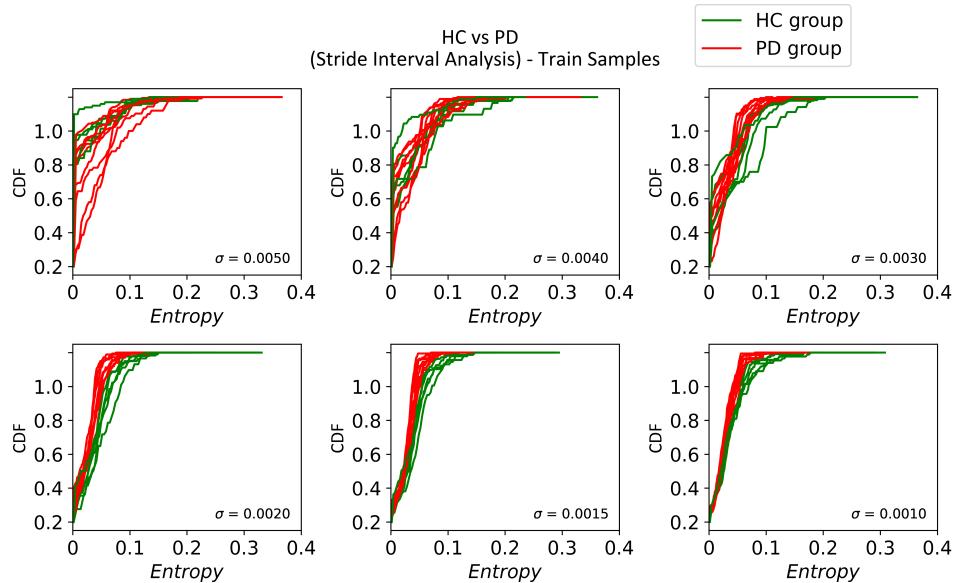


Fig. 3: Persistence entropy plots for stride interval analysis corresponding to the training set. $\sigma = 0.0020, 0.0015$, and 0.0010 offer separation between HC and PD groups. We choose $\sigma = 0.0015$ for checking validity with the validation set.

to a digital signal using 12 bit A/D converter using an ambulatory, ankle-worn micro-computer that also recorded the data. Subsequently, the time between foot-strokes was automatically computed. Data from the healthy subjects collected as subjects walked in roughly circular path for 15 minutes. Data from the subjects with PD was collected for 6 along a long hallway. For the purpose of subsequent analysis, the first 5 minutes of data from each subject was used.

B. Analysis

The reconstruction of the configuration space from the raw time-series involves the use of invoking Taken's theorem [5] which requires evaluation of the time-delay parameter τ and embedding dimension d as described in §II of the main manuscript. In the main manuscript, we assume $d = 2$ and use the left and right foot time-series ($\eta_L(\cdot), \eta_R(\cdot)$ respectively) to construct the configuration space as $M := ((\eta_L(n), \eta_R(n))_{n=1}^N$ which subsumes and exploits the inherent time-delay between

the feet movements. For this dataset, however, only the time-series data for a single foot was available. Therefore, we fixed $d = 2$ and determined an optimal time-delay τ for the dataset.

To determine the optimal τ for configuration space reconstruction, we adopted an empirical strategy. We determined an optimal τ_s for each subject s by minimizing the mutual information [6]. Once the optimal τ_s was determined for each subject, the average of all τ_s values was used as the optimal time-delay τ for the entire dataset.

Thus, if the raw time-series of a subject s is denoted by $\mathbb{N}^* \in n \mapsto y_s(n) \in \mathbb{R}$, then the configuration space $M := ((y_s(n), y_s(\tau+n))_{n=1}^N$, where N is the number of strides used for analysis.

C. Results

The results post implementation of the TGA framework on this new dataset have been encapsulated in Figure 6. CDF

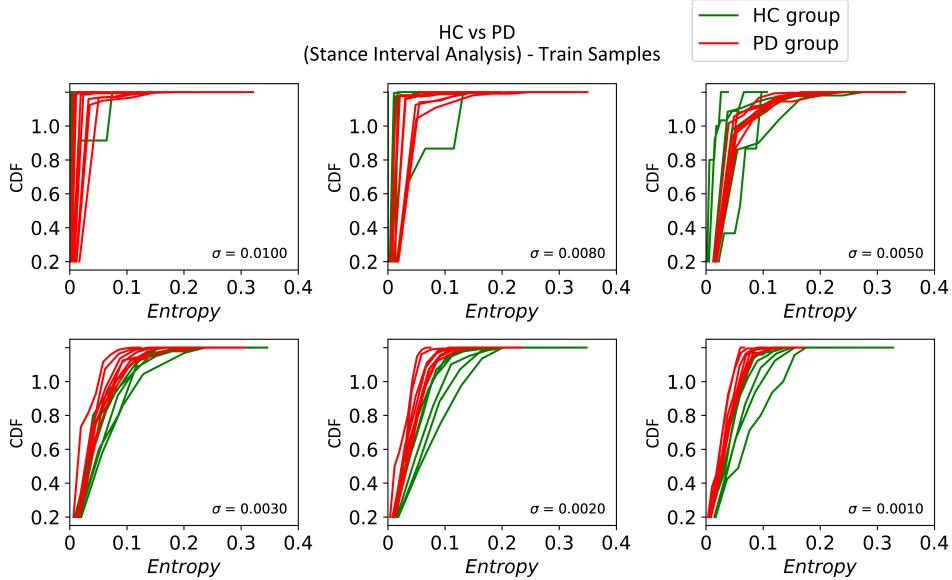


Fig. 4: Persistence entropy plots for stance interval analysis corresponding to the training set. $\sigma = 0.0030, 0.0020$, and 0.0010 offer separation between HC and PD groups. We choose $\sigma = 0.0020$ for checking validity with the validation set.

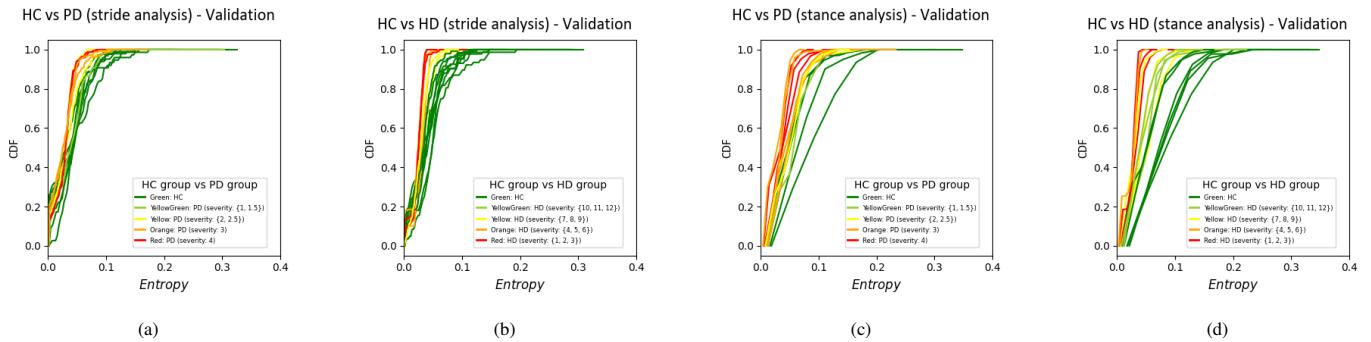


Fig. 5: Validation comparison for stride and stance interval analysis between PD, HD, and HC groups using $\sigma = 0.015$ for stride analysis and $\sigma = 0.0020$ for stance analysis. The chosen σ values effectively identify trends based on disease severity in the validation set, which is entirely disjoint from the training set. This demonstrates that the identified patterns are true trends and not a result of parameter overfitting.

TABLE V: Best-fit α_s and E_s values for healthy young adults.

Subject	Age	α_s (stride)	E_s (stride)
Y-1	23	20.9	0.039
Y-2	29	30.2	0.038
Y-3	23	28.8	0.045
Y-4	21	34.8	0.030
Y-5	26	40.6	0.035
Mean \pm SD		31.06 ± 7.31	0.037 ± 0.006

plots of persistence entropy reveal stark differences between PD and HC gait. Specifically, there appears to be a trend with O-HC curves closer to PD curves as opposed to Y-HC curves, indicating deterioration in gait with aging. Clearly this shows that the framework captures the deterioration in gait that occurs due to aging.

Tables VI-VII demonstrate that the scalars α_s and E_s not only distinguish between healthy and neurodegenerative gait

TABLE VI: Best-fit α_s and E_s values for healthy old adults.

Subject	Age	α_s (stride)	E_s (stride)
O-1	76	42.7	0.026
O-2	74	39.5	0.018
O-3	75	26.4	0.022
O-4	77	36.6	0.025
O-5	71	33.6	0.040
Mean \pm SD		35.76 ± 6.11	0.026 ± 0.008

TABLE VII: Best-fit α_s and E_s values for PD subjects.

Subject	Age	α_s (stride)	E_s (stride)
PD-1	-	43.1	0.020
PD-2	-	41.6	0.017
PD-3	-	53.7	0.012
PD-4	-	48.1	0.022
PD-5	-	40.4	0.025
Mean \pm SD		45.38 ± 5.10	0.019 ± 0.005

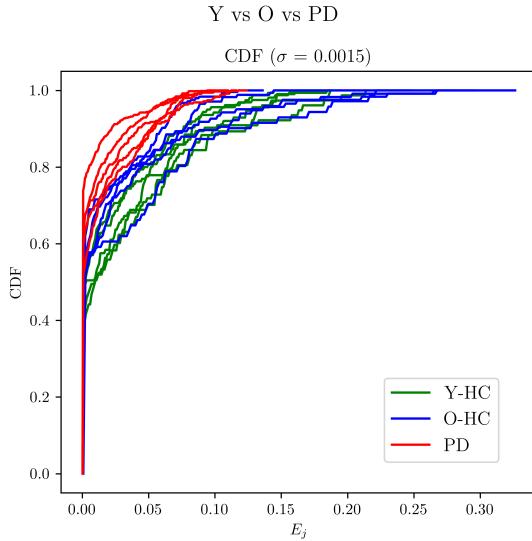


Fig. 6: Persistence entropy plots clearly distinguish the healthy group from PD subjects. Additionally, the plots show a trend where curves for older adults are closer to those of PD subjects, while curves for younger adults are further apart. This indicates that the method effectively captures gait deterioration due to aging as well as changes due to neurodegeneration.

but also reveal significant differences between young and old healthy adults. This indicates that the TGA framework generalizes well across different gait parameters and datasets and can effectively identify gait deterioration due to aging.

VI. DATA LENGTH AND STRIDE REQUIREMENTS FOR TGA FRAMEWORK

Sections V shows that the TGA framework generalizes to different datasets and distinct gait parameters. However, both the datasets [1] and [3] analysed had a key advantage of comprising of long trials with a sufficiently large number of strides for the TGA framework to discern between healthy and neurodegenerative gait and identify clear trends with disease severity.

The practical challenges faced by patients in advanced stages of diseases, such as PD, where long trial durations may not be feasible, are recognized. Constructing a density estimate from time-series data typically benefits from longer trials, as they provide more data points, resulting in a more accurate density estimate and better overall analysis. The framework naturally extends to accommodate longer trials.

To determine the minimum trial length required for reasonable analysis, a heuristic strategy is adopted. The fraction of the total number of strides used for analysis is defined as ρ . For example, $\rho = 0.4$ means that only the first 40% of the total strides are used for analysis across all subjects. The final output for various values of ρ is then compared with the baseline where $\rho = 1$ (i.e., all available strides in the dataset are used). The value of ρ at which the results begin to closely resemble those obtained with $\rho = 1$ serves as an estimate of the minimum fraction of the trial length required for the

framework to produce reliable results.

As shown in Figure 7, for $\rho = 0.2$ and $\rho = 0.4$, there was no clear separation between the CDF of persistence entropy for healthy subjects and those with neurodegenerative conditions across all three groups. For $\rho = 0.6$, the plots begin to resemble those obtained from the entire dataset (i.e $\rho = 1$) for most subjects, while the results for $\rho = 0.8$ are nearly identical to the analysis of the complete dataset. Hence, we contend that a minimum of 60% of the 5-minute walking trial (typically comprising of 300 strides), i.e., 3 minute walking trial comprising of 180 strides are required to make reasonable claims using the proposed TGA framework. For subjects unable to walk for this duration independently, this framework may not be suitable and could lead to erroneous assessments. However, it is important to note that the recorded data used for analysis need not be part of a single trial comprising of continuous strides. Instead, the strides can be collected across several trials with appropriately spaced breaks to aid patients with severe gait impairment. This could enable in extending the utility of this framework to severely affected subjects as well.

REFERENCES

- [1] J. M. Hausdorff et al., "Dynamic markers of altered gait rhythm in Amyotrophic Lateral Sclerosis," *Journal of Applied Physiology*, vol. 88, no. 6, pp. 2045–2053, 2000.
- [2] A. L. Goldberger et al., "Physiobank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," pp. 215–220, 2000. [Online]. Available: <https://physionet.org/content/gaitndd/1.0.0/>
- [3] ——, "Physiobank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," pp. 215–220, 2000. [Online]. Available: <https://physionet.org/content/gaitdb/1.0.0/>
- [4] D. N. Kozlov, *Organized Collapse: An Introduction to Discrete Morse Theory*, ser. Graduate Studies in Mathematics. American Mathematical Society, 2021, vol. 207.
- [5] E. Bradley and H. Kantz, "Nonlinear time-series analysis revisited," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 9, 2015, paper no. 097610.
- [6] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical review A*, vol. 33, no. 2, p. 1134, 1986.

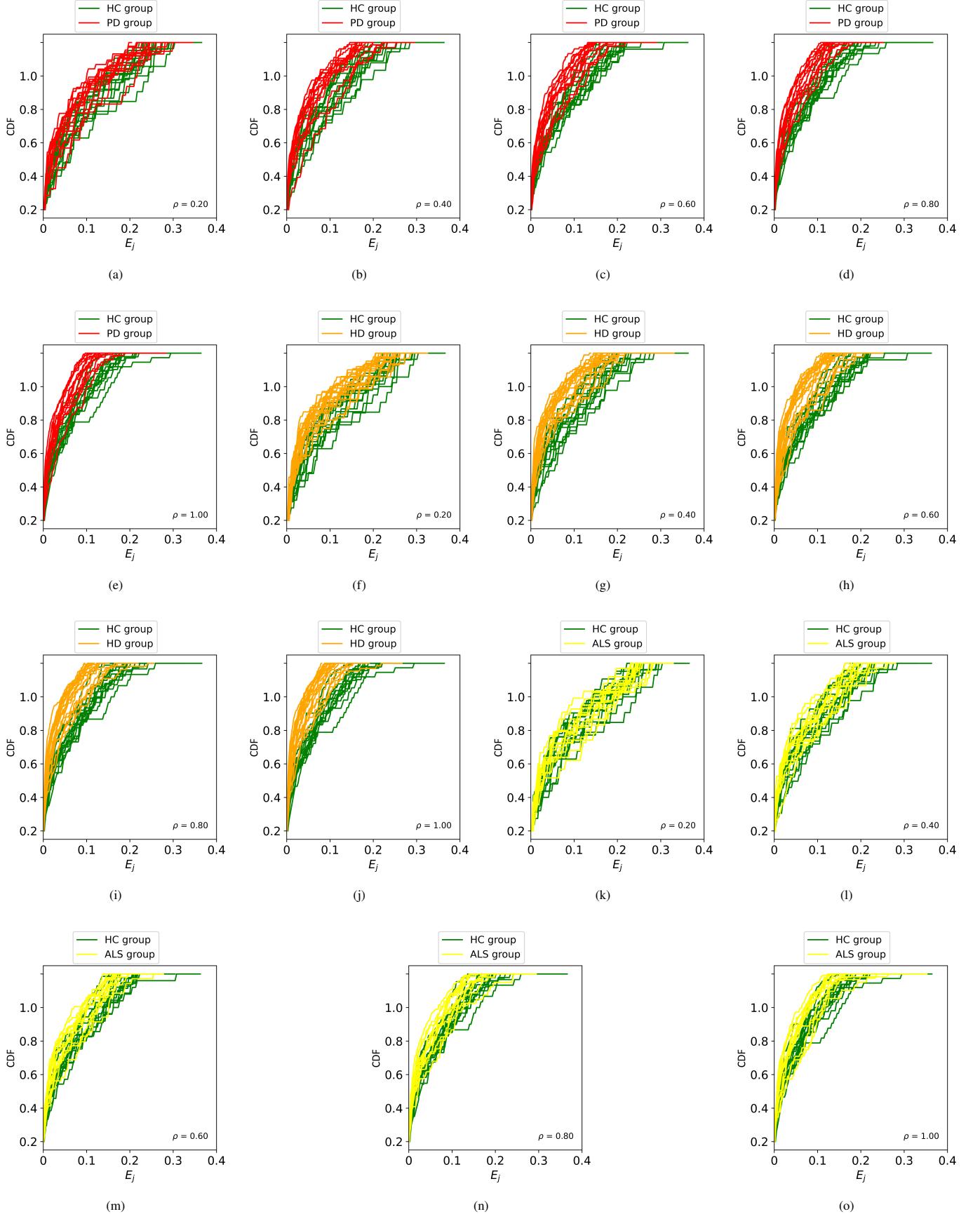


Fig. 7: Persistence entropy plots from the TGA framework using various dataset fractions. Clear trends emerge for $\rho = 0.6$, but not for $\rho = 0.2$ or 0.4 . Thus, at least 60% of a 5-minute walking trial (3 minutes or 180 strides) is required for accurate TGA results.