

Preprocessing of general stenotic vascular flow data for topological data analysis

Christopher L. Bresten^{1,✉}, Jihoon Kweon², Xinjuan Chen^{1,4}, Young-Hak Kim³, Jae-Hun Jung^{1,*}, ✉

1 Department of Mathematics & POSTECH MINDS (Mathematical Institute for Data Science), Pohang University of Science & Technology, Pohang 37673, Republic of Korea.

2 Department of Convergence Medicine, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul, 05505, Republic of Korea.

3 Division of Cardiology, Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul, 05505, Republic of Korea.

4 School of Science, Jimei University, Xiamen, Fujian Province, 361021, P.R. China.

✉ These authors contributed equally to this work.

* Corresponding author

Abstract

A new analysis and classification method of vascular disease based on topological data analysis (TDA) has been proposed in [1]. The proposed method utilizes the application of persistent homology to hemodynamic variables. Particularly, 2D homology is obtained from the velocity field of the flow projected onto the unit sphere, known as so-called the S^2 projection. It was shown that such homology is closely related to the degree of vascular disease. The original method was developed based on the computational fluid dynamic (CFD) solutions of the straight stenotic vessels.

In this paper, we develop a preprocessing method that enables the proposed TDA method to be applied to general stenotic vessels of irregular geometry. The velocity field is subject to a coordinate transformation correcting for orientation and curved geometry. The preprocessed data is projected onto S^2 and the corresponding homology is calculated. We show that this preprocessing is necessary for the proposed TDA method to be successfully applied to general types of stenotic vessels. Validation was performed on a set of clinical data including reconstructed vascular geometry with corresponding diagnostic indices.

Introduction

Cardiovascular disease is the leading cause of death worldwide. Accurate diagnosis and prediction is crucial. The diagnosis methods today are based on the anatomical approach by investigating the geometry of the diseased vessel using image analysis and the functional approach by measuring the pressure drops across the stenosis. The gold standard diagnostic method uses fractional flow reserve (FFR) values, which are calculated by dividing a distal pressure measurement by a pressure reading proximal to the patient's stenosis [2, 3]. Newer approaches utilize computational fluid dynamics (CFD). Patient-specific CFD is also a possible analysis tool [4] to estimate FFR based on the simulated data. Recently, machine learning approaches were also proposed as an

alternative diagnosis methodology [5, 6, 7]. For these methods, clinicians need a canonical index that can represent the collective knowledge of the diseased vessel. Especially for the machine learning approach, one needs to extract features of the vascular disease for the given patient's data.

In [1] a new method of providing a measure of vascular disease has been proposed based on topological data analysis (TDA) [8]. In order to apply TDA to vascular flows for the proposed method in [1], the velocity field was projected onto a unit sphere (S^2), referred as the S^2 projection. It was shown that this simple projection method is highly efficient and powerful for visualizing and quantifying the vascular disease, with which the flow complexity is measured by observing the directional information of the velocity fields. For example, the overall flow pattern is laminar for the healthy vessel and is more turbulent when there is stenosis. On S^2 , the laminar flow is projected as a point cloud centered around one pole on the sphere while the fully developed turbulent flow provides a point cloud all over the sphere including two poles. Then the proposed TDA method computes the persistent homology of the point cloud on S^2 . It was shown that the two dimensional homology of S^2 projection of vascular flows is closely related to the degree of disease and can be used along with the value of FFR for diagnosis. The original development of the proposed method was based on the 3D spectral computational fluid dynamic (CFD) data of the regular straight stenotic vasculature.

In this paper, we verify that the proposed TDA method can be applied effectively to curved vessels with more general irregularity in a similar way to the case of the straight vessel. In order to apply TDA to general vessel type, we develop a projection method which preprocesses the hemodynamic data such that it becomes suitable for TDA. The proposed projection algorithm is based on the singular value decomposition (SVD) and principle component analysis (PCA). After the preprocessing, we decompose each vessel into small segments and TDA is performed to the velocity fields on S^2 in all the segments. We use the real patient MRI data provided by Asan Medical Center, Seoul, Korea. The numerical results clearly show that the persistent homology after the preprocessing projection is significantly correlated to the FFR values, hence can be used a meaningful tool for the prediction of cardiovascular disease.

Methods and materials

Clinical data: Two invasive angiographic images were selected at the end of the diastolic phase of the cardiac cycle, then combined to build three dimensional geometry using CAAS Workstation 7.5 (Pie medical Imaging, Maastricht, NL). This was done in compliance with the Declaration of Helsinki and approval from the IRB of Asan Medical Center was granted by waiver of research consent.

CFD simulation The boundaries of the vasculature were assumed rigid and discretized using Ansys ICEM CFD 15.0 (Ansys Inc, Canonsburg, PA, USA). The vascular flow was modeled as an incompressible fluid with density $1060 \frac{kg}{m^3}$ and Newtonian $0.0035 \frac{kg}{m \cdot s}$. The inflow rate was set at $3 \frac{ml}{s}$ and outflow boundaries on the opposite face of the vessel set. The steady state solution was solved for using Ansys Fluent 15.0. Bifurcated vascular geometry was not included. A first order finite element method was used with a mixed mesh of tetrahedral and hexahedral elements.

TDA Software: For the TDA computations JavaPlex [9] was used through its MATLAB interface. The Lazy witness complex was used, with landmark points selected by max-min selector; 75 landmark points were used. Pre-processing was done in Python.

Notation: In this paper, we let matrices be denoted in capital letters, and indexed using subscripts, so that $X_{i,j}$ represents the ij th element of the matrix X . Also we let the subscript $*$ in this context mean an entire row or column, so that $X_{*,j}$ will denote the j^{th} column and vice-versa. For the given X , let X^H denote the Hermitian

(conjugate transpose) of X .

For the hemodynamic variables, let $\rho = \rho(\mathbf{x}, t)$ be the density, $p = p(\mathbf{x}, t)$ the pressure, $\mathbf{u} = (u, v, w)^T$ the velocity vector for the position vector $\mathbf{x} = (x^0, x^1, x^2)^T \in \Omega$ and time $t \in \mathbb{R}^+$. Here Ω is the closed domain in \mathbb{R}^3 .

Preprocessing

The preprocessing happens in several steps. First the orientation of the vessel is regularized by PCA on the spatial coordinate point cloud. Then the vessel is segmented along the first principal axis, followed by an iterative process of segmentation and fitting of a cubic spline. This results in a curvilinear coordinate system that reflects the curvature of the vessel, which can be used to project the velocity field and segment the vessel.

The process assumes that the geometry of the vessel is contiguous, non-bifurcating, and has high aspect ratio.

Let \mathbf{x}_k be the coordinate vector, $\mathbf{x}_k = (x_k^0, x_k^1, x_k^2) \in \mathbb{R}^3$ where the solutions to the incompressible Navier-Stokes equations are calculated and $X \in \mathbb{R}^{n \times 3}$ be the set composed of all \mathbf{x}_k , $k = 1, 2, \dots, n$, $X_{k,*} = (x_k^0, x_k^1, x_k^2)$. \mathbf{x}_k constitute the point cloud we use for the analysis. Here x_k^0, x_k^1, x_k^2 denote the x -, y - and z - coordinate of each point cloud. To each \mathbf{x}_k , the hemodynamic variables are assigned, e.g. the 3D velocity vector $\mathbf{v}_k = (v_k^0, v_k^1, v_k^2) \in \mathbb{R}^3$ and the pressure $p_k \in \mathbb{R}^+$. \mathbf{v}_k and p_k are all functions of \mathbf{x}_k and t . We define $W \in \mathbb{R}^{n \times 3}$ as the velocity field, composed of \mathbf{v}_k such that $W_{k,*} = (v_k^0, v_k^1, v_k^2)$.

In order to discover this principle axis, effectively regularizing orientation, we employ PCA. The principle axis is the direction of highest aspect ratio, along which the blood flows. Despite the local and global curvatures of the vessel, the principle direction should unidirectional and this direction will be mapped into a straight line along the principle axis. To perform PCA, we first compute the mean coordinate values for each dimension of \mathbf{x}_k , μ_0, μ_1 and μ_2 such that

$$\mu_i = \frac{1}{n} \sum_{k=1}^n x_k^i, \quad i = 0, 1, 2. \quad (1)$$

Let T_1 be the affine transformation that translates X by μ as below

$$T_1 : X_{*,i} \longrightarrow (X_{*,i} - \mu_i), \quad i = 0, 1, 2. \quad (2)$$

The principle axis can then be found using the SVD.

In the following SVD plays a crucial role. Each point in space can be thought of as a sample of a three dimensional random variable in our setting. After centering the data, the matrix $T_1(X) \cdot T_1(X)^T$ is then the covariance matrix of the sampled random variable, which projects a linear combination to a vector of covariances. The eigenvectors of this operator are the linear combinations which give a multiple of themselves as covariances. These eigenvectors can be interpreted as the directions in the space where the sample of the random variable has the most variation. These eigenvectors are the right singular vectors of $T_1(X)$. This gives us a basis that we can use to represent the data, the one corresponding to the largest singular value having the largest covariance, and therefore the accounting for the more variation in the data than any other possible choice of basis vector. Since the points are a point cloud defining the geometry of a three dimensional object, this gives us the direction where the object is longest, effectively orienting the geometry in the way that is ideal linear approximation for us to chop it up in the direction of the flow field.

Let the following be the SVD of $T_1(X)$

$$U \cdot \Sigma \cdot V^H, \quad (3)$$

where $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{3 \times 3}$ are unitary matrices and $\Sigma \in \mathbb{R}^{n \times 3}$ is a rectangular diagonal matrix with non-negative real values on its diagonal. Since $T_1(X)$ is real, U and V are orthonormal, i.e. $U^H = U$ and $V^H = V$. The principle axes/directions are the columns of V . Let T_2 be the linear transformation of X projected onto the right principle vectors V

$$T_2 : X \longrightarrow X \cdot V.$$

Let \tilde{X} be the following

$$\tilde{X} = T_2 \circ T_1(X). \quad (4)$$

In practice, one can use a sparse sampling of X for the SVD as it also provides a good approximation to V . This is the case particularly because we are dealing with only 3 dimensions while there is a large number of points, so the shape of the vessel is well represented by a relatively sparse random samples. Figure 1 shows the schematic images of $T_2 \circ T_1$. The original vessel in finite element mesh domain (left figure) is mapped into the new domain along the principle axes through T_2 (right figure). The colored intervals are the segments of the vessel (see below for the domain decomposition).

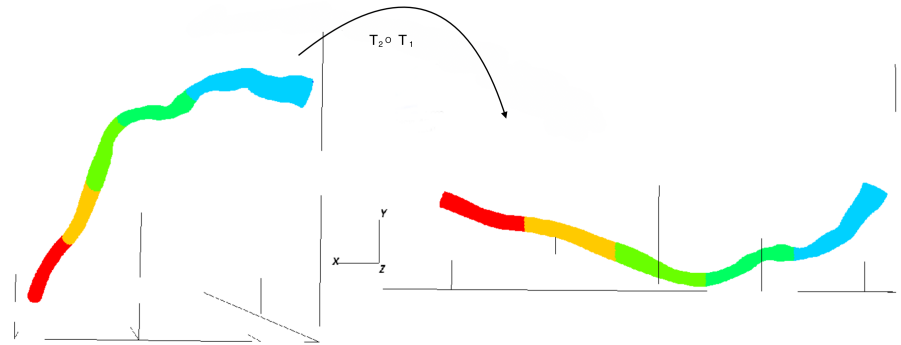


Fig 1. Orienting the vessel. Coordinate transformation $T_2 \circ T_1$ of X . The original coordinates of the vessel (left) are transformed into the new coordinate along the major axis (right). The colors show the initial five segments made along the principal axis.

Now we want to partition the vessel into m parts along the principal axis. First we find the maximum coordinate and minimum coordinate values in the first principle direction, $x^- = \min(\tilde{X}_{*,0})$ and $x^+ = \max(\tilde{X}_{*,0})$. We split the vessel into m non-overlapping intervals by dividing the total interval $L = [x^-, x^+] = \bigcup_{j=1}^m \tilde{X}_j$. Let X_j be the j^{th} segment corresponding to \tilde{X}_j in the original coordinate system. Similarly W_j is the corresponding velocity data corresponding to X_j .

Center line projection

The process above results in a set of initial conditions to a nonlinear parametric fit to the vascular geometry, which will be used to project the velocity field. The next step, is to begin the nonlinear fit by calculating the centroids of the segments X_j :

$$\bar{x}_j = \frac{1}{n_j} \sum_{\mathbf{x} \in X_j} \mathbf{x}, \quad (5)$$

where n_j is the number of elements in X_j . Now with m centroids, we can fit a parametric curve through them. The fitting can be done exactly in a collocation sense,

but we choose cubic splines with a small smoothing parameter to keep the derivative continuous. We will denote this as $\mathbf{P}(t)$, where $t \in [0, 1]$ is the parametric variable. We first use this curve \mathbf{P} to re-partition the vessel into a larger number of points M , this is done by nearest neighbors to each of the M points on the curve.

The centroids are calculated again, and a new curve is fit to these centroids. This can be done several times to get a better fit. The following Figure 2 shows the resulting bins after several iterations, with a final value of $M = 80$:



Fig 2. Segmented vessel. Partitioning of the vessel into 80 segments along cubic spline fit. Each color is a different segment.

Now that we have a good curve fit, we proceed to construct the projection operator for each point in X . This is done by sampling both \mathbf{P} and $\frac{\partial \mathbf{P}}{\partial t}$ at a larger number of points $N_p = 300$, $t_i \in T \subset [0, 1]$. This sampling is the one used to project the velocity field. Then, like the above segmentation process, we solve the nearest neighbor problem for each $\mathbf{x} \in X$:

$$t_{\mathbf{x}} = \arg \min_{t \in T} \|\mathbf{P}(t) - \mathbf{x}\|. \quad (6)$$

$t_{\mathbf{x}}$ allows us to address the curve $\mathbf{P}(t)$ for each point \mathbf{x} . The normalized derivative is then calculated:

$$\mathbf{v}_0(\mathbf{x}) = (\bar{p}_x, \bar{p}_y, \bar{p}_z) = \frac{1}{\|\frac{\partial \mathbf{P}(t_{\mathbf{x}})}{\partial t}\|} \frac{\partial \mathbf{P}(t_{\mathbf{x}})}{\partial t}. \quad (7)$$

Now, we find the null space of this vector and call them $\boldsymbol{\nu}_1(\mathbf{x})$ and $\boldsymbol{\nu}_2(\mathbf{x})$. Then we stack them into a projection matrix

$$V_p(\mathbf{x}) = (\boldsymbol{\nu}_0(\mathbf{x}), \boldsymbol{\nu}_1(\mathbf{x}), \boldsymbol{\nu}_2(\mathbf{x})). \quad (8)$$

This is used to project the velocity vector corresponding to point \mathbf{x} . This is done for all $\mathbf{v} \in W$.

Algorithm

In this section, we summarize the preprocessing algorithm and provide the pseudo codes. The algorithm is broken into several functions and several stages for clarity. The first stage finds the longest axis along the geometry and partitions the data along this axis. This is a linear approximation to seed the process. The centroids of each partition are used as points to fit a parametric cubic spline. This cubic spline is then used to

partition the geometry into more segments, again calculating the centroids, and fitting another spline. 148
149

After several rounds of this, the cubic spline fit is used as the centerline, approximating the first dimensional geometry of the vessel. This is used in a coordinate transform on the velocity field. The result is a velocity field represented in terms of its flow with respect to the geometry of the vessel. Figure 3 shows the velocity fields before (left) and after (right) the curvilinear projection. Notice how the sign of the velocity flips due to the curvature of the vessel in the left image (red area), while this is absent in the right image. 150
151
152
153
154
155
156

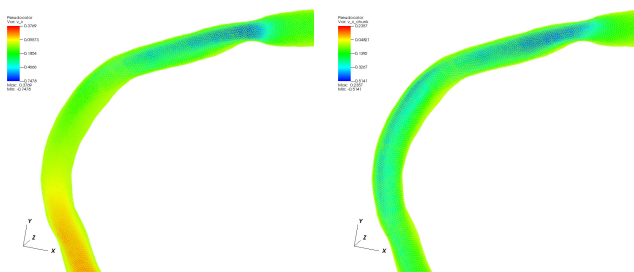


Fig 3. Transformed velocity field. Left: $v_0(x-$ direction) before projection, Right: v_0 after curvilinear projection. 157
158

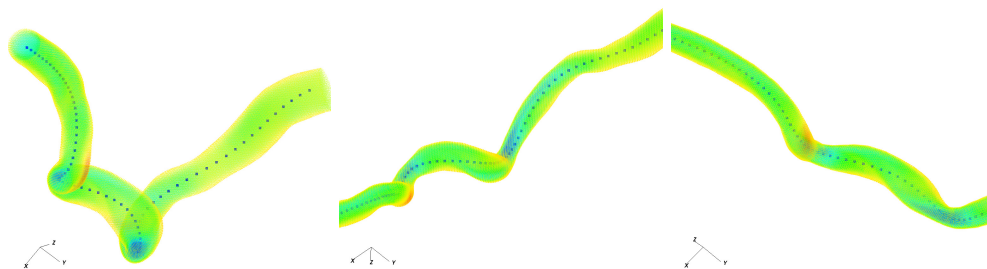


Fig 4. Parametric cubic spline fit. The least-squares fit parametric spline curve plotted to the vessel, also referred to as the centerline. 159
160

Algorithm 1: Initial linear partitioning. This serves as the initial conditions for the cubic spline fit.

Data: Spatial coordinates stacked row-wise in matrix X , velocity vectors stacked row-wise as matrix W , m ($\sim 4 - 15$) Chosen to maintain high aspect ratio and contiguity of partitions

for $i \in 0, 1, 2$ (each dimension in space) **do**
 | † $T_1: X_{*,i} \rightarrow (X_{*,i} - \text{mean}(X_{*,i}))$ (Center the domain about 0)
end

† Decompose $T_1(X)$ with SVD: $T_1(X) = U \cdot \Sigma \cdot V^H$

† Project onto the right singular vectors $T_2: \tilde{X} := T_1(X) \cdot V$

† Partition along the principal axis into m equal length intervals; sorting X by $\tilde{X}_{*,0}$ and binning into partitions, $\{X_j\}_{j=1}^m$

Algorithm 2: Segmentation. Partitioning along nonlinear axis.

Data: Spatial coordinates stacked row-wise in matrix X , M (about 10 – 100), partitions $\{X_j\}_{j=1}^m$ from Algorithm 1

† Calculate the centroids for each partition $\bar{x}_k = \frac{1}{n_k} \sum_{x \in X_k} x$

† Fit a parametric curve $\mathbf{P}(t) = (p_x(t), p_y(t), p_z(t))$ through the set of centroids $\{\bar{x}_j\}_{j=1}^m$. Cubic splines is a good choice.

† Sample $\mathbf{P}(t)$ on a large number of points $T = \{t_i\}_{i=1}^M$

for each $x \in X$ **do**

| † Find the nearest point on the sample of \mathbf{P} : $\mathbf{P}(t) | \min_{t \in T} \|\mathbf{P}(t) - \mathbf{x}\|$

end

† Partition X in this manner, so that each point in the partition belongs to a point $\mathbf{P}(t_i)$. Call each partition $\{X_i\}_{i=1}^M$.

161

Algorithm 3: Projection of velocity field. Transform of velocity field onto curvilinear co-ordinates.

input : Spatial coordinates stacked row-wise in matrix X , Velocity vectors stacked row-wise as matrix W , N_p (larger, 300+). Partitions from Algorithm 2 $\{X_k\}_{k=1}^M$

output : \tilde{W}

† Fit a parametric curve $\mathbf{P}(t) = (p_x(t), p_y(t), p_z(t))$ through the set of centroids $\{x_k\}_{k=1}^M$. Cubic splines is a good choice.

† Sample $\frac{\partial \mathbf{P}}{\partial t}$ and $\mathbf{P}(t)$ on a large number of points $T = \{t_i\}_{i=1}^{N_p}$.

for each $t \in T$ **do**

| † Normalize the derivative $\nu_0(t) = \frac{1}{\|\frac{\partial \mathbf{P}(t)}{\partial t}\|} \frac{\partial \mathbf{P}(t)}{\partial t}$.

| † Find two orthonormal complements to $\nu_0(t)$ and stack into a matrix $V(t) = (\nu_0(t), \nu_1(t), \nu_2(t))$ (column-wise).

end

for each row x **in** X **do**

| † Find the nearest neighbor on $\mathbf{P}(t)$: $t_x = \arg \min_{t \in T} \|\mathbf{P}(t) - \mathbf{x}\|$ and its associated projection operator $V(t_x)$.

| † Project \mathbf{v} , the element of W , corresponding to \mathbf{x} onto basis V : $\tilde{\mathbf{v}} = \mathbf{v} * V(t_x)$. Append to \tilde{W} .

end

Algorithm 4: Main loop. Main loop connecting all previous components.

Data: Spatial coordinates stacked row-wise in matrix X , velocity vectors stacked row-wise as matrix W , $m = 5$ (or similar, depending on aspect ratio), $N_p = 300$ (or similar)
† Do the initial linear guess on m points.
for $M \in \{10, 20, 30, 80\}$ **do**
| † Do the curve partitioning Algorithm 2 with M points.
end
† Do the curve projecting routine with N_p points Algorithm 3 to transform W to \tilde{W} .

Topological data analysis

Once all the necessary preprocessing and transformations are carried out, we apply TDA to the transformed velocity field. As mentioned in Introduction, the main purpose of this paper is to show the applicability of the TDA analysis method proposed in [1] to clinical data of general type of vasculature and to show that the proposed method is efficient in describing stenotic disease. First we must briefly explain the key element of TDA that we apply to the flow field. For more details, we refer readers to [1, 10].

Let \mathcal{X} be the topological space. The topological space \mathcal{X} that we consider in this paper is the space constructed by the velocity field. Note that, however, it is possible to construct \mathcal{X} from more hemodynamic variables than the velocity field such as the pressure variable. Although more general space is interesting, such cases are not considered in this paper. In order to construct a meaningful topological structure given the point cloud, we first explain singular homology. Homology is a topological invariant that describes the holes of various dimensions of \mathcal{X} .

The assumption is that there exists \mathcal{X} and the point cloud is a set of sampling out of \mathcal{X} . Singular homology is on \mathcal{X} rather than the point cloud. An n -simplex is a convex set composed of $n + 1$ vertices (e.g. points in velocity field in our case). For example, 0-simplex is a point, 1-simplex is an edge, 2-simplex is a filled triangle and 3-simplex is a filled tetrahedron. The n -simplex is the n -dimensional version of the triangle. The standard n -simplex is the convex hull of the standard basis in \mathbb{R}^{n+1} . Singular n -simplex in \mathcal{X} is a continuous map σ from the standard n -simplex to \mathcal{X} . Let $C_n(\mathcal{X})$ be the free abelian group whose basis is the set of singular n -simplices in \mathcal{X} . Then we can define the boundary map, δ_n from $C_n(\mathcal{X})$ to $C_{n-1}(\mathcal{X})$ where $C_{n-1}(\mathcal{X})$ is constructed from $C_n(\mathcal{X})$ by removing the vertices of $C_n(\mathcal{X})$ one at a time. Then the n -dimensional homology group H_n is defined by $H_n(\mathcal{X}) = \text{Ker } \delta_n / \text{Im } \delta_{n+1}$ where Ker and Im are the kernel and image groups, respectively. Roughly we interpret the rank of H_n as the number of holes residing in the n -dimension of \mathcal{X} . Thus if we can find H_n for $n = 0, 1, 2, \dots$, we can obtain a rough idea of the structure of \mathcal{X} in terms of holes in it. However, it is difficult to find H_n of \mathcal{X} because \mathcal{X} is arbitrary in general. Thus, instead of dealing with the homology of \mathcal{X} , we use rather the homology of the roughly imitated space of \mathcal{X} .

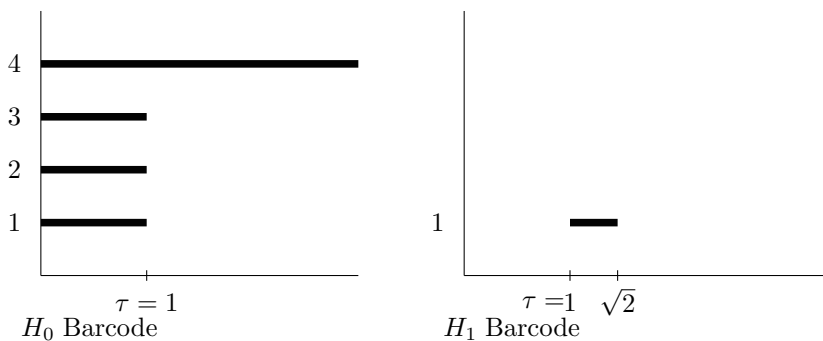
To imitate the original space \mathcal{X} algorithmically, we consider a simplicial complex K which is a collection of simplices. The collection satisfies the conditions that K contains all lower-dimensional simplices of σ if σ is a simplex of K and that the intersection of two simplices in K is a simplex in K if the intersection exists. Then in the similar way, the homology group of K , $H_n(K)$ is defined as $H_n(K) = \text{Ker } \delta_n / \text{Im } \delta_{n+1}$ where δ_n is defined similarly as above. K is a collection of simplicies, so given the point cloud, we basically connect simplices hierarchically. The number of generators of $H_n(K)$ is known as the Betti number, β_n . Roughly speaking, β_0 is the number of the connected

components of the resulting K , β_1 is the number of the one-dimensional holes of K , e.g. the loop or cycle structures of K , β_2 is the number of the two-dimensional holes of K , and so on.

In this work, we use the Vietoris-Rips simplicial complex, for which we consider the metric space (\mathcal{X}, d) where d is a usual Euclidean metric in this paper. Then the Vietoris-Rips complex, $V(\tau)$ is built on the point cloud out of \mathcal{X} having a k -simplex for every collection of $k + 1$ vertices within in a distance τ of each other in the Euclidean metric. Here, τ is known as the filtration parameter. Thus the Vietoris-Rips complex is constructed with the scale of τ . Once constructed, we can compute H_n of $V(\tau)$. However, it is difficult to know which value of τ can generate the most similar simplex to \mathcal{X} . Thus, instead of finding H_n for a fixed value of τ , we construct $V(\tau)$ for various values of τ in ascending order, say, $V(\tau_i)$, $i = 0, 1, 2, \dots$ and $\tau_i < \tau_j$ for $i < j$. Then we have various $H_n(V(\tau_i))$ for various τ_i . Clearly given N number of vertices, if $\tau = 0$, there are N vertices isolated and not-connected and obviously $\beta_0 = N$, $\beta_1 = 0$, $\beta_2 = 0$, and so on.

Persistence Barcode: The variation of homology with respect to τ yields the concept of persistent homology. The collection of such homology versus τ is known as the barcode. The barcode provides the information of the Betti number and the information of the persistence, the length of the generating barcode, across τ . The persistence tells how long the given geometric hole structure maintains with τ . The persistence, the interval of each barcode, plays an important role in our method. Particularly, the persistence of the 1D and 2D barcodes are closely related to the degree of flow complexity when the velocity field is mapped onto the unit sphere. The zero-dimensional barcode of H_0 is the graph of β_0 versus τ . For the given N vertices, $\beta_0 = N$ for $\tau = 0$. Thus all the 0-dimensional barcodes of H_0 starts from $\tau = 0$ while those barcodes of higher dimension than 0 starts from τ where the corresponding hole forms.

The following diagram shows the barcodes for H_0 (left) and H_1 (right), respectively, for the point cloud which is composed of the four vertices of the unit square. From the left barcode we know that there are four vertices given (4 bars in y -axis) when $\tau = 0$. We also know that when $\tau = 1$ all the four vertices are connected and there remains only one connected component afterwards, which is visualized as the single barcode labeled 4 in y -axis. As all the four points are connected while no vertices are connected diagonally because $\tau = 1$ is smaller than the diagonal distance of two vertices. Thus we know that one hole forms at $\tau = 1$ and it lasts until $\tau = \sqrt{2}$ where two vertices are connected diagonally and two triangles form. Note that once triangle is formed, inside is filled as a 2-simplex. Then the hole also disappears, which is the end point of the one-dimensional barcode in the right figure.



Persistent homology of vascular flows: TDA of vascular disease is to construct a simplicial complex based on the velocity fields. For the given point cloud (velocity fields in our case), a simplicial complex is created by gluing a finite number of simplices together, for which a gluing condition is needed. For the condition, the topology is

described with a metric with the parameter τ explained above. The distance between two simplices is determined using the Euclidean metric d . For example, two 0-simplices must be connected forming an edge if their Euclidean distance is less than or equal to the value of τ . Here we note that it is hard to use the raw data of vascular flows for TDA. In [1] it was shown that the raw data needs to be transformed into a point cloud that is suitable for a meaningful TDA. We use the S^2 projection proposed in [1] also described in the following, but we also note that this is not a unique transform. There could be a better transform for the construction of the simplicial complex.

Figures 5 and 6 show the sample S^2 projection of the curved vessel obtained from the patients' data. The color represents the pressure value, which is absent from our analysis but helpful in this image. As shown in the figures, the S^2 projection shows how the velocity fields are distributed. It is interesting to observe the existence of the circular pattern on S^2 . We hypothesize that those circular patterns are related to the prediction of the disease development, which we do not aim to investigate but will be considered in our future work. In TDA such circular patterns are summarized in 1D homology H^1 , while existence of voids due to multidirectional flow are described by H_2 .

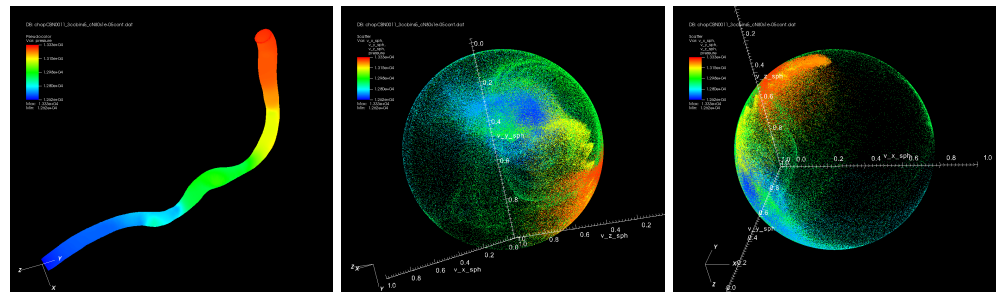


Fig 5. Spherical projection Sample spherical projection with pressure as color variable(middle, right), and pseudocolor plot of pressure (left) on the actual morphology of the stenotic vessel.

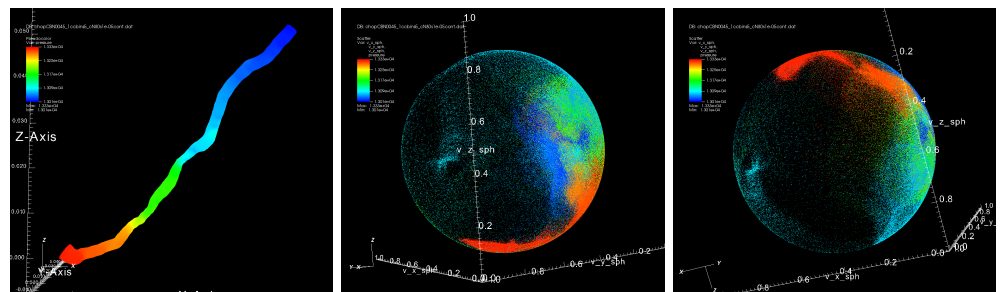


Fig 6. Spherical projection Sample spherical projection with pressure as color variable(middle, right), and pseudocolor plot of pressure (left) on the actual morphology of the stenotic vessel.

Application to the preprocessed vessel

Calculating the persistent homology on the entire vessel is prohibitively expensive, so we perform out TDA on small segments of each vessel, as depicted in Figure 2. Even with the centerline projection, there is a large amount of noise in the boundary regions of the vessel. Due to the spherical projection, this noise appears as any other velocity vector to the TDA steps. Operating on small segments of the vessel helps alleviate inaccuracies caused by this noise.

According to [1], the 2D persistence homology of the velocity fields projected on S^2 can be used to determine whether there is a stenosis. Therefore, by decomposing the

whole vessel into smaller segments and calculating the 2D persistences homology within each one, we can get the identity of the segment where stenosis occurs. With the centerline projection methodology described previously, the segmentation of vessels is obtained by chopping up the 3D space of vessels into M small segments along with the parametric curve fit. One can see that the size of each segment is not necessarily the same since it depends on the slope of the centerline in the segment region.

We work with the maximum persistence in both H_1 and H_2 . This is the length of the longest interval in the persistence diagram, which can be visually understood as the longest line in the bar code. The maximum H_2 persistence is used to determine the segment closest to the stenosis.

Results

Here we validate the proposed method using clinical data, which is the primary result of this paper. We show a statistically significant correlation between the FFR and the topological persistence calculated in the manner described in the previous sections. To be specific, after segmentation of vessels with preprocessing, we consider the maximum H_1 and H_2 persistences in each segment of vessels, which are the maximal length of the generated 1D and 2D barcodes of the given point cloud, i.e. the velocity field projected on S^2 in this work. We find that there is highly weak correlation in the case where the preprocessing steps are omitted. This is due to the confounding factor introduced by the arbitrary orientation of the vessel in the spatial domain, which results in a misrepresentation of the velocity field, where the direction of the flow changes with respect to the coordinate system, but not with respect to the vessel. Preprocessing with centerline projection alleviates this unwanted factor, and provides an appropriate segmentation of the vessel along its natural curvature.

The results are given in Table 1. The column labeled *Segment number* is the number of the segment in the vessel where the largest maximum H_2 persistence occurs, which is presumably the location of the stenosis. The column labeled *FFR* is the FFR value, which is estimated by dividing a distal pressure by proximal pressure, both of which are measured after a probable lesion is located using X-ray angiography. And the column labeled *H_1 Persistence* is the maximum H_1 persistence in the segment where the largest maximum H_2 persistence occurs. And the column labeled *H_2 Persistence* is the value of the largest maximum H_2 persistence, i.e., the maximum H_2 persistence in the corresponding *Segment*. The first column is the vessel id.

Vessel Id.	FFR	H_1 Persistence	H_2 Persistence	Segment number
CBN0759	0.62	1.01179	0.22	54
CBN2122	0.66	0.41905	0.3	21
CBN0012	0.69	0.82216	0.16	17
CBN3355	0.69	0.53659	0.18	18
CBN0006	0.7	1.05928	0.14	21
CBN1290	0.7	0.52892	0.12	21
CBN3222	0.7	0.71805	0.18	52
CBN0047	0.72	0.49718	0.18	34
CBN3366	0.72	0.67079	0.28	22
CBN0045	0.73	0.46359	0.42	45
CBN0018	0.74	0.75791	0.5	38
CBN1255	0.74	0.55535	0.22	36
CBN1353	0.75	0.64994	0.22	8
CBN1711	0.76	0.79459	0.16	32
CBN3265	0.77	0.4601	0.14	12
CBN1960	0.78	0.77871	0.32	7
CBN1492	0.78	0.65907	0.12	61
CBN3284	0.78	0.32153	0.56	33
CBN2034	0.79	0.90161	0.18	49
CBN0754	0.79	0.64229	0.14	33
CBN1368	0.8	0.87848	0.1	37
CBN1203	0.8	0.33451	0.16	71
CBN0874	0.8	0.39672	0.52	27
CBN1741	0.8	0.79987	0.2	50
CBN0569	0.8	0.60789	0.36	39
CBN1088	0.81	0.60526	0.14	18
CBN2043	0.81	0.72963	0.24	53
CBN3254	0.81	1.1454	0.12	36
CBN0022	0.81	0.2421	0.5	30
CBN1185	0.81	0.45952	0.2	70
CBN1345	0.82	0.49201	0.38	32
CBN1510	0.82	0.34708	0.44	3
CBN3400	0.82	0.35935	0.14	42
CBN1103	0.83	0.60585	0.12	39
CBN1484	0.83	1.01112	0.1	50
CBN0252	0.83	0.38384	0.38	51
CBN1084	0.84	1.24371	0.12	41
CBN1089	0.84	0.38813	0.26	48
CBN1868	0.85	0.35258	0.38	44
CBN0003	0.85	0.46665	0.12	49
CBN0540	0.85	0.719	0.24	41
CBN0014	0.85	0.30592	0.44	41
CBN1419	0.86	0.34468	0.26	28
CBN0918	0.86	0.41882	0.2	65
CBN1731	0.86	0.59057	0.16	79
CBN0460	0.86	0.43247	0.44	45
CBN1245	0.87	0.34222	0.52	39
CBN0011	0.87	0.37225	0.2	41
CBN0046	0.87	0.39283	0.48	59
CBN0048	0.87	0.78032	0.16	12
CBN0520	0.87	0.36278	0.18	42
CBN1404	0.88	0.61015	0.22	62
CBN0559	0.88	0.8052	0.16	44
CBN1379	0.89	0.39916	0.42	49
CBN1594	0.89	0.26515	0.54	5
CBN3446	0.9	0.71575	0.18	46
CBN0582	0.91	0.05394	0.0	64
CBN1602	0.91	0.55928	0.18	39
CBN0443	0.92	0.3045	0.56	34
CBN1096	0.92	0.07802	0.0	24
CBN0013	0.93	0.01654	0.0	5
CBN0051	0.94	0.40696	0.38	21
CBN0038	0.95	0.33592	0.06	55
CBN1258	0.96	0.91225	0.18	64

Table 1. Results: For each vessel, the FFR value, maximum H_1 and maximum H_2 persistences in the segment where the largest maximum H_2 persistence occurs.

We observe that for each vessel, the maximum H_2 persistences are close to zero in most segments. When a segment is not near the stenosis, the variation of the flow directions in this segment after preprocessing is reduced, i.e., the point cloud on S^2 concentrates near one point, hence no significant two-dimensional topological structure in it. Figures 7 and 8 show the maximum H_1 and H_2 persistences in each segment for a few vessels as an example. The y -axis of all top images is the maximum H_1 persistence, the y -axis of all images on the bottom is the maximum H_2 persistence, and the x -axis of all images is the segment number. We can see that the vessels with ID CBN0013, CBN0582 and CBN1096 in Figure 7, which have FFR values 0.93, 0.91 and 0.92, respectively, implying a healthy state, have close-to-zero maximum H_2 persistences in all segments. However, for the vessels with ID CBN0018, CBN0045 and CBN2122, which have FFR values 0.74, 0.73 and 0.66 in Figure 8, there are segments in which the maximum H_2 persistences are larger than 0.3.

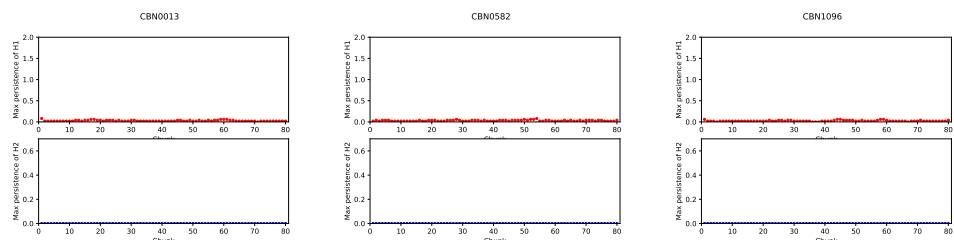


Fig 7. Maximum H_1 (Top) and H_2 (Bottom) persistences in each segment for vessels CBN0013, CBN0582, CBN1096

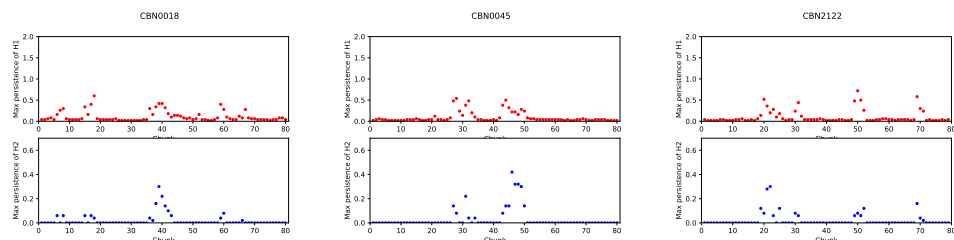


Fig 8. Maximum H_1 (Top) and H_2 (Bottom) persistences in each segment for vessels CBN0018, CBN0045, CBN2122

Figures 9 and 10 show the H_0 (top figure), H_1 (middle) and H_2 (bottom) barcodes for the corresponding vessels in Figures 7 and 8, respectively. The barcodes are calculated using the point cloud collected from the whole vessel domains after the preprocessing. As shown in the figure, the H_2 persistences for the vessels of CBN0013, CBN0582 and CBN1096 are non-existent as shown in Figure 7 while the H_2 persistences for the vessels of CBN0018, CBN0045 and CBN2122 are significant as shown in Figure 8. We also observe that the H_1 persistences of the vessels of CBN0013, CBN0582 and CBN1096 are smaller than those of the vessels of CBN0018, CBN0045 and CBN2122. Note that H_1 persistences in Figure 9 are existent while those in Figure 7 are not. This is due the fact that the H_1 in Figure 7 is only from a segment of the given vessel but the H_1 in Figure 9 is calculated using the point cloud from the whole vessel.

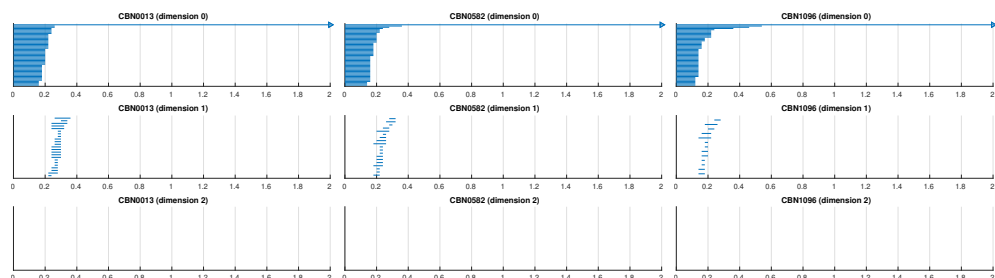


Fig 9. Barcodes for CBN0013, CBN0582, CBN1096 for H_0 (top), H_1 (middle) and H_2 (bottom).

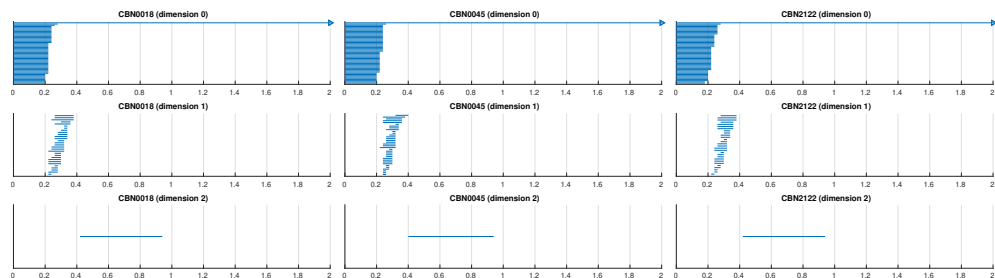


Fig 10. Barcodes for CBN0018, CBN0045, CBN2122 for H_0 (top), H_1 (middle) and H_2 (bottom).

In Figure 11, we show the correlation between the FFR values and the maximum H_1 persistences from Table 1. In both images, the x -axis is the FFR value and the y -axis is the maximum H_1 persistence in the segment where the largest maximum H_2 persistence occurs for each vessel. The straight lines are fitting lines after applying linear regression with least squares method to the corresponding points in each image. We can see that the p -values for the linear fit are 0.13852 and 0.00248 for the non-preprocessed and preprocessed data, respectively. This demonstrates that we do not have a statistically significant correlation between FFR and our potential diagnostic index without the preprocessing scheme, but have a potentially useful diagnostic index with TDA if the proper preprocessing is applied.

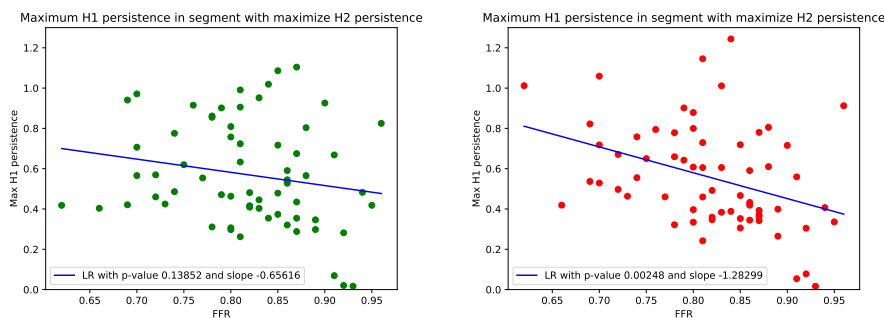


Fig 11. Correlation with FFR. Correlation between maximum H_1 persistence in the segment where the largest maximum H_2 persistence occurs and the FFR values for both before (Left) and after (Right) preprocessing with the centerline projection scheme.

Conclusion

Vascular disease is the leading cause of death worldwide. Accurate measurement of the degree of vascular disease is crucial for diagnosis and treatment. In [1] a new method based on TDA has been proposed, employing topological features as a potential diagnostic index. The proposed method was tested with the CFD solutions calculated in the straight stenotic vessels without curvatures in [1]. In this paper, we developed a preprocessing method that enables the TDA method to be applied to the curved stenotic vessel. We applied the proposed method to clinically sourced data, showing a statistically significant correlation to the gold standard diagnostic index, the FFR. It was also shown that the preprocessing method was necessary to achieve a statistically significant correlation. This validates the proposed method as a potentially powerful diagnostic tool.

In future work we plan to refine the preprocessing method and make adaptations for application to more general vascular geometries such as bifurcated vessels, multiple stenoses. There is also potential for improvement of computational efficiency, parameter selection and other methodologies in the preprocessing scheme.

S1 Appendix.

S1 Figures: Maximum H_1 and H_2 versus segment number In this appendix, we provide the maximum H_1 and H_2 persistences calculated in each vessel segment used in preprocessing steps versus the segment number for those vessels in Table 1. These figures show not only the flow complexity but also the possible locations of stenosis.

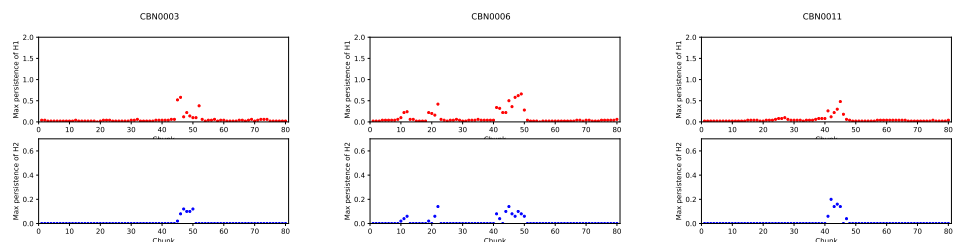


Fig 12. Maximum persistences for CBN0003, CBN0006, CBN0011

369
370

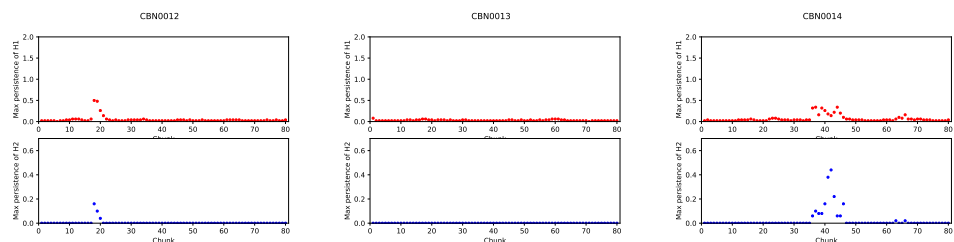


Fig 13. Maximum persistences for CBN0012, CBN0013, CBN0014

371
372

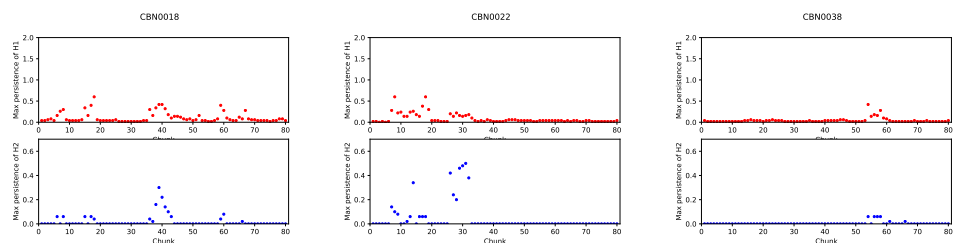


Fig 14. Maximum persistences for CBN0018, CBN0022, CBN0038

373
374

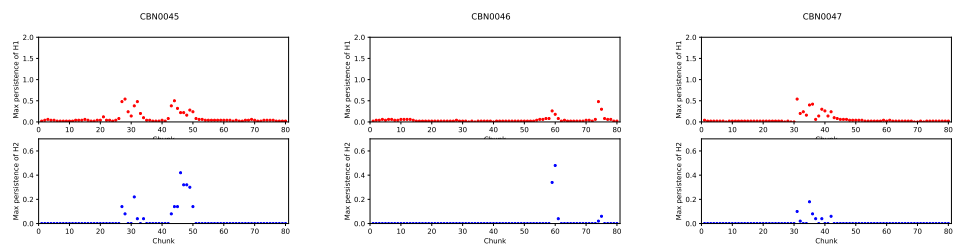


Fig 15. Maximum persistences for CBN0045, CBN0046, CBN0047

375
376

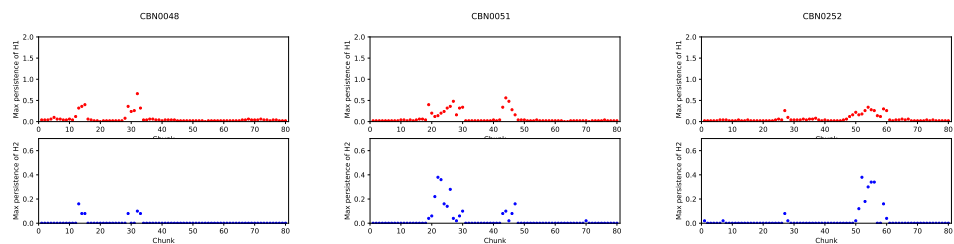


Fig 16. Maximum persistences for CBN0048, CBN0051, CBN0252

377
378

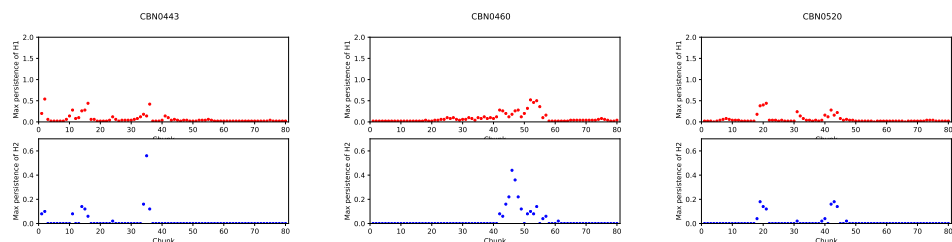


Fig 17. Maximum persistences for CBN0443, CBN0460, CBN0520

379
380

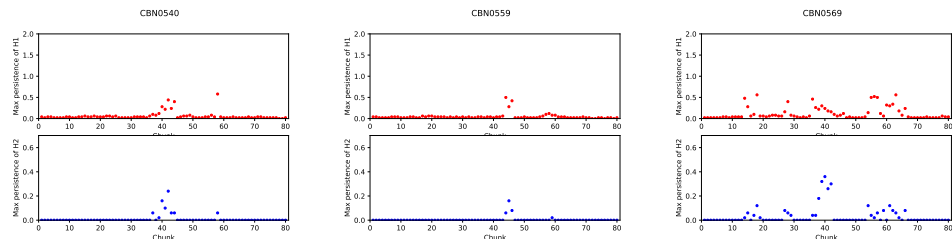


Fig 18. Maximum persistences for CBN0540, CBN0559, CBN0569

381
382

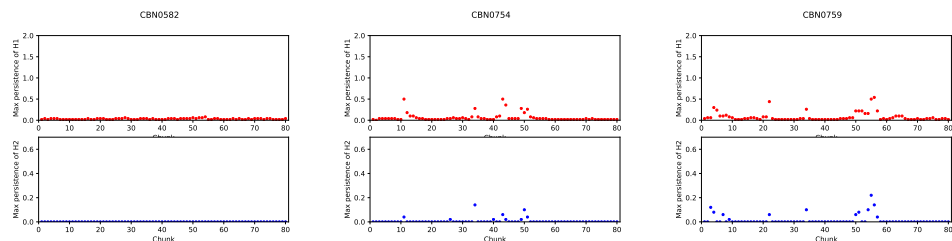


Fig 19. Maximum persistences for CBN0582, CBN0754, CBN0759

383
384

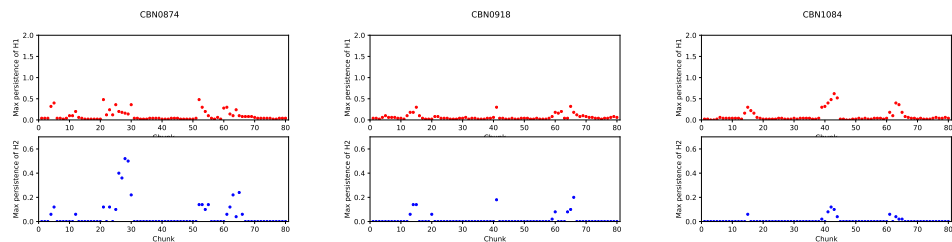


Fig 20. Maximum persistences for CBN0874, CBN0918, CBN1084

385
386

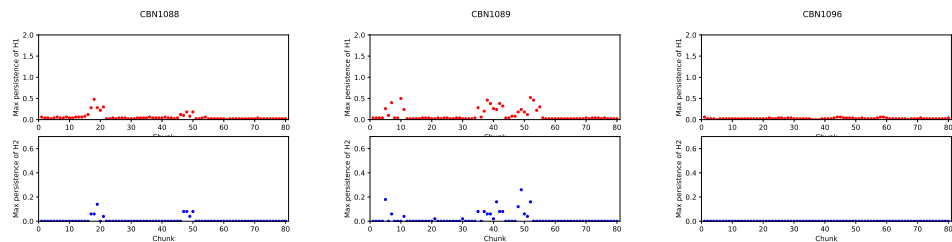


Fig 21. Maximum persistences for CBN1088, CBN1089, CBN1096

387
388

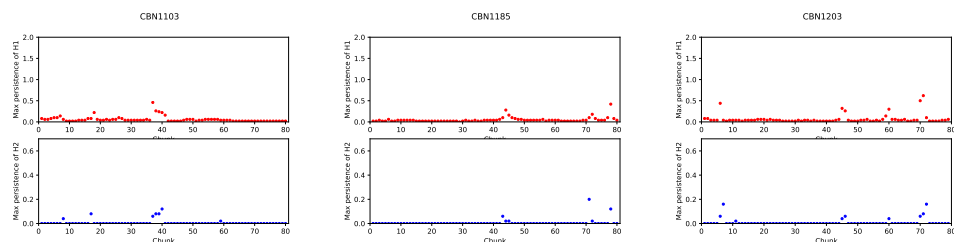


Fig 22. Maximum persistences for CBN1103, CBN1185, CBN1203

389
390

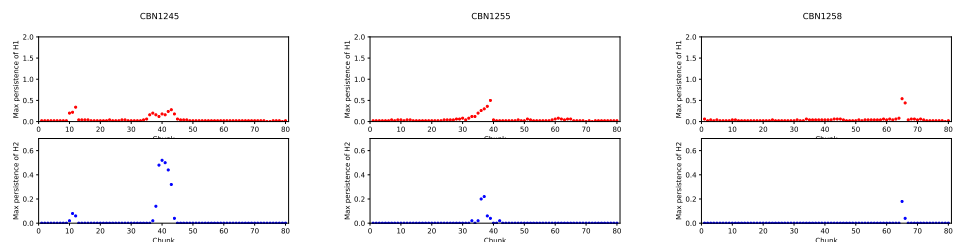


Fig 23. Maximum persistences for CBN1245, CBN1255, CBN1258

391
392

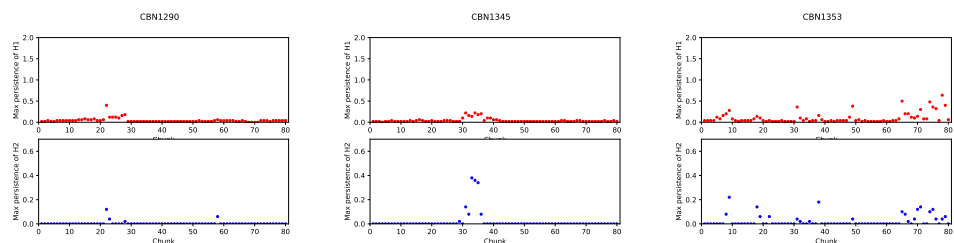


Fig 24. Maximum persistences for CBN1290, CBN1345, CBN1353

393
394

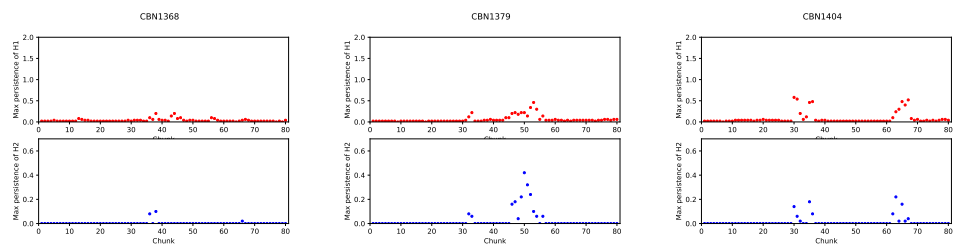


Fig 25. Maximum persistences for CBN1368, CBN1379, CBN1404

395
396

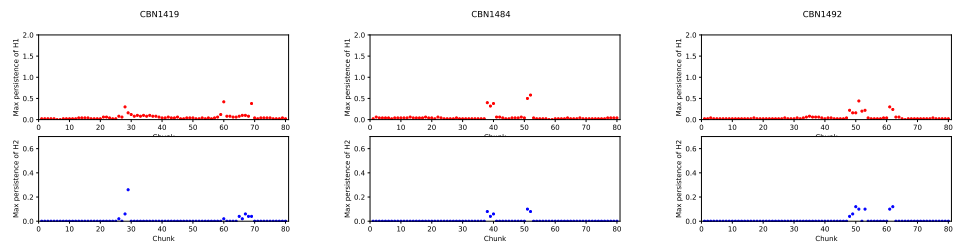


Fig 26. Maximum persistences for CBN1419, CBN1484, CBN1492

397
398

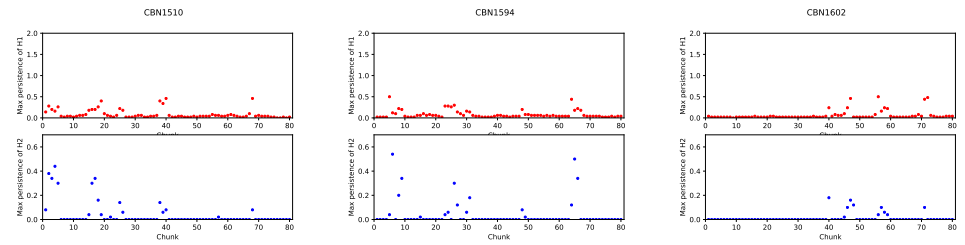


Fig 27. Maximum persistences for CBN1510, CBN1594, CBN1602

399
400

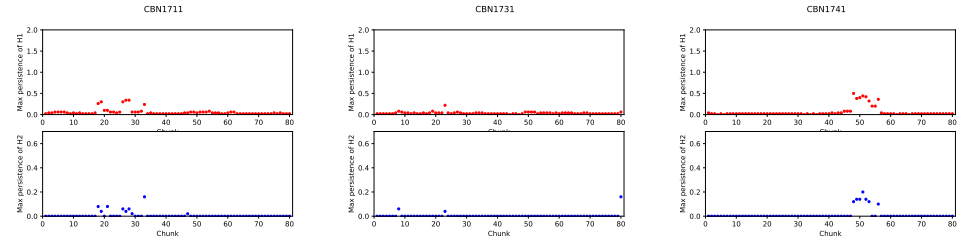


Fig 28. Maximum persistences for CBN1711, CBN1731, CBN1741

401
402

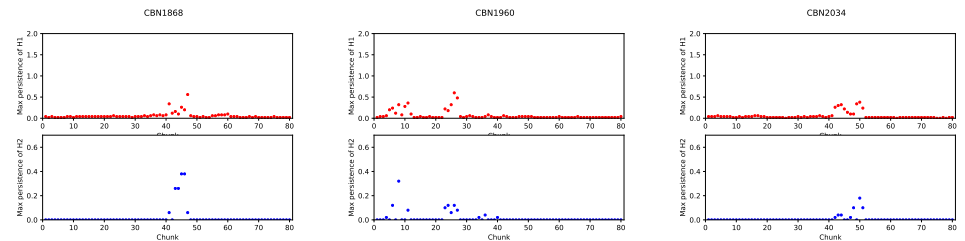


Fig 29. Maximum persistences for CBN1868, CBN1960, CBN2034

403
404

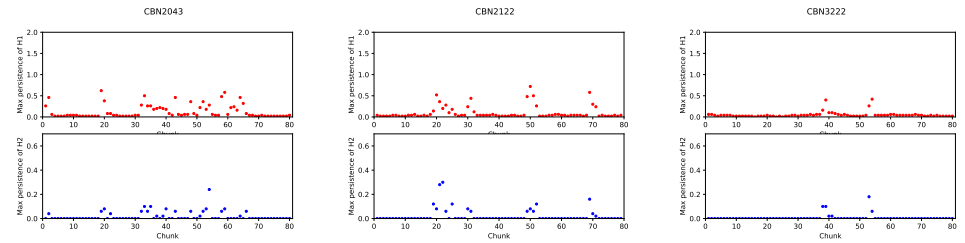


Fig 30. Maximum persistences for CBN2043, CBN2122, CBN3222

405
406

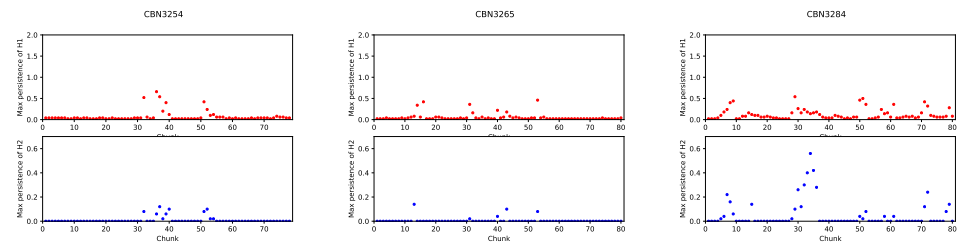


Fig 31. Maximum persistences for CBN3254, CBN3265, CBN3284

407
408

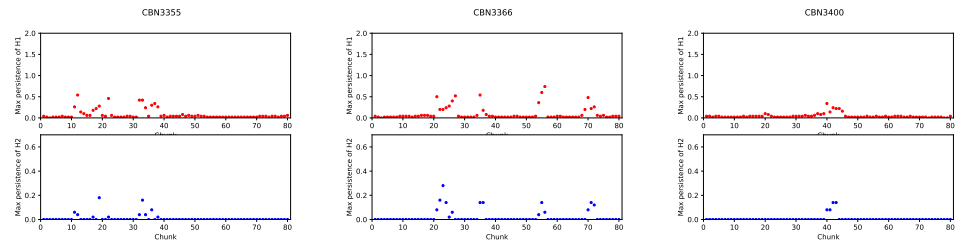


Fig 32. Maximum persistences for CBN3355, CBN3366, CBN3400

409
410

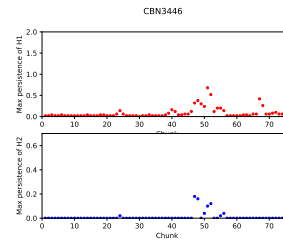


Fig 33. Maximum persistences for CBN3446

411
412

Acknowledgments

413

This work has been supported by Samsung Science & Technology Foundation under grant number SSTF-BA1802-02. Christopher Bresten is supported by KRF program by National Research Foundation under NRF-2018H1D3A1A01074679.

414
415
416

References

1. Nicponski J, Jung JH. Topological data analysis of vascular disease: I A Theoretical framework. *Frontiers in Applied Mathematics and Statistics*. 2020;6:34:1–22.
2. Islam Y Elgendy AAB C Richard Conti. Fractional flow reserve: an updated review. *Clin Cardiol*. 2014;37(6):371–380.
3. Tonino P, Fearon W, De Bruyne B, Oldroyd K, Leesar M, Ver Lee P, et al. Angiographic versus functional severity of coronary artery stenoses in the FAME study fractional flow reserve versus angiography in multivessel evaluation. *J Am Coll Cardiol*. 2010;55 (25):2816–2821.
4. Zhong L, Zhang JM, Su B, Tan RS, Allen JC, Kassab GS. Application of Patient-Specific Computational Fluid Dynamics in Coronary and Intra-Cardiac Flow Simulations: Challenges and Opportunities. *Frontiers in Physiology*. 2018;9:742.
5. Leiner T SAea Rueckert D. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *J Cardiovasc Magn Reson*. 2019;21:61.
6. Martin-Isla C, Campello V, Izquierdo C, Raisi-Estabragh Z, Baeßler B, SE P, et al. Image-based cardiac diagnosis with machine learning: a review. *Front Cardiovasc Med*. 2020;24:1.

7. Shameer K, Johnson K, Glicksberg B, Dudley J, Sengupta P. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *Heart*. 2018;104:1156–64.
8. G C. Topology and data. *Bull Am Math Soc*. 2009;46:255–308.
9. Adams H, Tausz A, Vejdemo-Johansson M. JavaPlex: A research software package for persistent (co) homology. In: *International Congress on Mathematical Software*. Springer; 2014. p. 129–136.
10. Carlsson G. Topology and data. *Bulletin of the American Mathematical Society*. 2009;46(2):255–308.
11. Jung JH, Lee J, Hoffmann KR, Dorazio T, Pitman EB. A rapid interpolation method of finding vascular CFD solutions with spectral collocation methods. *Journal of Computational Science*. 2013;4(1-2):101–110.
12. Wall ME, Rechtsteiner A, Rocha LM. Singular value decomposition and principal component analysis. In: *A practical approach to microarray data analysis*. Springer; 2003. p. 91–109.
13. Bendich P, Marron JS, Miller E, Pieloch A, Skwerer S. Persistent homology analysis of brain artery trees. *The annals of applied statistics*. 2016;10(1):198.
14. Gholizadeh S, Zadrozny W. A Short Survey of Topological Data Analysis in Time Series and Systems Analysis. *arXiv preprint arXiv:180910745*. 2018;.
15. Maria C, Boissonnat JD, Glisse M, Yvinec M. The gudhi library: Simplicial complexes and persistent homology. In: *International Congress on Mathematical Software*. Springer; 2014. p. 167–174.
16. Gudhi :: Anaconda Cloud; 2018. <https://anaconda.org/conda-forge/gudhi>.
17. de Silva V, Skraba P, Vejdemo-Johansson M. Topological analysis of recurrent systems. In: *Workshop on Algebraic Topology and Machine Learning, NIPS*; 2012.
18. Abbott BP, Abbott R, Abbott T, Abernathy M, Acernese F, Ackley K, et al. GW151226: observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Physical review letters*. 2016;116(24):241103.
19. Hofer C, Kwitt R, Niethammer M, Uhl A. Deep learning with topological signatures. In: *Advances in Neural Information Processing Systems*; 2017. p. 1634–1644.
20. Bendich P, Marron JS, Miller E, Pieloch A, Skwerer S. Persistent homology analysis of brain artery trees. *The annals of applied statistics*. 2016;10(1):198.