Topological Detection of Alzheimer's Disease using Betti Curves

Ameer Saadat-Yazdi, Rayna Andreeva, and Rik Sarkar

School of Informatics, University of Edinburgh. Edinburgh, UK.
s1707343@sms.ed.ac.uk, r.andreeva@sms.ed.ac.uk, rsarkar@inf.ed.ac.uk

Abstract. Alzheimer's disease is a debilitating disease in the elderly, and is an increasing burden to the society due to an aging population. In this paper, we apply topological data analysis to structural MRI scans of the brain, and show that topological invariants make accurate predictors for Alzheimer's. Using the construct of Betti Curves, we first show that topology is a good predictor of Age. Then we develop an approach to factor out the topological signature of age from Betti curves, and thus obtain accurate detection of Alzheimer's disease. Experimental results show that topological features used with standard classifiers perform comparably to recently developed convolutional neural networks. These results imply that topology is a major aspect of structural changes due to aging and Alzheimer's. We expect this relation will generate further insights for both early detection and better understanding of the disease.

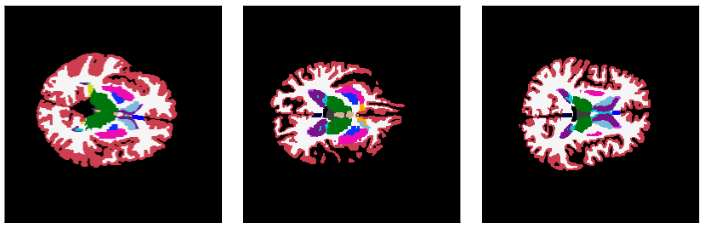
Keywords: Topological data analysis · Alzheimer's disease · MRI.

1 Introduction

Alzheimer's disease (which is often abbreviated as AD) and other neurodegenerative diseases are closely associated with aging. As average life expectancy increases worldwide, the number of patients with brain aging and associated diseases will rise rapidly. The current estimated number of Alzheimer's patients is around 47 million, which is projected to increase to 152 million by year 2050 [25].

Deterioration of the brain manifests as several complex structural, chemical and functional changes, making it challenging to distinguish diseases from aging. For example, with age, cerebral ventricles expand and cortical thickness decreases; lesions and atrophies arise [13]; the brain volume itself contracts with old age, while gray and white matter volumes are known to expand and contract at different times in the life cycle. These degenerative changes interact in complex ways with progression of Alzheimer's disease and the atrophy induced by it [18]. See Figure 1 for MRI scans of brains showing degeneration in brain tissue. Thus, understanding these changes and their different manifestations will be crucial to the prevention and management of the diseases.

Aging related symptoms – atrophied regions, lesions etc – affect connectivity in the brain; and aberrations in connectivity are also closely related to neural disorders including Alzheimer's, Autism Spectrum Disorder and many others [23].



(a) Normal (b) Aging (c) AD

Fig. 1: Topological changes with aging and Alzheimer's, showing the effects in gray matter (red) and white matter (white). Both aging and AD result in the loss of white matter volume and a thinning of the gray matter. Gray matter of the Aging brain breaks down into smaller connected components.

In this paper, we study the changes in *connectivity* of the brain by observing the changes in its topology. Using data analysis on structural MRI scans, we identify signatures of deterioration in connectivity, and develop techniques to isolate the signs of Alzheimer's from ordinary aging.

Our contributions. We take the approach of persistent homology – which is a way to describe topology at multiple scales or function values. Persistent homology produces artifacts called Betti Curves [7] that count the number of disconnections and holes. Our approach computes persistent homology separately for different functions for white and gray matter, and we find that the resultant Betti curves make accurate predictors of the chronological age. These ideas are described in Section 4.1.

Since changes in connectivity are common features of aging as well as of Alzheimer's, we develop a method to factor out the topological signature of aging to identify cases of Alzheimer's (Section 4.2). This method works using the relation between chronological age and Betti curves. It first predicts a Betti curve given the chronological age of the patient. Next, it computes the Betti curve of the MRI scan, and finds the *residual Betti curve* – the difference between the computed and predicted curve. This residual curve is found to produce significant gains in identification of Alzheimer's disease from MRI scans.

Experimental results (Section 5) show that topological features are good predictors of both age and Alzheimer's, and the newly defined residual Betti curves factor out the aging affect from Alzheimer's fairly successfully. In particular, we obtain an R^2 score of 66% and Mean Absolute Error (MAE) of 4.47 years for the age prediction task (Section 5.1) and $F1$ score of 0.62 and 0.74 balanced accuracy for detecting Alzheimer's. These results are comparable to recent CNN based methods [30] (Section 5.2). Thus, our results suggest that the topological changes are an important descriptor of the structural changes in the brain and should be investigated in greater detail for understanding of aging and Alzheimer's.

2 Related Works

A multitude of automated methods to estimate brain age [14,8] and detect Alzheimer's have been suggested in the literature in recent years, with their success highly dependent on the set of participants, image preprocessing procedures, cross-validation procedure, reported evaluation metric, number and modality of brain scans [30]. For detection of AD, they generally fall into two main categories. The first one is those that classify AD based on biomarkers(e.g. [24,18]), where the biomarkers have been generally based around describing brain atrophy [4]. The second one is for studies based on deep learning and convolutional neural networks (CNNs) [1,3,22].

The studies in the first group mainly focus on measuring brain atrophy from a structural MRI, which is an important biomarker for determining the status and strength of the neurodegeneration from AD [29]. However, the existing methods are known to be incomplete. Though the deep learning approaches from the second group show promising results, there are challenges in terms of generalisation [26,2], bias and data leakage [30]. A recent summary of CNN-based methods and a number of possible issues can be found in [30].

Topological data analysis (TDA) has been used to study other static anatomical data (e.g. retina images [5,17]) as well as dynamic data such as functional MRI (fMRI) scans [27]. fMRI represents activity in different parts of the brain by detecting blood flow. In contrast, MRI (or structural MRI) produces a static anatomical image and not of the temporary activity. To the best of our knowledge, the current paper is the first TDA-based analysis of AD on MRI data of the brain, showing that Topology is a significant feature of aging and AD.

3 Preliminaries

3.1 Homology and Betti Curves

Betti numbers are *topological invariants* that count numbers of holes in each dimension. The dimension n Betti number is written as β_n. For example, β_0 counts the number of connected components, β_1 is the number of one-dimensional or "circular" holes and β_2 counts the number of enclosed spaces or voids. We present here only a basic description needed for our exposition. For a detailed mathematical treatment, we refer the reader to various excellent texts on algebraic topology [19,9].

The Betti number is sensitive to noise and small perturbations in data. Topological data analysis (TDA) is made robust through the use of topological persistence [31,6] – working off the assumption that structures which persist over multiple scales of data, or multiple values of a relevant *filtration function*, are important. In order to use them in our analysis, starting with an image I, we build a representation of the data as cubical complexes F_r, according to the sub-level sets of some carefully selected filtration function $f : I \to \mathbb{R}$. All pixels which when used to evaluate f fall below some threshold parameter r are included in the cubical complex, and, by increasing r, we obtain a sequence of nested cubical

complexes $F_{r_0} \subset F_{r_1} \subset F_{r_2} \subset \ldots$. The *Betti curve* [7] is the sequence consisting of the Betti numbers of these complexes: $B_n(I, r) := \beta_n(F_r)$.

3.2 Dataset

The dataset used in this study is a publicly available dataset of MRI scans OASIS-3 [21]. It consists of scans of 1098 individuals, taken over a period of 10 years as part of a longitudinal neuroimaging study. We have used a portion of the dataset consisting of 888 scans, where 733 are healthy and 155 with AD. Table 1 shows the gender and age distribution. The dataset was chosen for the availability of precomputed FreeSurfer [10] files which include skull-stripped scans, brain segmentations [12], cortical thickness measurements [11] and 3D reconstructions. This allows for a much more streamlined analysis of the images, and means that we can relate the results directly to the precomputed measurements of brain volumes and thicknesses. These measurements form the feature sets for the *Baseline model* used in experiments.

	Healthy	**AD**
Age (mean ± SD)	66.9 ± 9.3	74.6 ± 8.1
Gender (M/F)	258/475	74/81

Table 1: Demographic data of the subset used in experiments.

4 Algorithms and methods

4.1 Betti curve based features on MRI data

To compute the Betti curves of MRI scans, we need an appropriate filtration function. Given the different qualitative properties we seek to capture between gray and white matter, we isolate the two regions and compute the persistent homology using two different filtration parameters.

The gray matter is processed using density filtration [16]. This function smooths out the image by assigning to each pixel the number of non-zero neighbours within a 3 pixel radius. For white matter, we wish to detect discolourations associated with lesions, thus the filtration is performed directly on the pixel intensity values.

Persistence diagrams are computed on filtration values. These are then normalized to $[0, 1]$ and used to compute the Betti curves (E.g. see Fig. 2). Additionally, Freesurfer's preprocessing procedure produces point clouds outlining a 3D reconstruction of the outer surface of the brain. We compute the Betti curves for these point clouds from Vietoris-Rips filtration [6]. The persistent homology computation produces Betti curves for dimensions $0, 1$ and 2 using 100 filtration values in each dimension, resulting in three 100-dimensional vectors.

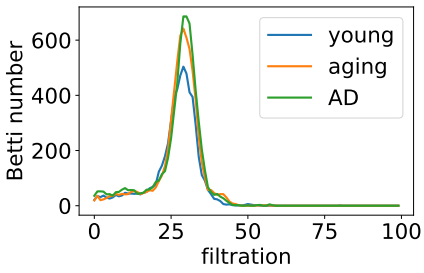


Fig. 2: Example of a dimension 1 Betti curve computed from the gray matter. The age of subjects are 56 for young, 65 for AD and 83 for aging.

For the machine learning algorithms, we concatenate these curves from different dimensions into a single vector, and call this the Betti curve $B(s)$ for subject s. Concatenation of curves from different combination of regions of the brain (gray matter, white matter and surface reconstruction) produces different instances of $B(s)$.

4.2 Aging vs Alzheimer's

We first develop techniques to correlate age with the Betti curve features, and then compute *Residual Betti Curves* that represent Betti Curves modulo the effect of age.

Predicting age from Betti curves. The age prediction model is built using a random forest regressor [20] on the Betti curves. Several models are trained for each set of the Betti curves (white matter, gray matter and point cloud) independently. We train an additional model consisting of the various Betti curves concatenated together. We compare the performance of all these models to a baseline model consisting of a random forest regressor trained on the FreeSurfer volumetric statistics.

Residual Betti curves: Factoring out the effect of age. Next, we consider the inverse problem of predicting the combined Betti curves from the age of a person. For a person of age a, we call this the expected Betti curve $E(a)$. The true combined Betti curve $B(s)$ can be regarded as a sum:

$$B(s) = E(a) + R(s),$$

where we call $R(s)$ the residual Betti curve for subject s of age a. $R(s)$ is the difference between $B(s)$ and $E(a)$ and represents how much the brain structure deviates from the expected (healthy) brain. Values in R will be usually small for healthy subjects and large for AD patients.

The classification model is then trained on the set of residuals to obtain the distribution $P(A|R(s))$ where A is a binary random variable indicating the presence of Alzheimer's.

This training for Alzheimer's prediction operates as follows. Normalized Betti curves for the entire dataset are computed. E is estimated using a linear regression on a healthy subpopulation. A Support Vector Machine (SVM) model is trained on the remaining data to predict the presence of Alzheimer's given the residuals $R(s) = B(s) - E(a)$

Example residual curves are shown in Figure 3. We report the test set performance of various combinations of Betti curves as in the age prediction model.

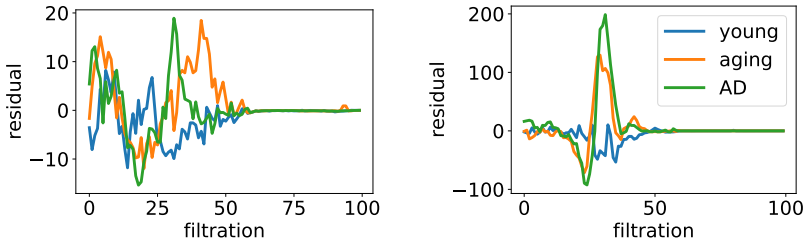


Fig. 3: Comparison of gray matter residual curves in the dimensions 0 and 1. The first dimensional residual (right), shows clearer distinction between the old subject and and the AD patient. The age of subjects is 56 for young, 65 for AD and 83 for aging.

4.3 Implementation details

The source code is available on `GitHub`[1], where more implementation details can be found. Persistent homology was computed with `giotto-tda` library [28]. Baseline comparisons were conducted with a set of standard features. The random forest age prediction model, as well as the SVM and linear regressor models of the AD classification model are instantiated with the default hyperparameters in `scikit-learn`. For the former, sampling procedure is stratified to preserve the distribution of ages in the data, resulting in 666 train curves and 466 test curves. For the latter, the linear regressor is trained on 200 healthy curves and the SVM is trained on 566 curves (100 more curves are added to the remaining 466 by oversampling due to imbalance between subjects with and without AD).

5 Experimental results

5.1 Brain age prediction

. Initial experiments explore the utility of Betti curves as predictors of brain age. This is done by training several random forest models on the Betti curves

[1] https://github.com/ameertg/BrainAgingTDA

obtained from the various regions, white matter, gray matter and point cloud surface reconstructions. We investigate the performance of the regression when restricted to the Betti curves of a particular region as well its performance when the Betti curves are concatenated together. The *Baseline* model consists of a random forest regressor trained on the FreeSurfer volumetric statistics.

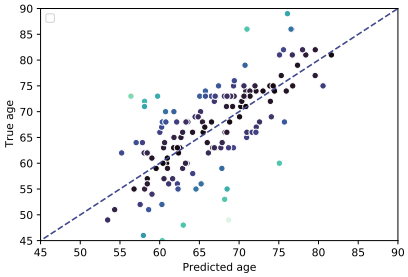


Fig. 4: Predicted versus actual age for healthy subjects.

The correlation between the true and predicted ages can be seen in Figure 4. For evaluation, we consider the R^2 score on the testing set. This is a common metric used in the evaluation of regression models and describes the variance of the target variable explained by the model. An R^2 score of 1 indicates perfect fit while 0 indicates no correlation between the model output and the target labels.

Table 2 indicates that for raw MRI the combined white and gray matter features performs well with an R^2 score of 0.51, while the addition of point cloud information provides a large increase to an R^2 score of 0.66 and Mean absolute error of 4.47 years.

Model	R^2 **score**	**MAE**
Freesurfer Baseline	0.46 ± 0.02	5.43 ± 0.21
White matter	0.43 ± 0.02	5.80 ± 0.22
Gray matter	0.49 ± 0.03	5.56 ± 0.17
White + gray matter	0.50 ± 0.02	5.46 ± 0.24
Surface point cloud	0.53 ± 0.03	5.13 ± 0.32
All combined	$\mathbf{0.65 \pm 0.02}$	$\mathbf{4.47 \pm 0.23}$

Table 2: Mean test set model performances (mean and standard deviation) on 5 runs (with random train and test sets) in terms of R^2 score and Mean Absolute Error (MAE) for various feature sets. Based on combined Betti curve features.

5.2 AD detection with age correction

Here we compare two approaches to training an AD classifier. The first method uses the Betti curve obtained from the MRI scan while the second method incorporates information about aging by considering the residual Betti curves. As above, we test both methods on individual Betti curves and also on the concatenated curve which includes the white, gray and point cloud Betti curves. Since the number of healthy and AD patients are different, we compute F1 score and balanced accuracy which are different methods for interpreting unbalanced data. Table 3 shows that the initial model performs poorly on all feature sets but sees significant improvements in F1 scores when we introduce the residual curve. The residual approach performs particularly well on the gray matter features.

	Raw Betti curves		**Residual Betti curves**	
	F1 score	**Balanced acc**	**F1 score**	**Balanced acc**
Gray matter	0.00 ± 0.0	0.50 ± 0.0	$\mathbf{0.61 \pm 0.02}$	$\mathbf{0.74 \pm 0.03}$
White matter	0.13 ± 0.01	0.51 ± 0.02	0.50 ± 0.03	0.65 ± 0.04
Surface Point cloud	$\mathbf{0.42 \pm 0.02}$	0.76 ± 0.03	0.45 ± 0.03	0.57 ± 0.03
Combined	0.30 ± 0.02	$\mathbf{0.76 \pm 0.02}$	0.47 ± 0.04	0.60 ± 0.03
Baseline	0.36 ± 0.02	0.66 ± 0.02	0.22 ± 0.03	0.47 ± 0.02

Table 3: Mean and standard deviation test set performance of the SVM classifier on 5 runs (with random training and test sets) trained on the raw Betti curves (left); and on the residual Betti curves (right).

6 Discussion

The most striking observation from our experiments is that simple classification on topological features produces a balanced accuracy of 0.76 on Raw Betti Curves and 0.74 on Residual Betti curves, while in comparison, state-of-the-art CNN-based methods (3D subject-level CNN, 3D ROI-based CNN, 3D patch-level CNN, 2D slice-level CNN) achieve average balanced accuracy varying between 0.61 and 0.73 (see Table 6 in [30], last column, row Baseline). Our experiments use about half of the same dataset OASIS-3 [21], and thus the comparison needs some further study. However, it is clear from the results that topology can play a significant role in the study and detection of Alzheimer's disease.

Various other results in our study also point toward specific relevance of topology in the study of the brain. The residual Betti curves on Gray matter were particularly accurate, suggesting the need for further study of the topology of gray matter. Similarly, the topological age prediction was explored here only to the extent it was useful for subsequent detection of AD. Its combination with other techniques such as BrainAge [15] and relation to the Freesurfer and CNN features remain to be explored.

Acknowledgements. RA is supported by the UKRI (grant EP/S02431X/1).

References

1. Aderghal, K., Benois-Pineau, J., Afdel, K., Gwenaëlle, C.: Fuseme: Classification of smri images by fusion of deep cnns in 2d+ ε projections. In: Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing. pp. 1–7 (2017)
2. Andreeva, R., Fontanella, A., Giarratano, Y., Bernabeu, M.O.: Dr detection using optical coherence tomography angiography (octa): A transfer learning approach with robustness analysis. In: International Workshop on Ophthalmic Medical Image Analysis. pp. 11–20. Springer (2020)
3. Bäckström, K., Nazari, M., Gu, I.Y.H., Jakola, A.S.: An efficient 3d deep convolutional network for alzheimer's disease diagnosis using mr images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 149–153. IEEE (2018)
4. Beheshti, I., Demirel, H., Initiative, A.D.N., et al.: Feature-ranking-based alzheimer's disease classification from structural mri. Magnetic resonance imaging **34**(3), 252–263 (2016)
5. Beltramo, G., Andreeva, R., Giarratano, Y., Bernabeu, M.O., Sarkar, R., Skraba, P.: Euler characteristic surfaces. arXiv preprint arXiv:2102.08260 (2021)
6. Carlsson, G.: Topology and data. Bulletin of the American Mathematical Society **46**(2), 255–308 (2009)
7. Chung, Y.M., Lawson, A.: Persistence curves: A canonical framework for summarizing persistence diagrams. arXiv preprint arXiv:1904.07768 (2019)
8. Cole, J.H., Ritchie, S.J., Bastin, M.E., Hernández, M.V., Maniega, S.M., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., et al.: Brain age predicts mortality. Molecular psychiatry **23**(5), 1385–1392 (2018)
9. Edelsbrunner, H., Harer, J.: Computational topology: an introduction. American Mathematical Soc. (2010)
10. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
11. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences **97**(20), 11050–11055 (2000)
12. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron **33**(3), 341–355 (2002)
13. Fjell, A.M., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., et al.: High consistency of regional cortical thinning in aging across multiple samples. Cerebral cortex **19**(9), 2001–2012 (2009)
14. Franke, K., Gaser, C.: Longitudinal changes in individual brainage in healthy aging, mild cognitive impairment, and alzheimer's disease. GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry **25**(4), 235 (2012)
15. Franke, K., Gaser, C.: Ten years of brainage as a neuroimaging biomarker of brain aging: what insights have we gained? Frontiers in neurology **10**, 789 (2019)
16. Garin, A., Tauzin, G.: A topological" reading" lesson: Classification of mnist using tda. In: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). pp. 1551–1556. IEEE (2019)
17. Giarratano, Y., Pavel, A., Lian, J., Andreeva, R., Fontanella, A., Sarkar, R., Reid, L.J., Forbes, S., Pugh, D., Farrah, T.E., et al.: A framework for the discovery of

retinal biomarkers in optical coherence tomography angiography (octa). In: International Workshop on Ophthalmic Medical Image Analysis. pp. 155–164. Springer (2020)
18. Habes, M., Janowitz, D., Erus, G., Toledo, J., Resnick, S., Doshi, J., Van der Auwera, S., Wittfeld, K., Hegenscheid, K., Hosten, N., et al.: Advanced brain aging: relationship with epidemiologic and genetic risk factors, and overlap with alzheimer disease atrophy patterns. Translational psychiatry **6**(4), e775–e775 (2016)
19. Hatcher, A.: Algebraic Topology. Cambridge University Press (2002)
20. Ho, T.K.: Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition. vol. 1, pp. 278–282. IEEE (1995)
21. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A., et al.: Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. MedRxiv (2019)
22. Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., Wang, J.: Applications of deep learning to mri images: A survey. Big Data Mining and Analytics **1**(1), 1–18 (2018)
23. Ouyang, M., Kang, H., Detre, J.A., Roberts, T.P., Huang, H.: Short-range connections in the developmental connectome during typical and atypical brain maturation. Neuroscience & Biobehavioral Reviews **83**, 109–122 (2017)
24. Papakostas, G.A., Savio, A., Graña, M., Kaburlasos, V.G.: A lattice computing approach to alzheimer's disease computer assisted diagnosis based on mri data. Neurocomputing **150**, 37–42 (2015)
25. Patterson, C., et al.: World alzheimer report 2018 (2018)
26. Qayyum, A., Qadir, J., Bilal, M., Al-Fuqaha, A.: Secure and robust machine learning for healthcare: A survey. IEEE Reviews in Biomedical Engineering **14**, 156–180 (2020)
27. Rieck, B., Yates, T., Bock, C., Borgwardt, K., Wolf, G., Turk-Browne, N., Krishnaswamy, S.: Uncovering the topology of time-varying fmri data using cubical persistence. Advances in Neural Information Processing Systems **33** (2020)
28. Tauzin, G., Lupo, U., Tunstall, L., Pérez, J.B., Caorsi, M., Medina-Mardones, A.M., Dassatti, A., Hess, K.: giotto-tda:: A topological data analysis toolkit for machine learning and data exploration. J. Mach. Learn. Res. **22**, 39–1 (2021)
29. Vemuri, P., Jack, C.R.: Role of structural mri in alzheimer's disease. Alzheimer's research & therapy **2**(4), 1–10 (2010)
30. Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O., et al.: Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. Medical image analysis **63**, 101694 (2020)
31. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete & Computational Geometry **33**(2), 249–274 (2005)