scientific reports

Check for updates

OPEN

Persistence diagrams for exploring the shape variability of abdominal aortic aneurysms

Dario Arnaldo Domanin[1,6], Matteo Pegoraro[2,6]✉, Santi Trimarchi[3,4], Maurizio Domanin[3,4] & Piercesare Secchi[5]

Abdominal aortic aneurysm consists of a permanent dilation in the abdominal portion of the aorta and, along with its associated pathologies like calcifications and intraluminal thrombi, is one of the most important pathologies of the circulatory system. The shape of the aorta is among the primary drivers for these health issues, with particular reference to all the characteristics which affect the hemodynamics. Starting from the computed tomography angiography of a patient, we propose to summarize such information using tools derived from Topological Data Analysis, obtaining persistence diagrams which describe the irregularities of the lumen of the aorta. We showcase the effectiveness of such shape-related descriptors with a series of supervised and unsupervised case studies.

One of the main drivers of contemporary medical research is the idea of *personalized medicine*, which aims at providing better targeted treatments to improve therapy outcomes. The achievement of such a challenging task relies primarily on two factors: the constantly increasing amount of data made available by modern data collecting pipelines, like medical imaging, and the joint efforts of clinicians and statisticians, trying to unpack the valuable information contained in the provided data.

The statistical understanding of any phenomenon is always limited by the interpretable methods and algorithms that the analyst can resort to. As a consequence, there is an increasing need of new and original statistical methods aimed at the analysis of different and heterogeneous kinds of complex data, validated and interpreted by the expert clinician. In many medical situations, especially those related to imaging data, many complexities arise because of the difficulty in decoupling interesting variability between statistical units from that which the clinician considers as ancillary. This problem is at the root of *object oriented data analysis*[1], whose foundational principles focus on data representation, by embedding the atoms of the statistical analysis in a suitable mathematical space where their variability of interest can be captured and explored.

In this work we consider the problem of analysing and characterising *abdominal aortic aneurysms* (AAAs) eliciting the information contained in the shape of blood vessels with tools from algebraic topology.

Abdominal aortic aneurysm consists of a permanent dilation of at least 1.5 times the expected diameter of the abdominal portion of the aorta (see Fig. 1a), the biggest artery of the human body[2], characterized by chronic inflammation processes in the inner layers with degradation or redistribution of elastin and collagen[3] under the haemodynamical stress[4]. Such dilatation can evolve asymptomatically towards the progressive enlargement of the vessel up to its rupture which is frequently fatal, making it one of the most prominent pathologies of the circulatory system. AAA is the most common aortic pathology. The incidence of ruptured abdominal aortic aneurysm is up to 17.5 per 100,000 person/year in Western countries[5]. It is mainly located between the renal and the iliac arteries, which can also present dilatations on their owns called *iliac aneurysms*. The healthy segment of the abdominal aorta above the dilatation is defined as *aortic neck*.

Two main features characterize the aneurysm: *intraluminal thrombus* (ILT), thrombus for short, and wall *calcification*. The former is the stratification of several blood particles in the inner layers, the latter is the accumulation of calcium in the aortic wall responsible of its stiffening. Both negatively affect the aorta changing its physiological behavior by modifying the *lumen*'s structure, the region where the blood flows, with bumps and irregularities.

Since vessel's morphology plays a major role in local haemodynamics which, in turn, determines the development of the AAA[6–9], in this study we focus on the aortic lumen. Hence we are only interested in the

[1]Moxoff, 20127 Milan, Italy. [2]Department of Mathematical Sciences, Aalborg University, 9220 Aalborg, Denmark. [3]Dipartimento di Scienze Cliniche e di Comunità, Università degli Studi di Milano, Milan, Italy. [4]Dipartimento CardioToracoVascolare, S.C. di Chirurgia Vascolare, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy. [5]MOX, Department of Mathematics, Politecnico di Milano, 20133 Milan, Italy. [6]These authors contributed equally: Dario Arnaldo Domanin and Matteo Pegoraro. ✉email: matteop@math.aau.dk

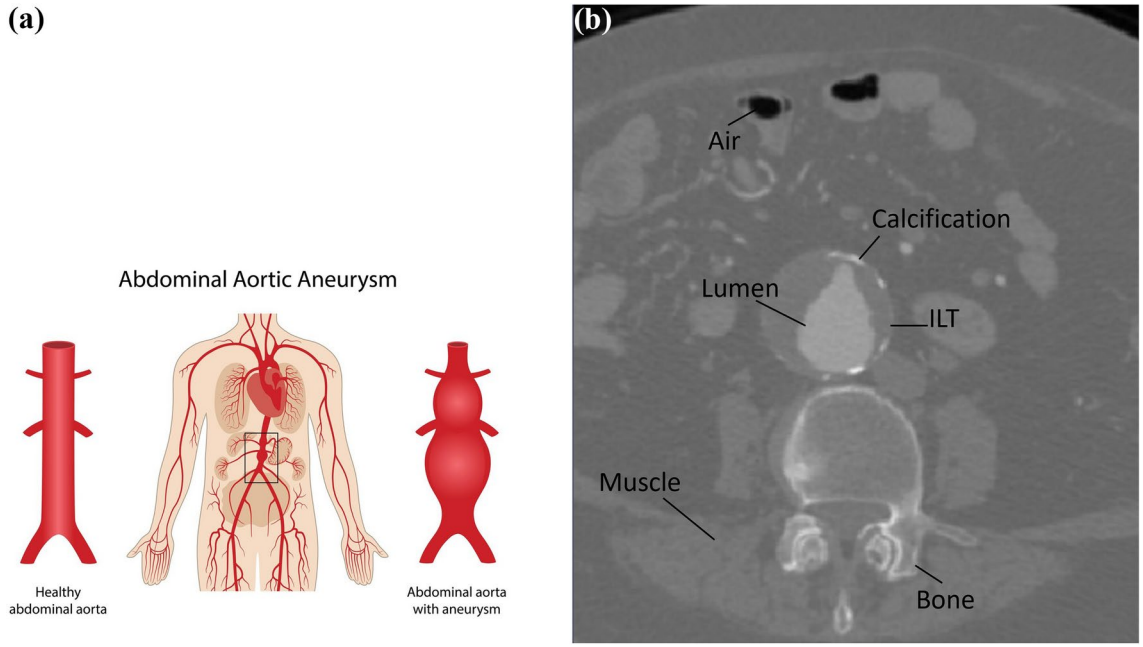

Figure 1. (**a**) A schematic view of the circulatory system with highlighted the abdominal portion of the aorta, with both an healthy and aneurysmatic example. Image from https://www.vascularcures.org/abdominal-aortic-aneurysms; (**b**) A coronal slice of a CTA scan. Different shades of grey differentiate between different organs and tissues. This allows both the clinicians to see the entire anatomy of the patient and the algorithm to execute the segmentation process.

information directly associated to the shape of the lumen. For this reason, calcium formations which could occur on the external part of a thrombus will be disregarded. We point out that also clinical practice generally gives more relevance to lumen-exposed calcifications, due to the protective role of the thrombus[10,11].

Most of the literature dealing with the analysis of AAAs focuses on the growth rate of small AAAs[12,13], in order to prevent their rupture and identify the necessity of surgery, leveraging on shape-related or biomechanical numerical features. See, for instance,[14–18]. Differences between these works appear in the statistical methodology employed for prediction, but, mostly, in the kind of features collected: AAA and intraluminal thrombus' diameter, volume, axes are considered in most studies;[16,18] and others include also biomechanical indices - like peak wall stress and peak wall rupture indices - and clinical variables, showing improved predictions.

The aim of our work is different and is more in line with[19]; indeed, we want to estabilish a new mathematical representation of the geometric complexity of the aorta and of the surrounding blood vessels. A deep dive on the heuristic power of this new representation is the main aim of this paper; we leave to future work the assessment of its predictive power in terms of growth rate of small AAAs. Nevertheless, most of the information conveyed by morphological and shape related variables is also contained in our novel representation, which, however, has the extra advantage of summarizing additional information which is harder to convey in terms of numerical variables; like, for instance, a quantification of the calcifications along the blood vessels.

To obtain such representation we resort to algebraic topology[20] and, in particular, to *persistent homology*[21], one of the most diffused frameworks in Topological Data Analysis (TDA). By introducing a carefully studied and application-driven filtering function, we are able to synthesize the Computed Tomography Angiographies (CTAs) of patients - a particular kind of medical images, see Fig. 1b - capturing the statistically sufficient features of their *shapes*. Moving from a preliminary segmentation pipeline, already well established and whose goal is to turn CTA images into 3D meshes, we transform them into mathematical objects called *persistence diagrams*. Thus, the statistical analysis is moved from the initial CTA images to their representations embedded in the space of persistence diagrams. To demonstrate the effectiveness of this representation we propose a number of analyses characterized by the simplicity of the pipeline, but heavily relying on the power of the chosen representation.

The approach pursued in this work paves the way for more refined analyses - for instance on the growth rate of small AAAs - grounded on the same topological representation and aimed at providing clinicians with valuable insights and checkpoints to complement their current analyses. We also foresee applications of analogous filtering functions to other contexts needing similar shape-dependent characterizations, such as cerebral aneurysm and the development of the atheromatous plaque in arterial vessels.

Structure of the paper

The manuscript is structured in order to put emphasis on the methods section, which conveys the main novelties of the present work. These lead to two kinds of results: results regarding within-patient variability and results regarding between-patients variability. More in details: in "Data", we give a brief overview of abdominal aortic aneurysms and their related pathologies, and concisely describe the segmentation pipeline to obtain the meshes

from the CTA scans. "Methods: object representation through persistent homology" is the methods section, devoted to presenting the framework which leads to the representation of patients' meshes via the persistence diagrams of a radial filtration function. "Results: reading a persistence diagram" is entirely focused on reading patient-related information from persistence diagrams. In "Results: population analyses" we propose a number of data analysis situations which are easily tackled with the use of persistence diagrams. Finally, "Discussion and conclusion" ends the manuscript with a discussion and a conclusion. The Appendix contains some further technical details.

Data

In this section we describe our data set, outlining the pipeline used to obtain the final meshes.

CTA scans

The most used and effective method for anatomy visualization is Computed Tomography Angiography. CTA is used to get a visualization of the human body and is realized with emission of X-rays through a tomograph. A CTA scan consists of a 3-dimensional image of a region of interest in the form of a volume composed by voxels with different shades of grey. Each shade is associated to a different tissue, making it possible to distinguish not only between fat, bones and muscle but also artery, intraluminal thrombus and calcification - see also Figure 1b.

The data considered in this work consist of 48 CTA scans of the abdominal region, 24 of which picture a pathological aorta. To secure data consistency, the following criteria have been applied:

- non-minor and non-pregnant patients.
- in case of AAA, AAA greater than 3.5 *cm* in diameter;
- in case of AAA, infrarenal AAA not extended above the renal arteries with presence of aortic neck;
- CTA performed from the origin of the supra aortic branches to the inguinal fold;
- CTA with slice thickness ≤ 1 mm, with dye contrast volume of $70 - 90$ ml at 5 ml/s, approximately 1 ml of contrast per kg, scan delay of 8 seconds post injection for sufficient arterial filling CTA, to assure high quality data.

All the CTA scans are provided by Ospedale Maggiore Policlinico. Patients with healthy aortas have been collected from the Policlinico's archive while the patients affected by AAA come from the Vascular Surgery Unit. For privacy reasons, all data are anonymized.

The software used to visualize all the meshes and CTA scans shown in this work is *Paraview*[22] version 5.11.0.

Segmentation pipeline

The primary function of a CTA scan in this study is to enable the segmentation of the aortic lumen, rather than for visualization purposes. Aortic segmentation refers to the process of obtaining a three-dimensional reconstruction of the lumen via a finite number of points, as visually described in Fig. 2. The product of the segmentation is a mesh, consisting of a set of vertices, edges and cells that form a net of triangles.

The tract of the vessel considered in this study ranges between the distal renal artery and the common iliac arteries, including the area eventually affected by the AAA and iliac aneurysms, using manually selected points - see also Fig. 1b. The pipeline used to perform the segmentation relies mainly on an edge-based technique with the support of thresholding and region-based methods to improve performances and results. Following[23], thresholding makes an initial guess of the points belonging to the lumen. Then region-based and edge-based techniques[24] help with recognition of the rounded shape of the vessel, when projected on the coronal plane. In this way, a label is applied to each of the voxels of the CTA scan creating a volume whose surface is extracted and triangulated.

The resolution of the segmentation (i.e. the average length of the edges) is 1.0 mm, empirically chosen to balance the trade off between accuracy and computational costs.

A by-product of the segmentation is the *centerline*. Roughly speaking, the centerline is a continuous curve in space that represents the center of the vessel lumen and makes for a powerful descriptor of its shape. Among the several methods proposed in the literature for its identification, we use the algorithm *vmtk 1.4.0* (http://www.vmtk.org), a collection of libraries and tools for 3D reconstruction and surface data analysis of blood vessels[25,26]. It generates a mathematical object stable to perturbations of the vessel's lumen surface. Identification of the centerline is of greatest importance for singling out points on the lumen surface considered as peripheral.

Methods: object representation through persistent homology

We now describe the mathematical objects used to represent the atoms of our statistical analysis, that is to represent the meshes obtained from segmentation of the CTAs of the patients. The key element that drives the upcoming construction is a function defined on a mesh, called *filtering* function, with the property that the morphological features we are interested in appear as local minima and local maxima of this function. In particular, the filtering function should capture inward and outward bumps of a blood vessel, so it must be based on some sort of radial distance.

Filtrations and homology

Consider a triangular mesh $K \subset \mathbb{R}^3$ as the union of the set of its triangles, K^2, the set of its edges, K^1, and the set of its vertices K^0. This data amounts to a *simplicial complex*[20] which is the combinatorial starting point of persistent homology. Every element in K is called a simplex: *0-simplex* if it is a vertex, *1-simplex* if it is an edge and *2-simplex*, if it is a triangle.

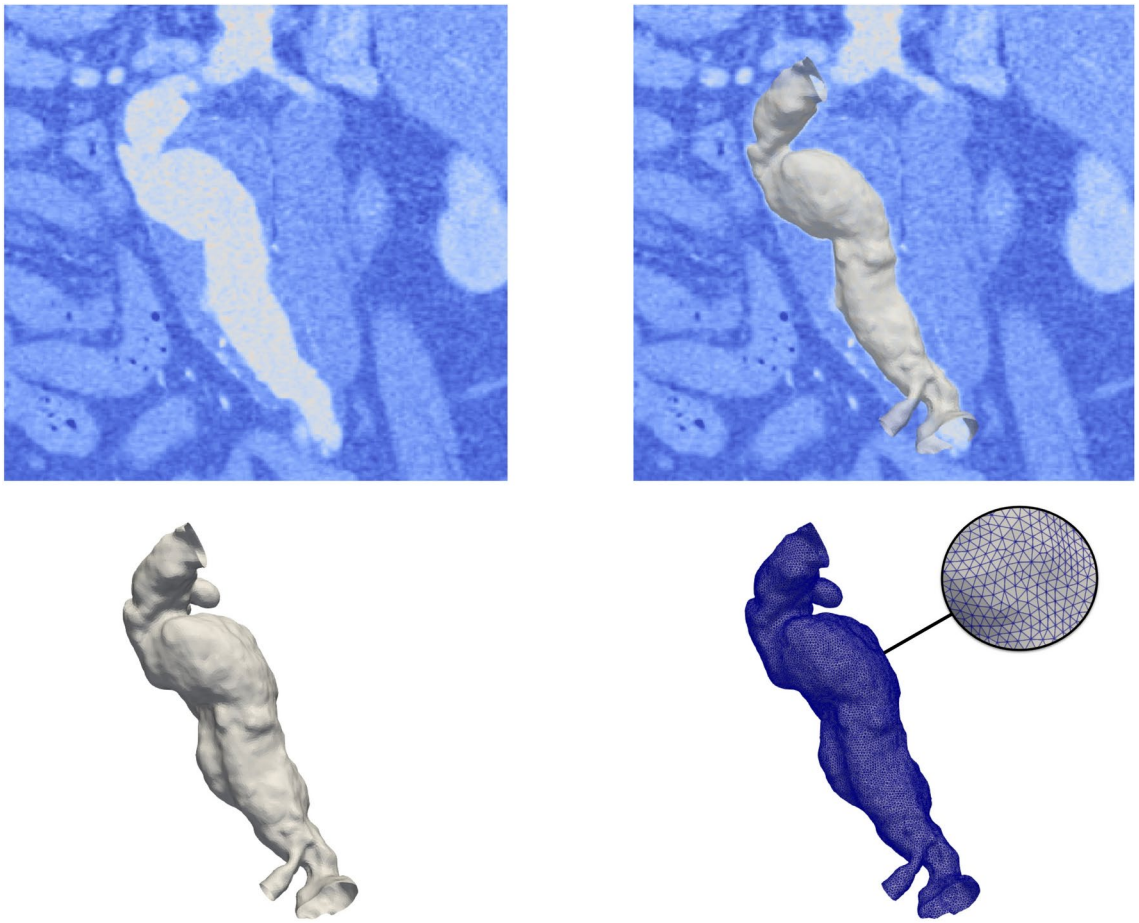

Figure 2. A schematic view of the segmentation process, starting from the CTA scan and obtaining the final mesh representing the portion of interest of the Aorta and the iliacs.

To define the *radial filtering function f*, we start with a function $f : K^0 \to \mathbb{R}$ defined on the vertices of the mesh, and we suitably extend it to the whole *K*. Let $\gamma \subset \mathbb{R}^3$ be the centerline of the blood vessel. For every vertex $v \in K^0$ define:

$$f(v) = \min_{p \in \gamma} \| v - p \|,$$

that is, the distance of the vertex v from the centerline. For every $\sigma \in K^1 \bigcup K^2$, being it an edge or a triangle, we extend *f* as follows:

$$f(\sigma) = \max_{v \in \sigma \cap K^0} f(v).$$

This extension is motivated by the following facts: it is easy to handle computationally and, by standard topology results[21], it is equivalent to the piecewise linear extension on the mesh of the *f* defined on its vertices.

We now use $f : K \to \mathbb{R}$ to order the vertices, edges and triangles of *K* into a filtration of simplicial complexes. For $t \in \mathbb{R}$, consider the simplicial complex K_t defined by the sublevel set

$$K_t = f^{-1}((-\infty, t]).$$

Since f can assume only a finite number of distinct values, let them be $t_0 < t_1 < \ldots < t_m$ to obtain the filtration $\{K_{t_i}\}$ of simplicial complexes, such that

$$K_{t_0} \subset K_{t_1} \subset \ldots \subset K_{t_m} = K.$$

Some pivotal observations are in order; they can be visualized by looking at Figure 3.

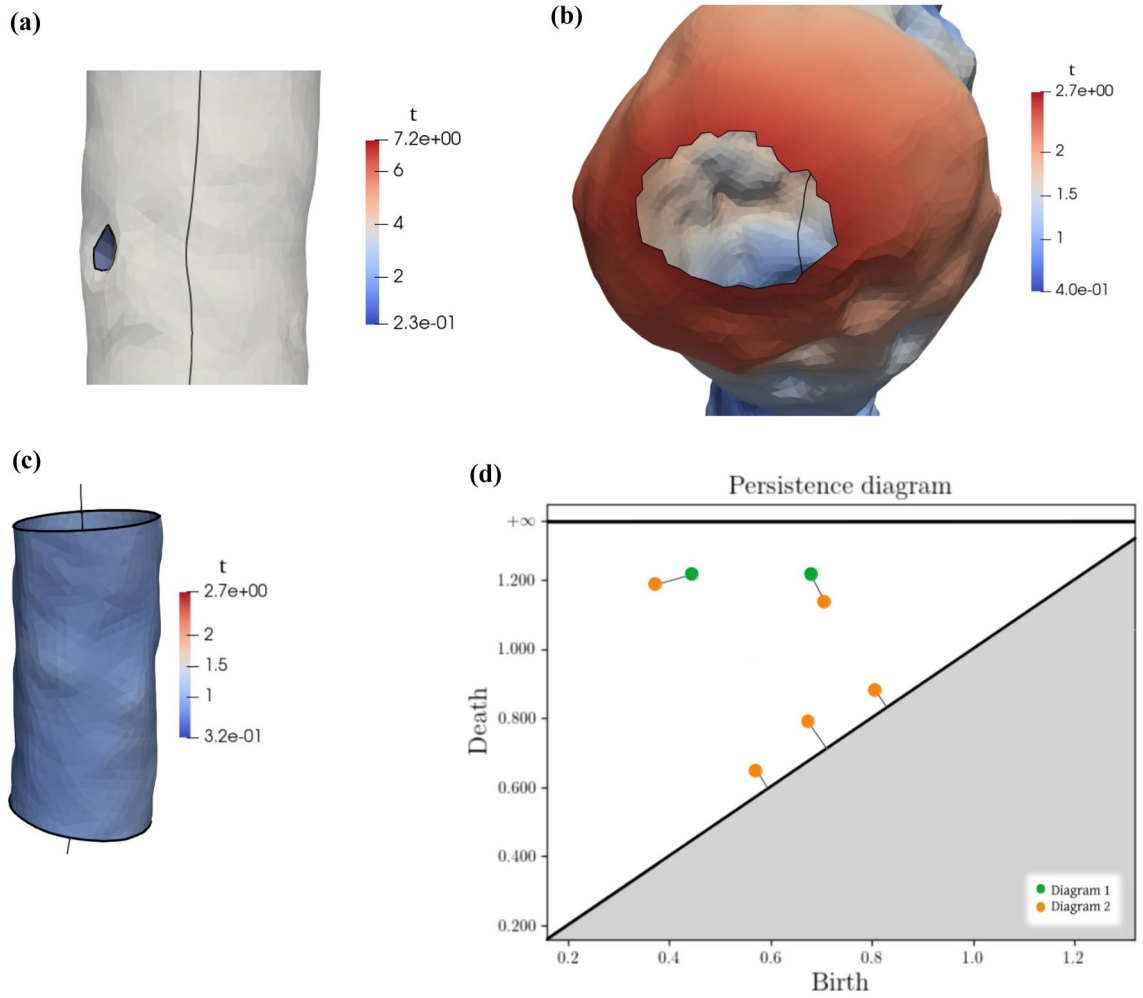

Figure 3. (**a**) A path connected component created by an inward bump, in the sublevel set filtration defined in "Filtrations and homology". Path connected components created by other inward bumps merge with each other at saddle points or local maxima. The color map of the sublevel set refers to the value of *t*: the radial distance from each point in the mesh to the centerline; (**b**) a loop created by an outward bump, in the sublevel set filtration defined in "Filtrations and homology". When the whole bulge is contained in the filtration, the loop disappears. The color map of the sublevel set refers to the value of *t*: the radial distance from each point in the mesh to the centerline; (**c**) loops generated by the tubular structure of the mesh. The color map refers to the value of t: the radial distance from each point in the mesh to the centerline; (**d**) two persistence diagrams and the optimal matching between them giving the bottleneck distance.

First we define the path connected components of f. An inward bump of the vessel corresponds to a local minimum of f; when it appears in the filtration $\{K_{t_i}\}$, a new path connected component is "born" in the sublevel sets of f; see Fig. 3a. That path connected component persists until it is merged with other path connected components, born at lower local minima. Visually, for this to happen, one needs that the whole inward bump generated by the local minimum is added to the filtration, so that different path connected components can meet at local maxima or saddle points. An outwards bump, instead, produces a loop which surrounds its boundary - see Fig. 3b. As the filtration value increases and the simplicial complexes grow, this loop, which encloses the local maximum of the outward bump, becomes smaller and smaller, eventually superimposing to the local maximum itself, and "dying" - see Fig. 3b. There is also another way in which loops are created, as showcased by Fig. 3c: the tubular structure of the blood vessels allows also loops that go around the centerline. Two such loops are equivalent if they can "slide" on the mesh, superimposing one with the other - like the top and the bottom black loops in Fig. 3c.

Thus, the mesh is transformed through f into the filtration of simplicial complexes $\{K_{t_i}\}$; identifying along the filtration "births" and "deaths" of path connected components and loops, characterizes the blood vessel in terms of the irregularities of its lumen.

Topologically speaking path connected components and loops are captured, respectively, by homology in dimension 0 and 1, which we now formally and briefly define. For more details see[21].

For notational coherence with the literature, we refer to (oriented) triangles and edges in a simplicial complex via the set of their vertices enclosed by squared brackets: e.g. $[x_0, x_1, x_2]$ for a triangle, i.e. a 2-simplex, and

$[x_0, x_1]$ for an edge, i.e. 1-simplex. Given a simplicial complex K_{t_i}, for $n = 0, 1, 2$ we generate the vector space $\mathrm{C}_n(K_{t_i})$ over the field $\mathbb{Z}_2 = \{0, 1\}$, by considering the set of all finite formal sums of the n-simplices belonging to K_{t_i}. For $n = 1, 2$, the boundary operators $\partial_n : \mathrm{C}_n(K_{t_i}) \to \mathrm{C}_{n-1}(K_{t_i})$ are then defined by setting

$$\partial_n(\sigma) = \sum_{i=0}^{n} \sigma_{-i},$$

when σ is an n-simplex, and by extending linearly to the whole vector space $\mathrm{C}_n(K_{t_i})$; here σ_{-i} is the (oriented) $(n-1)$-simplex obtained from σ by deleting its i-th vertex – e.g. $[x_0, x_1, x_2]_{-1} = [x_0, x_2]$. The boundary operator ∂_0 maps $\mathrm{C}_1(K_{t_i})$ in the trivial vector space whose only element is 0. We can now introduce the spaces of *n*-boundaries and *n*-cycles of K_{t_i}: $\mathrm{Z}_n(K_{t_i}) = \ker(\partial_n)$ are the *n*-cycles and $\mathrm{B}_n(K_{t_i}) = \mathrm{Im}\,(\partial_{n+1})$ are the *n*-boundaries of K_{t_i}. Finally, we define the *n*-dimensional simplicial homology groups $\mathrm{H}_n(K_{t_i}) = \mathrm{Z}_n(K_{t_i})/\mathrm{B}_n(K_{t_i})$. Note that the quotient $\mathrm{Z}_n(K_{t_i})/\mathrm{B}_n(K_{t_i})$ is well defined since $\partial_n \circ \partial_{n+1} = 0$. In particular, linearly independent vectors in $\mathrm{H}_1(K_{t_i})$ are independent loops which are not filled by triangles (see Fig. 3, right), and linearly independent vectors in $\mathrm{H}_0(K_{t_i})$ are points which are on different path connected components (see Fig. 3, left). Elements in $\mathrm{H}_n(K_{t_i})$ are referred to as homology classes or (equivalence) classes of *n*-cycles. Tracking down the evolution of $\mathrm{H}_1(K_{t_i})$ and $\mathrm{H}_0(K_{t_i})$ along the filtration $\{K_{t_i}\}$ returns the information on the vessel shape we are going to explore in this paper.

Normalization

One further step is needed to ensure a fair comparison between filtrations obtained from vessels of different patients.

As it is often the case with shape analysis, the magnitude or the size difference between patients can act as a confounding factor to the point of overshadowing the shape variability, which is the one we are interested in. For instance, different patients can have different healthy dimensions for the aorta and for the iliacs, making it harder to compare them by contrasting the filtrations and the homology groups generated by the their respective filtering functions. To avoid this potential bias in the analysis, we resort to a normalization process, expressing all distance values relative to the healthy size of the patient's aorta, i.e. the average neck radius. We recall that the aortic neck is in fact the first healthy tract of the aorta, right below the distal renal artery. Its diameter is equal to an healthy aortic diameter, it does not have high variability and it summarizes the size factor of the aorta, in accordance with the scientific literature[27]. So, for each patient, the filtering function values have been divided by its average neck radius, obtained as reported in Appendix A.

On top of that, since different portions of the blood vessel are characterized by a different relative radius, with respect to the neck's radius, we can in fact relate the normalized filtering function values to morphological features of the aorta, see "Results: reading a persistence diagram".

Persistence diagrams

Looking at the homology groups obtained from a filtration is not practical for applications, but there is a very interpretable and concise summary of the information contained in the sequence of $\mathrm{H}_n(K_{t_i})$, called the *persistence diagram* in dimension *n*. These are the objects that we are going to use to represent data.

A persistence diagram is a finite collection of points in $\mathbb{R}^2$ - also called *persistence pairs* -, with every point $p = (b, d)$ representing the birth and the death of a persistent homology class. In particular, the birth of an homology class is the first "time t" it appears along the filtration, while the death time happens when the class merges with another class born earlier. For example: a local minimum of the filtering function, indicative of a bump in the vessel, induces a path connected component which is born at the value of the minimum, see Fig. 3a; such path connected component carries on, or *persists*, along the filtration, and dies when it merges with a path connected component born earlier, therefore associated to a lower local minimum. The absolute difference between the times of birth and death is clearly related to the prominence of the bump, and is called *persistence* of the homological feature. Note that points can appear multiple times in the same persistence diagram.

Populations of persistence diagrams

Lastly, it is possible to compare and analyze the topological information carried by different persistence diagrams by defining suitable metrics, satisfying formal stability results; for an extensive review see, for instance,[21]. It is thus possible to quantify the dissimilarity of two different diagrams and perform classification, clustering and dimensionality reduction on populations of diagrams.

To simplify the upcoming formulas, let

$$D = \{(b_1, d_1), \ldots, (b_n, d_n) \mid n \in \mathbb{N}, b_i, d_i \in \mathbb{R} \cup \{\infty\}, b_i < d_i\} \cup \{(b, d) \mid b, d \in \mathbb{R}, b = d\}.$$

represent a persistence diagram. Given two diagrams D_1 and D_2 and $p \geq 1$, the *p*-Wasserstein distance between them is:

$$W_p(D_1, D_2) = \left(\inf_{\gamma} \sum_{x \in D_1} \| x - \gamma(x) \|_\infty^p \right)^{1/p} \tag{1}$$

where the $\inf$ is taken over all bijections γ between diagrams D_1 and D_2 and $\|\cdot\|_\infty$ is the sup norm in $\mathbb{R}^2$. In other words we measure the distances between the points of the two diagrams, pairing each point of a diagram either with a point on the other diagram, or with a point on $y = x$ (see Fig. 3d). Each point can be matched once and only once. The minimal cost of such matching provides the distance. The case $p = \infty$ is usually referred to as the bottleneck distance and has the following form:

$$W_\infty(D_1, D_2) = \inf_\gamma \sup_{x \in D_1} \| x - \gamma(x) \|_\infty .$$

Note that, to compare two persistence diagrams, we do not need to establish a relationship between the generating meshes. In particular this means that no alignment or registration is needed to proceed with the analysis; this is the ancillary source of variability between patients we are not interested in.

Finally, let us mention an important stability result[28]. Given two filtering functions $f, g : K \to \mathbb{R}$ and their respective persistence diagrams $D(f)$, $D(g)$:

$$W_\infty(D(f), D(g)) \leq \| f - g \|_\infty,$$

meaning that persistence diagrams are a faithful representation of the functions, in terms of the sup-norm.

Closing this section on persistent homology, we point out the existence of representations alternative to persistence diagrams, for instance persistence landscapes[29], persistence images[30], persistence silhouettes[31], accumulated persistence functions[32]. All of these come with their own properties and stability results and can be used for analyses which are possibly more refined or more tailored to the application of interest.

Results: reading a persistence diagram

To understand the descriptive power of the persistence diagrams obtained with the pipeline illustated in the previous section, a parallel reading of the original mesh and the associated diagrams has been made for every patient in the study, connecting the most important aortic wall features - such as AAA, calcifications and thrombus - to the diagram points.

All the results contained in the upcoming subsections have been manually verified by simultaneously looking at the persistence diagram, the simplicial complexes and the CTA scans, with the support of the collaborating clinicians.

Persistence diagrams and related distances are computed using the python library *gudhi*[33] version 3.8.0, while for clusterization and visualization purposes the libraries *scikit-learn*[34], *scipy*[35], *seaborne*[36] and *matplotlib*[37] have been adopted.

The effect of normalization

We have anticipated in "Normalization" that normalization is instrumental for comparing different patients since it allows for the appraisal of the distance from the centerline when a persistence pair is created or ceases to exist, relatively to the size of the mean aortic neck radius. Normalization is, in fact, also significant for the identification of the portion of the aorta where the change in homology occurs, locating it on a specific section of the vessel. Note that the diameter of the vessel differs along the aorta, especially after the aortic bifurcation. In fact, the diameter of a healthy abdominal aorta is tipically $17.5 \pm 2.1\,\mathrm{mm}$, larger than that of an healthy iliac artery, $10.85 \pm 1.69\,\mathrm{mm}$; see[38,39]. Although significant differences in the dimension of the aorta for males and females occur, the proportion between the diameter of the main aorta and that of the iliac branches are similar. Thus:

- when an homology class is born (or dies) at a value close to the mean neck's radius, i.e. a value close to 1, the change must be located at the same distance from the centerline as the aortic mean neck's radius, hence on the healthy portion of the aorta. The only exception being an iliac aneurysm, whose presence will be discussed later in "Notable pairs";
- in the cohort of the patients considered in this study, the diameter of healthy iliacs is around 40-70 % of the respective aortic mean neck's radius. Thus, changes in homology located on the iliac arteries occur when the normalized distance assumes values in the range of $[0.4 - 0.7]$;
- lastly, patients affected by AAA can instead have a more variable range, up to 7 and more in the most concerning cases.Figure 4 depicts a schematic visualization of these facts.

Notable pairs

By combining the information given by birth, death, persistence and type of pairs of points in the persistence diagram, it is possible to identify *notable pairs*, that is persistence pairs shared among patients with similar characteristics associated with a specific aortic region.

Structural pairs

The first class of points we consider are those associated to an infinite persistence, i.e. such that $d = \infty$. These points correspond to features which are born at a certain filtration value, but which never die, and so persist up to ∞. They indicate structural features shared by all aortas. In particular, each filtration coming from a correctly segmented aorta must have exactly three points with infinite persistence: one associated to 0-cycles and two

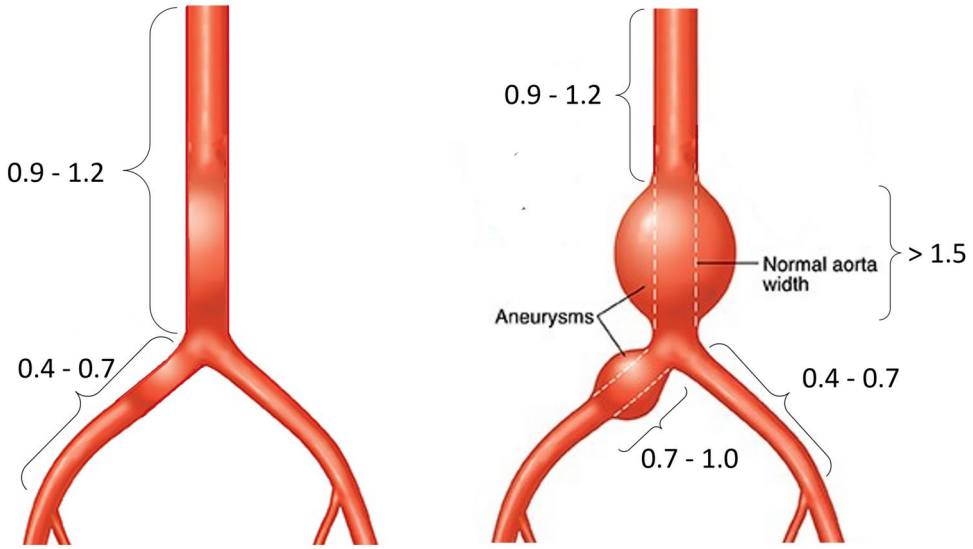

Figure 4. A schematic view of the aorta, with the normalized distances from the centerline made explicit. Image modified from https://prescrivere.blogspot.com/

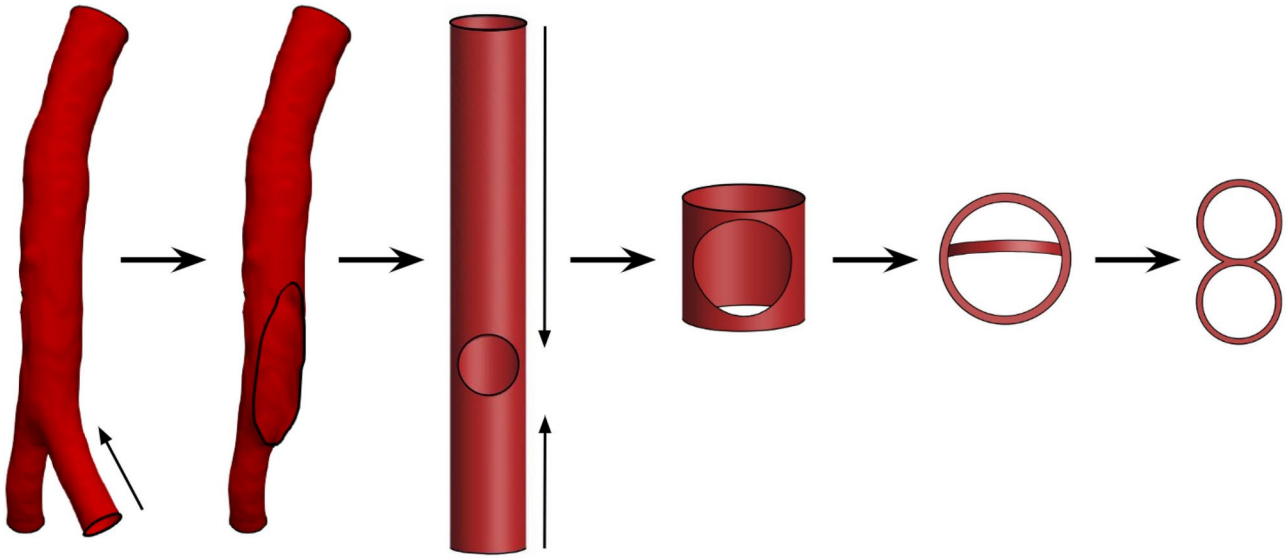

Figure 5. All the objects in this figure are homotopy equivalent.

associated two 1-cycles. Note that this ensures that the Wasserstein distances between the associated diagrams are finite.

Indeed, the mesh representing the aorta should be *homotopy equivalent*[21] to a finite cylinder with a hole, which, in turn, is equivalent to an eight figure: two circles glued in one point, see Fig. 5. Roughly speaking, two objects are homotopy equivalent if, starting from one, it is possible to obtain the other using stretching and bending but not tearing. Homotopy equivalence preserves homology and so each mesh should feature a connected component with two independent 1 dimensional loops as can be seen in Fig. 5.

These three points, representing the most persistent homological features, can also be used as indication of the correct representation provided by the mesh itself, proving the deep connection between the persistence diagram and the aorta. In fact, if a persistence diagram shows additional points with infinite persistence, this signals an error during the segmentation phase, such as the creation of non-connected points or the presence of artificial holes in the mesh.

Iliac aneurysm

Now we tackle the problem of identifying iliac aneurysms; the reader should refer to Figs. 6 and 7 to get a visual description of the content of this section.

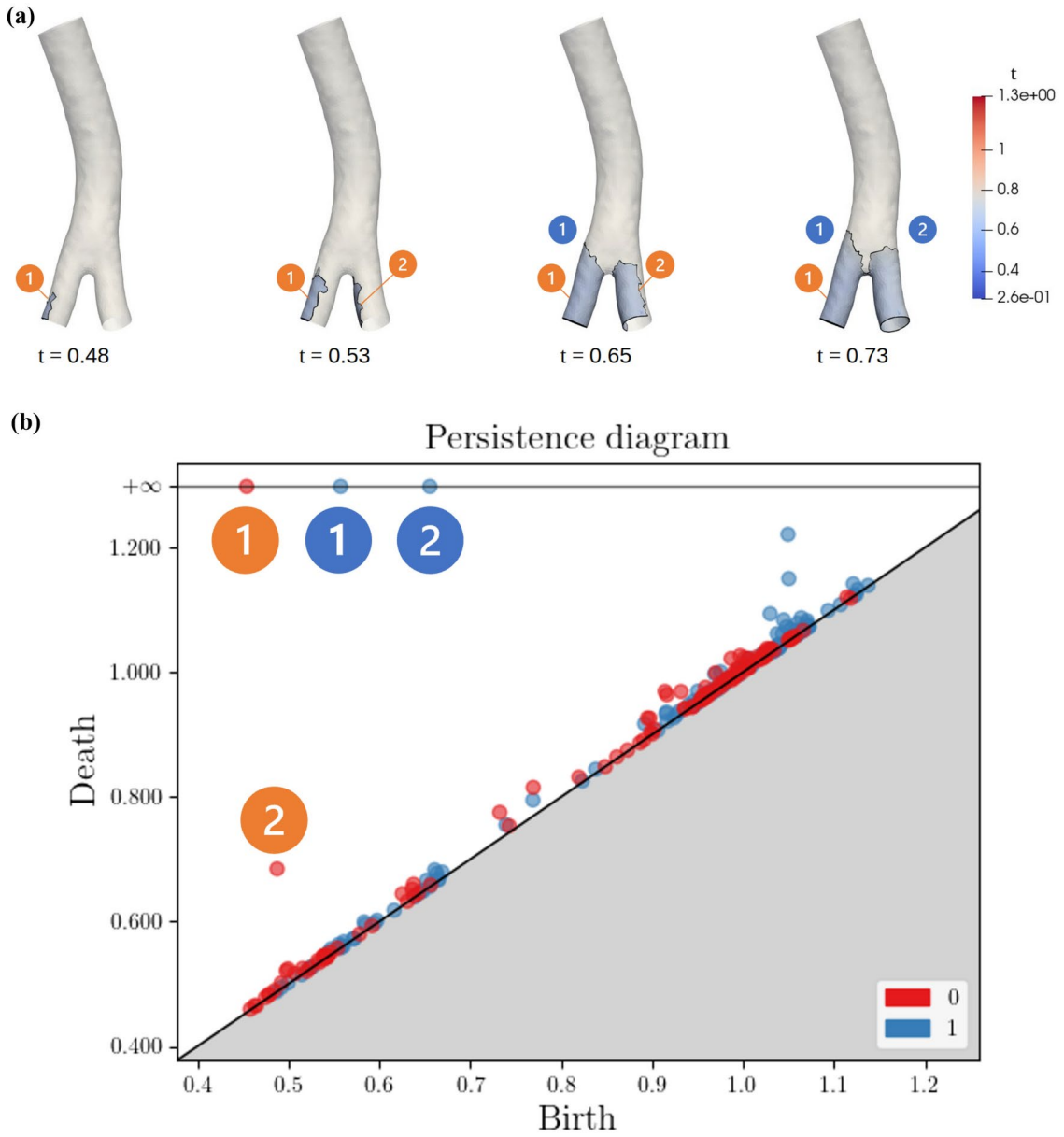

Figure 6. (**a**) Initial steps of the sublevel set filtration of a patient with healthy iliacs, showing the births of the iliacs-related persistence pairs: first appear the two local minima of the radial distance located on the two iliacs - orange (1) and (2) - and then the two loops going around the tubular structure of the iliacs - blue (1) and (2). The color map of the sublevel sets refers to the value of *t*: the radial distance from each point in the mesh to the centerline; (**b**) the persistence diagram, in dimension 0 and 1, of an healthy patient, with highlighted the persistence pairs related to the iliacs, whose labels are coherent with the ones in (**a**).

Aneurysms can affect one or both the iliac arteries, in the same way as they affect the abdominal aorta, resulting in a stretch of the wall - see Fig. 7c. An iliac aneurysm can also be associated with calcification (as in Fig. 7) and thrombus, just as the AAA.

When the iliacs have comparable diameters there are two connected components born at similar values, resulting in two classes of 0-cycles that later merge together - as in Fig. 6. The single class of 0-cycles with infinite persistence represents the closest point of the mesh to the centerline, the very first point appearing in the filtration. This class is always born at the iliacs, given their smaller radiuses. Moreover, the other iliac's first point induces the birth of a secondary early class of 0-cycles with high persistence. These components will be merged together as the radius value increases (see point "(1)" and "(2)" in Fig. 6). For these reasons, each persistence diagram, except the ones with iliac aneurysm, have a high persistence point with birth coordinate given by the radius at which the second component is born and death coordinate given by the merging radius.

When an iliac aneurysm is present, it significantly deforms the lumen and so the minimal radius of the aneurysmatic iliac is greatly increased - see Fig. 7. As a consequence the first point of the second iliac appears

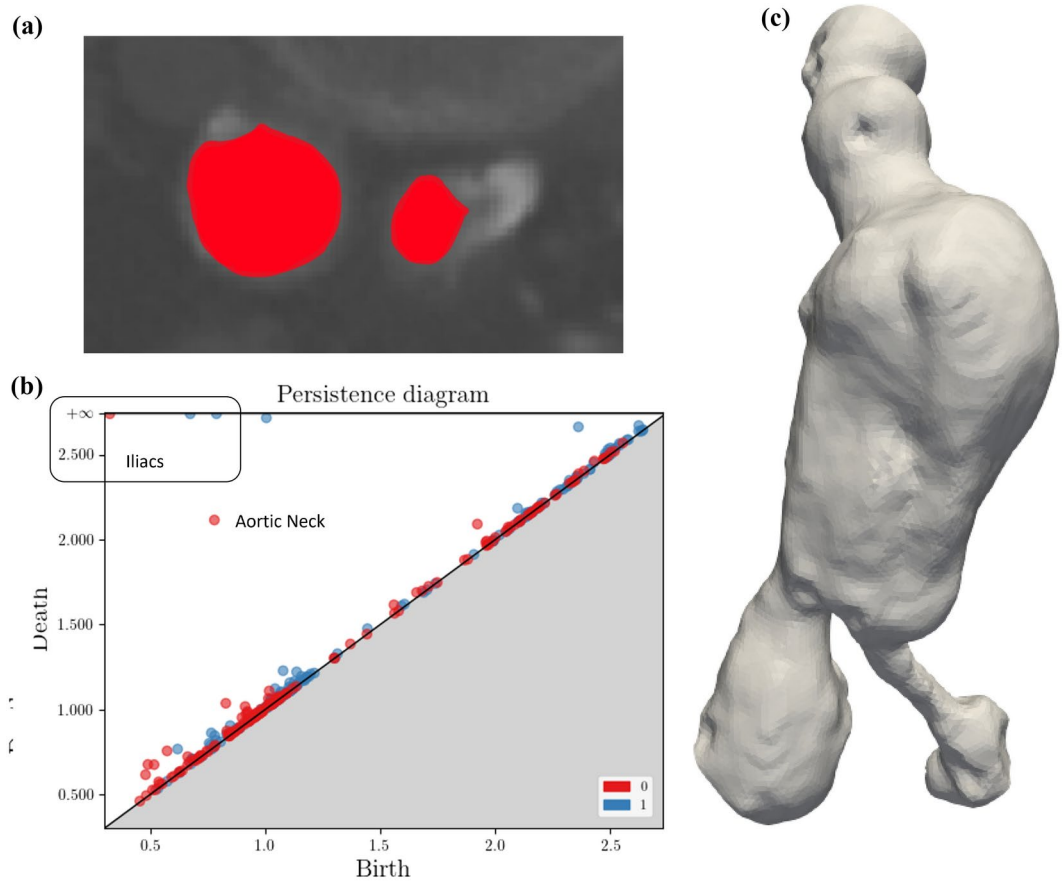

Figure 7. (**a**) CTA scan slice of iliacs presenting calcification in the right iliac and aneurysm in the left. The red portion shows the lumen restriction created by calcifications and the bulge created by the aneurysm; (**b**) the persistence diagram of a patient with an aortic aneurysm and an iliac aneurysm. The lack of an early *0*-cycle with high persistence is visible. As is the presence of one class of high persistence 1-cycles created by the AAA; (**c**) mesh of the lumen of a patient being affected by calcifications and aneurysm in the iliacs: the left one presents a bulge created by the aneurysm, while the right one is severely occluded by calcifications.

much later in the filtration, often after other points at the bifurcation region. Thus, instead of having two connected components born at similar times, we have a component that is expanding until it covers both iliacs. As a consequence, the absence of this second high persistence point in dimension 0 is a primary indicator for an iliac aneurysm.

Similarly, any irregularity in the iliacs, inflating their lumen, may have the effect of delaying the birth of 1-cycles with infinite persistence. For instance, not having two classes of 1-cycles with infinite persistence born at roughly the same times (i.e. having roughly the same diameters) implies that one iliac is inflated. The only reasonable cause for that inflation is the presence of an aneurysm. Thus, if the two classes of 1-cycles with infinite persistence have a pronounced difference between their birth coordinates or have birth coordinates which are high (note that a 1-cycle is always going to appear at radius 1, around the neck), this indicates the presence of one or more aneurysms in that region of the blood vessels.

Abdominal aortic aneurysm

We now focus on AAAs, which are obviously distinguished by large radial distances from the centerline and therefore by high values of birth and death coordinates of their associated points on the persistence diagram - see also Fig. 8. But more can be said; in fact, persistence diagrams of patients affected by AAA are characterized by the following persistence pairs:

1. A notable and highly persistent class of 0-cycles is usually present when patients are affected by an AAA. These cycles are associated to the sudden increase of the distance from the centerline when moving form the aortic neck region to that part of the vessel where the AAA occurs. In fact, the presence of the AAA "splits" the aorta in two parts, separated by the AAA - see Fig. 8. Along the filtration, each part generates a connected

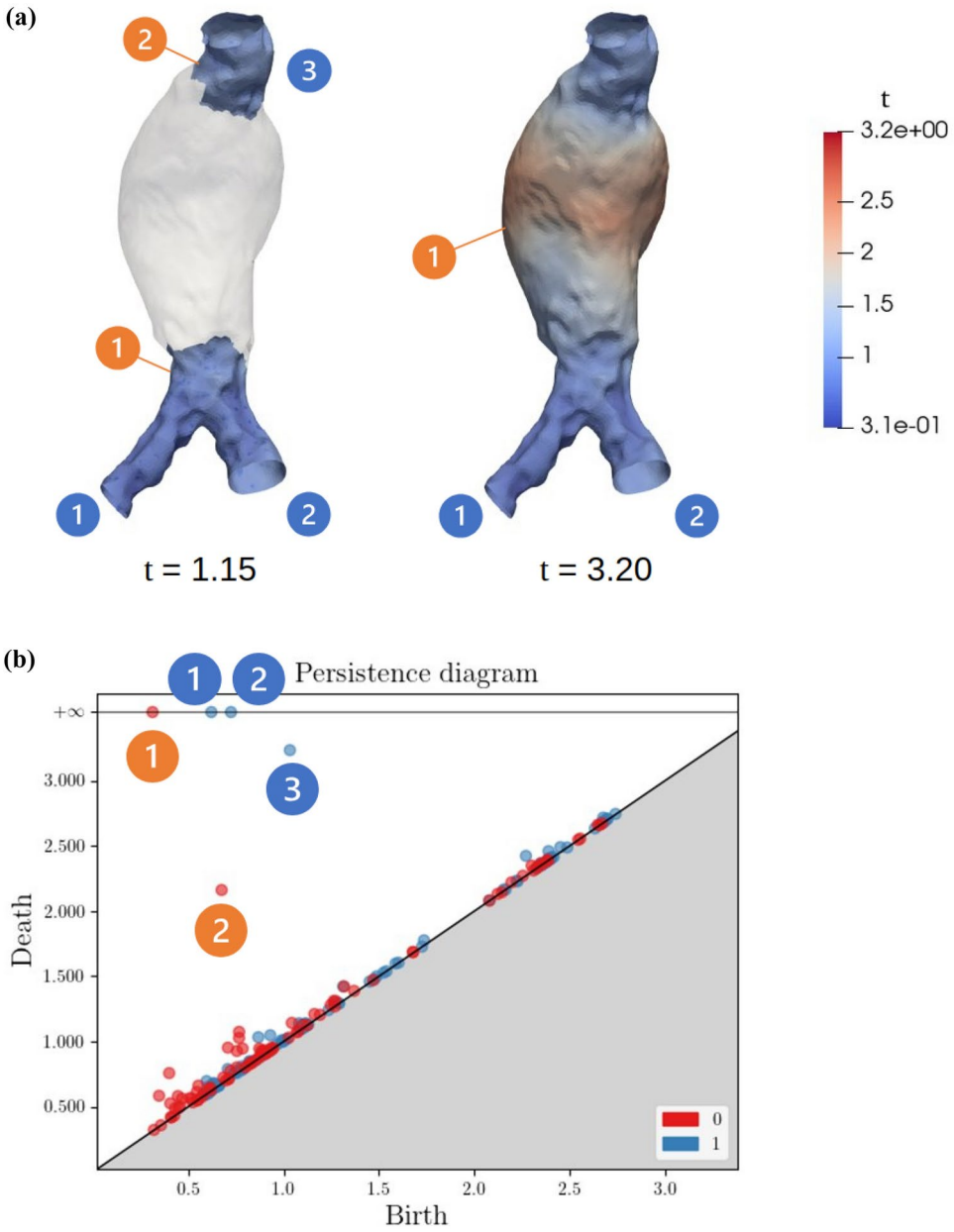

Figure 8. (**a**) Two different steps of the sublevel set filtration of a patient affected by AAA: in the first step (t=1.15, with t being the radial distance from each point in the mesh to the centerline) we clearly see the two path connected components - orange (1) and orange (2) - which are separated by the AAA. At very high filtration values the AAA is added to the filtration and the path connected components merge. A similar phenomenon involves also 1-cycles: on the left, the loop going around the upper portion of the aorta cannot "slide" down on the mesh and be equivalent to the loop going around the lower portion of the aorta - made by (1)+(2). This is instead possible on the right, causing the eventual death of loop (3). The colormap illustrates the distance of each mesh point from the centerline; (**b**) the persistence diagram of the filtration in (**a**). We have highlighted the high persistence path connected component and loop caused by the AAA splitting in two parts the aorta.

component; those two components merge only when it is possible to have a path across the aneurysm; see Fig. 8.

2. Similarly, the AAA generates also a notable highly persistent class of 1-cycles, associated with the tubular structure of the neck, which does not appear in healthy patients. As already mentioned AAA splits the blood vessel in two parts - see Fig. 8. Each part has a loop going around the aorta. Only when all the aneurysm appears in the filtration the loop around the blood vessel on one side of the aneurysm and the one on the other, can be merged with each other and become equivalent. One of them therefore dies and the other one persists as structural persistence pair.As a consequence, the corresponding point on the persistence diagram

is easily recognizable since its birth is around the value 1.0 and its death has the highest value among all the points with finite persistence. Again, all of this is clearly visible in Fig. 8.

To summarize, two main features make the persistence diagram of a patient with AAA easily distinguishable - see also Fig. 8:

1. The presence of persistence pairs borning and dying at high distance values. In fact, no healthy aorta has a region with a distance from the centerline higher than 1.4, while the regions of an aneurysmal aorta can exceed the value of 7.0.
2. The presence of at least one 0-cycle and one 1-cycle classes with high persistence in the neck area (birth around 1).

Calcifications

Calcifications are calcium deposits creating a local thickening of the aortic wall, on the internal side, resulting in indentations on the lumen's surface, when in contact with the calcified wall. Thus, they produce local minima in the filtering function, as showcased in Fig. 9. These inward bumps are picked up as 0-cycles with medium-persistence in the persistence diagram; see Fig. 9b.

Thrombus

Lastly we consider blood vessels affected by a thrombus. A thrombus is often made by a non-homogeneous permeable substance[40,41], which can flake off with the flow of blood creating wide ledges in the lumen (see

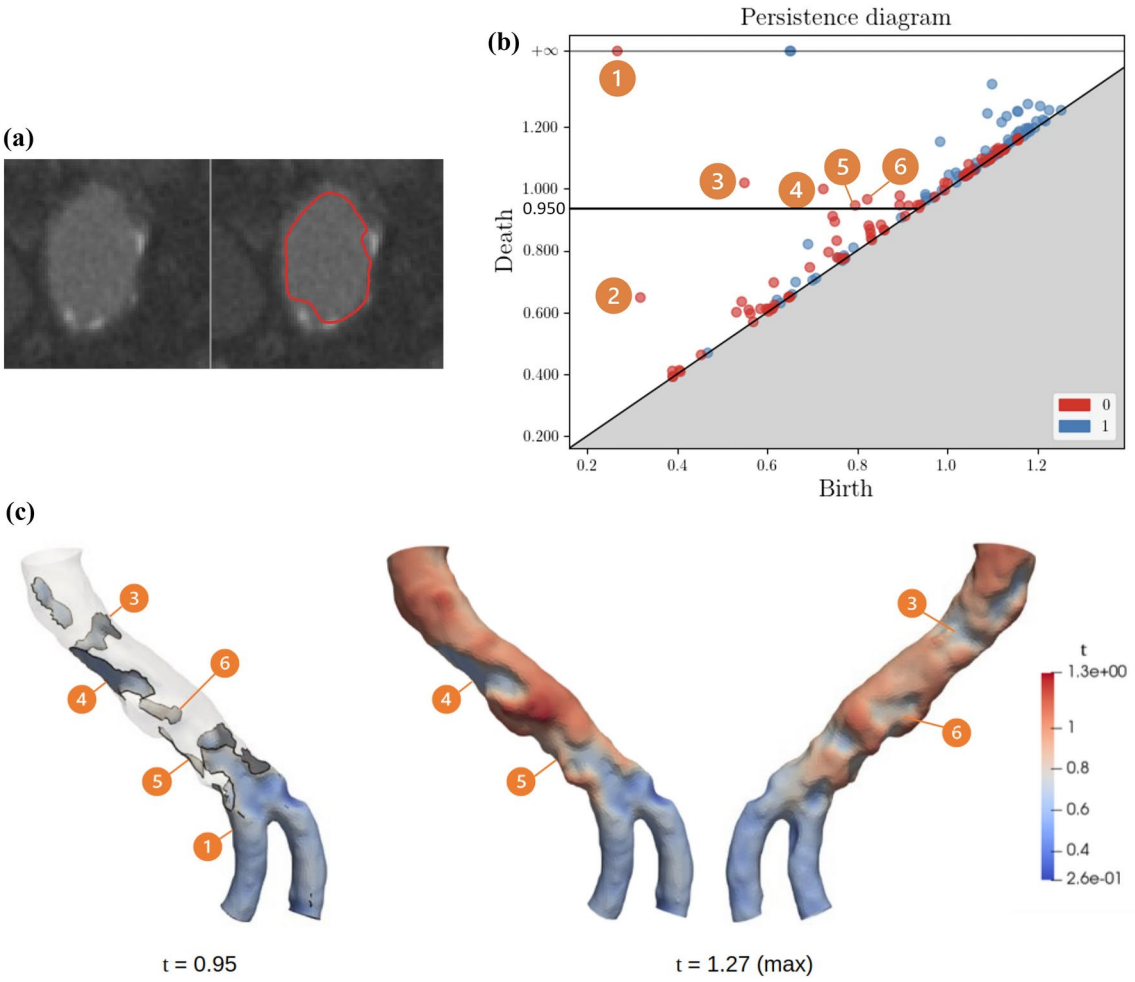

Figure 9. (**a**) A CTA slice of a non-aneurysmal aorta with presence of calcifications which creates bump in the lumen of the blood vessel; (**b**) the persistence diagram of the filtration in (**c**): we see a number of persistence pairs with medium persistence, being born in the main body of the aorta - i.e. with birth value higher than the aortic neck, as a consequence of the irregularities due to calcifications; (**c**) different time steps of the sublevel set filtration of an aorta without AAA but with calcifications: calcifications create many medium-sized irregularities and bumps in the lumen which are picked up by the filtration, creating a number of path connected components arising and persisting for some time. The color map of the sublevel set refers to the value of *t*: the radial distance from each point in the mesh to the centerline.

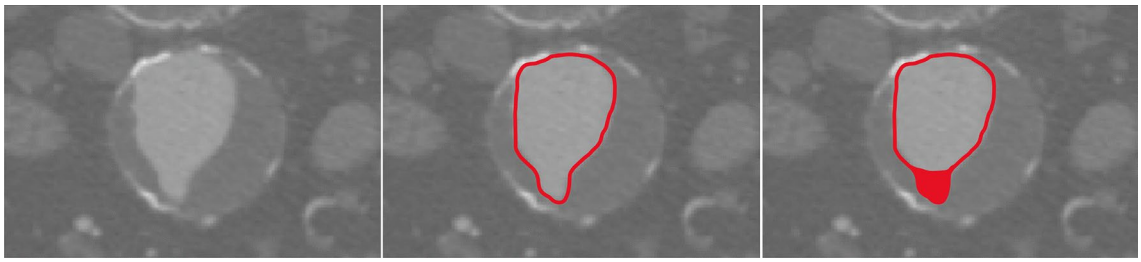

Figure 10. CTA slices of an AAA with thrombus occluding the lumen.

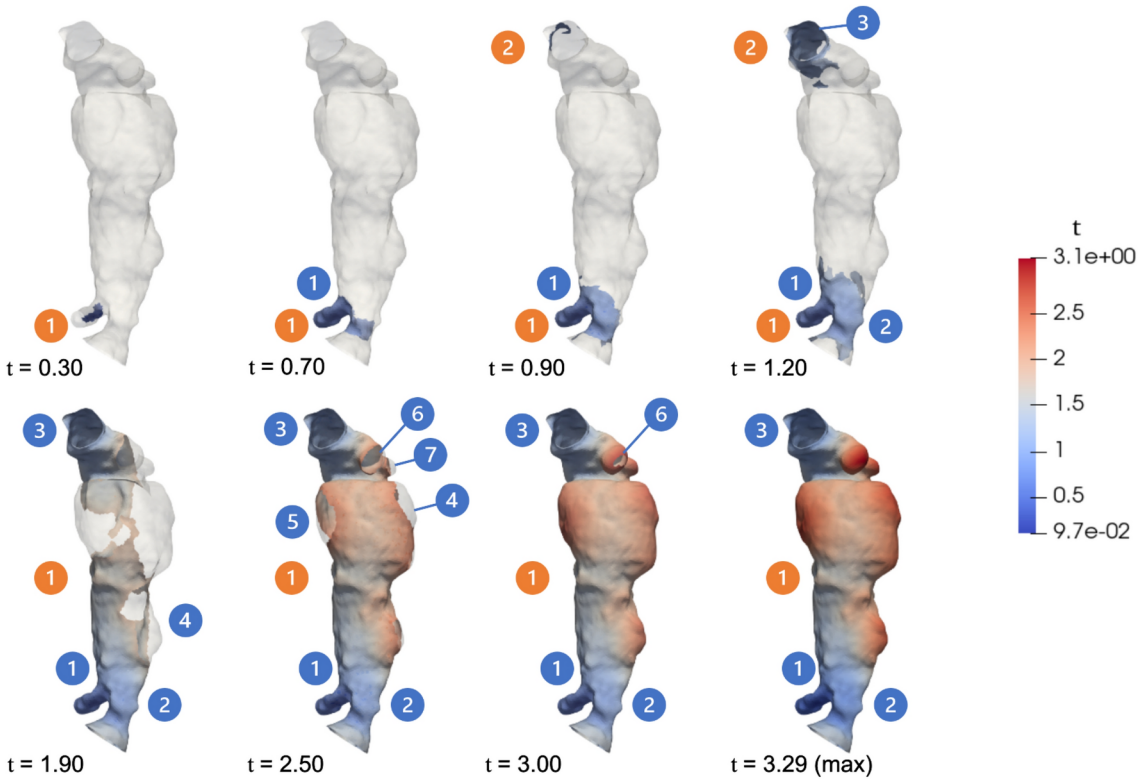

Figure 11. A patient affected by AAA, iliac aneurysm and thrombus. The filtration function values are reported in black on the bottom left corner of the aorta. The sublevel sets are represented by the higlighted portions of the mesh. The numbered features identify 0-dimensional (red) and 1-dimensional cycles associated to points in the diagram appearing in Fig. 12. The color map of the sublevel set refers to the value of *t*: the radial distance from each point in the mesh to the centerline.

Fig. 10). It is most likely to be located in the area affected by AAA, due to the concentration of metabolites associated with inflammation. Thus, the presence of a pronounced thrombus produces a series of local maxima and/or minima, which induce 1 and 0 cycles in the persistence diagrams with relatively high birth coordinates and moderate persistence, see also Fig. 12.

A detailed example

We now show an example of a concerning patient affected by AAA, iliac aneurysm and thrombus.

In Fig. 11 we report the main steps of the sublevel set filtration of the selected patient. While in Fig. 12 the associated persistence diagram is shown.

- At value 0.3 a 0-cycle of the first iliac appears.
- At value 0.7 a 1-cycle, around that iliac, appears. Note that the 0-cycles expands onto the other iliac without creating a new path connected component. There is in fact an aneurysm on the second iliac preventing the tapering of its extremity.
- At value 0.9 a novel 0-cycle appears on the aortic neck. The AAA separates this cycle from the one appeared in the beginning.

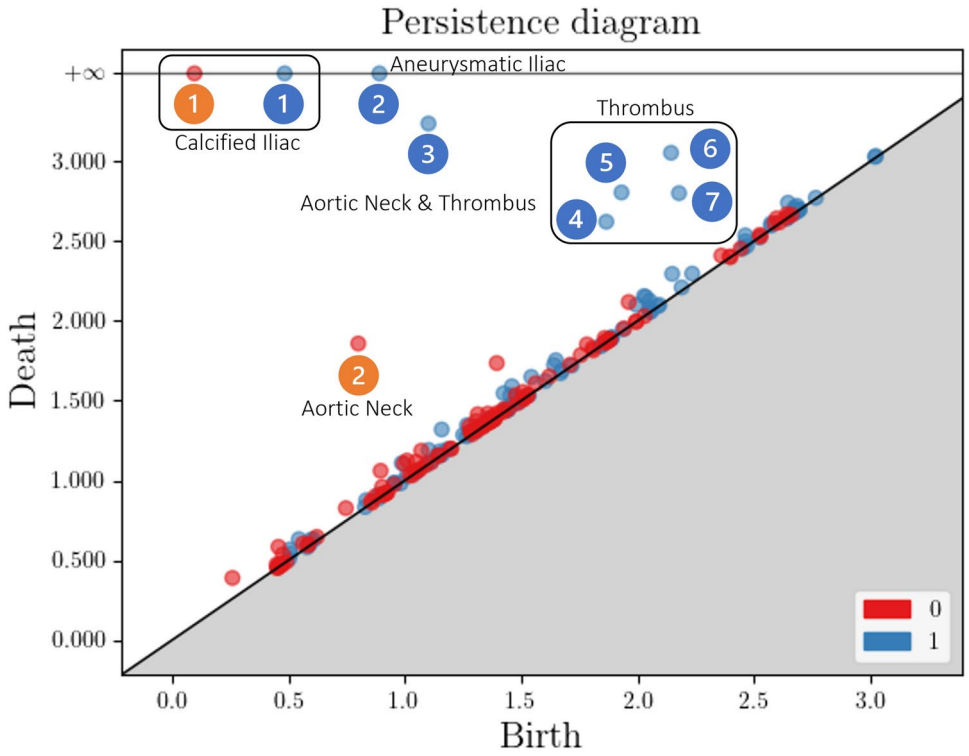

Figure 12. The persistence Diagram associated with the mesh shown in Fig. 11. Features labelled as (1) both in dimension 0 and 1 are the structural pairs associated to the iliac which does not present an aneurysm. In particular, path connected component (1) is born very early along the filtration and that is because of some inwards bumps due to some calcifications. Persistence pairs (2) - both in dimension 0 and 1 -, instead, are born very late, almost at the filtration value of the neck. This is a clear consequence of the iliac aneurysm. The 1-cycles' class (3) is generated because of the AAA splitting the blood vessel in two parts. An in fact it dies at the latest value of the filtration. Then we have a group of 1-cycles - from (4) to (7) - with medium persistence, which reflect the irregularities in the lumen caused by some heterogeneous thrombus.

- At value 1.20 a 1-cycle going around the blood vessel appears.
- As the distance form the centerline increases - that is, the value of the radial filtration function increases -, the initial path connected component expands enclosing gradually also the AAA, connecting to and killing the path connected component and the loop associated to the neck. On the way, several classes of 1-cycles are created with high birth times; they are associated to the local maxima of the filtration function created in the aneurysmatic portion of the aorta by the thrombus.

Results: population analyses

The previous section discussed how persistent homology translates some relevant aortic features into persistence pairs shown as points of a persistence diagram. We now explore the behaviour of these topological summaries at the population level, to understand which kind of between patients variability can be captured with persistence diagrams.

We present two paradigmatic problems, both characterized by their simplicity: clustering and discrimination or, with different words, unsupervised and supervised classification of persistence diagrams. In the initial clustering exercise, we look for a natural stratification of persistence diagrams embedded in a metric space endowed with a Wasserstein metric. Secondly, we exploit the analysis of "Results: reading a persistence diagram" to introduce a supervised classification pipeline aimed at the construction of classifiers identifying patients with: (1) AAA, (2) calcifications, (3) thrombus, (4) iliac aneurysm.

Clustering

Throughout the manuscript we argued that irregularities in the aortas are represented by inward and outward bumps, representing pathologies of different nature, which are captured separately by 0-cycles or 1-cycles represented as persistence pairs in a persistence diagram.

Since both the persistence diagrams for classes of 0-cycles and 1-cycles are relevant to capture shape differences between aortas, we introduce the following family of metrics between patients P_i and P_j:

$$d_p^p(P_i, P_j) = \lambda \cdot W_p^p(D_i^0, D_j^0) + (1-\lambda) \cdot W_p^p(D_i^1, D_j^1),$$

where $p \geq 1$, $\lambda \in [0,1]$ and, for $k = 0, 1$, the persistence diagrams D_i^k, D_j^k are those for the k-cycles of patient i and j respectively; W_p is the p-Wasserstein distance introduced in Eq. (1). The case $p = \infty$ amounts to:

$$d_\infty(P_i, P_j) = \lambda \cdot W_\infty(D_i^0, D_j^0) + (1-\lambda) \cdot W_\infty(D_i^1, D_j^1),$$

i.e. the weighted average of the bottleneck distances between the diagrams. Although different mixing weights are allowed, in the following we set $\lambda = 0.5$ for simplicity. Tailoring the choice of λ could benefit the analysis of specific classification problems.

Figure 13 depicts the matrices of pairwise distances between patients in the cohort of our study, when $p = \infty$ and $p = 2$ respectively. Statistical units have been ordered so that healthy patients come before those with an AAA: indeed, $H01, \ldots, H24$ index the healthy patients without AAA while $A01, \ldots, A24$ indicate the patients with AAA. By visual inspection, it is clear, both in Fig. 13b and a, that this grouping is very well captured by both distance matrices. In fact, visualizing a low-dimensional representations, obtained by MultiDimensional Scaling (MDS), of the metric spaces embedding the persistence diagrams – see Fig. 13d and c – we immediately notice how patients without AAA are clustered together, while the presence of an AAA, by introducing high persistence points in the persistence diagram, generates a larger variability in the MDS representation of the persistence diagrams of diseased patients - see "Notable pairs".

This grouping can indeed be captured by a clustering algorithm. For instance we report the dendrograms relative to agglomerative hierarchical clustering run with Ward linkage, in Fig. 14b and a. Leaves are coloured as in Fig. 13d and c.

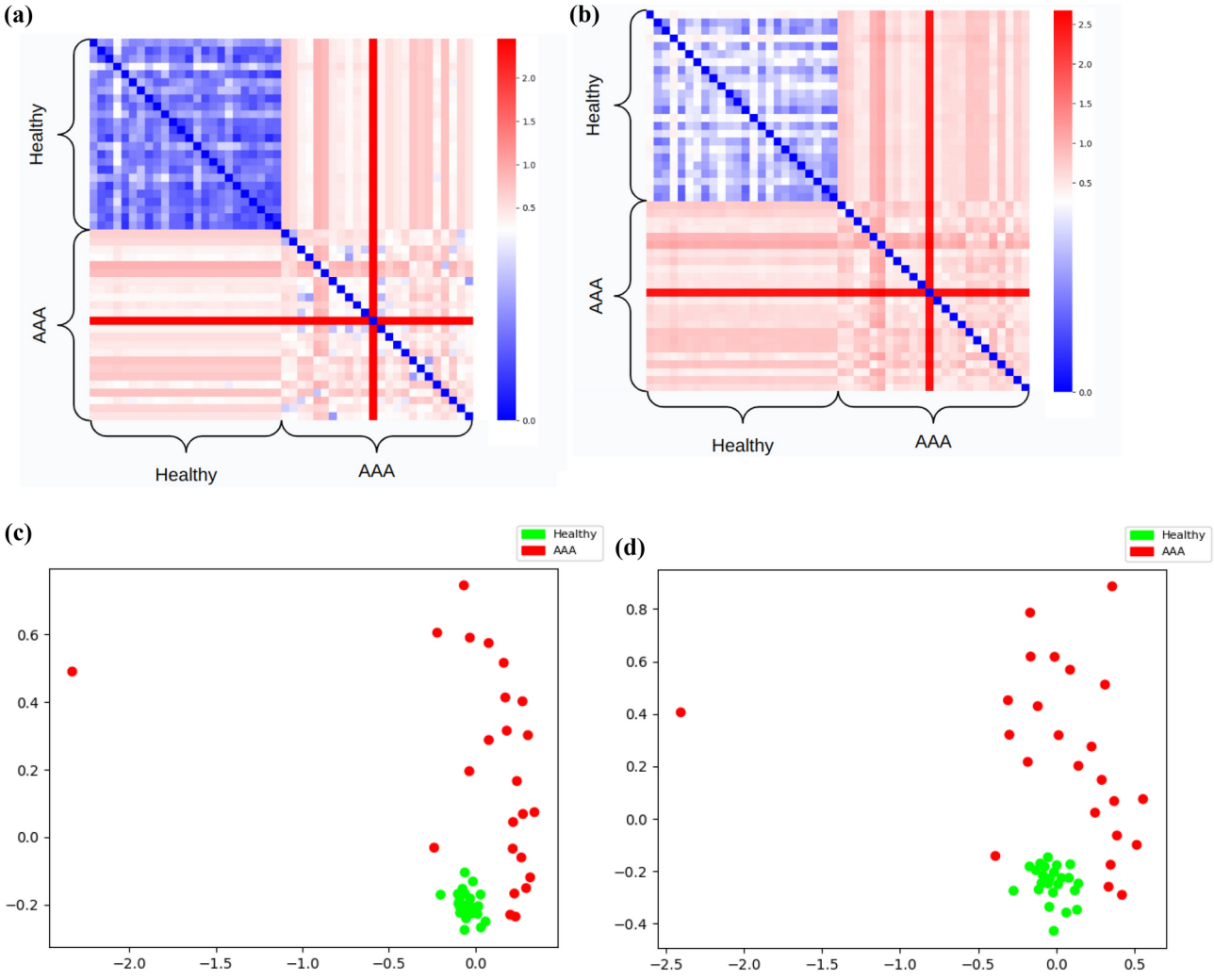

Figure 13. (**a**) The matrix of pairwise ∞-Wasserstein distances between the persistence diagrams computed in "Clustering". Patients are ordered so that healthy patients come before non-healthy ones. (**b**) The matrix of pairwise 2-Wasserstein distances between the persistence diagrams computed in "Clustering". Patients are ordered so that healthy patients come before the ones affected by AAA; (**c**) two dimensional MDS representation of the matrix in (**a**); (**d**) two dimensional MDS representation of the matrix in (**b**).

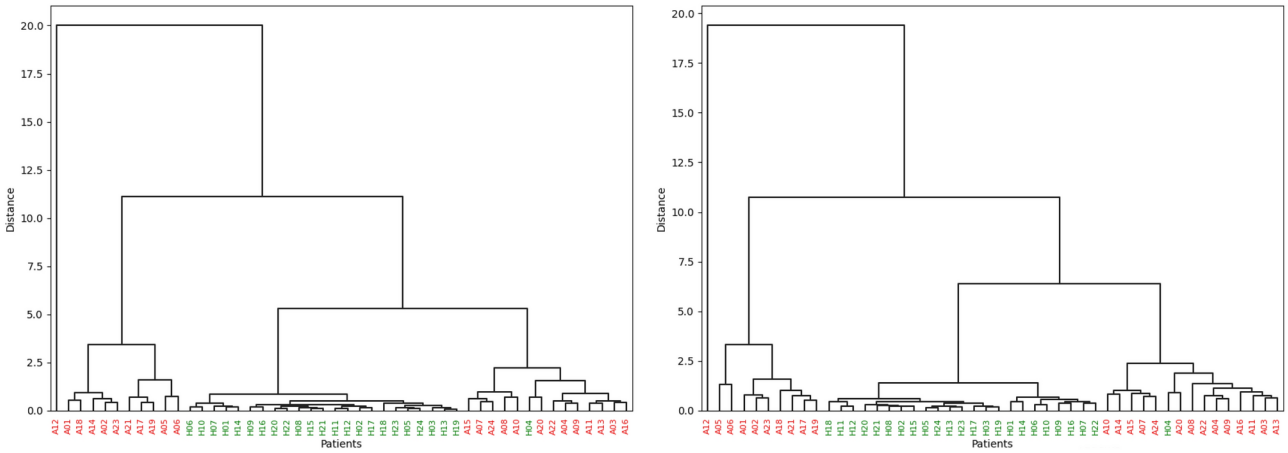

Figure 14. Hierarchical clustering dendrograms related to the clustering problem presented in "Clustering": for the ∞-Wasserstein distance (left) and for the 2-Wasserstein distance (right). Both with Ward linkage. Patients in red are affected by AAA, while healthy ones are drawn in green.

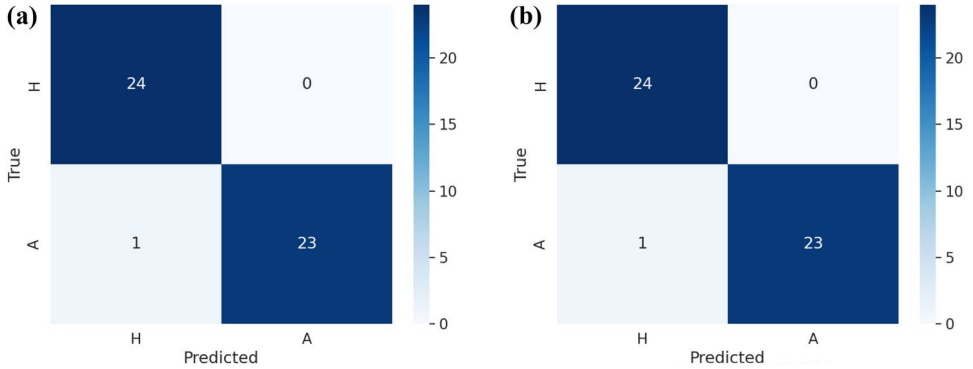

Figure 15. Confusion matrices for the classification case study presented in "Supervised classification-AAA": for the Bottleneck distance (**a**) and for the 2-Wasserstein distance (**b**). Label A is for patients with AAA, H for the "healthy" ones.

Supervised classification-AAA

In "Clustering" we reported an unsupervised algorithmic pipeline which generates a very clear clusterization of the topological representations of the patients in our study. We now setup a supervised classification algorithm aimed at discriminating patients with an AAA from the others.

The dendrograms in Fig. 14 justify the use of a very simple algorithm like k-nearest neighbors (KNN). We report the leave-one-out (l1out) confusion matrices in Fig. 15 for $k = 5$. We point out that for any choice of $k \in \{1, \ldots, 5\}$ we get 97% accuracy, with the only missclassified patient being always the same one, visibly close to the healthy patients also in Fig. 13. This patient has been checked by the collaborating clinicians and it has been regarded as having a relatively small AAA.

Supervised classification-calcifications

We now want to discriminate patients with calcifications from the others. To do so, we rely on "Notable pairs": therein we showed that calcifications appear as mid-persistence features in the zero dimensional persistence diagram, being local minima of the radial filtering function.

Thus, for all patients, we count the number of 0 dimensional persistence pairs with persistence greater than or equal to a threshold and less than ∞. Patients with or without calcifications have been identified and labelled by looking at their TAC, with the help of a collaborating clinician.

To quantify the model's performance, we employed a leave-one-out methodology alongside a grid search to optimize the thresholds τ_p for persistence and τ_c for the minimum number of cycles required to classify an aorta as calcified. The grid spans a range from 0.005 to 0.5, with increments of 0.005 for τ_p, while for τ_c it considers the natural numbers from 1 up to 20. The optimal threshold values turned out a value of $\tau_p = 0.1$, a value coherent with "The effect of normalization" and $\tau_c = 3$. Results with this final model are shown in Fig. 16: all patients whose count of filtered persistence pairs exceeds the threshold τ_c of 3 are classified as affected by calcification. Thresholds τ_p and τ_c have been selected to maximize accuracy; the optimal model results in

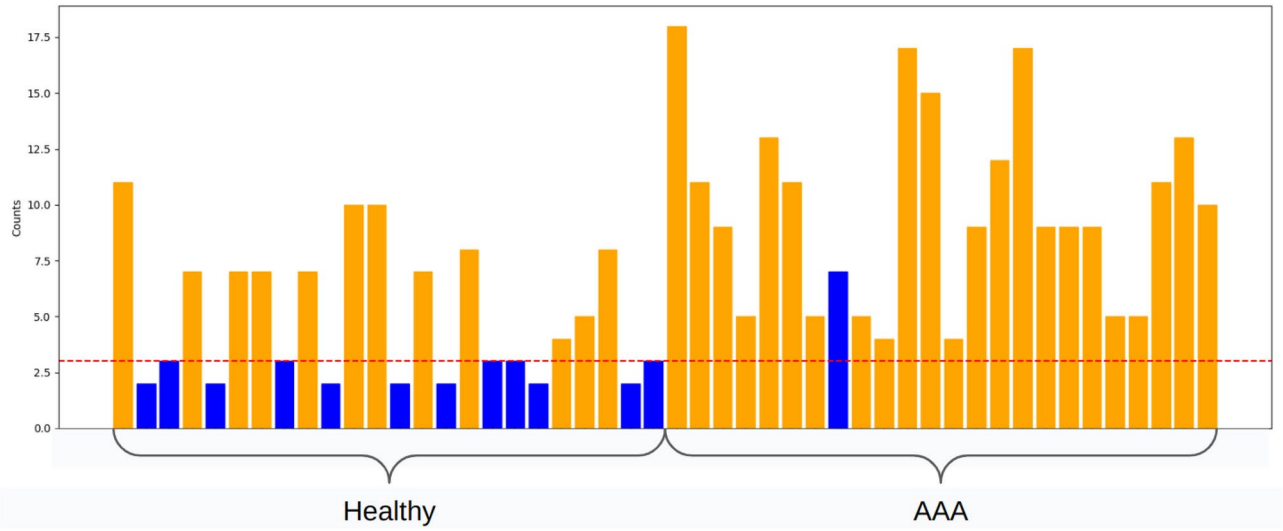

Figure 16. Plot related to "Supervised classification-calcifications", representing the counts of the persistence pairs in dimension 0 with persistence greater than 0.1 to recognize patients with calcifications. Patients which present calcifications are plotted in orange, the others in blue. There is clearly correlation between the presence of calcifications and irregularities of the lumen captured by 0 dimensional homology. We highlight that this holds true also for healthy patients, which is arguably the more interesting situation, as most of the patients affected by AAA present also calcifications. Patients are labelled and ordered as in previous figures.

only one misclassified patient. Notably, as τ_p increases, a better classification score is generally achieved with a lower τ_c, and vice versa as can be seen in Fig. 17. This relationship is intuitive; a lower τ_p retains more 0-cycles, necessitating a higher τ_c to compensate for the additional noise.

Note that, as shown by Fig. 16, identifying patients with calcifications is more of a problem for those without an AAA - the S patients in the left half of the figure - since almost all patients with an AAA show also the presence of calcifications.

Supervised classification-iliac aneurysm

We now deal with iliac aneurysms. As discussed in "Notable pairs", we expect an aneurysm in the iliac area to show up in the persistence diagram in two different ways: (1) the absence of a class of 0-cycles born on the affected iliac, separately from the one encompassing the healthy one; (2) one or two classes of 1-cycles with infinite persistence, born at high times. Since the absence of the second class of 0-cycles is redundant with respect to the information of the birth of the second class of 1-cycles, we just consider as features for this classification problem the birth coordinates of the two classes of 1-cycles with infinite persistence, ordered according to their birth. We call those two variables B_1 and B_2. The difference between patients with or without an iliac aneurysm, it's evident in Fig. 18a: large values of birth for both 1-cycles are associated to the presence of an iliac aneurysm. In particular the second coordinate, i.e. the largest birth, is the most discriminanting factor. This is coherent with the fact that having one iliac aneurysm increases the birth coordinate of the second class of 1-cycles with infinite persistence; while two aneurysms increase both coordinates. Thus the second feature is always altered by the presence of an aneurysm at the iliacs.

We fit a linear discriminant analysis model; by leave-one-out, this predicts correctly all patients but two. The confusion matrix is reported in Fig. 18b. Both misclassified patients are affected by iliac aneurysms, but classified as healthy by our pipeline: their aneurysms are indeed not severe, according to the collaborating clinicians.

Supervised classification-thrombus

Lastly, we turn to the problem of detecting thrombi, which have been discussed in "Notable pairs". We have argued that thrombi are characterized by irregularities in the lumen, most likely in the part of the blood vessel which is affected by AAA. Depending on the homogeneity of the obstruction, these irregularities might generate local maxima of the radial filtering function.

We approach thrombus in a manner similar to that described in "Supervised classification-calcifications", with the key difference of considering both 0 and 1 dimensional persistence pairs that retain a minimum finite persistence τ_p and birth coordinates bigger than τ_b. We conducted a leave-one-out analysis that makes use of grid search to optimize the thresholds τ_p and τ_b. The grid spans respectively from 0.05 to 0.18, with increments of 0.005 for τ_p and from 0.8 to 1.5, with increments of 0.2 for τ_b. This procedure resulted in three misclassified patients: A10, A20, A21 each having at least one retained class of cycles. The optimal classifier considers 0-cycles and 1-cycles with birth coordinate larger than or equal to 1.28 - ensuring that we are not on the iliac arteries or the neck - and persistence larger than or equal to 0.07 - to filter out smaller and noisy bumps. Disregarding the label declaring if they are 0-cycles or 1-cycles, the number of cycles retained for each patient is shown in Fig. 19; it is noteworthy that all healthy patients do not present any thrombi. Similarly to what noted in "Supervised

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0.005	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
0.010	35	35	35	35	35	35	35	35	35	35	35	35	35	35	36	36	36	36	36	36
0.015	35	35	35	35	35	35	35	35	35	35	35	35	36	36	36	36	36	36	37	40
0.020	35	35	35	35	35	35	35	35	35	35	36	36	37	38	41	40	42	43	42	41
0.025	35	35	35	35	35	35	35	36	36	37	40	41	42	42	43	43	43	43	39	38
0.030	35	35	35	35	35	35	36	37	38	43	45	45	44	42	42	41	41	37	36	33
0.035	35	35	35	35	35	36	38	41	42	44	43	43	42	40	38	35	32	31	30	27
0.040	35	35	35	35	36	39	42	43	44	44	42	42	39	37	36	32	31	29	27	26
0.045	35	35	35	36	39	41	41	43	44	42	41	40	34	34	33	32	28	28	27	24
0.050	35	35	36	37	40	43	43	43	43	41	39	36	34	33	29	27	25	24	22	18
0.055	35	36	36	40	43	44	45	43	42	41	40	35	32	29	28	26	23	21	18	17
0.060	35	36	37	43	44	43	45	40	38	38	36	33	30	27	25	22	20	19	17	16
0.065	35	37	38	42	44	44	44	39	38	37	33	32	30	23	22	20	18	16	16	16
0.070	35	38	42	43	44	43	39	38	37	34	30	29	25	22	20	18	17	16	16	15
0.075	35	38	43	45	44	42	39	38	36	31	29	24	22	19	18	18	16	16	16	15
0.080	35	38	46	44	44	40	40	38	35	30	24	21	20	19	18	18	16	16	15	15
0.085	35	38	46	44	43	39	39	36	31	28	23	21	20	19	18	17	16	16	15	15
0.090	35	38	46	45	42	39	37	33	29	28	22	20	19	18	17	17	16	15	15	15
0.095	35	40	46	44	40	38	34	32	29	26	20	19	18	17	16	16	15	14	14	13
0.100	35	42	47	44	38	38	34	32	27	24	20	19	17	17	16	16	14	13	13	13
0.105	35	43	46	41	38	37	33	29	24	21	20	17	17	16	16	15	14	13	13	13
0.110	35	45	46	41	38	35	30	26	24	21	18	17	17	15	15	14	13	13	13	13
0.115	35	45	46	41	38	33	29	26	23	21	18	16	16	15	14	13	13	13	13	13
0.120	35	46	46	40	36	31	28	25	23	18	18	16	16	14	13	13	13	13	13	13
0.125	36	47	46	39	35	31	28	24	23	17	17	16	15	14	13	13	13	13	13	13
0.130	36	47	46	37	34	29	28	23	21	17	16	15	15	14	13	13	13	13	13	13
0.135	36	47	46	37	33	28	26	22	19	17	15	15	14	14	13	13	13	13	13	13
0.140	36	47	44	36	30	28	25	22	19	17	15	14	14	14	13	13	13	13	13	13
0.145	36	46	42	34	30	28	24	22	19	17	14	14	14	13	13	13	13	13	13	13
0.150	36	46	42	34	30	27	22	22	17	17	14	14	13	13	13	13	13	13	13	13
0.155	36	46	42	33	29	27	22	22	17	17	13	13	13	13	13	13	13	13	13	13
0.160	36	46	40	31	28	26	21	18	17	14	13	13	13	13	13	13	13	13	13	13
0.165	36	46	38	30	27	24	21	18	14	14	13	13	13	13	13	13	13	13	13	13
0.170	36	46	37	29	24	24	20	16	14	14	13	13	13	13	13	13	13	13	13	13
0.175	36	46	35	29	24	22	19	15	14	14	13	13	13	13	13	13	13	13	13	13
0.180	36	45	35	28	21	22	18	15	14	14	13	13	13	13	13	13	13	13	13	13
0.185	36	45	34	27	21	22	17	15	13	13	13	13	13	13	13	13	13	13	13	13
0.190	36	45	34	26	21	22	16	14	13	13	13	13	13	13	13	13	13	13	13	13
0.195	36	45	32	24	21	22	16	14	13	13	13	13	13	13	13	13	13	13	13	13
0.200	38	45	31	23	21	21	15	14	13	13	13	13	13	13	13	13	13	13	13	13
0.205	38	43	31	23	20	19	14	14	13	13	13	13	13	13	13	13	13	13	13	13

Figure 17. A partial view of the table showing the aortas with calcifications correctly classified through the grid search of the optimal values for the parameters τ_p and τ_c. The rows represent different values of τ_p, while the columns indicate values for τ_c. The greener a cell is the more patients are correctly identified.

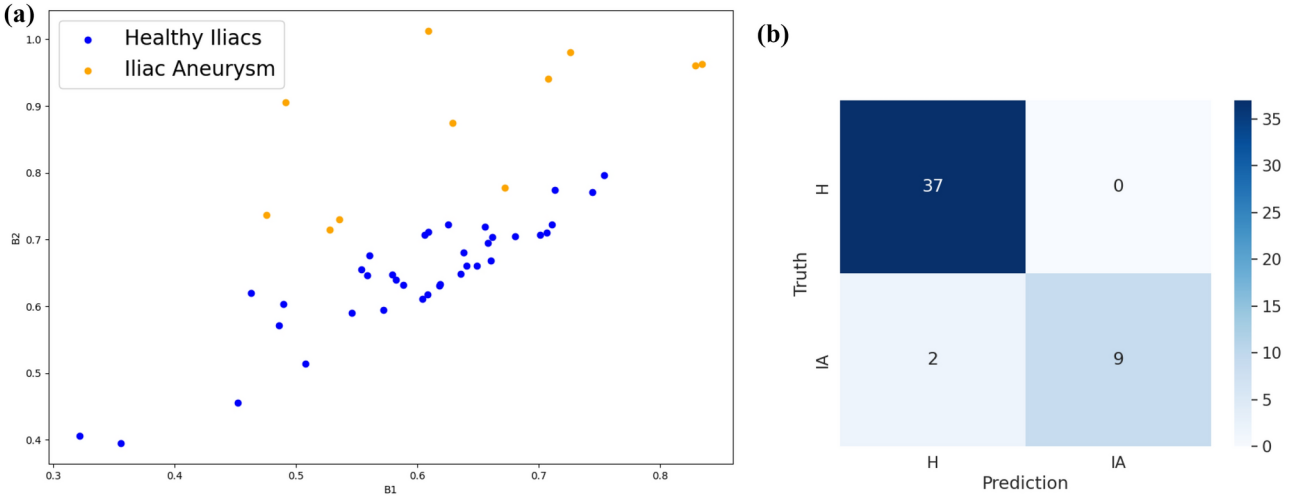

Figure 18. (**a**) The scatterplot of the variables B_1 and B_2 used for classifying patients with iliac aneurysm in "Supervised classification-iliac aneurysm"; (**b**) L1out confusion matrix for Iliac Aneurysm classification. Label AI is for patients with iliac aneurysm, H for the others.

classification-calcifications", a pattern can be observed where lowering τ_p necessitates higher values of τ_b, as can be seen in Fig. 20. This is due to the fact that aortas not affected by abdominal aortic aneurysms (AAA), and therefore not by thrombus either, have persistence diagrams with pairs born at lower filtration values.

Concerning the misclassified patients, in Fig. 21 we report data about one of these patients, to understand the roots of this classification errors. For instance, patient A10 is affected by thrombi but classified as not having

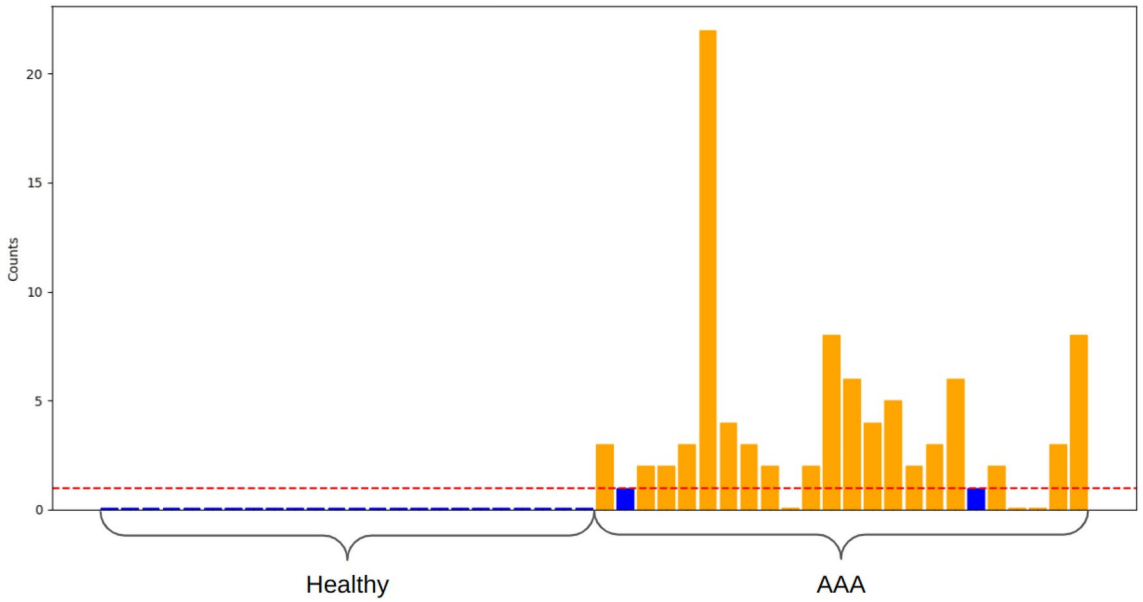

Figure 19. Plot related to "Supervised classification-thrombus", representing the counts of persistence pairs with medium persistence associated to irregularities of the lumen, trying to exclude bumps located on the iliacs and on the aortic neck. Patients which present thrombi are plotted in orange, the others in blue. Again we can see correlation between the presence of thrombi and irregularities of the lumen captured by the selected points in 0 and 1 dimensional homology.

	0.90	0.92	0.94	0.96	0.98	1.00	1.02	1.04	1.06	1.08	1.10	1.12	1.14	1.16	1.18	1.20	1.22	1.24	1.26	1.28	1.30	1.32	1.34	1.36	1.38	1.40
0.050	23	23	23	23	23	24	26	29	36	41	45	45	44	44	44	44	44	44	44	44	44	43	42	41	41	40
0.055	23	23	23	23	23	25	30	34	38	41	45	45	44	44	44	44	44	44	44	44	44	43	40	39	39	38
0.060	23	23	25	25	25	26	30	33	38	40	43	44	43	43	43	43	43	43	44	44	44	43	40	39	39	38
0.065	24	25	27	27	27	28	33	37	40	42	43	44	43	43	43	43	43	43	44	44	44	43	40	39	39	38
0.070	29	29	30	30	30	30	34	38	41	42	43	44	43	43	43	43	43	43	44	45	44	43	40	39	39	38
0.075	30	31	31	31	31	31	35	38	42	43	45	45	44	42	42	43	43	43	42	42	41	41	39	38	38	37
0.080	34	35	35	35	35	35	36	38	42	42	44	44	42	41	41	42	42	42	41	41	40	40	39	38	38	37
0.085	35	36	36	35	36	37	39	41	43	42	44	44	42	41	41	42	42	42	41	41	40	40	39	38	37	36
0.090	34	35	35	35	36	37	40	42	44	43	44	44	42	41	41	42	42	42	41	41	40	40	39	38	37	36
0.095	35	36	36	36	37	38	40	42	44	43	44	44	41	40	40	41	41	41	40	40	38	38	38	37	37	36
0.100	38	39	39	39	39	40	42	42	44	43	44	44	41	40	40	41	41	41	40	40	38	38	38	37	37	36
0.105	39	39	39	39	39	40	41	41	43	42	42	42	39	39	39	40	40	40	39	39	38	38	38	37	37	36
0.110	40	40	40	40	40	40	40	41	43	42	42	42	39	39	39	40	40	40	39	38	37	37	37	36	36	35
0.115	40	40	40	40	40	40	40	42	43	42	42	42	38	38	38	39	39	39	38	38	37	37	37	36	36	35
0.120	41	41	41	41	41	42	41	43	43	42	42	42	38	38	38	39	39	38	37	37	36	36	36	35	35	34
0.125	41	41	41	41	41	42	42	43	43	42	42	42	38	38	38	39	39	38	37	37	36	36	36	35	35	34
0.130	41	41	41	41	41	42	42	43	43	42	42	42	38	38	37	38	38	37	36	36	35	35	35	34	34	33
0.135	41	41	41	41	41	42	41	42	42	41	41	41	37	37	36	37	37	36	35	35	34	34	34	33	33	32
0.140	41	41	41	41	41	42	40	42	41	40	39	39	37	37	36	37	37	36	35	35	34	34	34	33	33	32
0.145	42	42	42	42	41	41	39	41	40	40	39	39	37	37	36	37	37	36	35	35	34	34	34	33	33	32
0.150	42	42	42	42	41	41	38	40	40	40	39	39	37	37	36	37	37	36	35	35	34	34	34	33	33	32
0.155	42	42	42	42	41	41	38	40	40	40	39	39	37	37	36	37	37	36	35	35	34	34	34	33	33	32
0.160	42	43	43	43	42	41	38	40	40	40	39	39	37	37	36	37	37	36	35	35	34	34	34	33	33	32
0.165	42	43	43	43	42	40	38	40	40	40	39	39	37	37	36	37	37	36	35	35	34	34	34	33	33	32
0.170	43	44	44	44	43	41	38	39	39	39	38	38	36	36	35	36	36	35	34	34	33	33	33	32	32	31
0.175	43	44	43	43	42	40	37	38	38	38	37	37	35	35	34	35	35	34	33	33	32	32	32	31	31	31
0.180	43	44	43	43	42	40	37	38	38	38	37	37	35	35	34	35	35	34	33	33	32	32	32	31	31	31

Figure 20. A partial view of the table showing the aortas with thrombus correctly classified through the grid search of the optimal values for the parameters τ_p and τ_b. The rows represent different values of τ_p, while the columns indicate values for τ_b. The greener a cell is the more patients are correctly identified.

any. The reason can be clearly seen in Fig. 21b: the thrombus outside the lumen is very homogeneous, resulting in a very straight and regular lumen, with no bumps. By construction, our methodology is good at capturing information contained in inwards and outwards bumps of the lumen, so the persistence diagrams of patient A10 are similar to those of a healthy patient. However, we do believe that segmenting also the thrombi from the CTA scans could solve this issue. We leave this further investigation to future works.

Discussion and conclusion

We introduced a sophisticated pipeline for analyzing data obtained from medical imaging of abdominal aortas, utilizing topological data analysis tools. Each reconstructed mesh of an abdominal aorta is represented by a

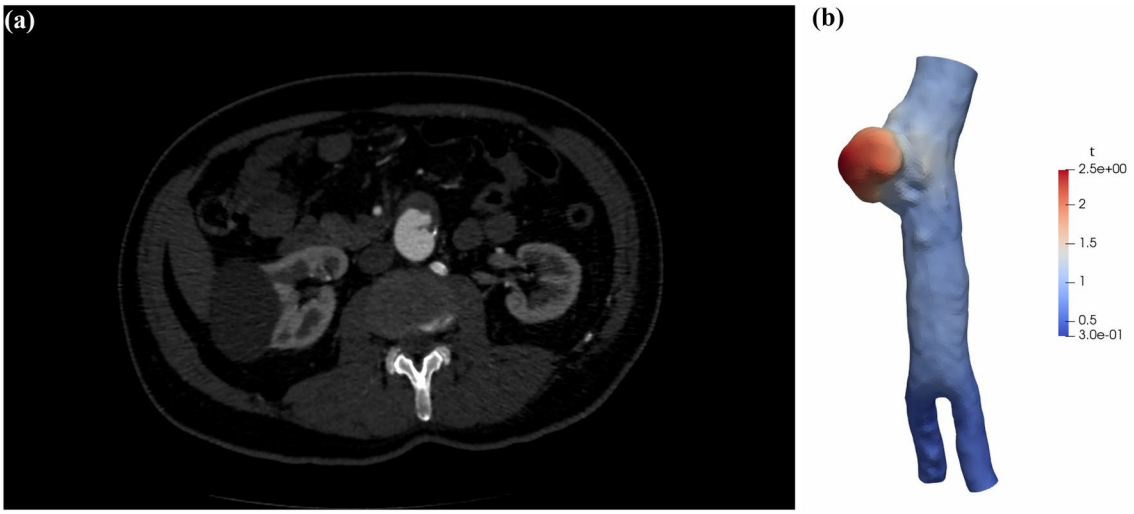

Figure 21. One of the patients (A10) missclassified by the analysis in "Supervised classification-thrombus": a CTA slice (left) and the reconstructed mesh (right). The color map refers to the value of *t*: the radial distance from each point in the mesh to the centerline.

persistence diagram, which highlights the 0-cycles and 1-cycles associated with a radial filtration function applied to the mesh. These diagrams effectively capture many shape-related features of the blood vessel, such as irregularities in the lumen caused by abdominal aortic aneurysms, as well as related phenomena like lumen-exposed calcifications and intra-luminal thrombi.

We argue that the proposed representation would significantly improve the analysis of abdominal aortic aneurysms (AAAs) and, to support this, we present the results of several classification tasks using a training set of persistence diagrams derived from 48 CTA scans, 24 of which depict aneurysmal aortas.

The take home message is that that the use of persistence diagrams greatly simplifies these classification tasks, due to the amount of information collected by these easy-to-handle mathematical objects. Additionally, because of the unsettling difference with more classical feature-gathering methods, we devoted a consistent part of this work to stress the interpretability of the rich information displayed by persistence diagrams.

We also want to discuss some limitations of the present work, which serve as points of improvement to be considered in future analyses.

Motivated by the importance attributed by literature to the geometry of the lumen of the blood vessels, we focused our attention on its segmentation and analysis. We stress that such relevance is supported both by haemodynamics simulations and studies, and from clinical practice. For instance, the calcifications that clinicians deem more relevant are the ones which are directly in contact with the lumen's surface, compared to the ones that are hidden behind some structures like thrombi. Still, further information could be segmented from the CTA scans, like the external walls of the vessels or the thrombi, and integrated into the pipeline. This additional segmentation step comes with a series of complexities that require a substantial effort to develop ad-hoc numerical methods, especially when applied to medium-to-large patient cohorts, which could make a standalone research project. The current segmentation procedure, in fact, leverages on the high contrast provided by arterial blood with the surrounding tissues in grayscale imaging, which applies only to the lumen.

Should this be achieved, we are confident that comparison of the radial function taken with respect to the lumen and with respect to the external aortic walls, or the direct segmentation of the thrombi, would further improve the already promising and relevant results we have obtained. In particular, it would provide an even more comprehensive view of all the morphological and clinical aspects of the aorta, such as the actual size of the aortic dilation and the distribution of thrombus and wall characteristics, which are only partially observable with the current setup.

Other further developments of this work could tackle the forecasting and regression problems which have already captured the attention of the medical and statistical communities, like estimating growth rates of AAA, indication to surgery and rupture's risks. To that end, persistence diagrams present a number of mathematical advantages, briefly mentioned in the manuscript, allowing for more refined object oriented data analysis which will be the focus of our future works.

Data availability

The data that support the findings of this study are available from Ospedale Maggiore Policlinico but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Ospedale Maggiore Policlinico.

Appendix: Average neck radius

Neck's average radius has been estimated with the following steps:

1.Linear interpolation of the centerline's points.
2.For a given parameter $s = 0.5mm$, empirically chosen, starting at the neck seed's height s_0 on the centerline, find the points s_i at distance *s* on the curvilinear abscissa.
3.For each of the s_i points find the planes P_i. orthogonal to the centerline in s_i. For each of the planes P_i identified, a *fitting ellipse* of the mesh points is built, following the Fitzgibbon's approach[42,43], and its semi-major and semi-minor axes M_i and m_i are computed. In some cases the slice can identify more regions; when this verifies, the ellipse is built using only the region of points that contains the point s_i.
4.Starting from the initial seed s_0, a check is made for all the planes P_j found. If the two proposition:

$$\frac{\mid M_i - M_{i+j} \mid}{M_i} < 0.10 \quad \forall j = 1, \cdots, k \quad \textit{with k=10 (5 mm)} \tag{2a}$$

$$M_i < 22.5mm \tag{2b}$$

hold, M_i and m_i are collected. The first and all the following s_i that do not respect the proposition are discarded. k is chosen empirically and the conditions are made in order to control both the raw AAA's diameter and its relative growth.

5.The result is the mean of the mean of M_i and m_i for each s_i that is not discarded. In this way we obtain an accurate approximation of the neck's radius.

Received: 16 April 2024; Accepted: 29 October 2024
Published online: 15 November 2024

References

1. Marron, J. S. & Dryden, I. L. *Object Oriented Data Analysis* (CRC Press, 2021).
2. Sonesson, B., Länne, T., Hansen, F. & Sandgren, T. Infrarenal aortic diameter in the healthy person. *Eur. J. Vasc. Endovasc. Surg.* **8**, 89–95 (1994).
3. Wilson, K. A., Lindholt, J. S., Hoskins, P. R. & et al. The relationship between abdominal aortic aneurysm distensibility and serum markers of elastin and collagen metabolism. *Eur. J. Vasc. Endovasc. Surg.* **21**, 175–178 (2001).
4. Dua, M. M. & Dalman, R. L. Hemodynamic influences on abdominal aortic aneurysm disease: Application of biomechanics to aneurysm pathophysiology. *Vasc. Pharmacol.* **53**, 11–21 (2010).
5. Mealy, K. & Salman, A. The true incidence of ruptured abdominal aortic aneurysms. *Med. Eng. Phys* **2**, 405–408 (1988).
6. Piccinelli, M. et al. Impact of hemodynamics on lumen boundary displacements in abdominal aortic aneurysms by means of dynamic computed tomography and computational fluid dynamics. *Biomech. Model. Mechanobiol.* **12**, 1263–1276 (2013).
7. Vergara, C., Le Van, D., Quadrio, M. & et al. Large eddy simulations of blood dynamics in abdominal aortic aneurysms. *Med. Eng. Phys*. **47**, 38–46 (2017).
8. Doyle, B., Cloonan, A., Walsh, M., Vorp, D. & McGloughlin, T. Identification of rupture locations in patient-specific abdominal aortic aneurysms using experimental and computational techniques. *J. Biomech.* **43**, 1408–1416 (2010).
9. Georgakarakos, E. et al. The role of geometric parameters in the prediction of abdominal aortic aneurysm wall stress. *Eur. J. Vasc. Endovasc. Surg.* **39**, 42–48 (2010).
10. Georgakarakos, E. et al. The influence of intraluminal thrombus on abdominal aortic aneurysm wall stress. *Int. Angiol.* **28**, 325–33 (2009).
11. Li, Z. et al. Impact of calcification and intraluminal thrombus on the computed wall stresses of abdominal aortic aneurysm. *J. Vasc. Surg.* **47**, 928–35 (2008).
12. Parr, A., Jayaratne, C., Buttner, P. & Golledge, J. Comparison of volume and diameter measurement in assessing small abdominal aortic aneurysm expansion examined using computed tomographic angiography. *Eur. J. Radiol.* **79**, 42–47 (2011).
13. Kauffmann, C. et al. Measurements and detection of abdominal aortic aneurysm growth: accuracy and reproducibility of a segmentation software. *Eur. J. Radiol.* **81**, 1688–1694 (2012).
14. Lee, R., Jarchi, D., Perera, R. & et al. Applied machine learning for the prediction of growth of abdominal aortic aneurysm in humans. *EJVES Short Rep*. **39**, 24–28 (2018).
15. Hirata, K., Nakaura, T., Nakagawa, M. & et al. Machine learning to predict the rapid growth of small abdominal aortic aneurysm. *J. Comput. Assist. Tomogr*. **44**, 37–42 (2020).
16. Kontopodis, N. et al. Prediction of abdominal aortic aneurysm growth by artificial intelligence taking into account clinical, biologic, morphologic, and biomechanical variables. *Vascular* **31**, 409–416 (2023).
17. Zhu, C. et al. Intraluminal thrombus predicts rapid growth of abdominal aortic aneurysms. *Radiology* **294**, 707–713 (2020).
18. Lindquist Liljeqvist, M. et al. Geometric and biomechanical modeling aided by machine learning improves the prediction of growth and rupture of small abdominal aortic aneurysms. *Sci. Rep.* **11**, 18040 (2021).
19. Kyriakou, F., Dempster, W. & Nash, D. A methodology to quantify the geometrical complexity of the abdominal aortic aneurysm. *Sci. Rep.* **9**, 17379 (2019).
20. Munkres, J. R. *Elements of Algebraic Topology* (CRC Press, 2018).
21. Edelsbrunner, H. & Harer, J. L. *Computational Topology: An Introduction* (American Mathematical Society, 2022).
22. Ahrens, J., Geveci, B., Law, C., Hansen, C. & Johnson, C. 36-paraview: An end-user tool for large-data visualization. *Vis. Handb.* **717**, 50038–1 (2005).
23. Fosbinder, R. & Orth, D. *Essentials of Radiologic Science* (Lippincott Williams & Wilkins, 2011).
24. Sonka, M. & Fitzpatrick, J. M. Handbook of medical imaging. In *Medical Image Processing and Analysis*. Vol. 2. (SPIE, 2009).
25. Antiga, L. et al. An image-based modeling framework for patient-specific computational hemodynamics. *Med. Biol. Eng. Comput.* **46**, 1097–1112 (2008).
26. Antiga, L. *Patient-Specific Modeling of Geometry and Blood Flow in Large Arteries*. Ph.D. Thesis, Politecnico di Milano (2002).
27. Kyriakou, F., Dempster, W. & Nash, D. A methodology to quantify the geometrical complexity of the abdominal aortic aneurysm. *Sci. Rep*. **9** (2019).

28. Cohen-Steiner, D., Edelsbrunner, H. & Harer, J. Stability of persistence diagrams. In *Proceedings of the Twenty-First Annual Symposium on Computational Geometry*. 263–271 (2005).
29. Bubenik, P. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16**, 77–102 (2015).
30. Adams, H. et al. Persistence images: A stable vector representation of persistent homology. *J. Mach. Learn. Res.* **18**, 1–35 (2017).
31. Chazal, F., Fasy, B. T., Lecci, F., Rinaldo, A. & Wasserman, L. Stochastic convergence of persistence landscapes and silhouettes. *J. Comput. Geom.* **6**, 140–161 (2015).
32. Biscio, C. A. & Møller, J. The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications. *J. Comput. Graph. Stat.* **28**, 671–681 (2019).
33. Maria, C., Boissonnat, J.-D., Glisse, M. & Yvinec, M. The gudhi library: Simplicial complexes and persistent homology. In *Mathematical Software–ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings 4*. 167–174 (Springer, 2014).
34. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
35. Virtanen, P. et al. Fundamental algorithms for scientific computing in Python. Scipy 1.0. *Nat. Methods* **17**, 261–272 (2020).
36. Waskom, M. L. Seaborn: Statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
37. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
38. Pedersen, O. M., Aslaksen, A. & Harald, V.-M. Numerically stable direct least squares fitting of ellipses. *J. Vasc. Surg.* **17** (1993).
39. Kim, H. et al. Aortoiliac diameter and length in a healthy cohort. *PLOS ONE* **17**, 1–13. https://doi.org/10.1371/journal.pone.0268077 (2022).
40. Adolph, R. et al. Cellular content and permeability of intraluminal thrombus in abdominal aortic aneurysm. *J. Vasc. Surg.* **25**, 916–26 (1997).
41. Tong, J. & Holzapfel, G. Structure, mechanics, and histology of intraluminal thrombi in abdominal aortic aneurysms. *Ann. Biomed. Eng.* **43**, 1488–1501. https://doi.org/10.1007/s10439-015-1332-5 (2015).
42. Fitzgibbon, A., Pilu, M. & Fisher, R. B. Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 476–480 (1999).
43. Halir, R. & Flusser, J. Numerically stable direct least squares fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.* **21** (2000).

Acknowledgements

Dario Arnaldo Domanin acknowledges the support of Lorena Corazza for the realization of the images. Dario Arnaldo Domanin acknowledges the support of Abele Simona for the help with the Segmentation Pipeline and *Moxoff*(https://www.moxoff.com/en/) for providing the code and the equipment used for the segmentations. Matteo Pegoraro acknowledges the support by Independent Research Fund Denmark, grant 1026-00037. Maurizio Domanin acknowledges the support of Irene Fulgheri for the help with Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico's ethical committee. Maurizio Domanin and Santi Trimarchi acknowledge the support by MUR, Italy, grant Dipartimento di Eccellenza 2023-2027. Piercesare Secchi acknowledges the support by MUR, Italy, grant Dipartimento di Eccellenza 2023-2027.

Author contributions

D.A.D.: Data Curation, Formal Analysis, Software, Writing; M.P.: Conceptualization, Methodology, Writing; S.T.: Resources, Validation, Reviewed and Approved The Manuscript; M.D.: Resources, Validation, Reviewed and Approved The Manuscript; P.S.: Conceptualization, Project Administration, Supervision, Writing.

Competing Interests

The authors declare no competing interests.

Ethical statement

The study has been conducted in accordance with the Declaration of Helsinki and subsequent modifications. It was approved by the ethics committee Comitato Etico Milano Area 2 at the Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milan, with protocol number 3457/2023. The need for a signature of a specific informed consent was waived by the same committee as the study was observational and retrospective. No further requirements of clinical trials was necessary for the same reason.

Additional information

Correspondence and requests for materials should be addressed to M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.