Uncovering the topology of the temporal region in Alzheimer's disease.

Philip Hartout[*]
Department of Biosystems Science and Engineering
ETH Zürich
Zürich, Switzerland
`phartout@ethz.ch`

Abstract

Topological data analysis is an exponentially growing field due to the insights that can be gleaned on real-world datasets using rigorously developed mathematical theory that can scale on today's computers. Here, we apply a topological data analysis pipeline to uncover novel insights in the Alzheimer's disease neuroimaging initiative dataset. In particular, we highlight the topological heterogeneity of the diagnostic categories in the dataset, and discuss how having knowledge of such heteogeneity can prevent misdiagnosis, either by humans or state-of-the-art trained machine learning classifiers.

1 Introduction

1.1 Alzheimer's disease

Alzheimer's disease (AD) is the most prevalent form of dementia in the world, with a forecasted 75 million cases in 2030 and 132 million by 2050 [WHO, 2017]. In EU member states and Switzerland, AD is already among the leading causes of death, and is projected to further accelerate in the future [Sleeman et al., 2019]. The associated costs are immense — in the United States alone, the cost of care of AD patients is expected to be \$2 trillion by 2030 —, and is poised to substantially burden the economic prosperity of developed countries in the future [WHO, 2017].

Crucial to the path of finding a working solutions for patients developing Alzheimer's disease is diagnosing the pathology early, so that nerve damage is minimized once treatment has been found [Yiannopoulou and Papageorgiou, 2020]. Although the definite diagnosis of a patient with Alzheimer's disease can only be done post-mortem, clinicians use a plethora of standardized tools to find indications of the developing pathology as early as possible, ranging from neuropsychological tests, blood and cerebrospinal fluid markers, and MRI images [McKhann et al., 2011, Lehmann and Teunissen, 2016, Smits et al., 2012].

All of these techniques rely on early manifestations of the disease. Although there are disputes on the root cause of the disease in the late onset form of Alzheimer's disease [Hur et al., 2020, Fulop et al., 2018, Tharp and Sarkar, 2013], there is still wide consensus that the presence of Amyloid β (Aβ), which stems from cleavage of amyloid precursor protein (APP), together with the aggregation of neurofibrillary tangles, stemming from hyperphosphorylated tau proteins, accumulate in the brain of patients with Alzheimer's disease, and leads to neural cell death [Dá Mesquita et al., 2016]. This cellular destruction leads to an cummulative effect: brain atrophy, particularly in brain regions involved in memory formation such as the medial temporal lobe (MTL), which contains the enthorinal cortex, the hippocampus, and the amygdala [Goedert and Spillantini, 2006]. These changes can be so

---

pronounced that they are observable using structural Magnetic Resonance Imaging (sMRI) [Frisoni et al., 2010].

Such images provide an extremely rich source of data, which can then be used for various purposes. One of them is classification, which has been shown to successfully discriminate between cognitively normal (CN) versus AD patients using deep learning techniques such as convolutional neural networks [Wen et al., 2020]. Additionally, this has lead to the identification of multiple regions being affected by the disease, and patterns have emerged as to which groups of regions are affected, leading to the definition of various subtypes of AD [Poulakis et al., 2018]. To gain even finer insights from the fine but observable alterations of brain shape in the context of Alzheimer's disease, the mathematical field of topology, which studies the properties of geometric objects under continuous deformations, such as stretching, twisting, crumpling and bending, which is exactly the type of deformations occurring in the context of Alzheimer's disease.

1.2 Topology

Topology has witnessed relentless theoretical progress since Henri Poincaré first addressed topological ideas as a distinct branch of mathematics in his 1895 publication of *Analysis Situs* [Poincaré, 1895, James, 1999]. Only recently, however, with the advent of modern computing, has the field of computational topology and topological data analysis (TDA) gained momentum to investigate complex patterns as shapes observed in physics, biology, and beyond [Ghrist, 2008, Dey et al., 1999, Amézquita et al., 2020]. While surveying the various applications of computational topology is beyond the scope of this report, we do want to focus on a number of techniques that are paramount to the workflow described in this report: cubical persistence, various vectorized representations of the persistence diagrams obtained from filtered cubical complexes, and the notion of pairwise distance between such representations. Note that we have not an extensive and formal introduction to topology and persistent homology. For such material, we refer to [Freedman and Chen, 2009, Edelsbrunner and Harer, 2010, Ghrist, 2008].

1.2.1 Cubical complexes and persistent homology

Prior to defining the notion of cubical persistence, *cubical complexes* need to be defined. For that, let us first assign to each elementary non-degenerate interval $[a, a+1] \forall a \in \mathbb{R}$ two degenerate intervals $[a, a]$ and $[a+1, a+1]$. For a d-dimensional space, a cube is then defined as a product of d elementary intervals $\Pi_{i=1}^{d} I_i$. The dimension of the cube is then equal to the number of non-degenerate interval in the aforementioned product, such that 0-cubes, 1-cubes, 2-cubes, and 3-cubes correspond to vertices, edges, squares, and 3D cubes, respectively. In this report, the 3D cube corresponds to a given voxel of our sMRI..

A cubical complex X of dimension d is then defined as a finite set of elementary cubes of at most dimension d, where X must be closed undertaking faces and intersections, i.e. for any cube in X, all of its faces must belong to X, and for any two cube in X, their intersection is either empty, or there is a common face between them.

Given a topological space $\mathbb{X}$ (in our case, the sMRI voxels), we use a filtering function $f : \mathbb{X} \to \mathbb{R}$, which here corresponds to the pixel density of each voxel, to study the topology of the sublevel sets $\mathbb{X}_t = f^{-1}(-\infty, t]$ of cubical complexes that arise. This is a method to study topological spaces referred to as *persistent homology*. A common representation of the evolution of topological complexes as a function of the value of the filtration function is the *persistence diagram* (PD), which is a multiset of points. For each homological dimension (here, $1, 2, 3$), we obtain a collection of points, with an associated x and y coordinate which corresponds to the birth and death of a topological feature in homology dimension $n = 1, 2, 3, \ldots$, respectively. We refer to a homological feature as *persistent* if the difference between its birth and death value with respect to the given filtration function is particularly high.

1.2.2 Persistence images and persistence landscapes

PDs are endowed with a metric space; the p-Wasserstein distances can be computed between any two persistence image, and these metrics are stable under small perturbations of the data. So, it is possible to perform a variety of ML techiques using PDs as a statistic for clustering. However, multiple ML algorithm require more than a metric, and the computation cost of the p-Wasserstein

distance increases linearly with the number of points in the PD. It is therefore desirable to have a stable, efficient-to-compute, interpretable (with respect to the PD) and tunable mapping from the PD to a vector space in $\mathbb{R}^n$.

One such representation is the *persistence image* (PI) of a PD, which as been proven to be stable upon small perturbations of data while still retaining the underlying features in the data useful for classification [Adams et al., 2017]. Computing the PI from a PD D consists of a two step process. First, the PD is mapped to an integrable function $\rho_D : \mathbb{R}^2 \rightarrow \mathbb{R}$ called a persistent surface. This surface is a weighted sum of Gaussian distributions, each centered around a point of the PD. The matrix of pixel values can be obtained from the computation of the integration of ρ_D on a grid overlaid on the surface [Adams et al., 2017].

Another representation associated to the PD is the persistence landscape (PL) [Bubenik, 2015, 2020]. This representation maps the PD into a Hilbert space, which is useful for ML applications. In order to define a persistence landscape, let us take a pair (b, d), which refer to the birth and death of a feature. We now define the piecewise linear function $f_{(b,d)} : \mathbb{R} \rightarrow [0, \infty]$ as

$$f_{(b,d)}(x) = \begin{cases} 0 & \text{if } x \notin (b, d) \\ x - b & \text{if } x \in (b, \frac{b+d}{2}] \\ -x + b & \text{if } x \in (\frac{b+d}{2}, d] \end{cases} \tag{1}$$

The PL of the birth-death pairs $\{b_i, d_i\}_{i=1}^n$ is the sequence of functions $\lambda_k : \mathbb{R} \rightarrow [0, \infty]$, $k = 1, 2, 3, \ldots$ where $\lambda_k(x)$ is the kth largest value of $\{f_{b_i, d_i}(x)\}_{i=1}^n$. We set $\lambda_k(x) = 0$ if the kth largest value does not exist, which results in $\lambda_k = 0$ for $k > n$.

1.2.3 Pairwise distances and means

A crucial element in our investigations is the concept of distance between vectorized topological representations, so let us examine this subject further. Intuitively, and as noted by [Berwald et al., 2018], it important to take the meaning of the points of the PD into account; namely that a point close to the diagonal $(c, c + \epsilon)$ represents a feature that lived for a short time ϵ. A diagram with this small lifetime point should therefore be close to the same diagram without that point, where the feature wouldn't appear at all. Hence, it makes sense to introduce the notion of minimal cost required to match up points of the two diagrams, either off-diagonal to off-diagonal, or off-diagonal to the nearest point on the diagonal (for small values of ϵ). In this context, distance functions usually applied to evaluate the distance between two probability density estimations are relevant: the bottleneck distance and the p-Wasserstein distance, where $p \geq 1$. The p-Wasserstein distance between two diagrams D_1 and D_2 is the infimum over all bijections $\gamma : D_1 \cup \Delta \rightarrow D_2 \cup \Delta$, where Δ is the multiset $\{(s, s) \mid s \in \mathbb{R}\}$ with multiplicity $(s, s) \mapsto +\infty$, of

$$\left(\sum_{x \in D_1 \cup \Delta} ||x - \gamma(x)||_q^p \right)^{1/p} \tag{2}$$

where we usually have $q = \infty$. When we let $p \rightarrow \infty$, we recover the bottleneck distance.

We also use the notion of an average persistence landscape, where given a collection of PD, we compute their associated PL, which is a matrix of fixed dimension $m \times h$ where m is the length of the vector, and h is the homology dimension. We compute the average PL by taking the mean over all samples for each cell in m for each homology dimension h, so we end up with a PL representative of the collection of the samples in the collection.

1.3 Research questions and outline

In this report, we answer the following research question: how can the topological feature extraction performed on a patch of the MTL help us avoid misdiagnoses, or characterize categories of patients at risk of misdiagnosis, either by a state-of-the art machine learning model or a human?

This report is structured as follows: after having introduced AD and fundamental concepts related to TDA in this introduction, we go on to present and justify methodological choices we have made regarding the topological data analysis conducted on sMRI data in section 2; in section 3, we report the findings extracted from the data, which we discuss in section 4.
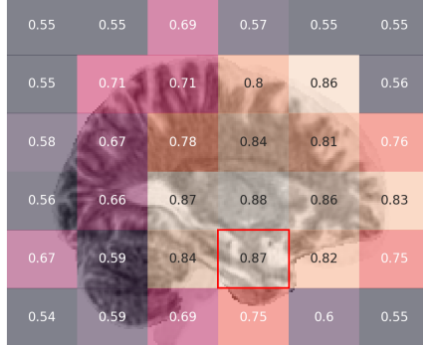
Figure 1: AUCPRC on each patch, achieved using a model described in earlier work. The chosen patch for the topological data analysis is boxed in red.

2 Methods

Here, we present and justify the specific methodological choices made for this pipeline. All of the code used to compute the findings presented in this paper is currently available upon request on GitHub.

2.1 ADNI data

AD and healthy controls of matched age groups (cognitively normal, CN) T1-weighted, 1.5 Tesla sMRIs were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [Jack Jr et al., 2008]. Further preprocessing steps to reduce noise and extract brains structures is highlighted in appendix A.1. Then scans were divided into 216 patches, each of dimension $(30 \times 36 \times 30)$, providing a possibility for a far more focused and computationally efficient investigation while preserving high resolution. In this report, the choice of working with a patch is also supported by the fact that an investigation of *local* changes in brain architecture may reveal more than a brain scan-wide analysis, and hasn't been performed before.

From earlier work attempting to classify CN subjects from AD patients using a convolutional neural network (CNN), we know that a given patch, shown in grey in figure 1, has a particularly high discriminatory potential, so we selected this patch for all our further analyses. Support for the use of this patch also comes from its anatomical relevance, since it contains regions that are relevant in the context of Alzheimer's disease such as the hippocampus, the enthorinal cortex, and the amygdala [Goedert and Spillantini, 2006].

2.2 TDA setup

To perform the topological analysis on the patch, we used `giotto-tda`, a library specifically made for the integration of TDA pipelines for ML [Tauzin et al., 2020]. Each filtration on the cubical complexes has been done in three homological dimensions $1, 2, 3$, which makes sense given the patch is a three-dimensional image. We otherwise used the default parameters provided by the giotto-tda documentation.

We also used this library to compute both the persistence images and persistence landscapes. To compute the persistence images, we used 0.05 as a standard deviation for the Gaussian kernel, no weight function, and the default dimension of 100×100. For the landscapes, we wanted to keep only the most prominent features, so we kept only λ_1, and set the PL vector lengths of 100.

Computing the average persistence landscape for each diagnostic category was done using the mean of each subject for each of the vector coordinate. The pairwise distance between two PLs was taken using the l^1 norm.

Note that we compute the distances in two settings:

Intra-patient distance : this allows us to assess the distance of the different PDs of the same patient, to see if there is any evolution over time.
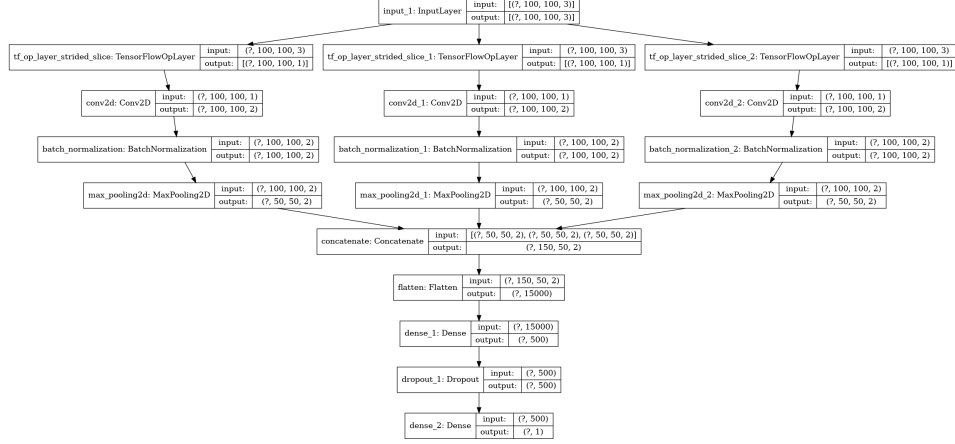
4

Figure 2: Computation graph to predict the phenotype of a given set of persistent images.

Intra-diagnostic category setting : here, we compute the distance of each PL computed from the PD available for a diagnostic category with respect to the mean PL of that image.

2.3 Model architecture

For the ML classification task of classifying AD vs CN patients, we used a parallel CNN network, followed by one dense layers containing 500 neurons and with dropout rates of 50% at training time together. The output of the last dense layer is redirected to a single sigmoid neuron for prediction. The model was trained using an exponential decay learning rate scheduler and early stopping, which monitored the validation loss. All of the layers and utilities to train the neural network were provided by the Keras library [Chollet et al., 2015] and are available on the repository, and a depiction of the computation graph is shown in Figure 2. We also note that the model was trained on commodity hardware, i.e. a Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz for less than a minute per model, highlighting the computational efficiency of the approach.

3 Results

Here we present the results obtained from the above mentioned pipeline, starting with a qualitative assessment of the topological representations obtained from the TDA on the patch from the MTL followed by a performance assessment of the deep learning model. We then turn our attention to the topological heterogeneity observed both within each diagnostic category and then within each patient. Finally, we try to identify patients, so-called topological outliers, which could be potentially misclassified by either the deep learning model or humans.

3.1 Qualitative assessment of the TDA

We start by examining an example of the persistence diagram, image and landscape for a randomly sampled CN, MCI, and AD patient, shown in Figure 3. No significant difference is immediately apparent from these diagrams, but when we compute their associated PI, we see differences emerging across homological dimensions, as we can observe in figure 4. These differences seem to be salient features in the context of machine learning, should these differences exist in multiple patients.

3.2 Model Performance

In this report, we trained the deep learning model three times on three different partitions of the data to get an accurate picture of the performance of the model. The performance metrics of the deep learning model is shown in Table 1. The performance is comparable to state-of-the-art models trained on similar data [Wen et al., 2020].
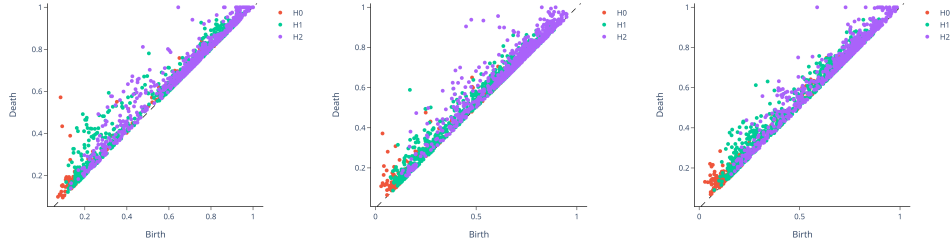
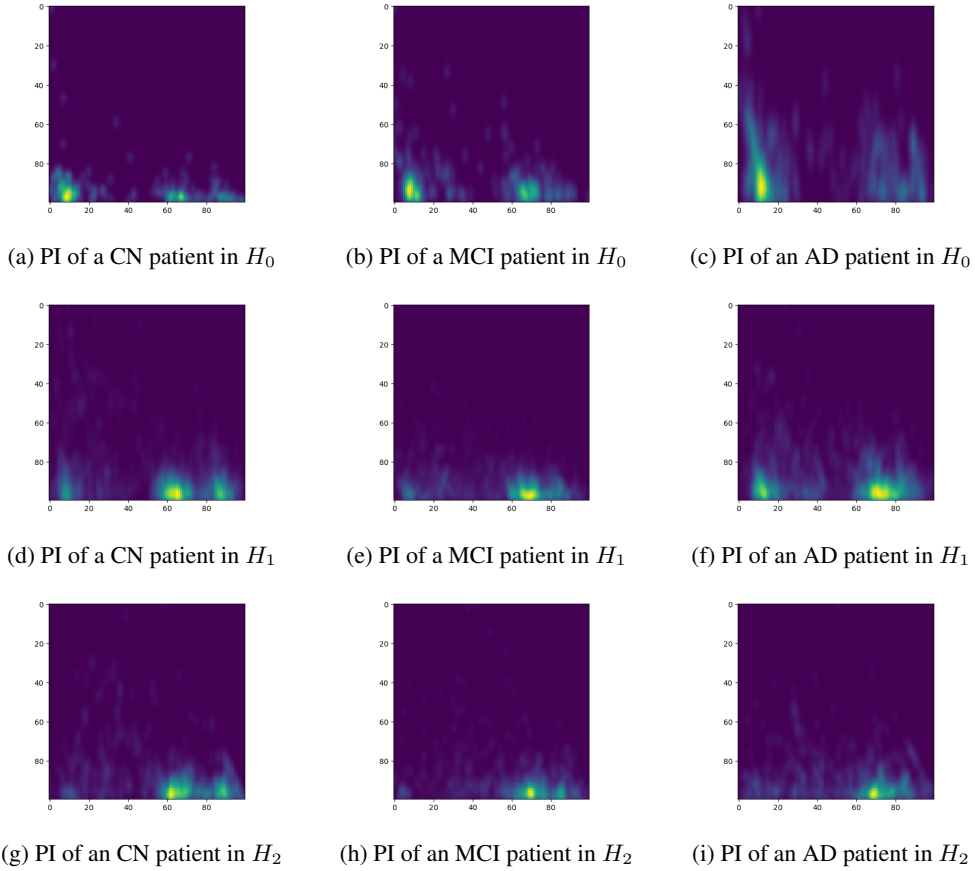Figure 3: Representative PD for each of the diagnostic categories.

(a) PI of a CN patient in H_0

(b) PI of a MCI patient in H_0

(c) PI of an AD patient in H_0

(d) PI of a CN patient in H_1

(e) PI of a MCI patient in H_1

(f) PI of an AD patient in H_1

(g) PI of an CN patient in H_2

(h) PI of an MCI patient in H_2

(i) PI of an AD patient in H_2

Figure 4: Representative PI for each of the diagnostic categories. Each column correponds to a diagnostic category whereas each row corresponds to a homological dimension.

Performance metric	DL model trained on PIs
Training accuracy	0.81 ± 0.01
Validation accuracy	0.78 ± 0.03
Precision	0.81 ± 0.04
Recall	0.77 ± 0.03
AUC	0.85 ± 0.03

Table 1: Performance metric of the deep learning model.

(a) Median PL for CN subjects   (b) Median PL for MCI subjects   (c) Median PL for AD subjects
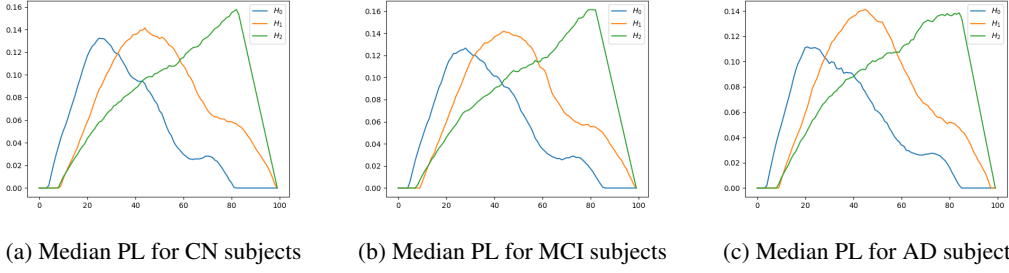
Figure 5: Median persistence landscapes for each of the diagnostic categories.

	Mean	Median	Standard deviation	Q3	Max	Skewness
CN H_0	2.16	2.00	0.78	2.50	7.41	1.78
CN H_1	2.61	2.27	1.17	2.93	9.47	1.92
CN H_2	2.38	2.23	0.88	2.79	7.19	1.39
MCI H_0	2.24	2.04	0.82	2.55	6.21	1.71
MCI H_1	2.57	2.19	1.29	2.80	11.87	2.57
MCI H_2	2.40	2.27	0.83	2.82	6.55	1.18
AD H_0	2.40	2.18	0.96	2.77	7.77	1.97
AD H_1	2.47	2.13	1.15	2.77	9.28	2.10
AD H_2	2.36	2.20	0.80	2.75	8.39	1.64

Table 2: Summary statistics of the distribution of distances shown in Figure 6

3.3 Topological outliers

3.3.1 Between patients

We now present our findings regarding the distribution of the distances between the PL of each image with respect to the median PL for each diagnostic category. The representative PL for each diagnostic category is shown in Figure 5, and the distribution of the l^1 norm between each patient and these median PLs is shown in Figure 6. As we can see, while the median PLs do not differ too much from one another in each of the homological dimensions, we see that some images seem to have a far greater distance to this median PL than the majority of PLs.

3.3.2 Within patients

As indicated in section , we can compute the distance between various PLs associated to the different timepoints available to a single patient to evaluate the distance between each of these topological representations at these timepoints, hence obtaining a representation of the topological evolution of that particular patient.

We find interesting qualitative results, with distance varying widely from one patient to the next. For instance, if we take a CN patient diagnosed as such throughout the time that that patient, has been monitored, we see relatively low distances between that patients and other timepoints, see Figure 7 for an example. However, this trend drastically changes when taking a patient who transitions from an MCI diagnosis to an AD diagnosis, as can be seen in Figure 8. Note that the colorscale is the same for Figures 7 and 8. However, these effects do not generalize: if we take the average distance for each homological dimension for each patient and compare the distribution of averages for patients who do change diagnoses versus those who don't, we don't see any quantitative difference, as can be seen in Figure 9. Surprisingly, however, we see that taking the average Wasserstein distance for each patient, a bimodal – similar for both deteriorating and stable patients – distribution arises, but dissapears as we let $p \to \infty$ as we recover the bottleneck distance. Additionally, while no significant differences are observes, we do see slight rightwards shifts in each mode for both Wasserstein and the L^1 norm of the persistent landscape, indicating a potential shift among some of the observed datapoints.
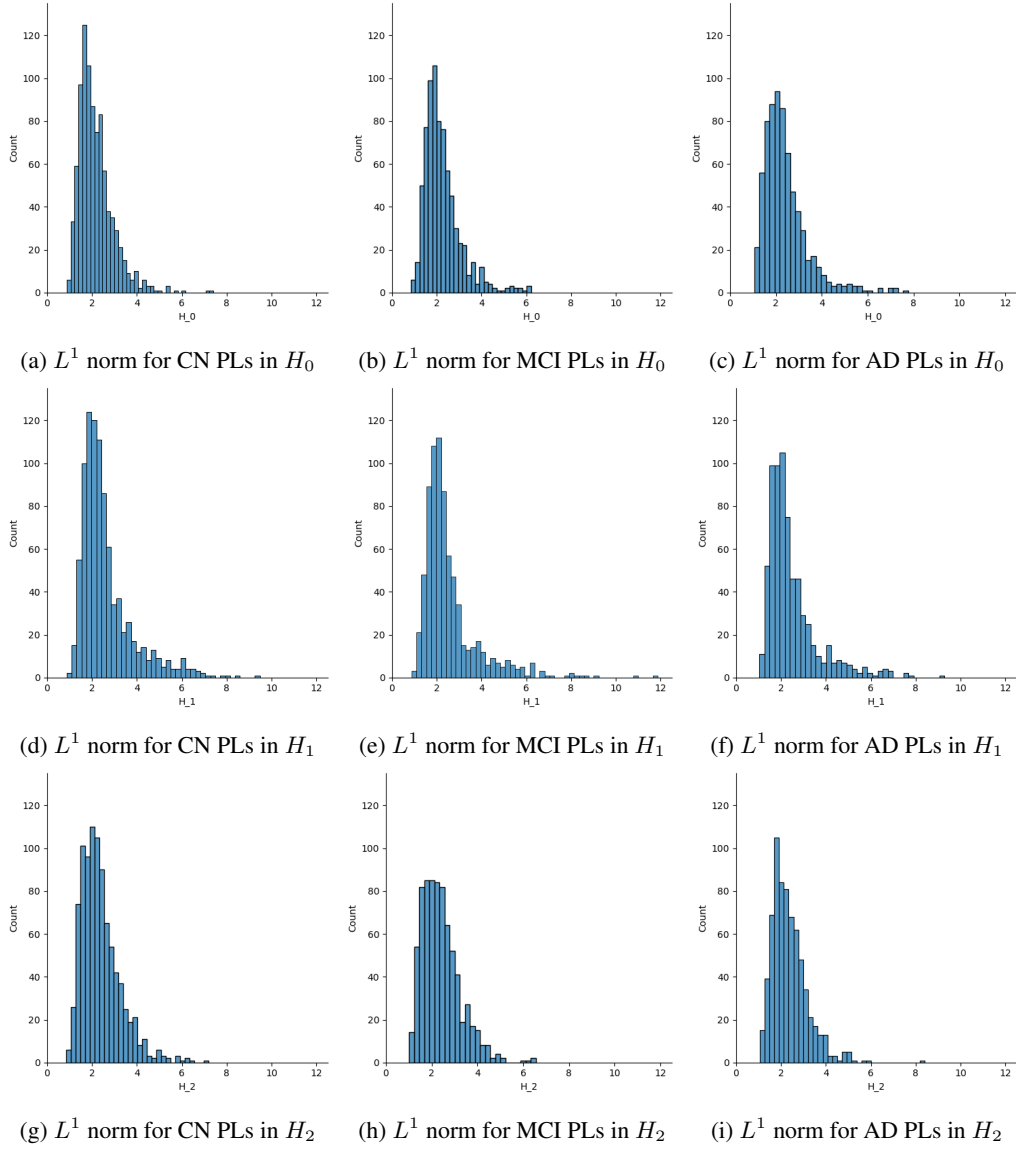
7

(a) L^1 norm for CN PLs in H_0 (b) L^1 norm for MCI PLs in H_0 (c) L^1 norm for AD PLs in H_0

(d) L^1 norm for CN PLs in H_1 (e) L^1 norm for MCI PLs in H_1 (f) L^1 norm for AD PLs in H_1

(g) L^1 norm for CN PLs in H_2 (h) L^1 norm for MCI PLs in H_2 (i) L^1 norm for AD PLs in H_2

Figure 6: Histogram showing the distribution of the L^1 norm taken between the median PL for a diagnostic categories in all homological dimesions.
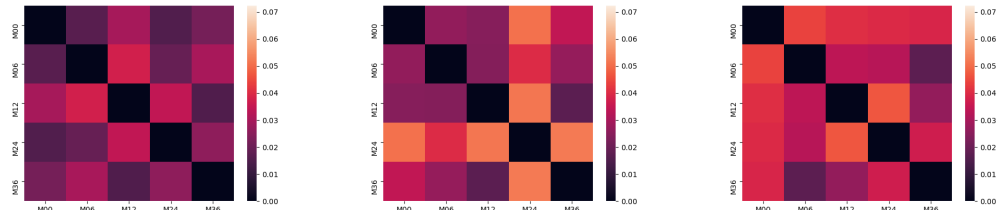

Figure 7: Topological evolution of a subject with an unchanging CN diagnosis.
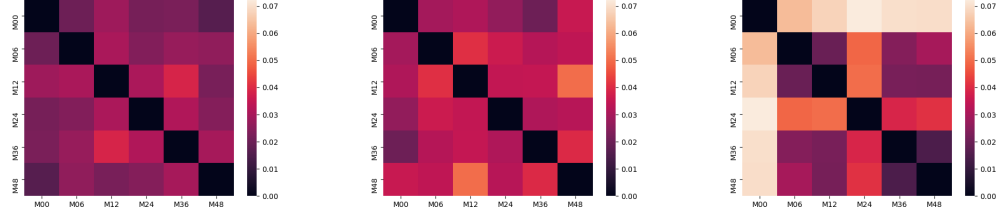
Figure 8: Topological evolution of a patient who transitions from MCI to AD in the course of the observation. For this particular case, the change in diagnosis occurred at $t = 24$, i.e. 24 months after the baseline diagnosis, which also corresponds with the highest distance from that baseline.

(a) Wasserstein distance with $p = 2$ in H_0

(b) Wasserstein distance with $p = 2$ in H_1

(c) Wasserstein distance with $p = 2$ in H_2

(d) Bottleneck distance with $p = 2$ in H_0

(e) Bottleneck distance with $p = 2$ in H_1

(f) Bottleneck distance with $p = 2$ in H_2

(g) L^1 landscape distance with 1 layer in H_0

(h) L^1 landscape distance with 1 layer in H_1
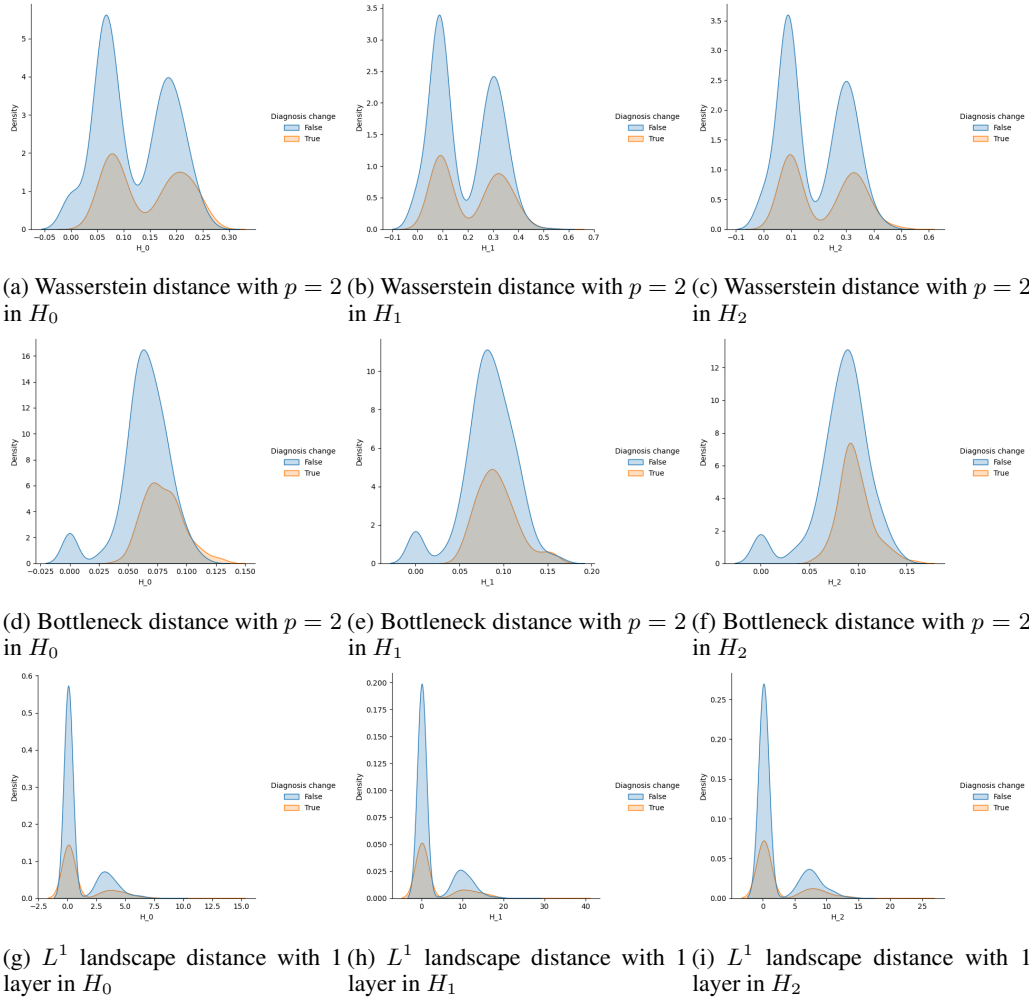
(i) L^1 landscape distance with 1 layer in H_2

Figure 9: Kernel density estimation of the average average distance between each image timepoint for each patient. The orange curve represents all those patients who have had at least one change in diagnosis over the course of the disease, whereas patients who haven't are within the blue curve.

9

(a) Distance from median CN PL inn H_0

(b) Distance from median CN PL inn H_1

(c) Distance from median CN PL inn H_2

(d) Distance from median AD PL inn H_0

(e) Distance from median AD PL inn H_1
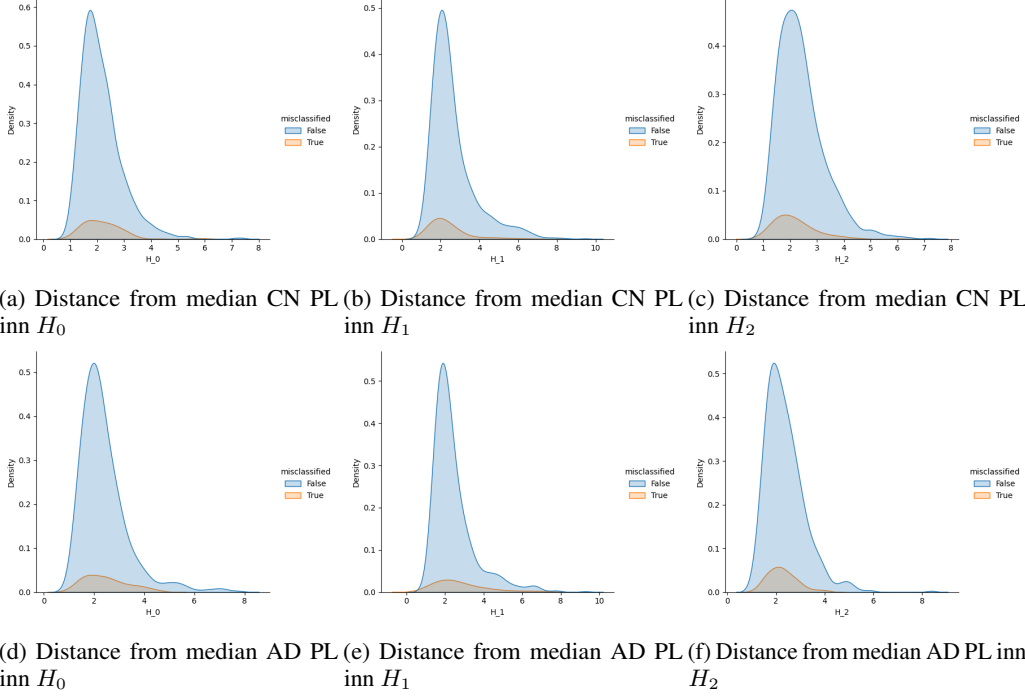
(f) Distance from median AD PL inn H_2

Figure 10: Kernel density estimation of distribution of the distance between the AD and CN median persistence image for images which have and haven't been misclassified.

3.4 Topological outliers and misclassified samples.

The distribution of distances with respect to the average persistent landscape was plotted for the patients who were correctly classified, and for those who were not correctly classified. The results are shown in Figure 10. We also examined the proportion of patients who switched diagnosis in the whole ADNI dataset. We found that 70% (64) of the misclassified patients had only one diagnosis versus 71% (323) in the whole dataset.

4 Discussion

References

WHO. Global action plan on the public health response to dementia 2017–2025. 2017.

Katherine E Sleeman, Maja de Brito, Simon Etkind, Kennedy Nkhoma, Ping Guo, Irene J Higginson, Barbara Gomes, and Richard Harding. The escalating global burden of serious health-related suffering: projections to 2060 by world regions, age groups, and health conditions. *The Lancet Global Health*, 7(7):e883–e892, 2019.

Konstantina G Yiannopoulou and Sokratis G Papageorgiou. Current and future treatments in alzheimer disease: An update. *Journal of Central Nervous System Disease*, 12:1179573520907397, 2020.

Guy M McKhann, David S Knopman, Howard Chertkow, Bradley T Hyman, Clifford R Jack Jr, Claudia H Kawas, William E Klunk, Walter J Koroshetz, Jennifer J Manly, Richard Mayeux, et al. The diagnosis of dementia due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):263–269, 2011.

Sylvain Lehmann and Charlotte Elisabeth Teunissen. Biomarkers of alzheimer's disease: The present and the future. *Frontiers in Neurology*, 7:158, 2016.

Lieke L Smits, Yolande AL Pijnenburg, Esther LGE Koedam, Annelies E van der Vlies, Ilona EW Reuling, Teddy Koene, Charlotte E Teunissen, Philip Scheltens, and Wiesje M van der Flier. Early onset alzheimer's disease is associated with a distinct neuropsychological profile. *Journal of Alzheimer's Disease*, 30(1):101–108, 2012.

Ji-Yeun Hur, Georgia R Frost, Xianzhong Wu, Christina Crump, Si Jia Pan, Eitan Wong, Marilia Barros, Thomas Li, Pengju Nie, Yujia Zhai, et al. The innate immunity protein ifitm3 modulates γ-secretase in alzheimer's disease. *Nature*, pages 1–6, 2020.

Tamas Fulop, Jacek M Witkowski, Karine Bourgade, Abdelouahed Khalil, Echarki Zerif, Anis Larbi, Katsuiku Hirokawa, Graham Pawelec, Christian Bocti, Guy Lacombe, et al. Can an infection hypothesis explain the beta amyloid hypothesis of alzheimer's disease? *Frontiers in aging neuroscience*, 10:224, 2018.

William G Tharp and Indra Neil Sarkar. Origins of amyloid-β. *BMC genomics*, 14(1):1–15, 2013.

Sandro Dá Mesquita, Ana Catarina Ferreira, João Carlos Sousa, Margarida Correia-Neves, Nuno Sousa, and Fernanda Marques. Insights on the pathophysiology of alzheimer's disease: The crosstalk between amyloid pathology, neuroinflammation and the peripheral immune system. *Neuroscience & Biobehavioral Reviews*, 68:547–562, 2016.

Michel Goedert and Maria Grazia Spillantini. A century of alzheimer's disease. *science*, 314(5800): 777–781, 2006.

Giovanni B Frisoni, Nick C Fox, Clifford R Jack, Philip Scheltens, and Paul M Thompson. The clinical use of structural mri in alzheimer disease. *Nature Reviews Neurology*, 6(2):67–77, 2010.

Junhao Wen, Elina Thibeau-Sutre, Mauricio Diaz-Melo, Jorge Samper-González, Alexandre Routier, Simona Bottani, Didier Dormont, Stanley Durrleman, Ninon Burgos, Olivier Colliot, et al. Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. *Medical Image Analysis*, page 101694, 2020.

Konstantinos Poulakis, Joana B Pereira, Patrizia Mecocci, Bruno Vellas, Magda Tsolaki, Iwona Kłoszewska, Hilkka Soininen, Simon Lovestone, Andrew Simmons, Lars-Olof Wahlund, et al. Heterogeneous patterns of brain atrophy in alzheimer's disease. *Neurobiology of aging*, 65:98–108, 2018.

Henri Poincaré. *Analysis situs*. Gauthier-Villars, 1895.

Ioan Mackenzie James. *History of topology*. Elsevier, 1999.

Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

Tamal K Dey, Herbert Edelsbrunner, and Sumanta Guha. Computational topology. *Contemporary mathematics*, 223:109–144, 1999.

Erik J Amézquita, Michelle Y Quigley, Tim Ophelders, Elizabeth Munch, and Daniel H Chitwood. The shape of things to come: Topological data analysis and biology, from molecules to organisms. *Developmental Dynamics*, 2020.

Daniel Freedman and Chao Chen. Algebraic topology for computer vision. *Computer Vision*, pages 239–268, 2009.

Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102, 2015.

Peter Bubenik. The persistence landscape and some of its properties. In *Topological Data Analysis*, pages 97–117. Springer, 2020.

Jesse J Berwald, Joel M Gottlieb, and Elizabeth Munch. Computing wasserstein distance for persistence diagrams on a quantum computer. *arXiv preprint arXiv:1809.06433*, 2018.

Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.

Guillaume Tauzin, Umberto Lupo, Lewis Tunstall, Julian Burella Pérez, Matteo Caorsi, Anibal Medina-Mardones, Alberto Dassatti, and Kathryn Hess. giotto-tda: A topological data analysis toolkit for machine learning and data exploration, 2020.

François Chollet et al. Keras. `https://keras.io`, 2015.

A   Supplements

A.1   Preprocessing of MRI data

We included all T1-weighted MRI images from ADNI1, 2, 3, and GO, which were captured and preprocessed by ADNI. Results included in our work come from preprocessing performed using `fMRIPrep` 20.1.1, a Nipype 1.5.0 based tool. All MRIs were corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection`, distributed with ANTs 2.2.0, and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a `Nipype` implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using `OASIS30ANTs` as a target template. Volume-based spatial normalisation to a standard coordinate space (`MNI152NLin2009cAsym`) was performed through nonlinear registration with `antsRegistration`, using brain-extracted versions of both T1w reference and the T1w template. We slected the template 'ICMB 152 Nonlinear Asymmetrical Template Version 2009c' for spatial normalisation.

Many internal operations of `fMRIPrep` use the `Nilearn` library, version 0.6.2, mostly within the functional processing workflow. For more details of the pipeline, please refer to the official documentation of `fMRIPrep`. Preprocessing was finalised by intensity normalisation of the extracted and MNI space registered brain images.