POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Topological Data Analysis of the topology of the Abdominal Aortic Aneurysm through the use of Persistent Homology

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Author: **Dario Arnaldo Domanin**

Student ID: 969332
Advisor: Prof. Piercesare Secchi
Co-advisors: PhD Matteo Pegoraro, Prof. Santi Trimarchi
Academic Year: 2022-23

Abstract

Abdominal Aortic Aneurysm (AAA) is a vascular disease that affects the lower portion of the aorta, the biggest artery of the human body. AAA is a dangerous pathology that, in an asymptomatic way, can evolve towards the progressive enlargement of the aortic diameter finally leading to its rupture with fatal consequences.

Yet, given the number of available studies, statistics seems to be little considered in the study of AAA, especially if compared with the large amount of mechanical and haemodynamical studies.

This thesis aims to fill the main gaps in statistical literature with two different kind analysis regarding the AAA's morphology.

The first part of this work consists of a cluster analysis using global aortic descriptors such as AAA's dimensions (maximum AAA diameter, tortuosity and, volume) and angles, obtaining five groups of patients characterized by different levels of expected AAA severity.

These results provide an initial overview of the different deformations that can affect the aorta when AAA develops. Despite that, the initial dataset has been put aside, as well as the initial aim, given the lack of other features to better describe the aortic morphology. Then, it was chosen to opt for another type of data: instead of using global descriptor, a new dataset has been built combining data derived from Computed Tomography Angiography (CTA) scans, the most used imaging source for the diagnosis of arterial diseases, and Persistent Homology. This method is widely used in Topological Data Analysis to study an object's topological features.

After a process called Segmentation of the CTA scans, that allows to retrieve a digitalization of the aorta in a 3-dimensional way, the data have been processed into Persistence Diagrams, a fundamental tool of Persistent Homology, obtaining an accurate summary of its topological features.

Then, a comparison between the original mesh and the diagram obtained has been made for every patient, linking the most important aortic wall features such as calcification and thrombus to the diagram's ones. Finally, the capacity to enclose and summarize the

topological information has been tested using particular distances obtaining promising results.

Topological Data Analysis allows to expand the use of statistics to this clinical problem in a mathematically rigorous and effective way, retrieving a low-dimensional representation of the aorta that is effective for research purposes.

This brings to an accurate summary of the aortic morphology, making Persistent Homology a promising mathematical tool that could be adopted in future studies on AAA's growth. It could also be used by clinicians in the always growing awareness and knowledge of AAA and other vascular pathologies.

Keywords: Abdominal Aortic Aneurysm, AAA, Persistent Homology, Topological Data Analysis

Abstract in lingua italiana

L'aneurisma dell'aorta addominale (AAA) è una malattia vascolare che colpisce la porzione inferiore dell'aorta, l'arteria più grande del corpo umano. L'AAA è una pericolosa patologia che, in modo asintomatico, può evolvere verso il progressivo allargamento del diametro aortico portando infine alla sua rottura con conseguenze fatali.

Tuttavia, nonostante il numero di studi disponibili, la statistica sembra essere poco considerata nello studio dell'AAA, soprattutto se confrontata con la grande quantità di studi meccanici ed emodinamici.

Questa tesi si propone di colmare le principali lacune della letteratura statistica con due diverse analisi riguardanti la morfologia dell'AAA.

La prima parte di questo lavoro consiste in una cluster analysis utilizzando descrittori aortici globali come le dimensioni dell'AAA (diametro massimo dell'AAA, tortuosità e volume) e angoli, ottenendo cinque gruppi di pazienti caratterizzati da diversi livelli di gravità prevista dell'AAA.

Questi risultati forniscono una prima panoramica delle diverse deformazioni che possono interessare l'aorta quando si sviluppa AAA. Nonostante ciò, il dataset iniziale è stato messo da parte, così come lo scopo iniziale, data la mancanza di altre caratteristiche per descrivere meglio la morfologia aortica. Successivamente, si è scelto di optare per un altro tipo di dati: invece di utilizzare descrittori globali, è stato costruito un nuovo dataset che combina i dati derivati dalle scansioni a Tomografica Assiale Computerizzata (TAC), la fonte di imaging più utilizzata per la diagnosi delle malattie arteriose, e l'Omologia Persistente. Questo metodo è ampiamente utilizzato nella Topological Data Analysis per studiare le caratteristiche topologiche di un oggetto.

Dopo un processo chiamato Segmentazione delle TAC, che permette di recuperare una digitalizzazione dell'aorta tridimensionale, i dati sono stati elaborati in diagrammi di persistenza, uno strumento fondamentale dell'Omologia persistente, ottenendo un accurato riassunto della sua topologia caratteristiche.

Quindi, per ogni paziente è stato effettuato un confronto tra la mesh originale e il dia-

gramma ottenuto, collegando le caratteristiche più importanti della parete aortica come calcificazioni e trombi a quelle del diagramma. Infine, la capacità di racchiudere e sintetizzare le informazioni topologiche è stata testata utilizzando distanze apposite ottenendo risultati promettenti.

La Topological Data Analysis permette di estendere l'uso della statistica a questo problema clinico in modo matematicamente rigoroso ed efficace, recuperando una rappresentazione a bassa dimensionalità dell'aorta efficace ai fini della ricerca.

Questo porta ad un accurato riassunto della morfologia aortica, rendendo l'Omologia Persistente un promettente strumento matematico che potrebbe essere adottato in futuri studi sulla crescita dell'AAA e dai clinici nella sempre crescente consapevolezza riguardo l'AAA e di altre patologie vascolari.

Parole chiave: Aneurisma dell'Aorta Addominale, AAA, Omologia Persistente, Topological Data Analysis

Contents

1 | Introduction and State of the Art

1.1. Introduction to the medical Problem

1.1.1. The Circulatory System

The circulatory system is the set of organs responsible for the circulation of blood in the human body and it consists of blood, heart, and blood vessels. Its main role is to allow the blood to flow through the whole body enabling many essential functions for the functioning of the organism such as maintenance of the homeostasis, regulation of the temperature, transport of molecules like oxygen, nutrients, metabolites, and catabolites.

The dense and complex network of tubes in which the blood flows is the vascular system, composed by arteries, veins and capillaries.

Capillaries are the smallest blood vessels in human body. Placed at the farthest terminal of an artery or a vein, they serve the last part of the cell supply of nutrients. In fact, their structure consists of a single layer of cells which promotes their permeability allowing the metabolic exchange with the cells it supplies [1].

Arteries instead transport the oxygenated blood from the heart to the cells while the veins make the opposite path, bringing the "used" blood exhausted of oxygen and rich of CO_2 to the heart, which will send it back to the lungs restarting the cycle. The main difference between the two, other than their function is their structure, due to the difference of pressure that they have to endure [1, 2].

If veins are not subjected to constant pressure and have inelastic and thin walls with valves in order prevent the blood to stagnate, arteries on the other hand have thicker and elastic walls, with muscular tissue, in order to resist and maintain stable the blood pressure.

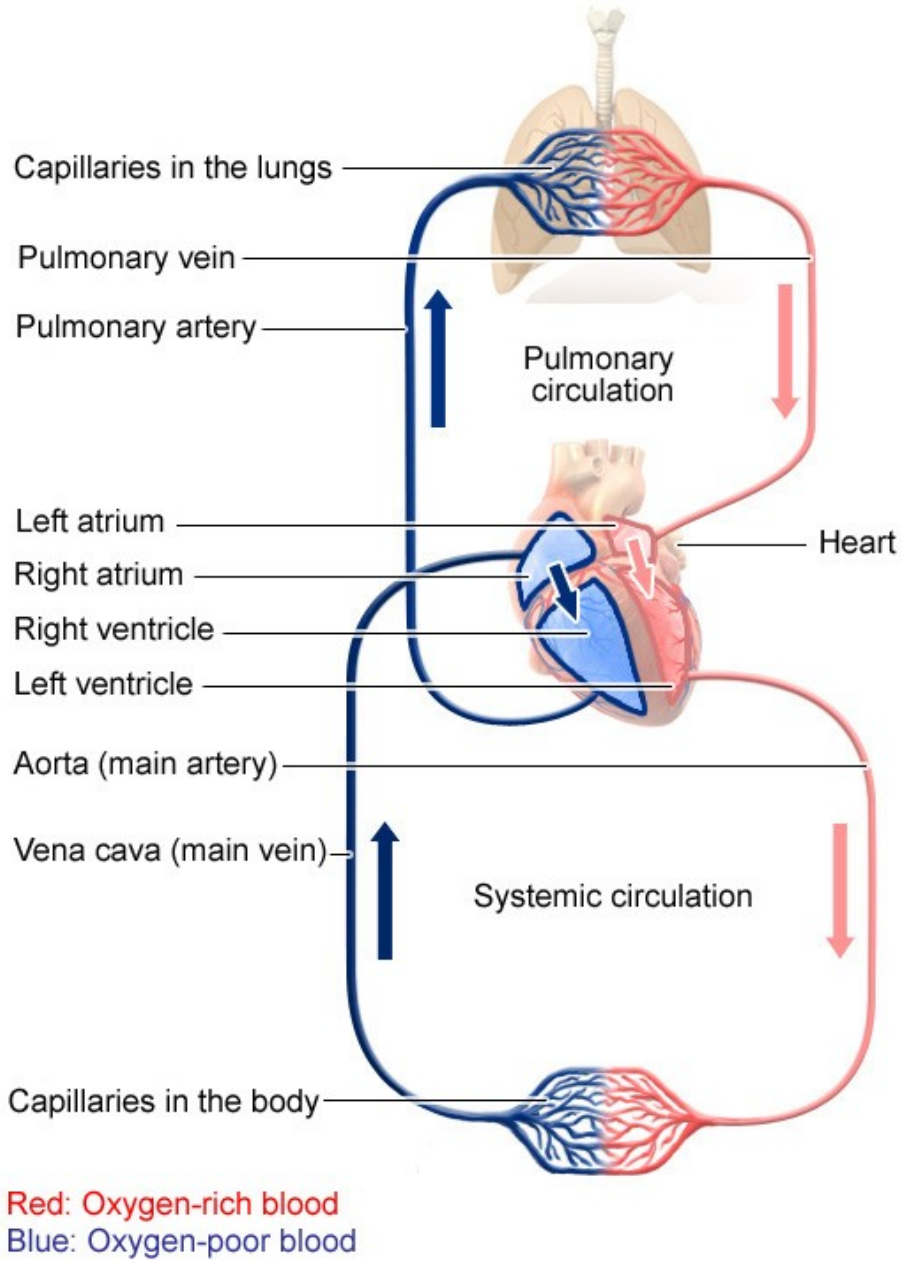


Figure 1.1: A graphical rapresentation of the circulatory system. Image from [3].

The aorta is the biggest and most important blood vessel of human body. It serves as an "highway", the first canal where the blood flows after exiting the heart and is distributed along all the body.

Aorta arises from the aortic Root at the left ventricle of the heart and ascends becoming the ascending aorta to create the aortic arch, above the heart. Then continues down,

behind the heart forming the descending aorta and goes through the thoracic cavity, below the diaphragm. Here, in the abdomen, it takes the name of abdominal aorta, the tract subject of this study. At the end of the abdominal cavity approximately at the fourth lumbar vertebra the aorta ends, splitting in two branch called iliac arteries.

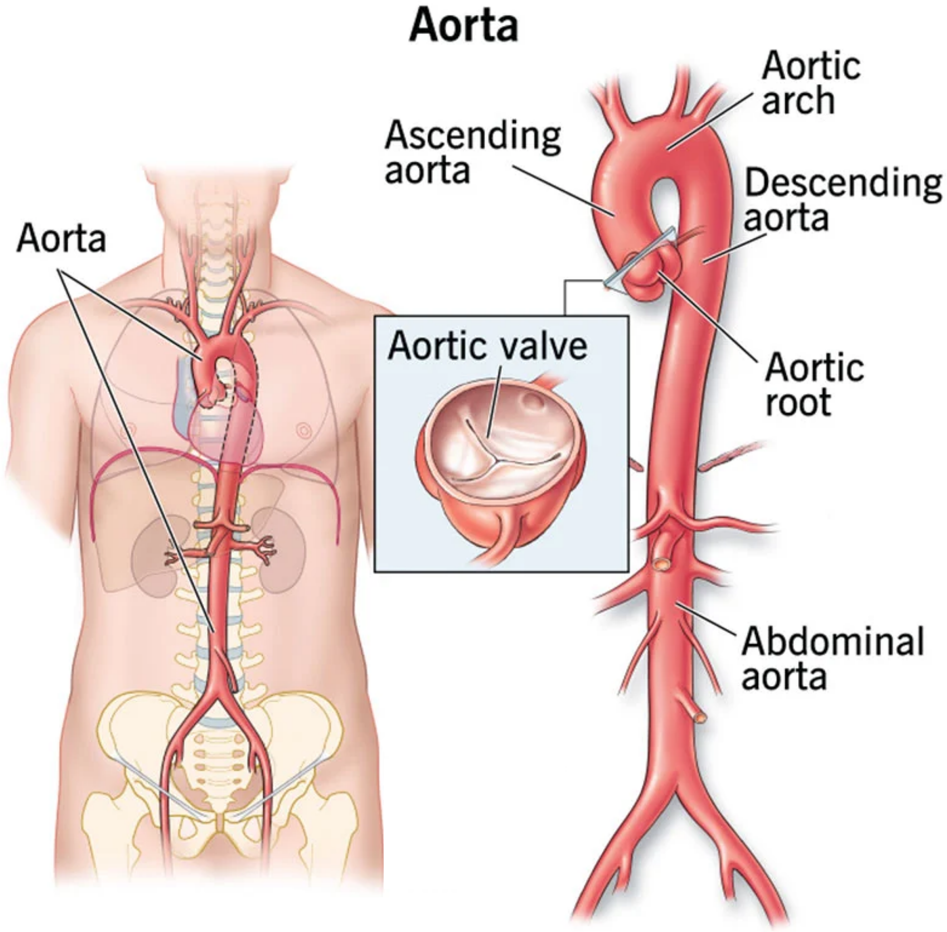


Figure 1.2: The aortic vessel. Image from [?].

The abdominal aorta has several branches. Just below the diaphragm there are three anterior branches to feed the internal organs: celiac trunk, superior and inferior mesenteric artery. There are then three pair of lateral branches: suprarenal, renal and ovarian artery, followed by other minor branches. At its end it divides into the iliac arteries that feeds the pelvic walls and the lower limbs and divide each one in turn in internal iliac artery and External iliac artery.

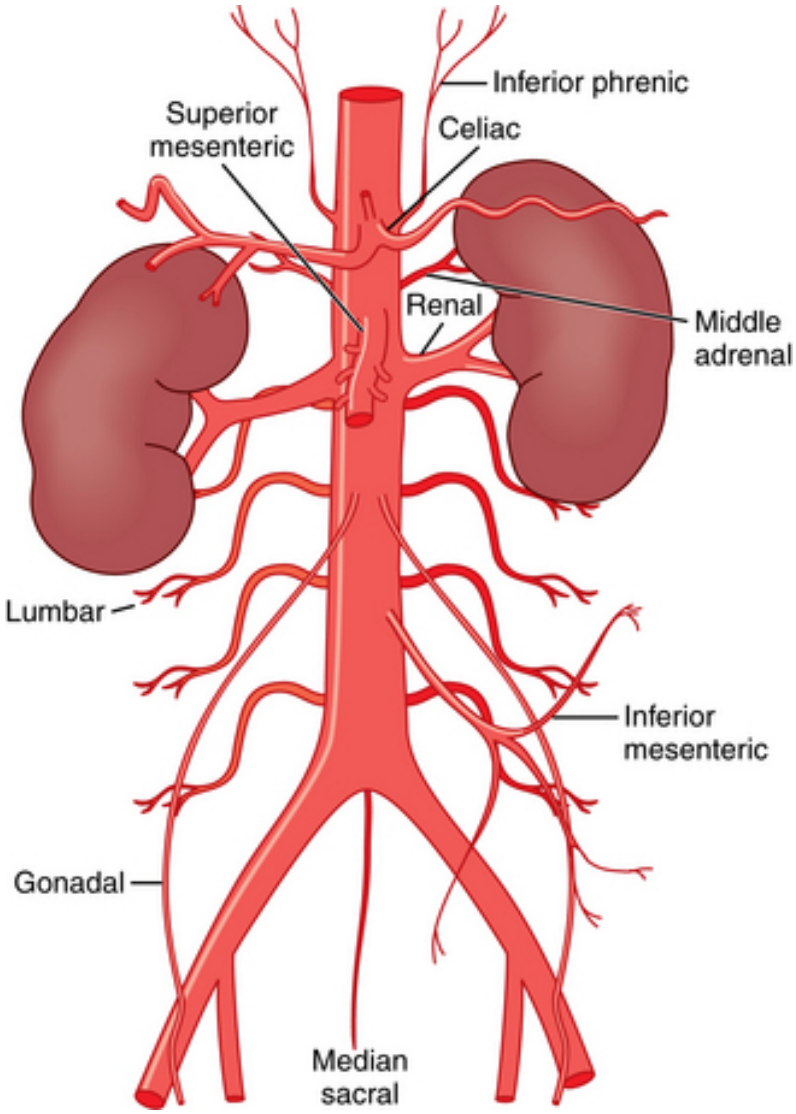


Figure 1.3: Anatomy of the abdominal aorta. Image from [4].

As well as acting as a mere conductor, the aorta plays another important role for the vascular system. Aorta have in fact to endure the continuous rise of pressure caused by the heartbeat which pumps "new" blood into the smaller branches that could be damaged and at the same time to avoid that the pressure levels get too low. For this reason, the bigger arteries, especially the aorta, are dynamical structure capable of expand during the systole, when the heart's ventricles contract and "release" the blood, and shrink during the diastole, when they relax and are filled with blood.

This is possible thanks to pressure receptors (baroceptors) and the aorta's wall needs to be strong enough to allow these multiple functions and to the particular histology of this artery. Aorta is in fact made by three distinct layers called tunica: an inner layer called intima, a middle one called media and adventitia, the outer layer.

The intima is the inner tunica of the aorta, in direct contact with blood. It is made up of one layer of endothelial cells and is supported by an internal elastic lamina [5].

Despite being the thinnest coat of the three tunic it plays important roles. It is in fact the site of the formation of growth-regulatory molecules, the regulation of the vascular tone and creates a surface for the blood flows that prevent the formation of thrombus (explanation in the next pages). Being very thin and exposed to the continuous flow of blood is particularly weak and its damage can bring to fatal pathologies such as dissection and aneurysm.

The tunica media consists mainly of smooth muscle, elastic collagen fibres and it is the thickest layer. The particular disposition of the structural fibres provides high mechanical strength, and this tunica is responsible for the properties of the whole artery. It also gives support to the vessel and the collagen fibres grants elastic properties to endure the high stress to which the vessel is subjected and maintenance of its homeostasis.

The outermost stratus is called tunica adventitia, a system of nerves, elastic fibres and contains the biggest concentration of collagen, that gradually decreases getting closed to the perimeter. The collagen serves a very important role thanks to their low extensibility and high stiffness giving strength to the external wall preventing excessive dilation when the pressure reaches its peak.

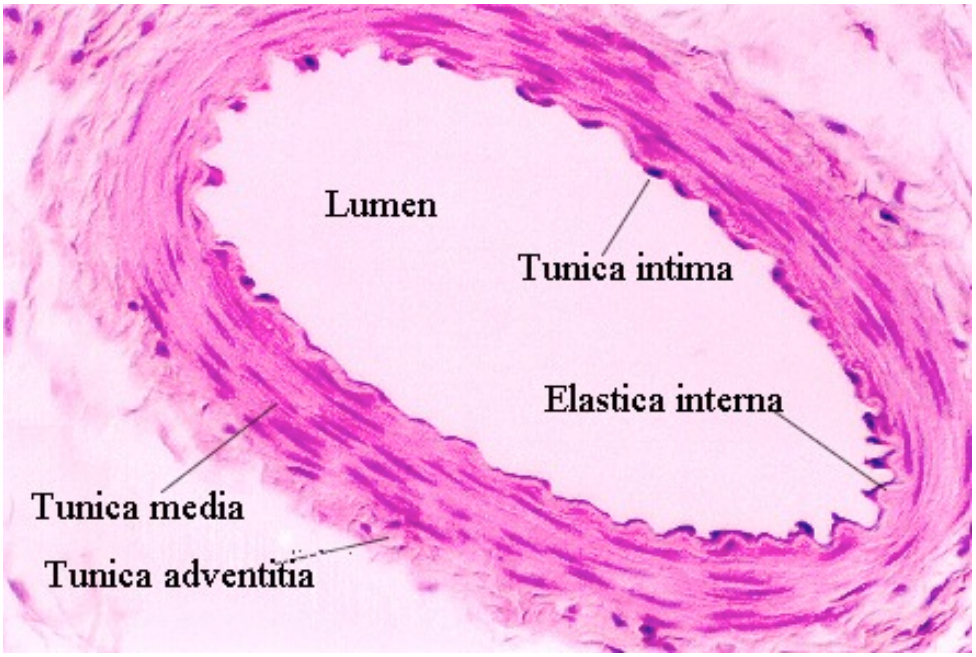


Figure 1.4: The histology of an artery, cut in cross section. Image from [6].

1.1.2. The Aneurysm

The size of the abdominal aorta depends on several factors such as gender, age, height, weight, and blood pressure, but generally its diameter is approximately 2 cm, varying from a range of 1.4 and 3 cm, that decreases gradually reaching the Bifurcation. When the diameter of aorta exceeds the threshold of 3 cm, or there is a dilatation, a bulging, or a ballooning of at least 50% increase in diameter respect the expected one, the artery is affected by aneurysm. aneurysms can occur anywhere throughout the circulatory system, but most commonly develop along the aorta and in blood vessels of the brain and if untreated can be fatal when a rupture occurs [7, 8].

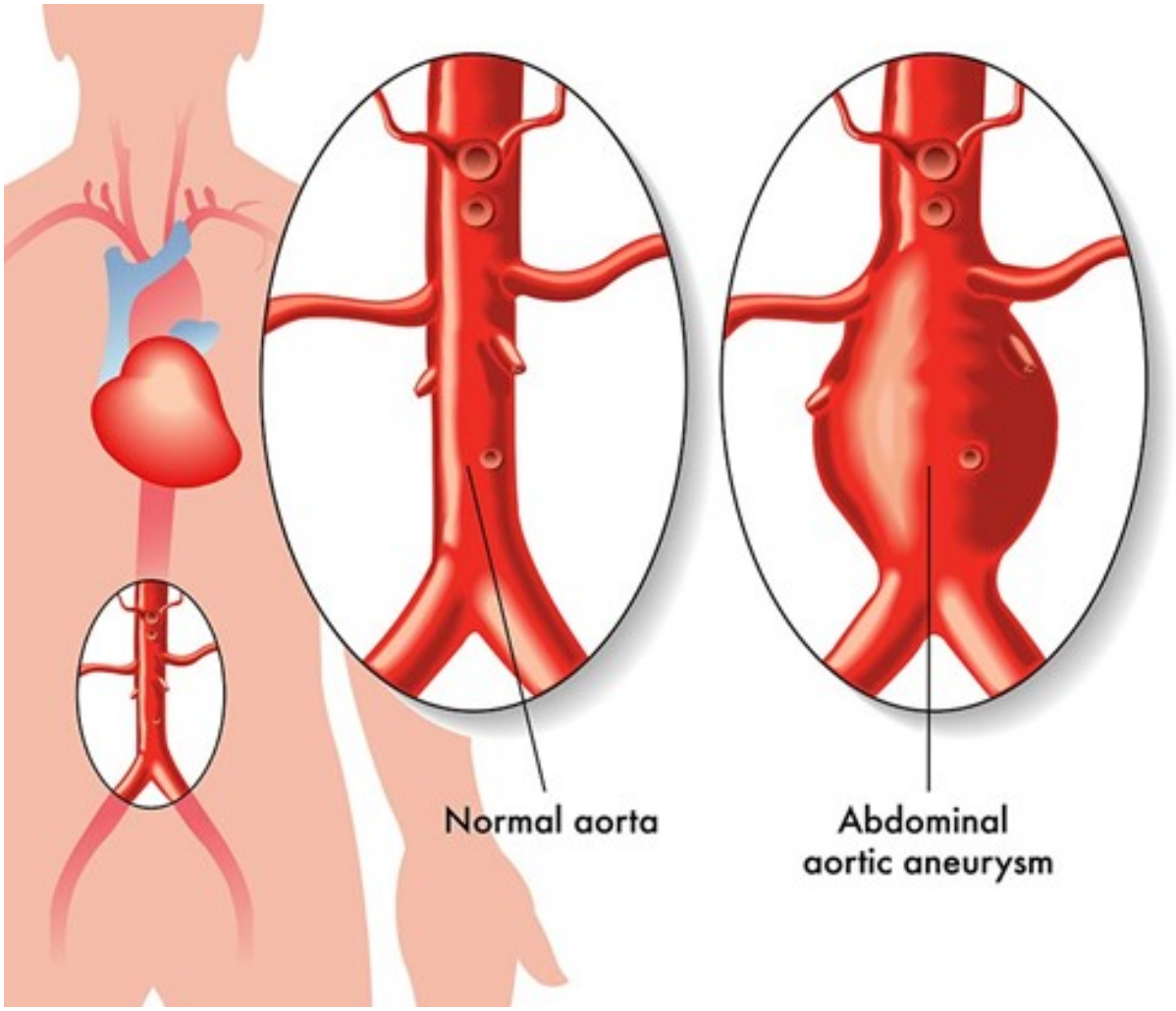


Figure 1.5: Comparison between a healthy and an aneurysmatic aorta. Image from [9].

Multiple factors can induce the formation of an Abdominal Aortic Aneurysm (AAA) and its evolution, so the aetiology is still not fully understood. They can vary from genetical factors, unhealthy lifestyle, bad habits like smoking and vitamin deficiency, hypertension,

age, and trauma [10]. It seems to be linked to other pathologies like diabetes mellitus, hypercholesterolemia, and obesity. Studies suggest that other possible causes can be infections and other pathologies of the vascular system and even hemodynamical factors. In the scientific literature AAA are classified as a particular form of atherosclerosis (or recently called atherothrombosis, to mark the strict link with atherosclerosis and formation of thrombus), a chronic disease that manifests during adulthood. Atherosclerosis is characterized by a chronic inflammation of the tunica intima on the inner wall of the artery that creates a set of multiple lesions that, if untreated, can bring to serious damages [11, 12]. Recent analysis have shown that AAA tissue, respect to normal aortic tissue, contains an excess cytokine, powerful regulator of the proteases, that have the capacity to degrade all components of the extracellular matrix in the arterial wall [13]. Proteases are enzymes responsible of the protein lysis that in normal conditions play the role of remodelling the tissues but, if in excess, can have a disruptive effect. Initiation of AAA is attributed to an excess of elastolysis to which follows a reduction of elastin that results in recruitment of inflammatory cells that raise the production of cellular proteases. This situation creates an immunoinflammatory chain reaction of biochemical factors that united to the constant flow of blood degrades over time the tunica intima and then the elastin in the tunica media and the collagen in the adventitia bringing eventually to the aneurysm's rupture [14, 15].

AAA often comes with two common factors:

- Thrombus: AAA are usually characterized by presence of intraluminal thrombus (ILT) that adhere along the wall reducing the lumen volume, in which the blood flows [16]. The two main components to a thrombus are aggregated platelets and red blood cells that form a plug, and a mesh of cross-linked fibrin protein that create a biologically active tissue. Thrombus in AAA can be caused both by an injury of the inner layer of the blood vessel, result of trauma, smoking and hypertension and hemodynamical changes in the blood flow like turbulences, or stasis generated by a deformation of the aorta, which greatly promote blood coagulation. The fully implications of ILT in AAA are currently unclear. Previous studies have demonstrated that ILT provides a biomechanical advantage by decreasing wall stress and acting like a mechanical buffer [17–20] , whereas other studies have associated ILT with aortic wall weakening [21]. ILT has also been shown to create an inflammatory environment, where given the high biological activity, neutrophils, cytokines, proteases are produced, and reactive oxygen species are sequestered creating hypoxia and increasing the inflammation even more. Given that seems that a great contribution to the protease activity originates from the ILT instead of being generated by the

AAA damaged wall.

- Calcification: another, less usual, phenomena that can verify within the aortic wall is the calcification, a gradual hardening of the soft tissue of the vessels caused by an accumulation of calcium salt precipitates. Calcification is one of the most common causes of arterial stiffening that comes with aging diminishing the aorta's elasticity [22–24]. This hardening can lead to a narrowing of the aorta impairing the regular haemodynamic and to congestive heart failure due to the consequent increased overload of work on the heart. As for the thrombus, clinical studies [25] are focused on the effect of calcification of the aorta, suggesting a correlation on AAA's rupture.

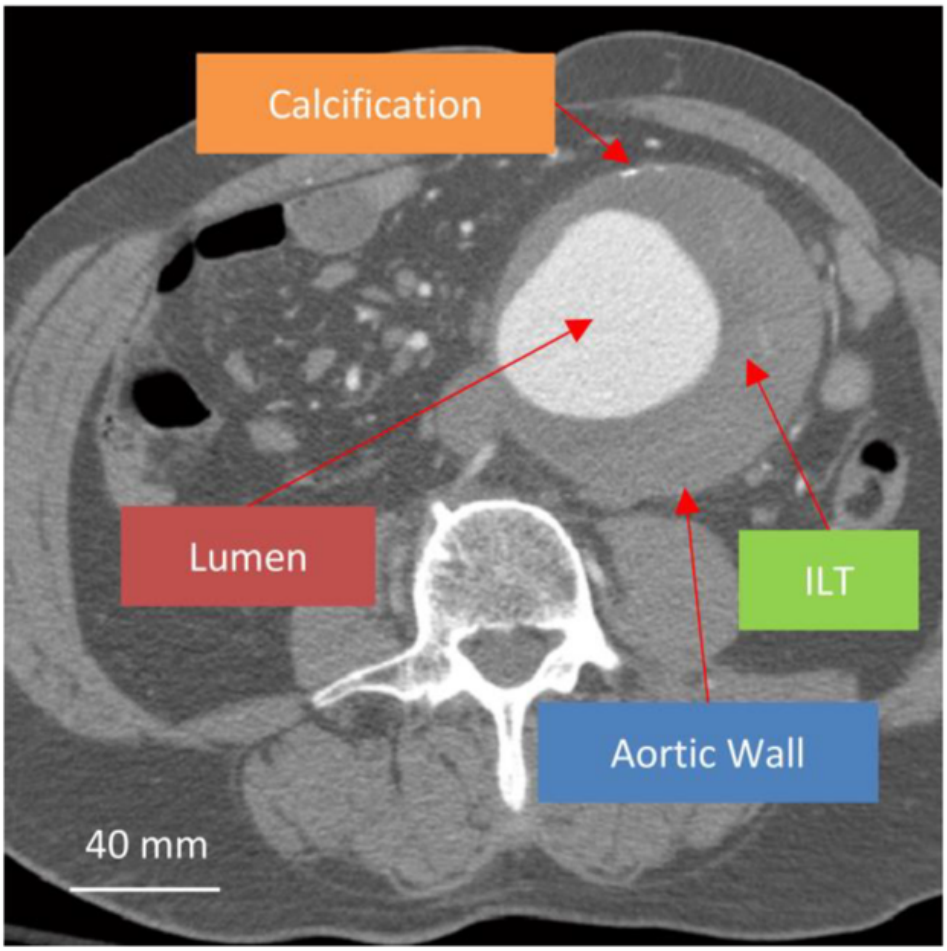


Figure 1.6: Calcifications adhere to the vessel's wall. Being made of calcium, if seen via Computed Tomography Angiography (CTA) scan they have similar brightness to bones. This can help the clinicians with the not trivial task of understand the real geometry of the AAA. Image from [26].

An AAA can have many different features and can be classified in different categories based on the aneurysm's conformation and positioning along the aorta. Concerning the shape, there are have mainly two different types of AAA: fusiform and saccular.

A fusiform AAA is recognisable by being a concentric bulge of the aorta, involving all the circular crown (all sides) of the artery. It is the most common kind between the observed patients.

On the opposite a saccular aneurysm, most common is the brain's vessels, is eccentric and more asymmetric compared to the former type. It involves only a local portion of circumference of the wall and is considered the most prone to rupture [27].

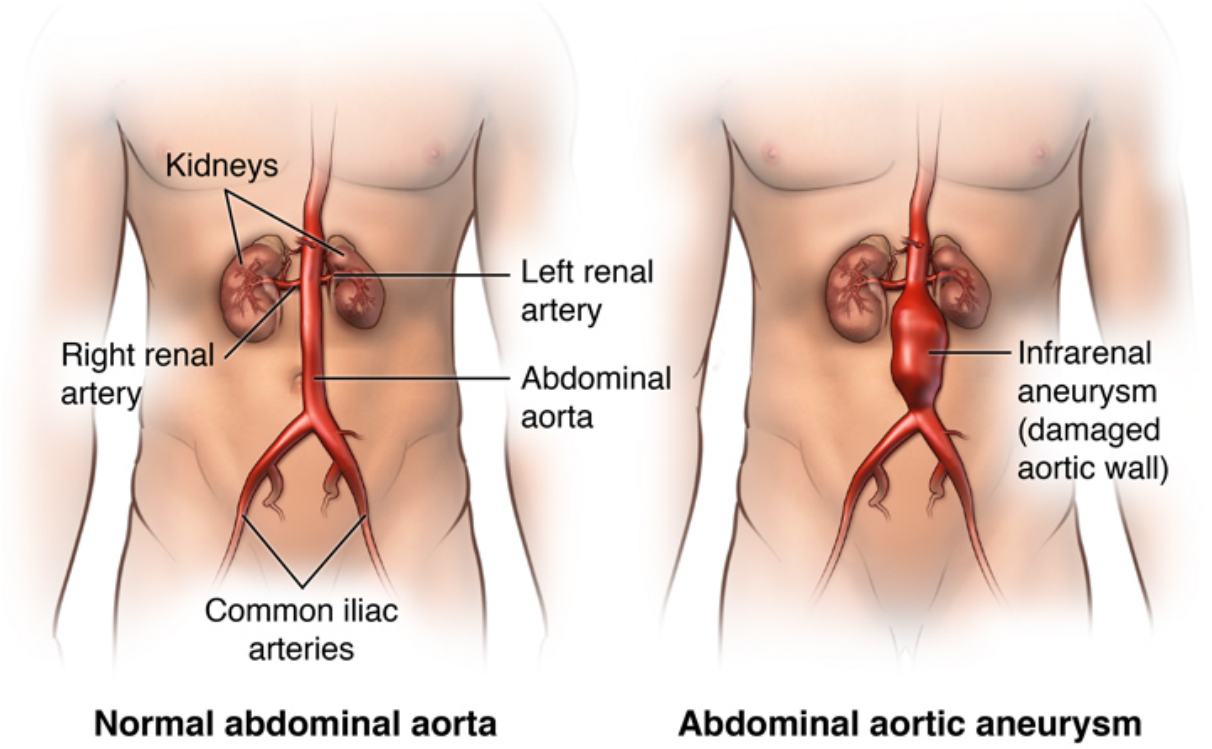


Figure 1.7: The two main types of AAA: fusiform and saccular. Image from [28].

A second distinction can be made by looking at the AAA position and divides the aneurysm in five classes. When the AAA involves the renal artery, it is called suprarenal or pararenal aneurysm, where in the former not only the renal arteries are involved but also part of the thoracic aorta. If renal vessels are not affected, we have instead juxta renal AAA if it stops at the renal height, infrarenal if there is a sane portion of aorta called *neck* long at least 10mm and a last group is the aortoiliac aneurysm if also the bifurcation or the iliac arteries are diseased.

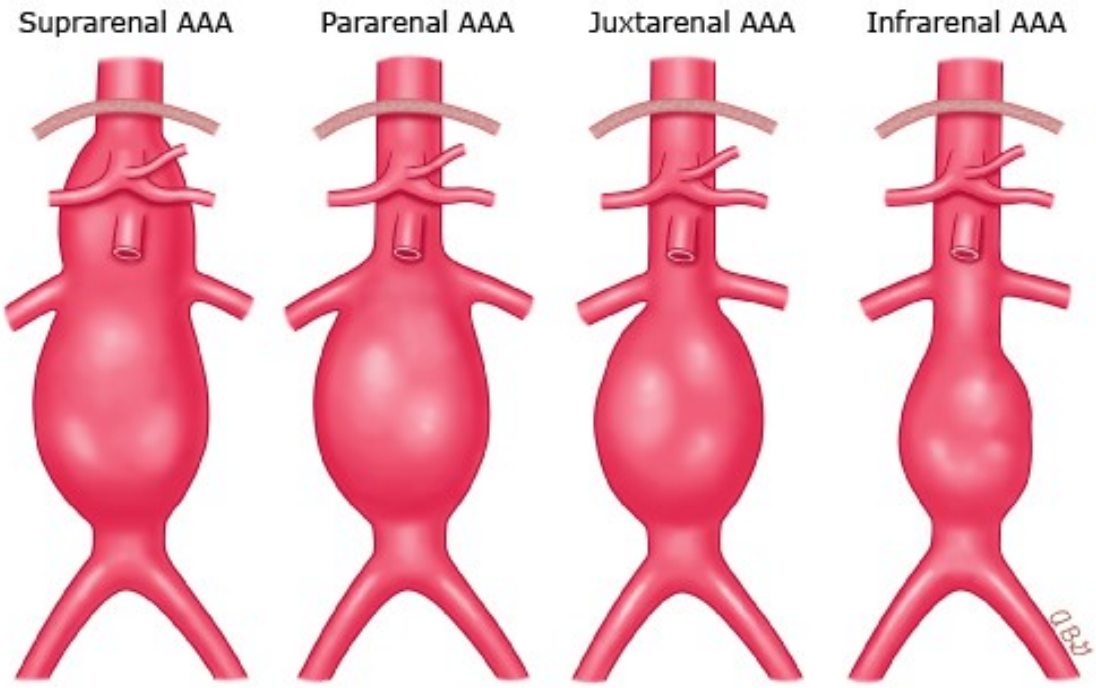


Figure 1.8: The four class of aneurysm. Image from [29].

1.1.3. Dangerousness and Repair

Cardiovascular pathologies, despite the great advances in diagnosis and surgery of the recent years, are one of the leading cause of death in the elderly population of the Western world [30].

Estimates state that each year more than 6 million of new cases are discovered in Europe and that more than a quart of them is related to the patients' death, costing over 210 billion Euros at year [31].

Aneurysm is the second most common disease of the abdominal aorta, and it can be found in around 3% of the population above 50 years, with an increasing incidence in older subjects.

One factor that contributes to make an aneurysm dangerous is its being asymptomatic. Even though AAA forms slowly, over the course of years or decades, this gradual expansion is usually asymptomatic, and its discovery and eventual diagnoses is often accidental, and occurs during medical check-ups related to other pathologies.

Rarely AAAs manifest with lower back pain, hypovolemic shock and in the most severe

cases can be recognised as a mass that pushes on the ventral area. Given that is not rare that an AAA evolves until its rupture making early discover and prevention very important. The mortality of patients with an aneurysm in rupture or already ruptured raises to 80% of the cases that either die before reaching the hospital or don't survive the surgery and is almost often fatal if the breach occurs frontally, causing a deadly haemorrhage.

Repair of AAAs is, over time, a necessary procedure since to carry out the surgery is the only solution to avoid the inevitable rupture if the patient does not die before due to other circumstances. The decision to carry out the operation is made weighing the risks and costs of the intervention and the dangers of the AAA.

At the moment, the main risk factor for rupture considered before the repair is the maximum diameter of the aneurysm. A patient becomes a possible candidate for the operation when its diameter exceeds 5.5 cm for males and 5 cm for females or there is a rapid growth rate.

Currently there two possible treatments for a AAA, Open Repair and Endovascular Aneurysm Repair (EVAR).

- Open repair: Open AAA repair is the standard surgical operation currently. Borne in 1950s it involves a highly invasive approach. The AAA is accessed through an incision in the patient's abdomen exposing the aorta. After clamping the aorta, the surgeon cuts the aorta at the aneurysm, removes the ILT and replace the damaged portion of the vessel by insertion of a graft, a fabric tube made of synthetic materials. In this way the graft forms a new artificial vessel inside the aorta. Even though open repair is a standardized treatment, it is associated with many complications, such as higher morbidity and mortality, higher lengths of stay, recovery period and blood losses, making this procedure not feasible for all patients, such as fragile or elderly people.
- Endovascular Aneurysm Repair: is a minimally invasive method for the treatment of AAA developed to reduce the risks associated with open surgery and to provide a treatment option for patients who are unsuitable to be surgical candidates. Invented in 1991 consists in using a stent graft, a collapsible graft that is inserted in the femoral artery through a small groin incision and guided into place with the support of X-rays to guide the graft into the site of the aneurysm.

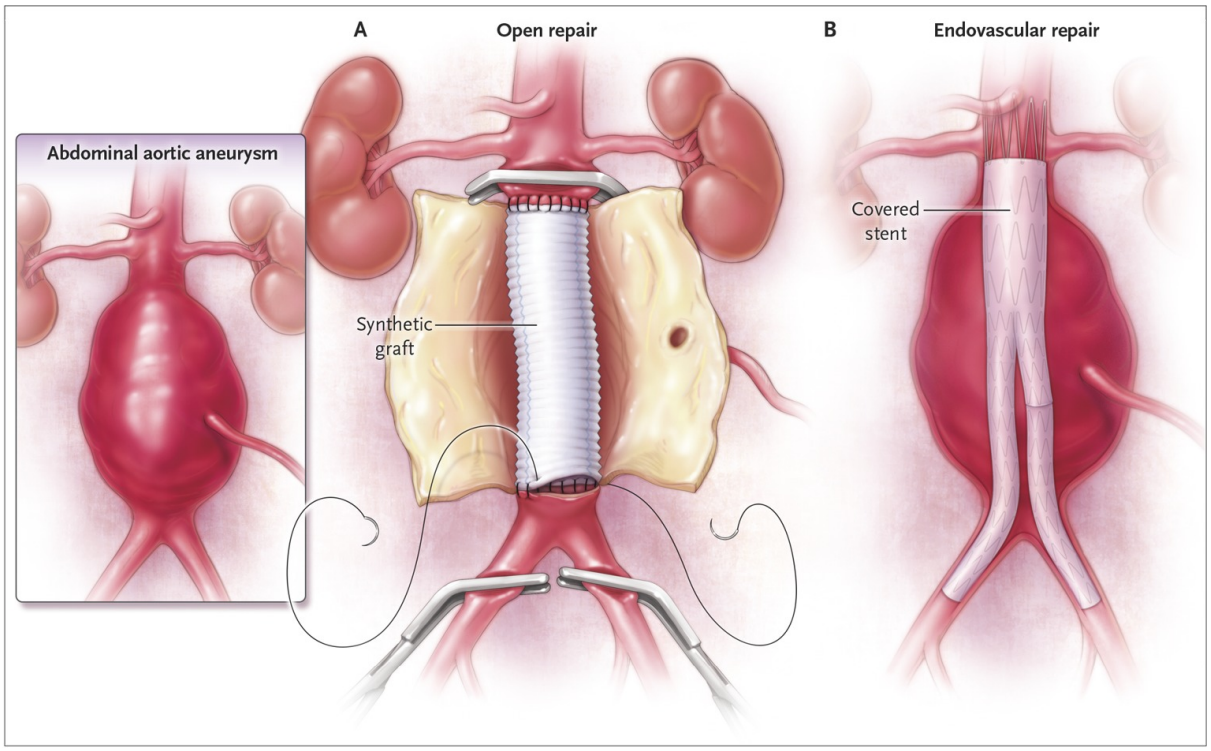


Figure 1.9: Comparative image of open repair and EVAR procedures. Image from [4].

The stent serves to anchor the graft in the blood vessel lumen and, in most cases, to provides structural support for the graft material which is made of PET or PTFE [32].

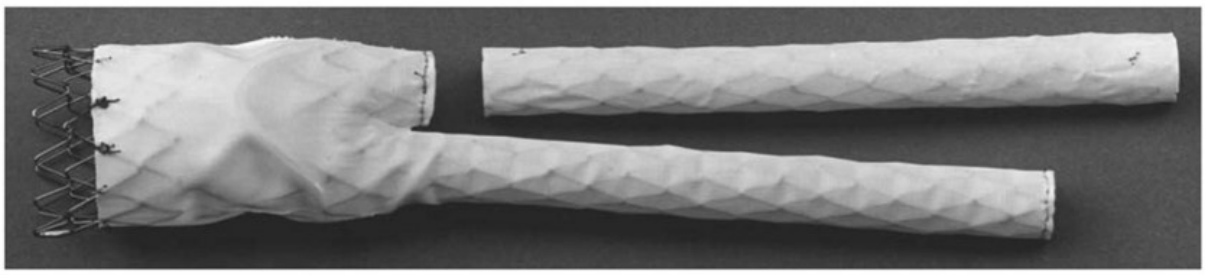


Figure 1.10: The two piece that compose the stent-graft. Image from [33].

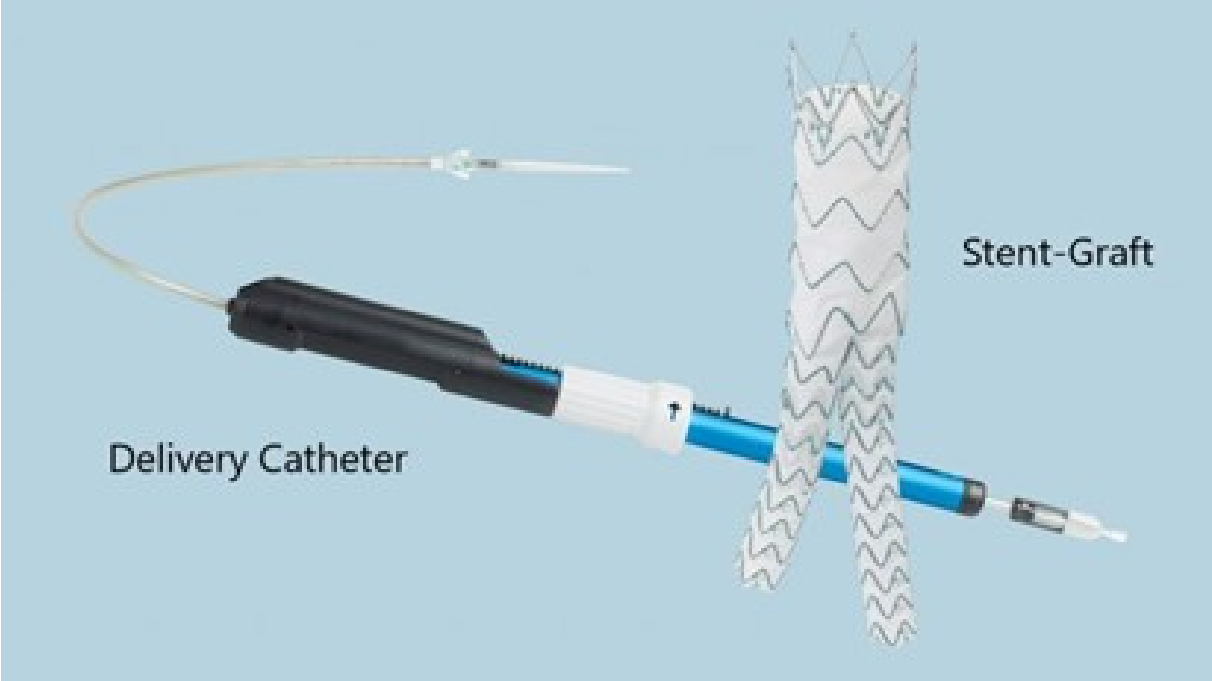


Figure 1.11: Stent-graft and delivery catheter. Image from [34].

After the positioning, the stent resumes its natural configuration due to a self-expandable construction morphology and its purpose is to exclude the aneurysm sac from the blood circulation (i.e., the lumen) carrying all the hemodynamic load preventing the rupture.

Despite being an innovative a minimally intrusive method, EVAR has some flaws when compared to open repair. First, not all the patients are suitable for this procedure, in fact EVAR to be properly placed needs an aorta with some characteristic. This comprehends non-acute angles between the aneurysm's neck and the AAA itself, a neck, and iliac arteries such as to create a robust landing zone for the stent to be placed and an aorta with generally regular shape. If these characteristics are not satisfied, there are two main complications: stent migration and endoleak. Stent migration happens when the stent fails to entirely adhere to the landing zone, so that the blood flow, gravity, and patient's movement lead to a displacement of the prosthetic downward, leading to a first type endoleak, the most severe one [35, 36]. Endoleaks are defined as the persistence of blood flow in the aneurysmatic sack despite the presence of a vascular prosthetic, failing the ultimate goal of the prosthesis itself.

It verifies in a range of 30%-40% of the cases and is often asymptomatic. Endoleaks may become evident both during surgery or even years after it, therefore lifelong imaging surveillance is necessary. There are 5 different types of endoleak and the causes are various. The type I endoleak is the most common and problematic one, the stent migration is

caused by an insufficient adhesion of the stent to the aortic wall.

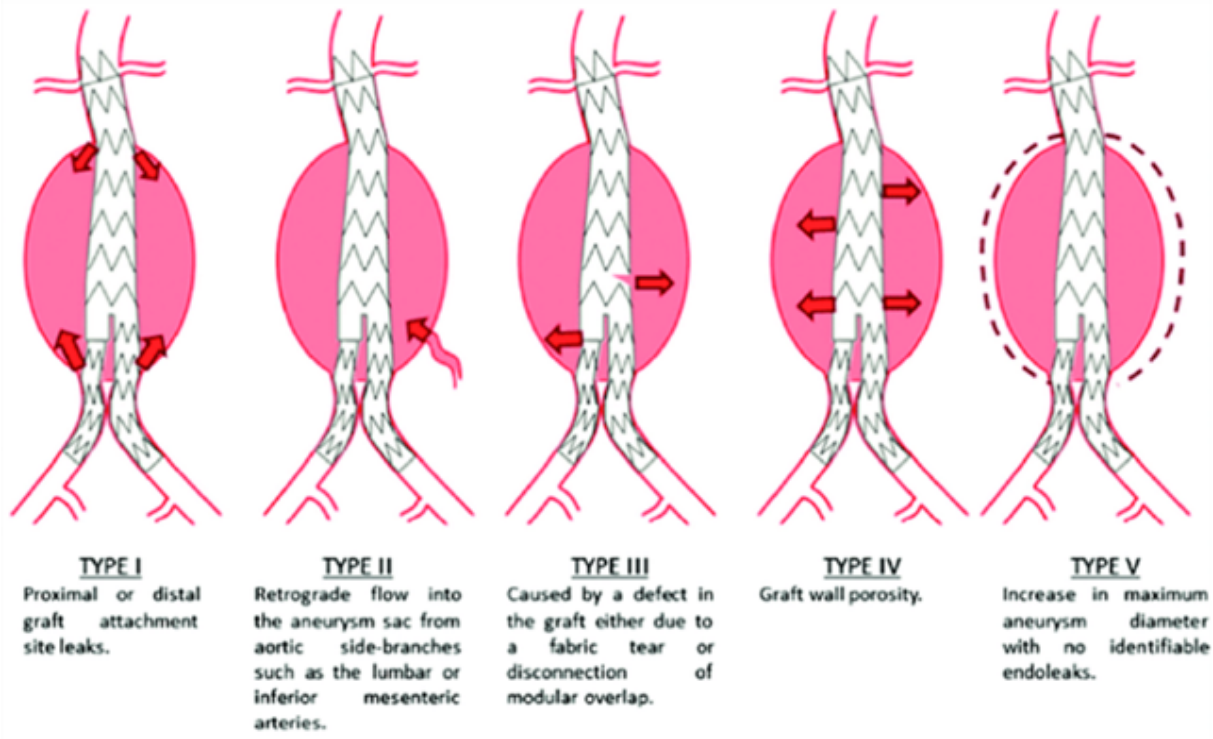


Figure 1.12: A brief presentation of the 5 types of endoleaks. Image from [37].

If not spotted in time endoleak can be as dangerous as not having a prosthesis, eventually bringing to the AAA's rupture and possibly the patient's death.

1.2. State of the Art

Although the AAA is still a not fully understood disease, there is a vast literature aimed to study it.

Between the quantity of scientific papers, the majority relegated to the medical and biological field, there is a slice of engineering and mathematical interest.

1.2.1. Clinical Studies

Scholarly efforts have been mainly addressed to the various clinical aspects that surround this pathology, as explained in the previous section, and can be breafly reassumed in:

- Main causes, such as genetic predisposition [38, 39], lifestyle and risk factors [40] like sex, age [41] or, smoking [42].

- Effects and causes of the presence of Calcification [43] and ILT [21, 44, 45] on the wall stress, growth rate and rupture. The scientific community still does not agree on the negative or positive effects of the ILT that, as explained earlier, could lower local wall stress or accentuate the inflammatory process, leading fastening the rupture.
- Surgery criteria [46] and comparison between EVAR and open repair techniques [47, 48] through mortality and secondary interventions. While EVAR seems to be the safest technique in term of survival, open repair may be preferred in patients who have particolarly advanced AAA and are deemed low risk.
- Post-surgical complications such as endoleak [49, 50] and mortality in patients with ruptured AAA [51, 52].

1.2.2. Studies of Engineering Interest

On the mathematical side, scientific literature can be divided into two groups: papers dealing with the computational side through mechanical and fluid dynamics studies and ones regarding statistical and machine learning studies.

Regarding the computational field, literature is focused on performing fluid dynamical simulation of the blood flow in a 3D digitalization of the aorta and computing indexes of clinical interest [53, 54]. One of the most important results of this studies is that indexes such as Oscillatory Stress Index (OSI) or Wall Sheer Stress (WSS) are found abnormal in aneurysmatic aortas, with a strict link between the aortic topology and the heamodynamical indexes. In particular it seems that high or low values of OSI are associated with local turbolences in the blood flow, bringing to an higher probability of ILT deposition, and thus higher inflammation.

Other papers are focused on estimate a rupture index, such as Rupture Potential Index (RPI). In fact, the rupture occurs when the wall tissue cannot withstand the locally acting wall stress exerted by the blood, making the time-average WSS and peak WSS some of the most important factors to keep in consideration [55–58]. WSS has also been used to prediction of the AAA's growth [57].

Even if this particular pathology is of special interest on the computational mathematical side, since the aneurysmatic aorta can be considered an object with a fluid flowing in it and thus model the problem using physical and mathematical laws, a smaller literature impronted on the statistical approach does exist.

The remaining slice of studies on AAA is focused on the use of machine learning and statistical tools to carry out the analysis and provide results. These groups seems to be

focused on two different aspects:

- Use of anatomical features, descriptors of the aorta such as AAA maximum diameter, ILT thickness and clinical risk factors such as sex, age and Body Mass Index (BMI) combined with classification and regression methods (SVM, KNN and trees) to predict AAA's growth, rupture [59–61] or, the surgery outcome and post-surgery mortality [62, 63]. These studies, despite being still in a preliminary state, agree with what discovered by clinical literature. In fact, maximum diameter and age seems to be positively linked to AAA's evolution and rupture.
- Use of classification techniques such as K-Nearest Neighbors (KNN) [64] or deep learning [65, 66] to perform a segmentation of the aorta using as training set manual segmentations. This is in order to speed up and improve the long process required for the segmentation, that is obtain a 3D reconstruction of the aortic vessel starting from the CTA scan (see Chapter 3 for additional details).

Statistical analysis seems to go no further than this point, without treating clustering or trying to use more complex data other than "traditional" numerical or global categorical descriptors of the aorta.

Despite the scarcity of statistical studies compared to other clinical applications, statistics and machine learning could be another way to study AAAs by succeding to manage an high number of data with a small effort in terms of time and computational power when compared to fluid dynamical studies, mainly based on onerous simulations.

This thesis fits in this context trying to fill the two gaps of the statistical approach on AAA: patient morphological aortic stratification and overcome of the global aortic descriptors in favor of other representations of the AAA.

In fact local descriptor of the vessel, as highlighted by the fluid dynamical studies where a strict connection between local geometry and WSS has been found, could enhance statistical analysis increasing the accuracy of results and helping to make light on this pathology.

2 | Morphological Clustering

2.1. Aim of the Project

The main scope of the initial part of this thesis is to understand if is possible to cluster and group patients using morphological features of the AAA. The final aims of clusterization are multiples and vary, the first is the finding of patients with similar aortic geometries, that could help and improve the making of ad-hoc prosthetics for that specific morphology. Other goals could be to help in providing classes of risk for clinicians or, together with fluids dynamic simulations, try to understand if is possible to link groups of patients to a possible range of fluid dynamical indexes or even the creation of an index of rupture's risk, useful for surgeons and different from what is already available in scientific literature.

2.2. Dataset Presentation

Given the difficulties faced in the search for data, a dataset available online was exploited. This dataset is the subject of study of a paper published on Nature Scientific Reports: "A Methodology to Quantify the Geometrical Complexity of the Abdominal Aortic Aneurysm" [67].

The article, published in 2019, aims to study and perform a sensitivity analysis on features regarding the anatomy of the aorta to help clinicians to decide whether carry out the EVAR procedure and avoid long-term complications, such as endoleak and stent migration, reducing the chance of secondary intervention to a minimum. This would lead to the creation of an extended set of geometrical features that, describing the AAA, could help the investigators to build a challenging case study to describe AAA's shape and create a protocol for the planning of EVAR. Clinicians need to know in depth AAA's geometrical data, since the feasibility of EVAR is bound not only to the landing zone's diameters but also to the angles that creates between neck and aneurysm and at the aortic bifurcation. In case of AAA's geometries that deviate from the instructions for use could led to unsuccessful clinical outcomes, obliging the surgeon to adopt alternative surgical strategies such as the more invasive open repair.

Despite this premise, few scientific papers discuss the clinical importance of angles or the way to correctly compute their values. This last concept is not trivial considered the anomalous geometry that an AAA can have and is a path less travelled. This happens since clinicians, if they decide to look at angles, stay on the bidimensional coronal or assail plane obtained via CTA scan, looking at a projection of the real angle that could result misleading. The dataset provided by M2S, Inc. West Lebanon, NH, USA that consisted of 258 patients, 222 (86%) of which were men, 34 (13%) were women and 2 (1%) were unspecified aged between 51 and 88 years old treated with Anaconda ™ stent graft. The numerical features were extracted from acrlongcta (acrshortcta) scans that were originally collected during the "Vascutek Anaconda stent graft system phase II IDE study" in the period May 2009 to July 2011 in U.S.A. and Canada. The inclusion criteria for the patients where:

- Infrarenal AAA greater than 4.0 cm in diameter, or growth rate greater than 1.0 cm/year.
- Iliac artery distal fixation sites greater than 20 mm in length.
- Ability to preserve at least one internal iliac artery.
- No AAA extended above the renal arteries.

Moreover, patients pregnant, patients with previous AAA repairs, leaking or with plaque, calcification, or thrombus thick more than 2 mm or with 50% or more coverage of the aorta's circumference were excluded.

Features can be distinguished in 4 different groups, after the exclusion of post-operatorial features that are not of interest for this thesis. The first group consists of 3 features: age, gender, and the combination with a thoracoabdominal aneurysm. Since the data were supplied anonymised, age and gender are the only descriptors of the patients. After the first analyses it was decided to discard this group of features to focus only on the morphology and shape of the vessels using these variables later in the result's interpretation phase.

The second group of features contains the dimensional features, all these lengths have been calculated cutting the aorta along one of the 3 main planes (coronal, sagittal and transverse) computing manually the distance between 2 chosen points. These features are:

- Aneurysm's neck length, i.e., the distance between the distal renal artery and the beginning of the aneurysm.

- Aneurysm's neck diameter on two different planes perpendicular to the centreline and their mean. Only the latter has been taken in order to avoid having too many correlated features. In fact, the 2 diameters have a Pearson's and Spearman correlation coefficient higher than 0.90 and having it higher than 0.97 when compared with the average value.
- Aneurysm's maximum diameter
- Distal diameter: the aorta diameter at the height of the aortic bifurcation.
- Length from renal artery to bifurcation: the distance between the bifurcation point and the distal renal artery.
- Aortic Volume
- Tortuosity index from aorta to left iliac artery, right iliac artery and their mean. As for the case of the aneurysm's neck diameter it was chosen to keep only the mean. Moreover, the difference between left and right tortuosity remains low and is never higher that 11% of the mean of the two which has a Pearson's and Spearman's correlation coefficient higher than 0.97 for both.

All diameters computed have been measured on CTA slices orthogonal to the centreline of the vessel and were calculated to the intima.

The third block of variables contains the normalized morphological features, such as the morphological features divided by the average neck diameter. These features are being considered to give importance not only to the raw number but also to the aneurysm's relative dimension respect to the patient using the average neck diameter, the initial part of the abdominal aorta still in health, as a baseline. In fact, despite the use of a fixed threshold usually considered by clinicians to decide whether operate a patient or not the aneurysm's relative dimensions to the healthy aorta are an important factor and should be considered.

The last category of features reported is consists of aortic and iliac angles. To describe the AAA morphology, a set of 10 different angles has been used to describe the pathological region, using the M2S v4.0.1 software. More specifically, 14 points were identified following the centreline of the vascular vessel as important to describe the AAA geometry. After a qualitative studying of a series of AAAs, these points were considered both strategic and adequate to draw the overall aneurysm shape. No points within the aneurysm are being considered since the endograft is unlikely to follow the angle generated with its insertion, making them irrelevant for the study considered.

P1	At the centre of the aorta, a few centimetres above the renals, usually at the superior mesenteric region.
P2	At the centre of the aorta, at the origin of the proximal renal artery.
P3	At the centre of the proximal renal artery, a few centimetres further from its origin (P2), at a suitable distance away from the junction.
P4	At the centre of the aorta, at the origin of the distal renal artery.
P5	At the centre of the distal renal artery, a few centimetres further from its origin (P4), at a suitable distance away from the junction.
P6	At the centre of the aorta, at the proximal end of the aneurysm.
P7	At the centre of the aortic bifurcation.
P8	At the centre of the right common iliac artery, at the most acute curve point.
P9	At the centre of the right femoral bifurcation.
P10	At the centre of the right external iliac artery, a few centimetres further from the bifurcation.
P11	At the centre of the left common iliac artery, at the most acute curve point.
P12	At the centre of the left femoral bifurcation.
P13	At the centre of the left external iliac artery, a few centimetres further from the bifurcation.
P14	The midpoint between points P2 and P4.

Figure 2.1: Listing and summarizing the 10 points used in [67].

For each of the patients in the dataset the 14 points have been selected manually on the CTA scans. Afterwards, using these points on a MATLAB algorithm 10 angles were computed in 3D with their projection on the coronal, sagittal and transverse planes.

More in details the 10 angles reported are:

1. Aorta-Right Renal Artery
2. Aorta-Left Renal Artery
3. Aorta-Neck
4. Neck-Body
5. Body-Right Common Iliac Artery
6. Right Common Iliac Artery Curve point
7. External Right Iliac Artery
8. Body-Left Common Iliac Artery
9. Left Common Iliac Artery Curve point
10. External Left Iliac Artery

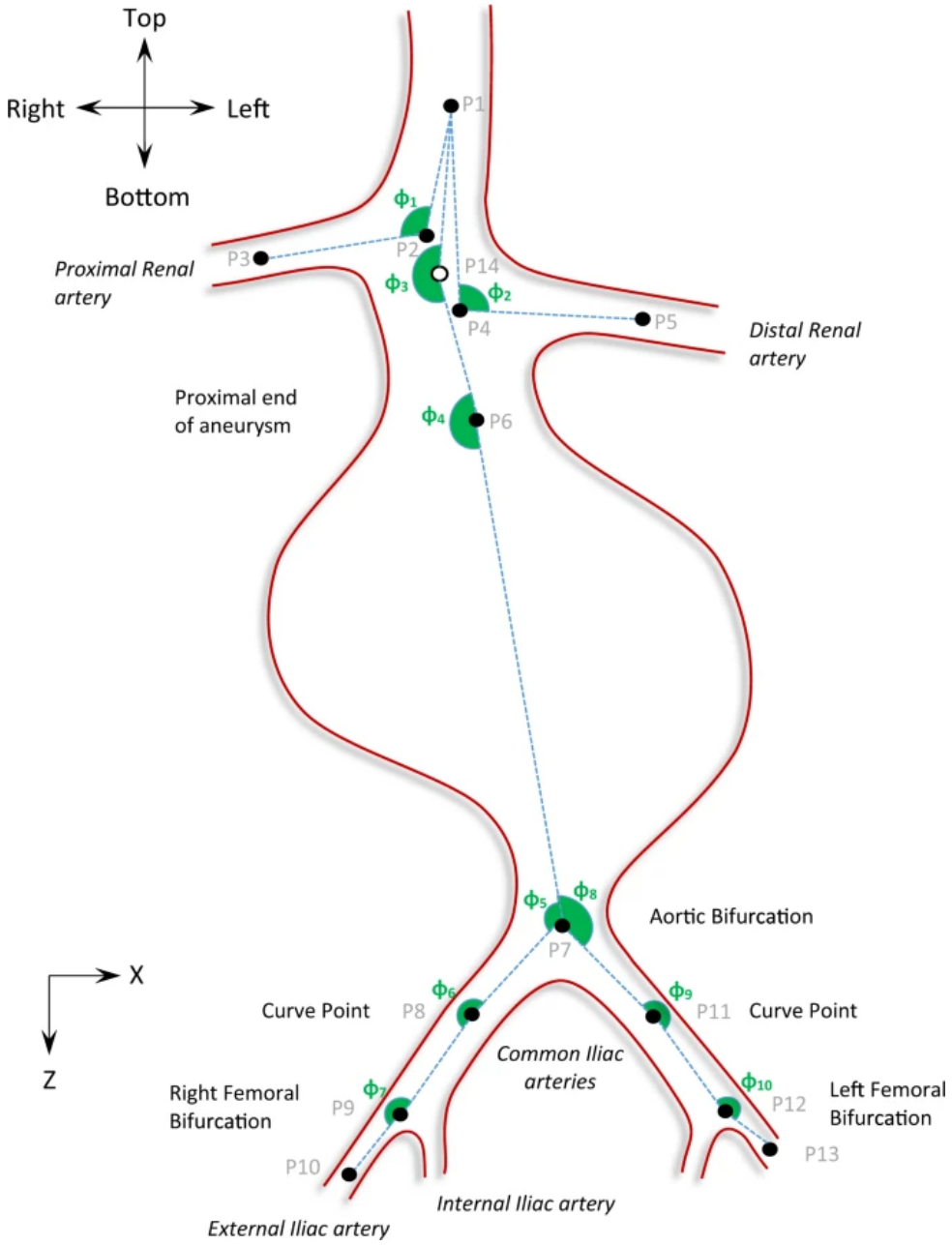


Figure 2.2: Points and angles used in [67] visualized on the coronal plane.

The main benefit of this procedure is given by the easiness of the identification of the points if interest on the CTA scans. In fact, rather than measure angles directly, a task that can be hard in space or when geometries are not ideal, the authors limit themselves into identifying material points reducing at the minimum the interpretation error.

This procedure, however, brings not only benefits but also disadvantages. First, the points are taken manually, giving more freedom to the authors at the cost of the introduction of the human error. The second, more important instead lies in the fact that for odd shapes

of the aortic vessel, common in patients with an aneurysm, the 3 points used to generate an angle can be too distant from each other and the angle can often be not representative of the real structure of the aorta.

In addition to the 10 angles mentioned above, the projection on the 3 main planes (coronal, sagittal, transverse) is available for a total of 40 angles: 10 in the 3D space, 30 in 2D.

2.3. Data Cleaning and Preprocessing

Despite the use of projected angles could have been useful and interesting since the clinicians are used to look at the CTA scan on the three main planes, it was decided to discard these 30 features. In fact, while the 3D angles have a standard deviation that varies from 18 to 30, projected angles can have a standard deviation up to 112, with a minimum of 0° to a maximum of 359° in some cases. This happens in the majority on angles on the transversal and sagittal plane. This is caused since a very small shift of one of the 3 points, especially the median one, can completely change the angle projected.

This variability is also highlighted by the authors of the article in a boxplot containing all the 40 angles.

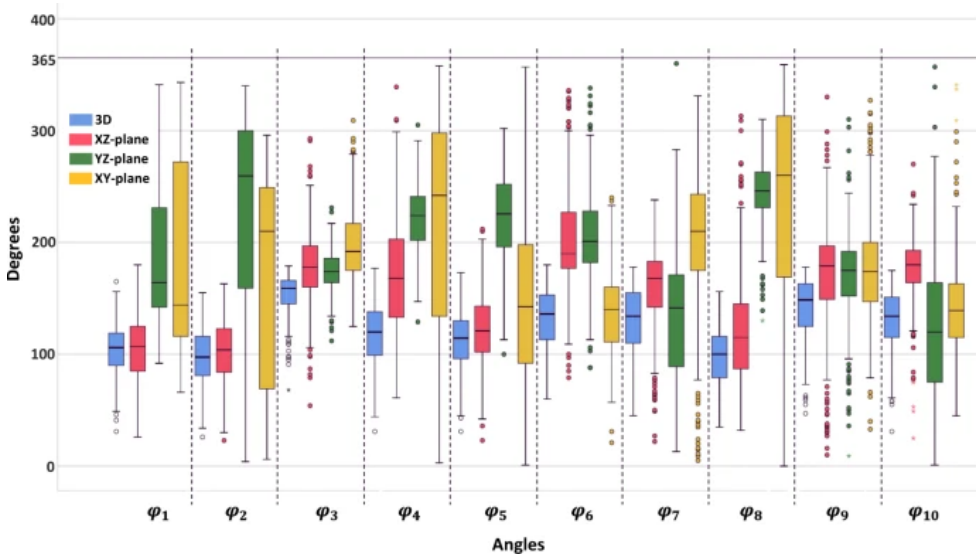


Figure 2.3: By this graph, shown in [67], is possible to see how certain angles projected in a specific plane maintain relatively close range and value to their 3D counterpart while on other planes loose robustness. It would have been possible to apply sinusoid transformations to some angles or subtract a certain value (180°) when the feature reaches a determined range.

It was also chosen to discard all the 2D angles from this analysis since it would bring to have a great number of features that are not representative of the real angle being not informative and introducing noise while having the original and robust variable available in the dataset. Moreover, the few projections that conserve robustness are very similar to the original angles, as can be seen with the Renal angles having a correlation higher than 0.8.

The only exception to this is the projection of the bifurcation angle on the coronal plane. This angle is not directly present in the dataset, so it was necessary to create it with the use of the data discarded. This decision was made since many studies, like [68] and [69], claim that complexity of the geometry, especially at the bifurcation, can bring to a disturbed blood flow resulting in disorganized, irregular secondary flow patterns that alters the aorta's normal functioning.

The coronal projection of angle ϕ_5 and ϕ_8 made possible to reconstruct the wanted variable with the formula: coronal bifurcation angle $= 360 - \phi_{5_{Cor}} - \phi_{8_{Cor}}$.

The choice of using the coronal projection was instead made since it's not possible to reconstruct the real 3D angle, since two 2D angles are not necessary in the same plane. Moreover, the coronal projections of the common iliac angles does not seem loose consistency and for this reason this angle should be a fair approximation of the real one.

After the removal of the missing values and the selected features the dimensions of the dataset used are 244 observations and 28 features. Data cleaning was not necessary since there are not outliers and the typos were already deleted by the authors of the article. The last part of the data pre-processing step is the data scaling. Since all the clustering algorithms used rely on the distance in order to group data in the same cluster so to give the same importance to all the variables, that have different unit of measure it was required to scale the data. The scaling was performed on *R* version 4.2.0 using a Z-score normalization: $\hat{x} = \frac{(x-\bar{x})}{\sigma}$, where x is the original feature vector, $\bar{x}$ is the mean of the feature vector and σ is its standard deviation. In this way the data have zero-mean and unit-variance and are ready to be used for the clustering.

2.4. Clustering Algorithms

For all the trials three different distances were used: Euclidean ($L^2 : d(x,y) = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2}$), Manhattan ($L^1 : d(x,y) = \sum_{i=1}^{p}|x_i - y_i|$) and Canberra (Weighted $L^1 : d(x,y) = \sum_{i=1}^{p}\frac{|x_i-y_i|}{x_i+y_i}$) since there all the variables are numeric. All the analysis and plots have been realized with *R* version 4.2.0 (2022-04-22 ucrt) and *Python* 3.9.9.

2.4.1. DBSCAN

The first clustering algorithm used is called Density-Based Spatial Clustering of applications with noise, also known by the abbreviation of DBSCAN and is one of the most cited algorithms in scientific literature. This popular method is a non-parametric algorithm that given a set of points in space groups together points that are close together while marking as noise points the observations that are far from high-density regions [70].

More precisely DBSCAN has two different parameters, ε and *MinPts*, and divides observations into *core* and *reachable* points. Points are core if there are at least $MinPts$ within a distance of ε from it and are reachable if are within a distance of ε from a Core point. If a point is a Core Point it creates a cluster with all the points that it can reach, on the other side if a point does not belong to a cluster is a noise point and a possible outlier. Usually, smaller *MinPts* and ε bring to more but smaller clusters, on the opposite side having a big *MinPts* and ε can bring to have only one or even no clusters.

To conduct this clustering the library *dbscan*, version 1.1-10 in *R* were used.

For each of the three distances a thick grid of 1000 elements for ε cointaining values from 0 up to 50 with an increment of 0.05 combined with a grid of *MinPts*, which contains all the natural numbers from 1 up to 50, was exploited. Even if, generally, *MinPts* should be greater than or equal to the features' number of the dataset it was decided to cover all possible combinations to avoid missing groupings.

To manage this many combinations of parameters the results were collected in one table (1000x48) for each distance and consider only the clustering that reach at least two groups with at least 10% of the patients each (25 observations) and less than 25% (61) of the noise points. This is done to avoid having a big part of the dataset considered only noise, since a pre-processing and outliers' analysis has already been made by the authors of the article, and to avoid having the patients distributed in an high number of non-significant clusters.

Using Euclidean distance, the algorithm does not create any clustering that satisfy the constraints. The same is true when using Manhattan and Canberra distance.

It was also made certain that for all the distances the ranges of the parameters cover both opposite cases, from only noise points (small ε and *MinPts*) to having only a cluster with few noise points (high ε). Given that, it was opted to not rely on DBSCAN for the analysis and use other clustering methods.

2.4.2. Hierarchical Clustering

The last algorithm exploited is agglomerative hierarchical clustering. This family of algorithms groups data into a tree of clusters. The algorithm starts treating all the observation points as different clusters and iteratively identifies the 2 closest clusters and merges them together. The algorithm ends when all the data are grouped in only one single cluster [71].

The output of this method is a dendrogram, a branching diagram representing the relationship of similarity among the clusters generating the tree of clusters mentioned above [71, 72]. Each branch represents a cluster, and the n terminal ends of each branch are called leaf and represents the singular observations.

The arrangement of the branches tells us which leaves are most similar to each other. The height of the branch points indicates how similar/close the clusters are from each other.

Hierarchical clustering does not utilize parameters but is necessary to define the distance and the linkage, which is the criterion by which the data are agglomerated.

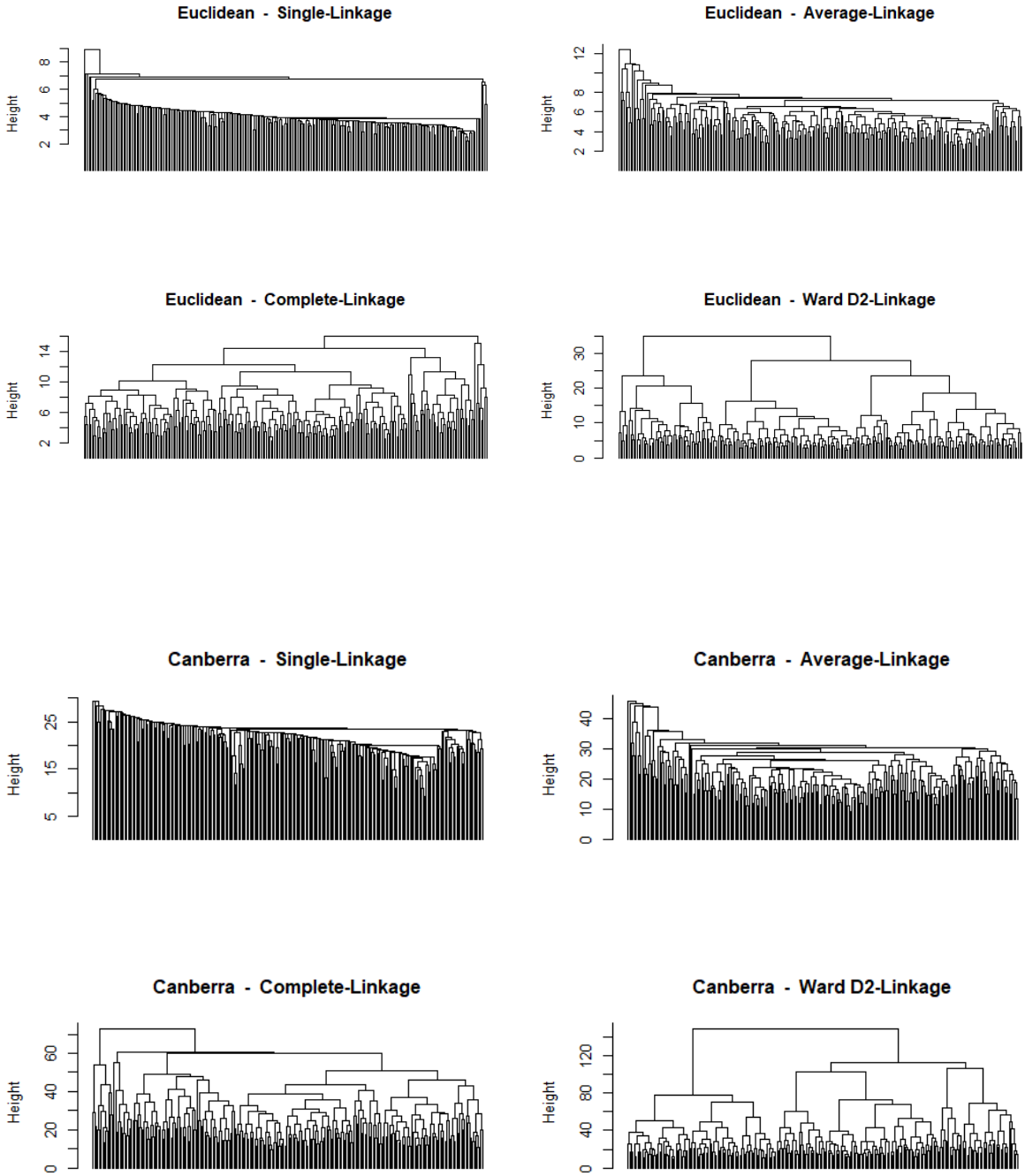
It was chosen to use four types of linkages for this analysis: single, medium, complete and ward D2 linkage.

- Single linkage: given the distance d, combines the two clusters X and Y that have the minimum distance $D(X,Y) = \min_{x \in X, y \in Y} d(x,y)$, i.e., that contain the closest pair of elements not belonging to the same cluster.
- Complete linkage: given the distance d, it combines the two clusters X and Y that have the minimum $D(X,Y) = \max_{x \in X, y \in Y} d(x,y)$, checking not the closest elements but the farthest.
- Average linkage: given the distance d, it combines the two clusters X, Y that have the minimum $D(X,Y) = \frac{1}{|X||Y|} \sum_{i=1}^{|X|} \sum_{j=1}^{|Y|} d(x_i, y_i) \quad \forall x_i \in X, \forall y_i \in Y$, i.e., the average pairwise distance between all possible couples made of elements of X and Y.
- Ward D2 linkage: Also known as Ward's minimum variance method, the groups are merged minimizing the growth of a given objective function, representing the dissimilarity within clusters.

When the distance used is Euclidean the objective function considered is the ESS, error sum of squares, to have the minimum information loss [71].

This clustering has been performed using the function *hclust* in the stats package version 4.2.0 on *R*.

To start the analysis, to visualize the results the dendrograms for any of the twelve combinations of distance and linkage were computed and plotted obtaining the following charts (Fig. 2.4):



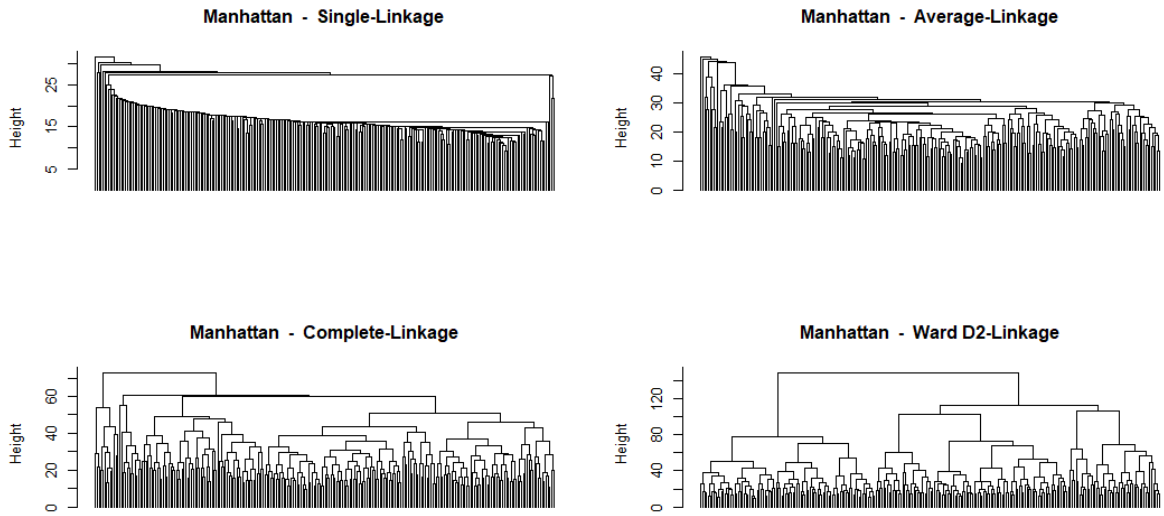


Figure 2.4: The dendrograms relative at the 12 combinations of distances and linkages. The method generates the same dendrograms Canberra and Manhattan distances with all linkages, the only exception being the single linkage.

In all these plots is clearly visible the main drawback of single-linkage, the tendency to create a "chaining effect", that verifies when big clusters are merged with single observations in the last iterations. Even if this linkage allows the detection of irregular and elongated shapes but in this case would bring to imbalance, with one set containing a big percentage of the data along with singletons.

The same effect is visible, in a less but still pronounced degree, in the dendrogram resulting from average-linkage.

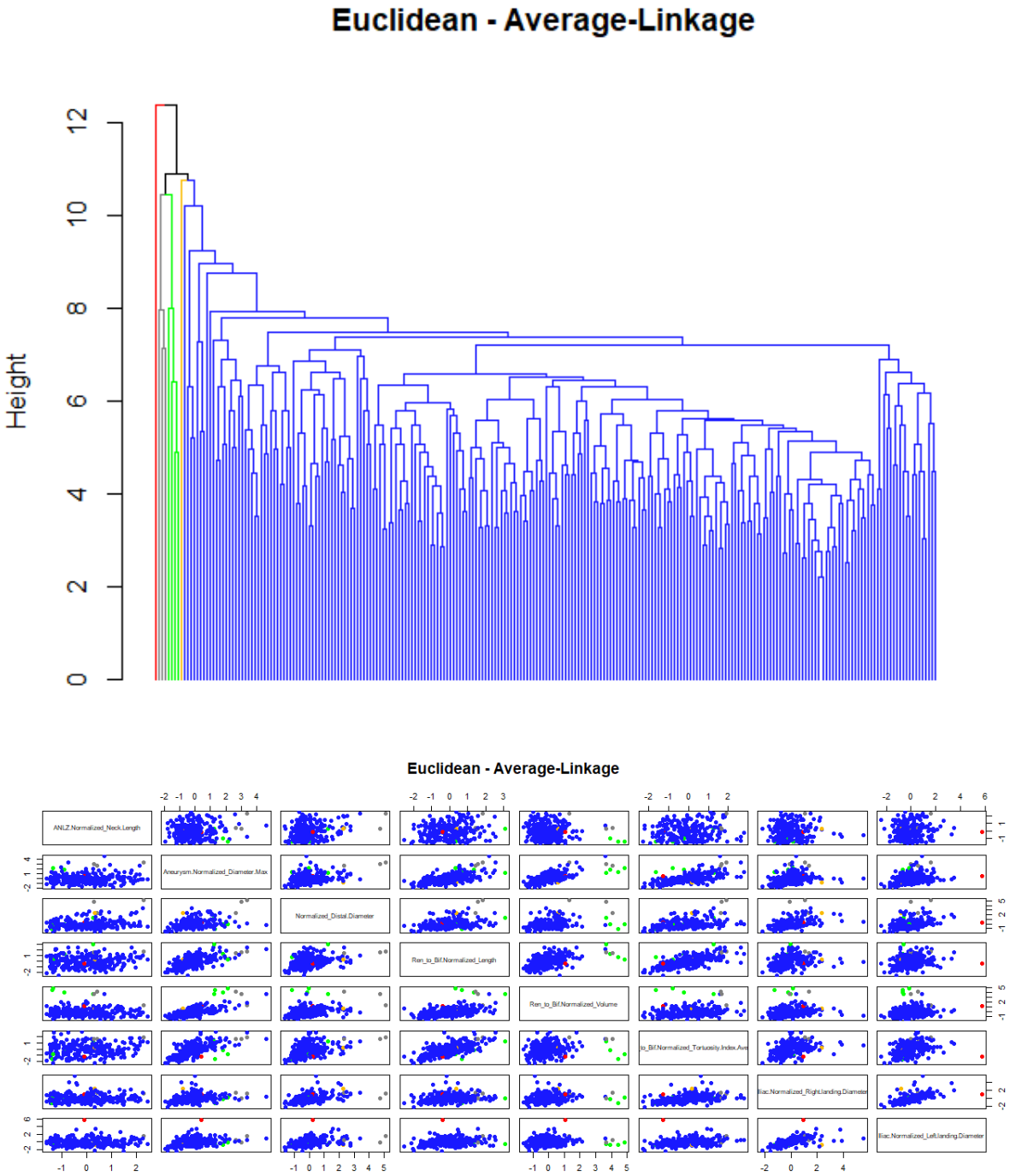


Figure 2.5: In this scatterplot can be seen that the singletons left alone could also be considered outliers. However, it was chosen not to discard them since they seem to be the most concerning patients, as can be seen by looking at the normalized volume of the patients coloured in green and grey compared to the other points.

For these reasons both single and average linkages were discarded from the analysis. Since

the most balanced clustering seems to be obtained using Ward D2 linkage, which generally produces the better results respect to the other linkages considered here, the efforts were focused on it [73].

The first thing that can be seen by looking at the dendrograms is that the plots generated by using Manhattan and Canberra distances are identical for most of the linkages, including Ward D2 Linkage. The next step of the analysis is to choose at which height cut the clusters, that determines the number of clusters K.

This step is crucial since cutting too early can bring to many non-significant clusters, while cutting lately can bring to a loss of information. it was decided to plot the difference of height at which the groups are merged, higher it is the more distant (i.e., different) the clusters produced are from each other. A common method to choose K is to look at the interval of "resolution", or height, in the dendrogram in which is possible to see a certain number of clusters. The wider the interval is, the more the algorithm suggests the associated number of clusters.

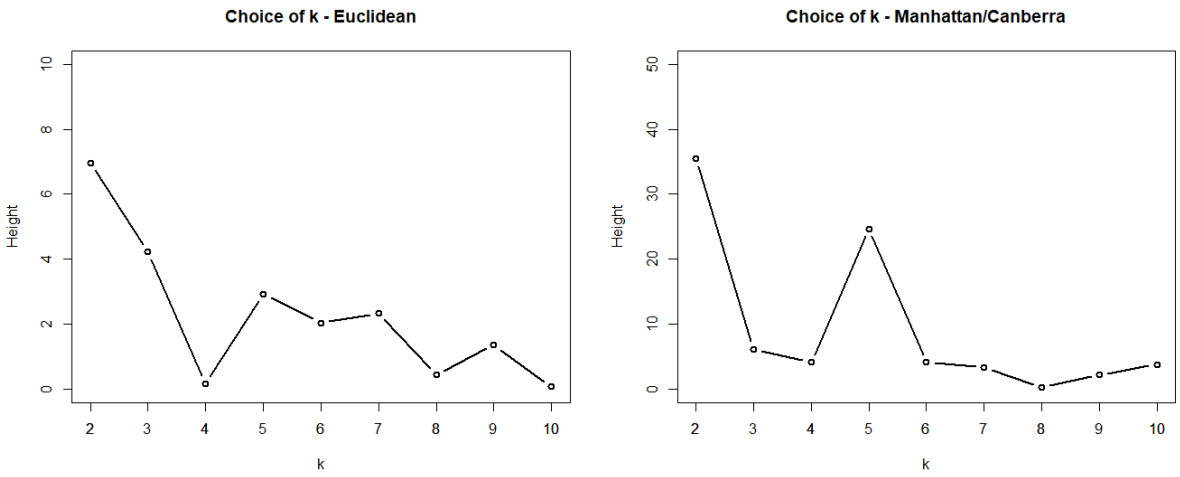


Figure 2.6: Here are the interval resolution for K between 2 and 10. Other values of K are discarded in advance since would bring to too many clusters for the numerosity of the dataset used.

By looking at the plots of the three distances is possible to see two main peaks at two values of K: 2 and 5.

In Figure 2.7 it is possible to see dendrograms and the scatterplot of the normalized features for K=2

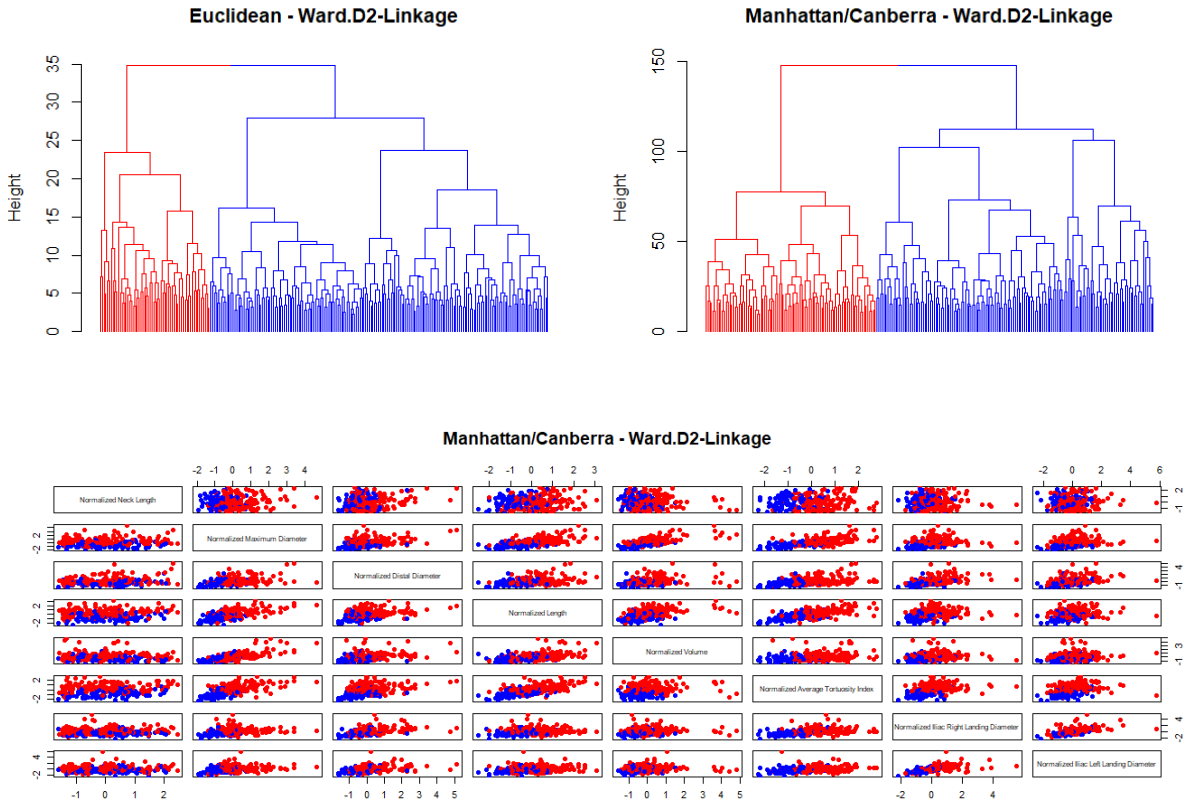


Figure 2.7: Dendrograms and pair plots produced by Single-Linkage.

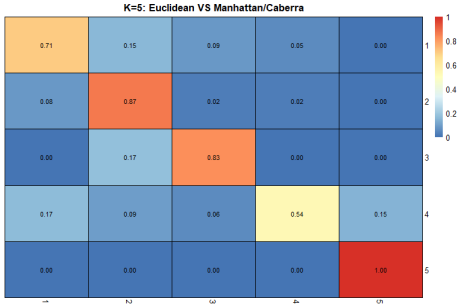
In both cases the patients seem to be divided, even if in different ways, mainly through the normalized features with a group of relatively big aneurysm and a relatively smaller one. This can be appreciated in the image above. The tendency to primarily differentiate the patients respect to the normalized aneurysm's dimension rather than respect to the regular ones could be an important indicator to start using these features also in the practical field.

It seems that the next most interesting number of clusters is obtained for K=5. Analysing this case, it will be also expose the case with K equal to 3 and 4, while bigger values of K would divide the dataset with smaller and less relevant groups.

Comparing the groups generated by Ward D2 linkage using the three definitions of distance mentioned above is possible to notice that, when the cut is performed at K=5 leading essentially the same clusters, with small variations. This suggest that this particular choice of clustering is, in a certain degree, robust to a chance of distance, since more than 74% of the patients are clustered in the same groups, an high value considering that five classes are used.

Euclidean \ Manhattan/Canberra	1	2	3	4	5
1	55	12	7	4	0
2	7	72	2	2	0
3	0	4	19	0	0
4	9	5	3	29	8
5	0	0	0	0	6

(a)



(b)

Figure 2.8: Here it is possible to see two tables. The first one is the confusion matrix with the cardinality of each combination of groups. If in the former is possible to appreciate the similarity of the two set of clusters, this is even more evident in the latter, where the relative heatmap with values standardized by rows can be seen.

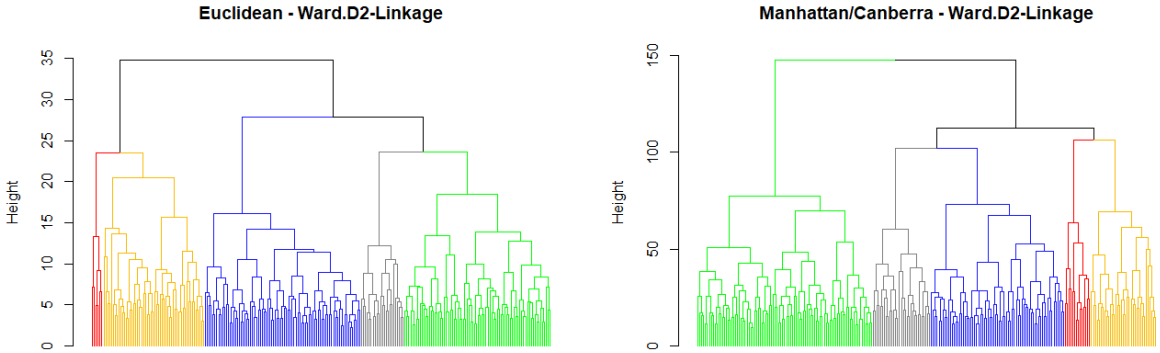


Figure 2.9: Here are instead the dendrograms with clusters coloured in pairs to match the corresponding group with the different distance. The main difference in the two clustering for K=5 is interestingly not much visible in the resulting groups as it is in the height at which clusters are merged. This is another factor that supports this choice of K, since for another value the clusters would be mixed, losing the strong similarity.

To evaluate how good was the algorithm to retrieve information from the data provided to it, it is possible to use the cophenetic matrix. It compares the matrix containing the distances at which two units are joint together with the dissimilarity matrix.

This evaluation has nothing to do with the number of clusters and the clustering structure in general, it only estimates how the hierarchical algorithm is good in translating the original distance matrix into the cophenetic matrix.

The higher the coefficient, the more this method indicates that one clustering is better than another. In particular, for Euclidean distance the cophenetic coefficient is 0.429 while

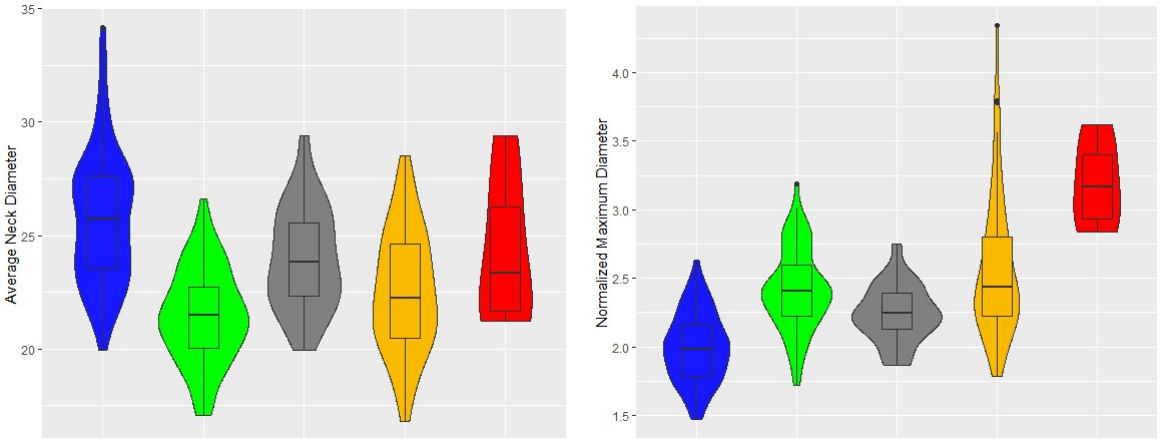
for Manhattan and Canberra is 0.291. Since a higher value of the cophenetic coefficient is obtained using L^2 distance, the following analyses have been made using this distance to quantify the dissimilarity of patients.

2.5. Clusters Interpretation

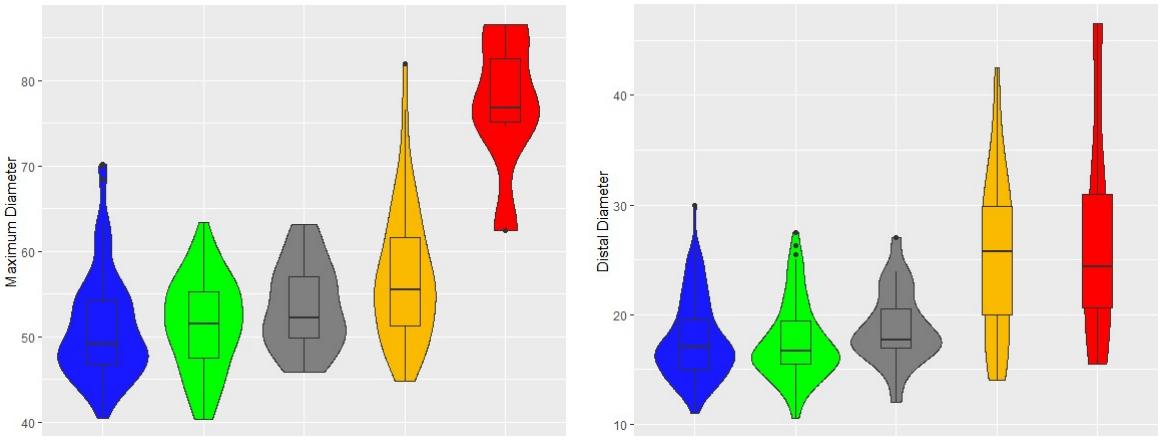
The groups found have been, after an interpretation phase with the clinicians, named based the expected severity of AAA. The five colors are Blue, Green, Grey, Orange, and Red.

Blue, Green, and Grey clusters are characterized by smaller aneurysm, with less than 50% of patients in all groups that would undergo to a surgery according to the criteria currently used in most of the hospitals of the world (maximum diameter higher than 55mm for males and higher than 50mm for females), while Red and Orange can be classified as bigger aneurysms with respectively 60% and 100% of patients in them.

The main difference between Blue and Green cluster can be seen not in AAA's dimensions but in the neck's shape. If Blue cluster is characterized by a long and wide neck, the Green one has a thinner and shorter neck. This reflects also into most of the normalized features, which are noticeably smaller for the blue cluster.



(a) (b)



(c) (d)

Figure 2.10: Here are four Boxplot and Violin-plot to better see the difference in distribution of the data in each group.

Even if neck is not directly involved in the Aneurysm its dimensions are important in order to decide whether or not a patient can be operated using EVAR technique. If the neck is too short or narrow the stent in the aortic prosthesis may not adhere properly to the vessel wall, resulting in post-surgery complications.

Blue cluster is the biggest set found and is composed by 83 patients and it can be considered the less concerning group with only 24% of the patients that satisfy the criteria for an immediate operation, on the other hand the Green cluster is the second largest group containing 78 patients. It can also be noticeable that the patients that constitutes the Green cluster are for more than 25% women, due to the neck's narrowness, and it includes the majority of the total women in the whole dataset. These two clusters have also a low tortuosity index, which contributes to make them preferable by the medics for the EVAR operation.

Orange and Red clusters are instead differentiated by the AAA dimensions rather than by the aortic neck.

In particular the patients belonging to the Orange cluster seem to have a bigger distal and iliac landing diameter and this reflects also into the normalized variables. A high distal and normalized distal diameter could mean that the aneurysm involves the aortic bifurcation and the iliac arteries. Again, even if the iliac arteries are not part of the abdominal aorta is important for the surgeons to know if the AAA is combined with an iliac aneurysm to avoid post-operatorial complications.

The Red cluster, composed of only 6 patients seems to collect observations that could

be considered outliers, with tortuous arteries with incredibly high aneurysm's maximum diameter and volume, as can be appreciated in Fig. 2.10. The totality of this group of patients would immediately undergo to surgery, having a mean maximum diameter of 77 cm and a normalized one more than 3 times bigger than the neck.

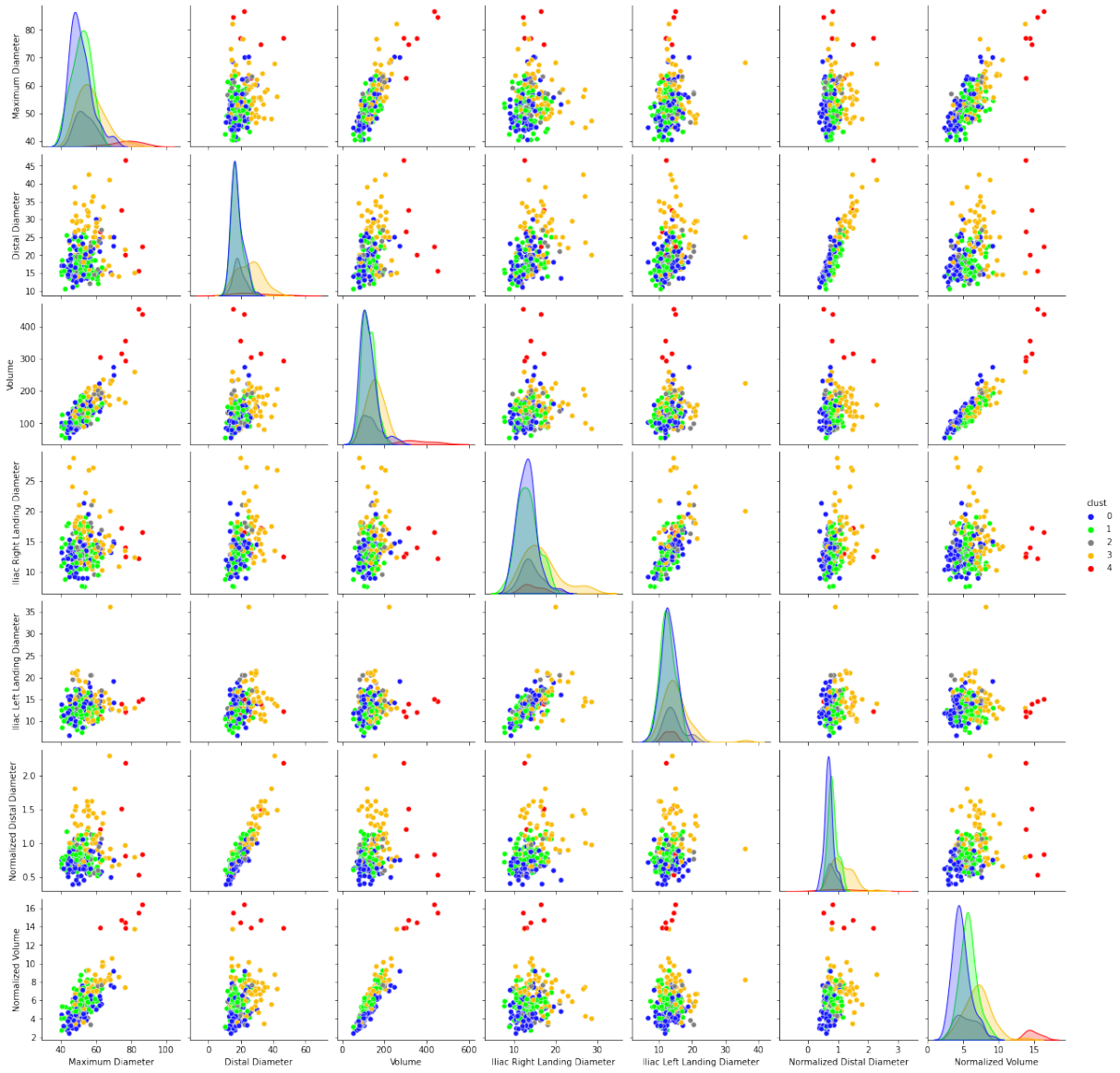


Figure 2.11: A scatterplot for the features related to the AAA, highlighting how much the Red and Orange clusters differ from the others.

So far none of the cluster described is characterized by any of the 11 features representing angles. This is not true for the Grey cluster, which on the opposite side is not particularly differentiated by the other groups on any dimensional feature. Grey cluster is the second smallest group, composed only of 22 patients, it seems to collect patients with small

aneurysm but unlike group Blue and Green is distinguishable by a smaller bifurcation coronal angle which reflects into acute curve iliac angles and on the higher tortuosity index.

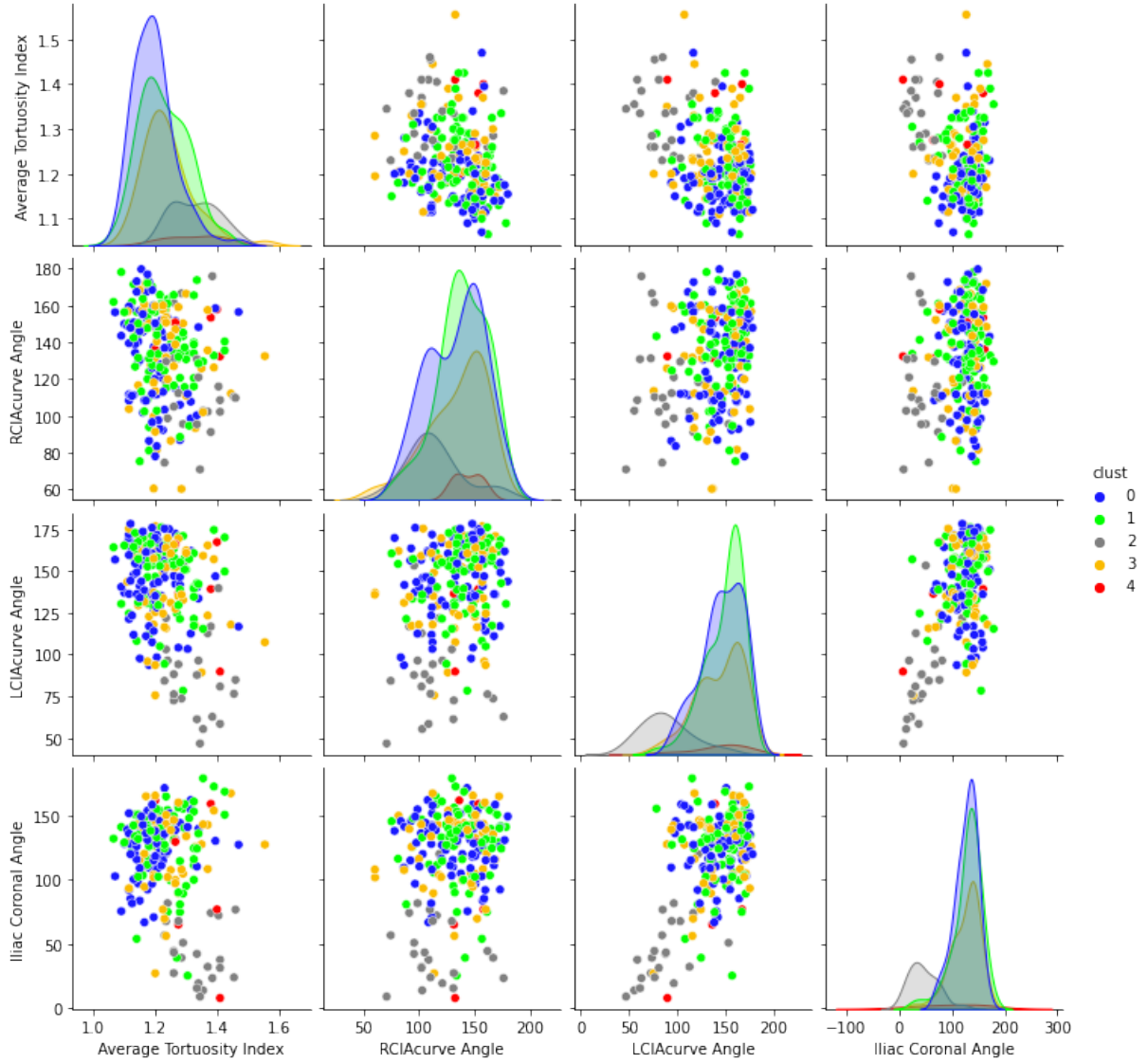


Figure 2.12: Scatterplot for the features related to the iliac angles. As is possible to see the Grey cluster has the most peculiar values.

These are the five groups found in the analysis. Since a big part of the angles that compose one third of the dimensionality of the dataset do not seem to contribute to the characterization it could be a good choice to remove them since these variables could bring noise and impoverish the division in groups. Moreover, the are no traces in literature that particular values of some angles such as renal angles could be related to the AAA.

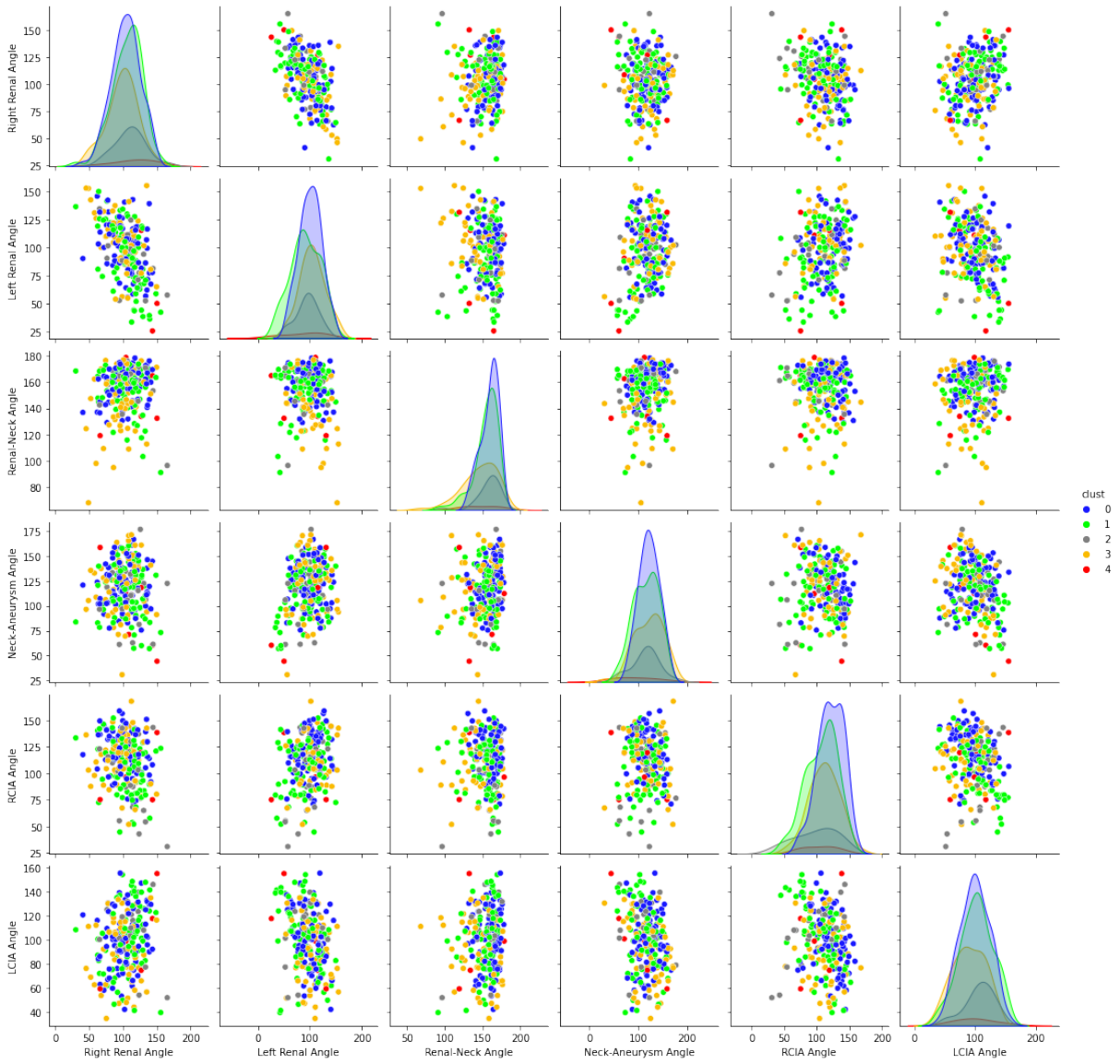


Figure 2.13: By this plot is evident how the points, observed on the feature space of some of the angles are randomly distributed with similar pattern and form a cloud of points. This could be a reason for the removal of these variables since could add noise to the grouping.

2.6. Cluster Refinement

In order to see if there are features that can be removed and select them to reduce the feature space, it was decided to perform a univariate Analysis of Variance (ANOVA) test for each variable.

ANOVA provides a generalization of the t-test. It is a statistical test to evaluate if the

means of two or more population are equal or not [74]. In this way, if the p-value associated which a certain feature is low there is statistical evidence to reject the null hypothesis and consider at least one of the five groups different in mean from one of the others. The data do not satisfy neither the assumptions of Gaussianity distribution for each factor level nor the same variances across the distributions, conditions required to correctly perform the traditional ANOVA test. The Gaussianity and homoscedasticity assumption have been tested respectively using the function *shapiro.test* and *bartlett.test* of package stats version 4.2.0 on R.

An alternative of the classical ANOVA which does not require assumptions is the permutational ANOVA test, that place the result's significance by comparing the actual F test result to the one obtained from random permutations of the data between the groups [75, 76].

To conduct this test, for each variable 5000 permutations were performed, bringing the minimum p-value to 0.00019996. To decide the variable to discard a criteria has been implemented: to delete all the features with a p-value higher than 0.01. Analysis have been conducted using the function *perm.anova* from package *RVAideMemoire* version 0.9-81-2.

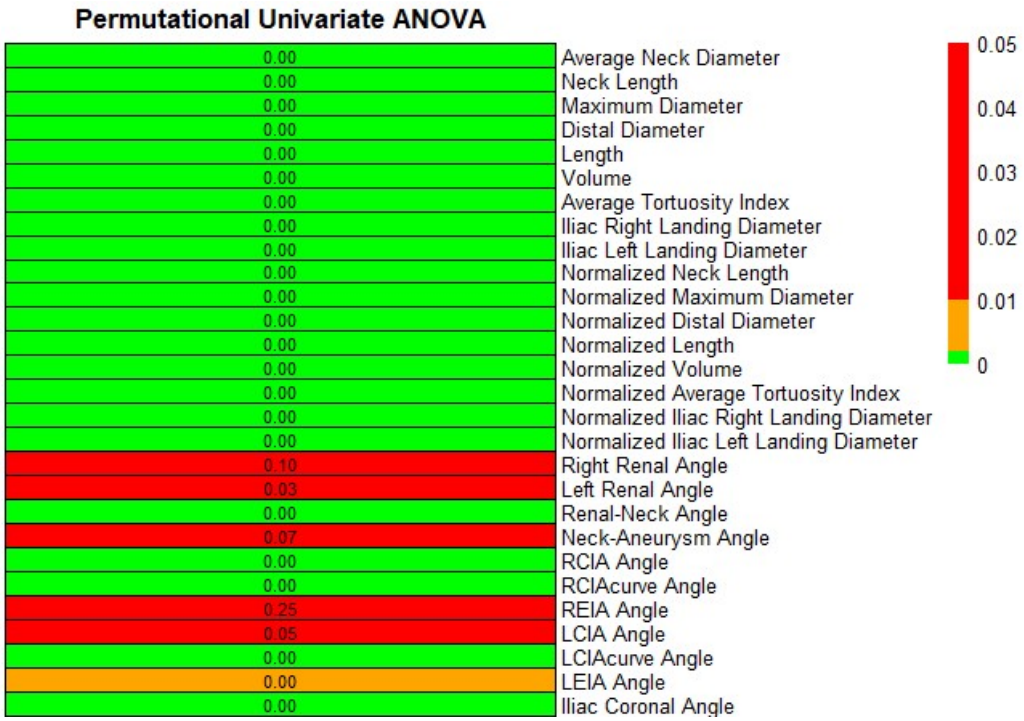


Figure 2.14: The results of Permutational ANOVA performed on each variable. Green values are reached with a p-value lower than 0.001.

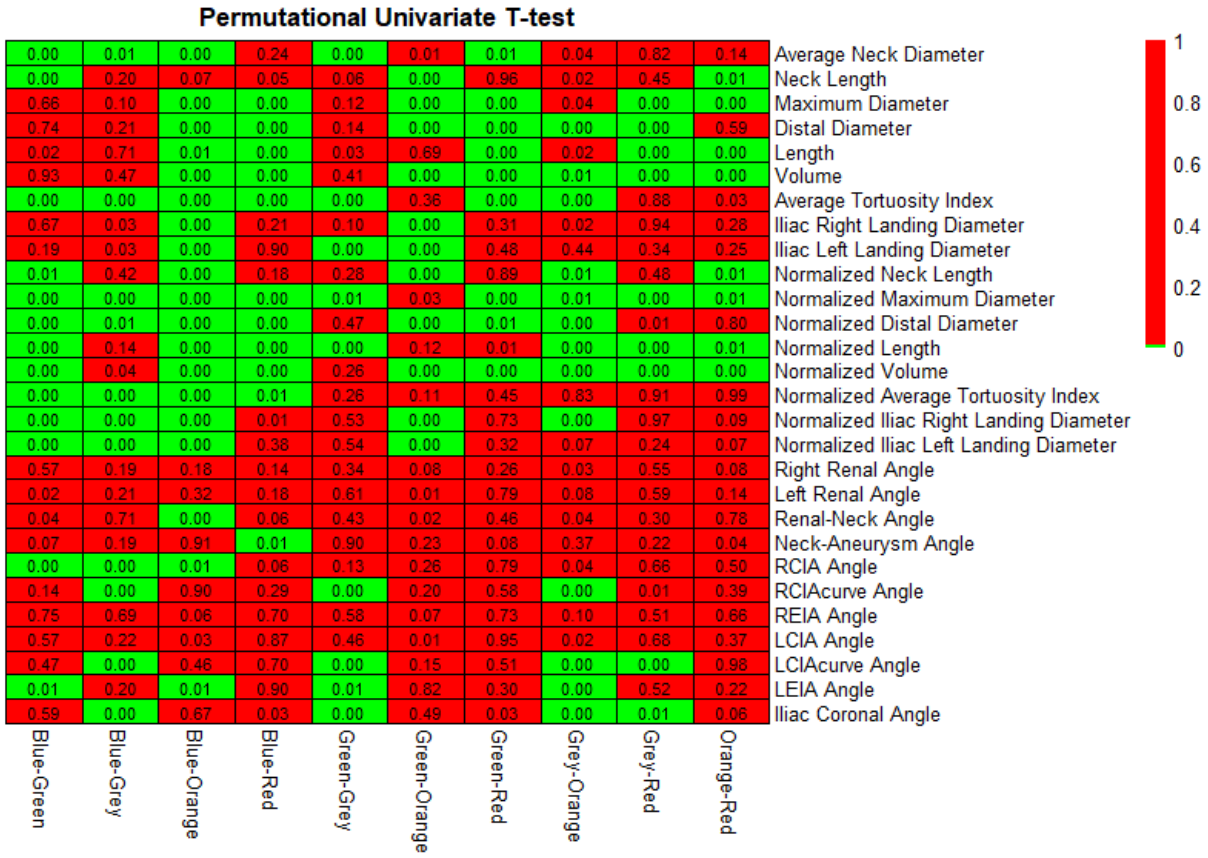


Figure 2.15: Here it is possible to see a table with the univariate t-test in order to see which variables of which groups differ the most, as expected little variations can be seen in the angle section, the lower part of the matrix.

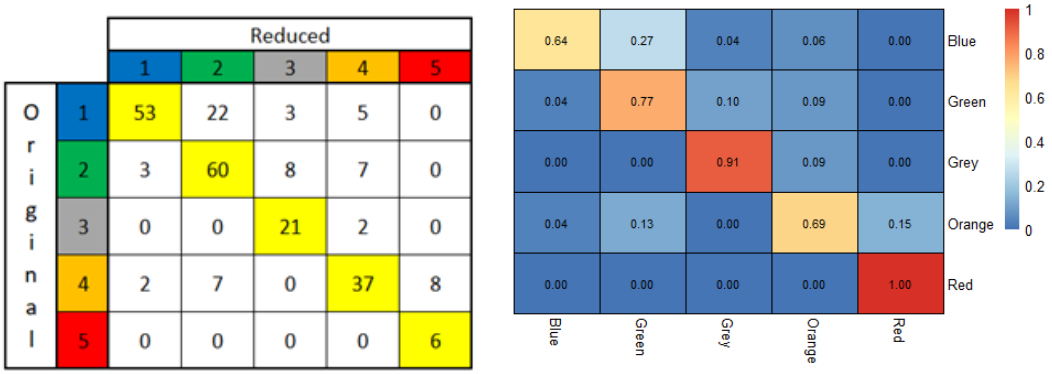
The variables chosen to be deleted are all angular features and are the right and left renal angle, the neck-aneurysm angle, right external iliac angle, and left common iliac angle. Given the symmetry of human body, the left external iliac and right common iliac angles were also discarded.

Removing these variables is not only useful to reduce the dimensionality of the feature space which is now 21 but the extraction of these features is not trivial and require additional work.

To evaluate if the removal of the variables considered not significant affects the grouping or not it was chosen to perform another clustering using the same method, distance and linkage used before with the new, reduced, dataset.

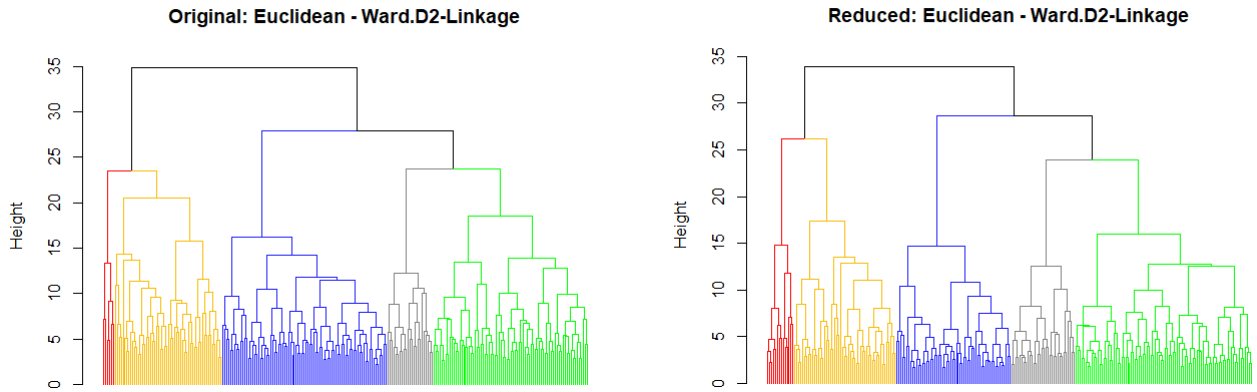
The new groups seem to resist the change of feature space remaining similar, with the main differences being in groups Blue and Orange which respectively fall into the Green and Red cluster, as showed in the heatmap below. The groups' interpretation is not

affected, on the contrary is enhanced.



(a) (b)

Figure 2.16: The first table is the confusion matrix with the cardinality of each combination of groups, while the former is the same table standardized by rows, to highlight the similarity of the two clustering.



(a) (b)

Figure 2.17: Here the two dendrograms for a comparison, changes are only in groups' dimensions, which merges in the same order, at similar heights.

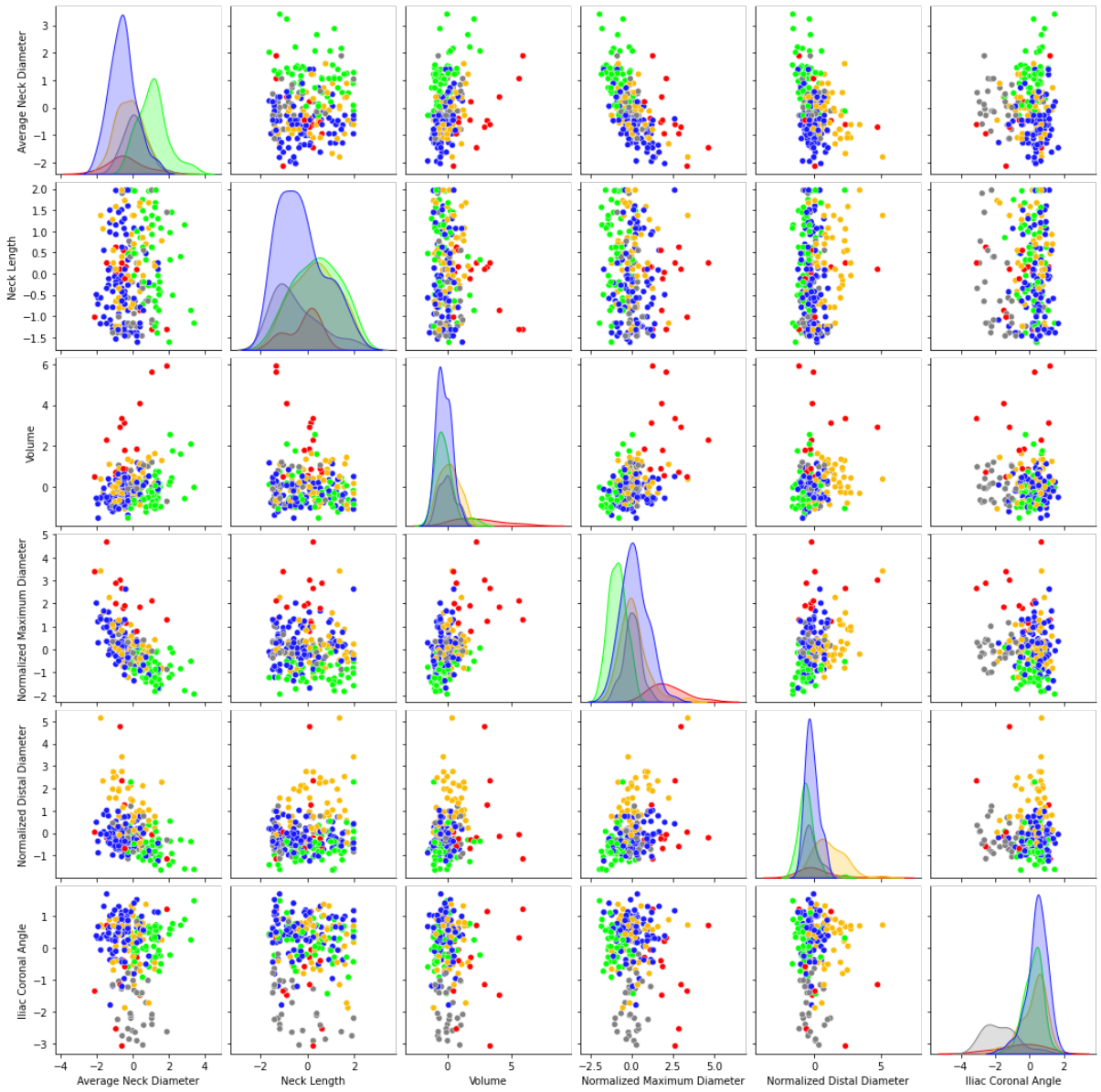


Figure 2.18: A final plot where is possible to see the groups' distribution for the variables considered most characterizing.

To summarize, there are two main clusters characterized by a small aneurysm and differentiated by neck's dimensions and consequently normalized dimensions, one in which the morphology of aorta is the main features, with acute and unnatural angles. The remaining two can be considered representing more concerning patients, one with an aneurysm that involves the aortic bifurcation and, in some cases, even involves iliac arteries and one small, clinically alarming, group with higher volume and maximum diameter.

These results were validated by Professor Maurizio Domanin and Santi Trimarchi of *Policlinico di Milano* and *Moxoff*'s data scientists.

2.7. Conclusions

Within this project it was possible to identify clusters that use different types of data, taking into account multiple aspects of the AAA such as absolute and relative dimensions and geometry, managing to describe different type of known AAA. At the same time this clusterization is robust with respect to a change of distance and dimensionality reduction, proving to be reliable even after being subjected to medical personnel specialized in the pathology.

However, this analysis has drawbacks, the main one being the fact that variables in the dataset consist only of global information with the only exception being the angles. In fact, a single observation can be associated to more aorta, that can be widely different from each other. This is in contrast with the results of fluid dynamical and mechanical studies, that simulate the patient's blood flow and whose results are extremely sensitive to the local geometry. For such reason having data representing even local properties of the aorta would be ideal.

There are two main ways to deal with this and are to directly segment CTA scans to build an ad-hoc dataset to meet the purposes of this research and to use advanced techniques to describe the aorta such as *Skeletal Representation* or *Persistent Homology*, as will be showed in the future chapters of this thesis.

To conclude, there is not a clear natural grouping between patients, with patterns and divisions. Not even all the algorithms used were capable to produce a notable clustering as in the case of DBSCAN. However, the groups found seems to have features that are worthy of consideration according to the clinicians of *Policlinico di Milano*, collecting different anomalies in abdominal aorta's geometry and, with further analysis and more data, could became a useful tool to better understand the pathology.

3 | Segmentation

3.1. CTA scans

In medical field, the most diffused and effective method used by clinicians to visualize the arterial anatomy is CTA. This method is used to get images of section of the human body realized with tomograph and is based on emission of X-ray. Patient is injected with an iodine contrast dye and the tomograph rotates around the patient emitting radiations that, reacting with the contrast dye, that is in different concentration in the various tissue of the body, allows a complete tracking of the vessel in which it flows. In this way is possible to capture up to 45 million data with different angulations and intensities. This data are then processed by a computer which allow a three dimensional image of the body in which the different organs can be recognised by different shades of grey.

Figure 3.1: A tomograph, the machine which allows the creation of CTA scans. Image from [77].

Radiation emitted, after passing through the patient, comes out attenuated. This attenuation depends on the absorption coefficient μ of the different organs. This gradient of grey is then quantified by the *Hounsfield scale*, used to describe radio intensity.

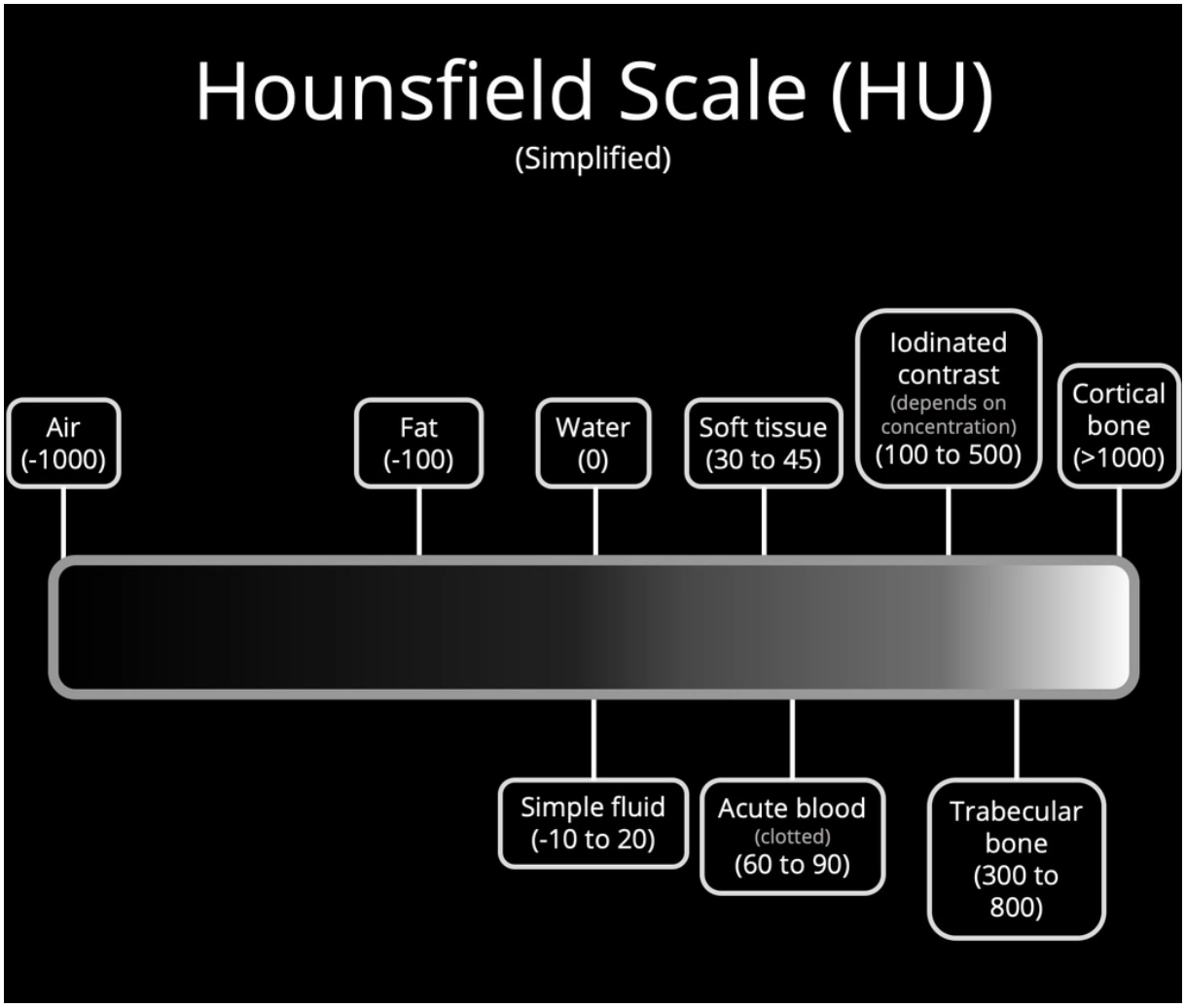


Figure 3.2: The Hounsfield Scale: attenuation values of some organs. It is important to note that there is not a precise value but a range since, based on the quality and quantity of the contrast fluid, CTA scans can have higher or lower intensities. Image taken from [78].

The human eye can distinguish up to twenty hues of grey, which would make impossible to evaluate all the variation of hues of grey contained in the image so obtained.

So, based on the region the clinician wants to study and the purpose of the exam, only a range of values is selected on the Hounsfield scale. This range is called window and all the values higher than the upper limit are represented as white, while the values lower than the lower limit are represented as black.

CTA scan, originally developed to diagnose anatomical injures of the brain, has now widely spread to analyse many regions of the human body, among which the abdominal one. The main advantage of this technique is that allows the acquisition of many images in a few seconds, making it suitable to study extended anatomical districts. The elevated precision also permits to trace the passage of the contrast material, useful in the study of blood vessels.

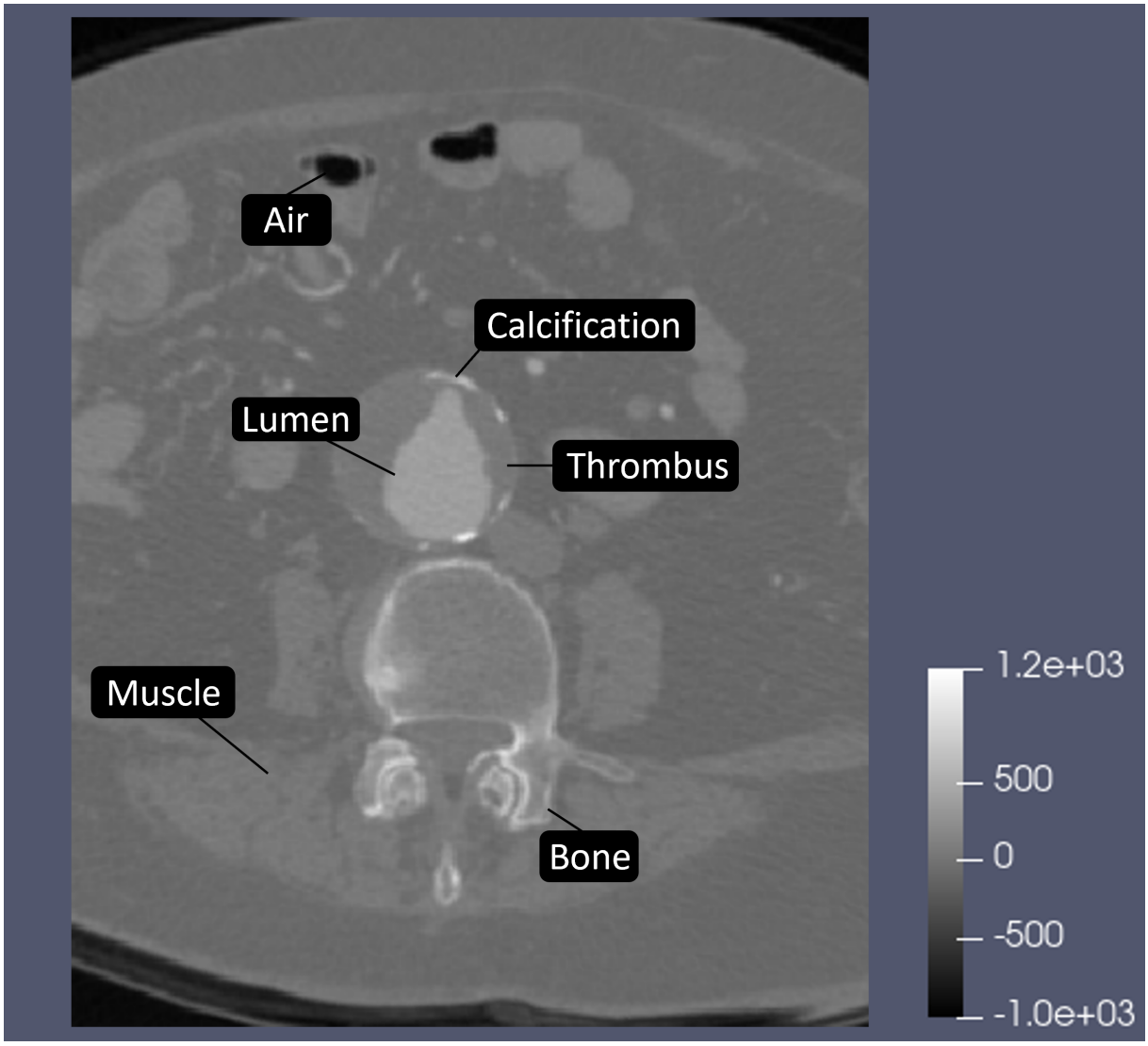


Figure 3.3: Coronal section of the abdominal cavity, with references to tissues intensities. The lumen of the aorta is visible surrounded by thrombus and calcifications.

Moreover, the images acquired through CTA provide a high spatial resolution with the examined region that is divided in single slices with the thickness of tenths of one millimetre and can be elaborated with a computer to obtain planar or three-dimensional projections.

Processing and sharing of CTA scan, contained in the Picture Archiving and Communication System (PACS) is regulated with the DICOM standard.

Digital Imaging and Communications in Medicine (DICOM) is a standard, used since 1993, for the management of digital images and correlated information in the clinical field and allows CTA scans to be visualized through apposite applications such as Slicer 3D.

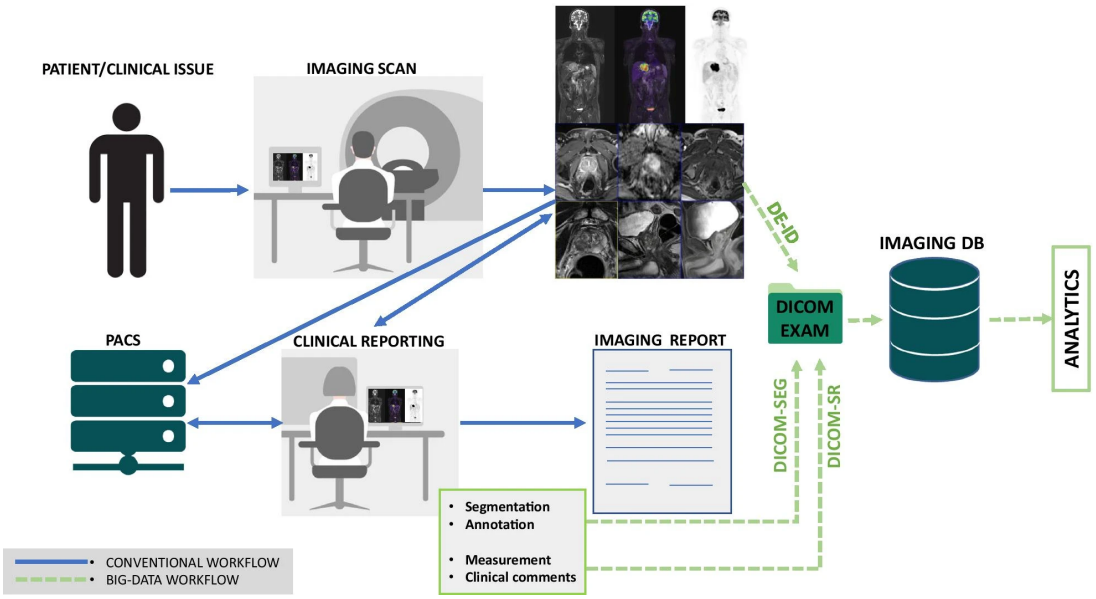


Figure 3.4: A scheme of the procedure that, starting from a patient, allows the acquisition of a CTA scan and its visualization. Image from [79]

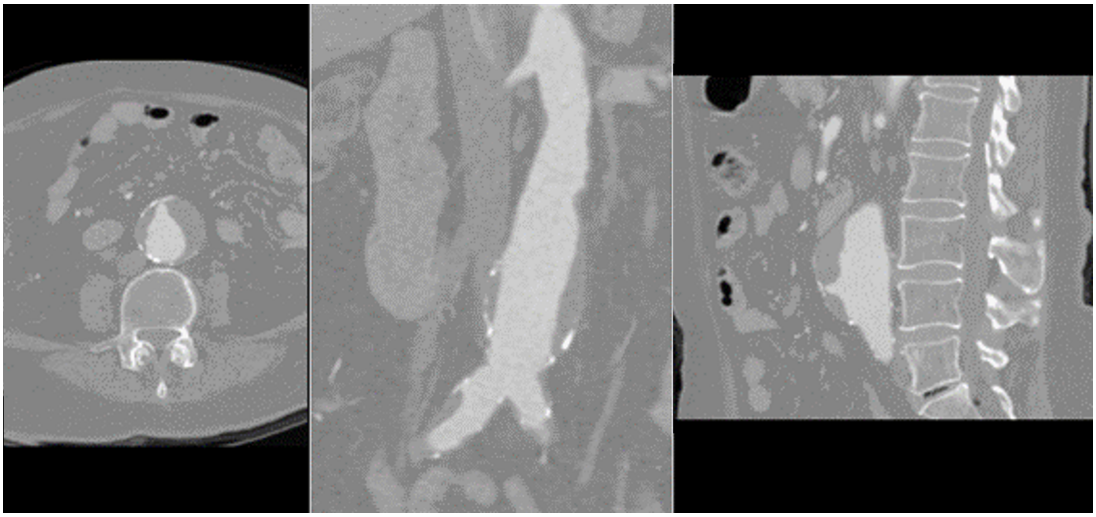


Figure 3.5: A pathological abdominal aorta affected by AAA visualized by CTA scan on the transverse, coronal, and sagittal plane.

Having a CTA scan means to have a digitalization of the part of interest. In this work the main utility of the CTA scan is not the visualization, as could be for a clinician, but the fact that can be used to create a segmentation of the aortic vessel.

3.2. Segmentation

In the medical field, arterial segmentation refers to the process that leads to the obtainment of a 3-dimensional reconstruction of the object of interest via a finite number of points. This can be done for a variety of purposes, including identifying specific areas that may be affected by diseases or injuries, studying blood flow through the artery and, planning surgical procedures.

Arterial segmentation is an important tool in the study of arterial diseases and injuries. In fact, it allows physicians and biomedical engineerds to better understand the specific areas of the artery that may be affected by pathologies, and to develop targeted planning of treatment.

The procedures and techniques used to perform the segmentaion explained in this chapter can be seen in detail by looking at [80].

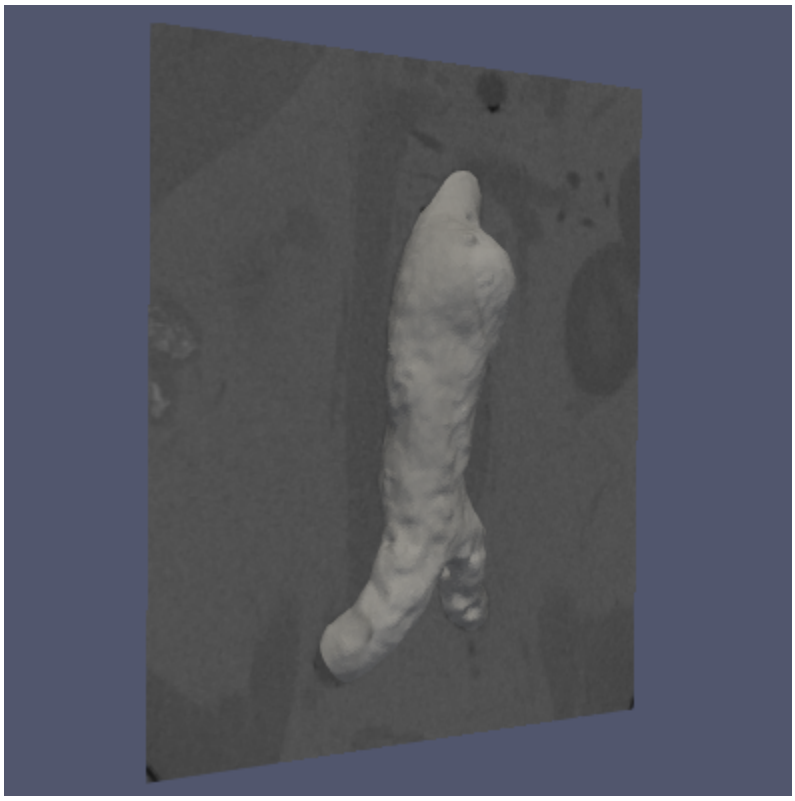


Figure 3.6: The segmentation of an aneurysmatic aorta along with the CTA scan from which it is extracted.

Image segmentation is the process of extraction of an area of interest from an image volume assigning a label to every pixel in the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as colour, intensity, or texture. There are different existing techniques which are used for image segmentation: thresholding, region-based methods, and edge-based methods.

- Thresholding: The simplest and most used method of segmentation, in thresholding pixels are allocated to categories according to the range of values in which pixels lies. Given a single threshold, t, the pixel located at position (i, j), with greyscale value f_{ij}, is allocated to category 1 (object to detect) if $f_{ij} \leq t$, otherwise the pixel is considered belonging to category 2 (background). Threshold t is often chosen manually by the operator who is performing image segmentation, by trying a range of values of t and seeing which one better identifies the object to detect. Several thresholding techniques have been proposed so far and distinguish in global and local techniques. Global methods apply one threshold to the entire image while local thresholding methods apply different threshold values to different regions of the image [81]. The threshold is set by the operator that performs the segmentation.

- Region-based method: This approach is based on algorithms that operates by grouping together pixels which are neighbours with similar values and splitting groups of pixels which are dissimilar in value. They are more immune to noise with respect to edge-based techniques, on the other hand they are expensive in terms of time and memory.
- Edge-based method: This pool of techniques is based on the identification of image edges. All edges are connected to form the object's boundaries required to segment the wanted regions. Edge detection is a series of actions whose purpose is to recognise points in an image where clear and defined changes occur in the intensity. There exist many algorithms able to detect edges and they can be grouped into two categories: search-based and zero-crossing based. Both methods have advantages and disadvantages: search-based algorithms have the advantage of having a very simple implementation and run extremely fast, but they have problems such as noise and discontinuity of the edges. Crossing-based methods involve more operations, but they allow a better quality in results.

The pipeline used to perform the segmentation relies on all the three methods in synergy, relying mainly on an Edge-based one with the support of thresholding and region-based to improve performances and results. Thresholding, the simplest one, makes an initial and more general guess on points belonging to the lumen. Then region-based and edge-based help instead with recognition of the rounded shape of the lumen when projected on the coronal plane.

This process applies a label to each of the voxels, 3D pixels, in the CTA. This labeling creates a three-dimensional volume, whose surface is extracted and triangulated.

The result of the segmentation is called mesh, a set of vertices, edges and cells. Two points can be linked together through edges and three edges that shares one single point between them build a face, a two dimensional simplex.

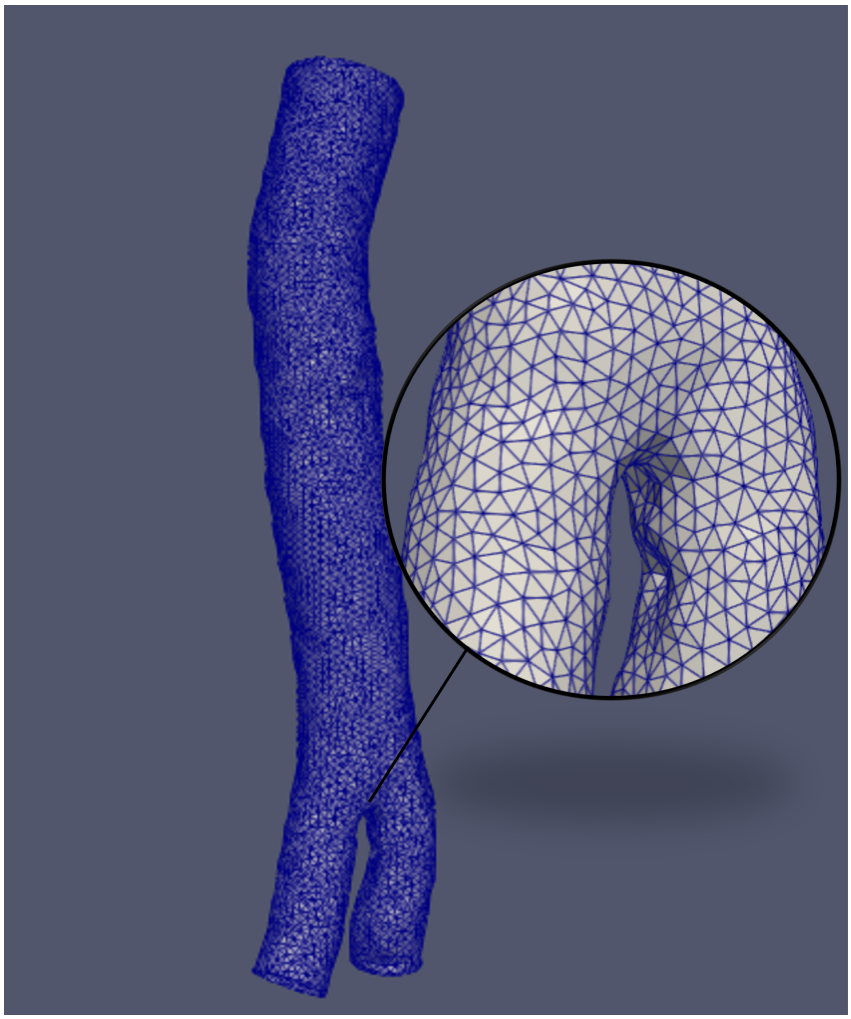


Figure 3.7: The mesh of a healthy aorta. Edges and cells composing it are highlighted in the zoomed detail.

Different cells can only share up to one edge between them. All edges share similar length, called mesh resolution. Mesh resolution refers to the level of detail in a mesh, which is a collection of vertices, edges, and faces that define the shape of an object in 3D space. The resolution of a mesh can be controlled by adjusting the number of polygons that make up the mesh.

A higher resolution mesh can produce more accurate and realistic results, but it also requires more computational resources to process and can result in larger file sizes. Lower resolution meshes are faster to process but may not be as accurate, especially when dealing with complex shapes or fine details. The union of hundreds of faces creates a representation of the object using a finite number of triangles.

Segmentation is made possible thanks to a complicated algorithm that make use of a

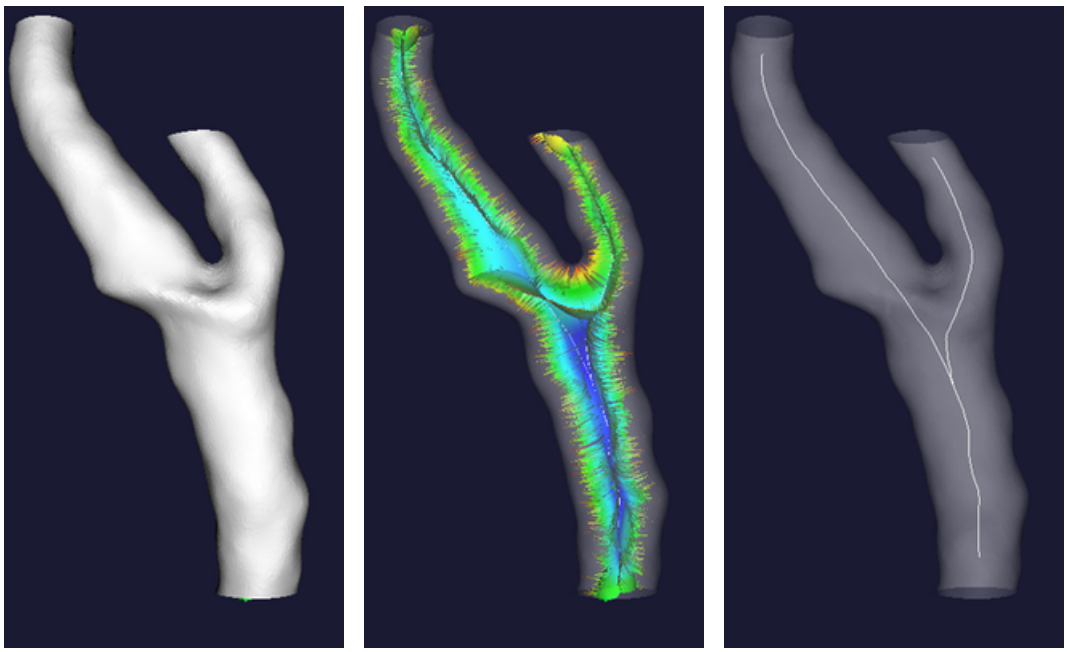
particular Python library called Vascular Modelling Toolkit (VMTK [82]), created by David Steinman and Luca Antiga, and appositely designed for this scope. VMTK is an open-source software that allows, starting from 3D images such as CTA scans, to realize 3D reconstructions and mesh generation for the study of blood vessels. The software is based on algorithms from other open-source libraries: Insight Segmentation and Registration ToolKit(ITK [83], Visualization ToolKit (VTK [84]) and Quality Tetrahedral Mesh Generator (Tetgen [85]).

3.3. Centerline

Another important tool is the centreline. The idea of centerline of a vessel is to find a continuous curve in space that represents the centre of the vessel itself, creating a powerful descriptor of its shape. Despite the fact that the concept of centerline is intuitive, its mathematical definition is not unique. Several methods have been proposed and described in the scientific literature for its computation from CTA scans or three-dimensional models (e.g. meshes).

Like segmentation, also computation of the centerline has been made possible thanks the algorithms implemented in VMTK. Being based on the surface model, has the advantages to be well characterized mathematically and be stable to perturbations on the surface.

In VMTK, Centerlines are determined as the path defined on the Voronoi diagram. In order to ensure that the final lines are in fact central, they are bound to run on the Voronoi diagram of the vessel model. That can be seen, as a first approximation, as the place in space where the centres of maximal inscribed spheres are defined. In this way, the path minimizes the integral of the radius of maximal inscribed spheres along the path, which is equivalent to finding the shortest paths in the radius metric. For additional information is possible to read Luca Antiga's PhD thesis [86] and the documentation of VMTK [87].



(a) (b) (c)

Figure 3.8: The central figure below shows a visualization of the Voronoi diagram associated with the shape of a carotid bifurcation. The colours are referred to the radius of maximal inscribed spheres (red: small, blue: large). The left figure is the surface object while in the right one is possible to see the centerline [87]

The object centerline is composed by two distinct sets of ordered points that are linearly interpolated, one for each iliac artery.

The concept of centerline, even if not belonging to the lumen's surface we are studying, is necessary to define how much points are peripheral, based on the minimum distance of each surface point to a point of the centerline. This distance will be used for normalize the distances, in order to have more coherent results between different patients.

3.4. Pipeline and Post-Processing

The pipeline used for the segmentation and centerlines has been provided by Moxoff and is written in *Python 3.8* using a patched version of *VMTK 1.5* made by Luca Azzolin. Other main contributors to the segmentation's pipeline are Abele Simone, Chiara Riccobene, Luca Azzolin and Teresa Babini.

The tract of the abdominal aorta used in this study ranges between the distal renal artery and hypogastric arteries, including the area eventually affected by the AAA in the

pathological cohort. The selection of the starting and ending points, also called seeds, has been done manually. For practical purposes only the lumen has been segmented in this analysis. In fact, thrombus segmentation would have been much harder to be segmented since it in CTA scans has a similar grey intensity to veins, which can be very close to aorta, bringing to inaccurate representations. Calcifications are also discarded since, being of the wall surface, are not always contiguous to the lumen.

Having the lumen, while not being the whole AAA, can still give us important information since lumen is the region where the blood flows and fluid dynamics takes place.

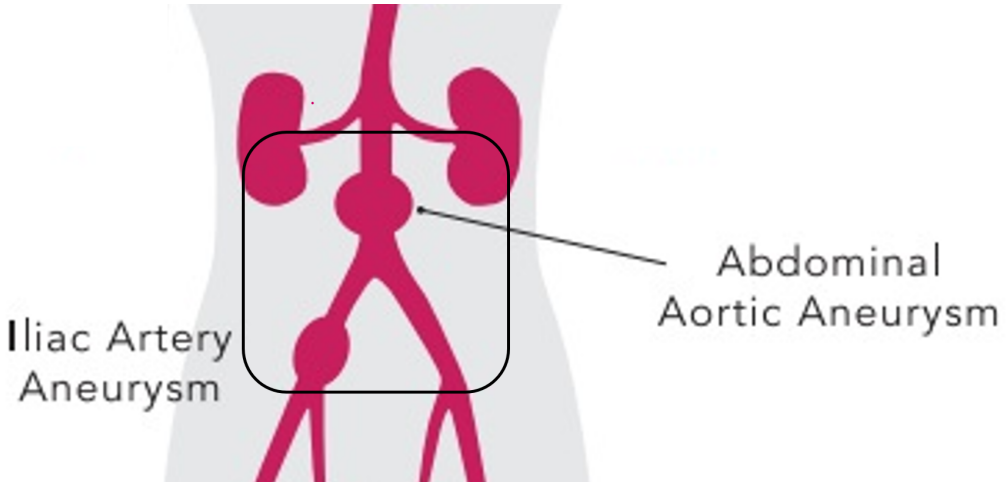


Figure 3.9: A schematic view of the area of interest segmented by the pipeline.

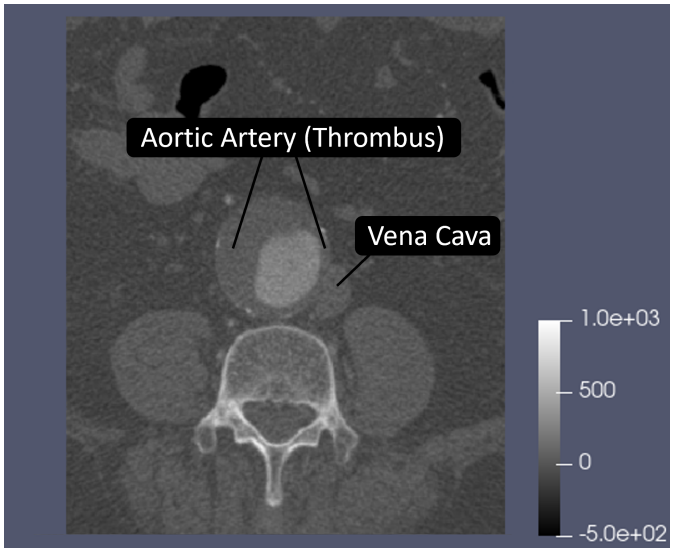


Figure 3.10: In this image, a case in which the aortic artery and the vena cava, the biggest vein of human body are close to each other. Without an accurate revision, a part of the vein would have been considered thrombus or vice versa.

The range selection for the thresholding method has been chosen by following the vast scientific literature [88], while also double checking each specific patient with the clinician's help. The reason of this is because the quantity and quality of contrast fluid used each time can vary, leading to different results and doing so, is possible to adjust the values for CTA scans that have higher intensity values or lower ones to obtain more consistent results.In the cases in which threshold adjustment was not enough manual correction was necessary, removing selectively unwanted part of veins, thrombus, or other arteries according to the clinician's opinion.

The segmentation's resolution is $0.7mm$, empirically chosen to balance the trade off between accuracy and computational costs.

3.5. Data and Aim

This work was developed in collaboration with *Fondazione IRCSS Ca' Granda, Ospedale Maggiore Policlinico*, Milan. Specifically, we collaborated with the vascular surgeons and Professors Maurizio Domanin and Santi Trimarchi, medical advisors for this project, segmentation revisor and provider of the diagnostic images. All the CTA scans, are provided by *Policlinico di Milano*. Patients with healthy aortas have been collected from the Policlino's PACS while the patients with a pathological aortas come from Vascular Surgery Unit of the hospital. For privacy reasons, all patients' data are anonymized.

At the end of the process, we obtain a dataset consisting of Segmented Aorta's lumen of 22 patients, 10 healthy and 12 aneurysmatic aortas, all approved by *Policlinico di Milano*'s vascular surgeons.

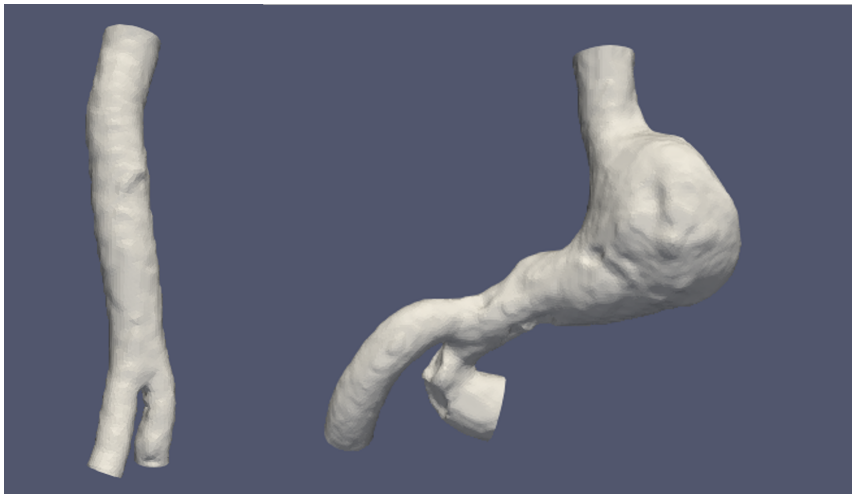


Figure 3.11: Two cases: on the left a healthy aorta, on the right a aneurysmatic one.

These 22 segmentations give us an accurate and complete digitalization of the abdominal aorta, included the first tract of iliac arteries, at complete disposal. In this way is possible either to extract dimensional features (e.g., maximum diameter, angles and, neck length) as done in the previous chapter, or use advanced method such as Skeletal Representation and Topological Data Analysis, which will be the core of this thesis. If the former summarizes the aorta by focusing on some of its features the latters make use of the entire aorta object allowing to also describe the local geometry.

4 | Persistent Homology

One of the fundamental methods for Topological Data Analysis (TDA) to analyse qualitative aspects of data that persist across many scales is called *Persistent Homology*, which helps to retrieve the essential topological features of an object. Persistent Homology offers a concise representation of the qualitative characteristics of the object in input, is resistant to perturbations of the input data and is independent of dimensions and coordinates [89], [90]. Numerous application fields, including biology, image classification, network analysis, chemistry and astrophysics are paying increasing attention to this method. As a result, a growing number of researchers are considering the use of Persistent Homology as a tool in their research.

4.1. Notions

One of the most important concept for the definition of Persistent Homology is called *Simplicial Complex*. Informally, a simplicial complex is a particular case of graph constructed by gluing together points, line segments, triangles, and their n-dimensional, counterparts called *simplicies*, and a common method to represent discretized objects.

Mathematically, a k-simplex σ is the convex hull of $k+1$ affinely independent points. The concept of simplex is the generalization of notion of triangle in an arbitrary dimension [91]. A k-simplex σ, is the simplest possible polytope of that dimension. So, a 0-simplex is a single point, a 1-simplex a segment, a 2-simplex a triangle, a 3-simplex a tetrahedron, and so on. The convex hull of any nonempty subset of the $n+1$ points that define an n-simplex is called a face of the simplex [91].

A simplicial complex Σ is a finite set of simplices where each face belong to Σ, and every non-empty intersection of any two simplices in Σ must be a face of both simplices. This is known as the gluing condition. Moreover, the dimension of a simplicial complex Σ as the largest dimension the simplices that compose it [90].

Given a complex Σ, the associated chain complex $C_*(\Sigma)$ can be defined as $C_*(\Sigma) := (C_k(\Sigma), \partial_k)_{(k\in\zeta)}$, with $C_*(\Sigma)$ the *Free Abelian group* [92] generated by the k-simplicies of

Σ. The boundary map $\partial_k : C_k(\Sigma) \to C_{k-1}(\Sigma)$ is a group homomorphism which links the k-simplicies and the $(k-1)$-simplicies of Σ such that $\partial^2 = 0$. Moreover, $Z_k(\Sigma) := ker(\partial_k$ can be defined as the group of k-cycles of Σ and $B_k(\Sigma) := Im(\partial_{k+1})$ the group of the k-boundaries of Σ. Note that $\partial^2 = 0$ implies that $B_k(\Sigma) \subseteq Z_k(\Sigma)$.

Finally, is possible to denote the k^{th} homology group of Σ as:

$$H_k(\Sigma) := H_k(C_*(\Sigma)) = \frac{Z_k(\Sigma)}{B_k(\Sigma)} \tag{4.1a}$$

In applications, mainly for computational reasons, homology groups are often considered with coefficients in a field, usually $\mathbb{Z}_2$. This means that $H_k(\Sigma)$ is a vector space over $\mathbb{Z}_2$.

Roughly Speaking, the dimension of the k^{th} homology group detects the presence of k-cycles in the object, also called *holes*. Holes in dimension 0 are path connected components in the graph representing the object, in dimension 1 are closed path surrounding a hole while in dimension 2 is the number of shells enclosing voids and cavities [90].

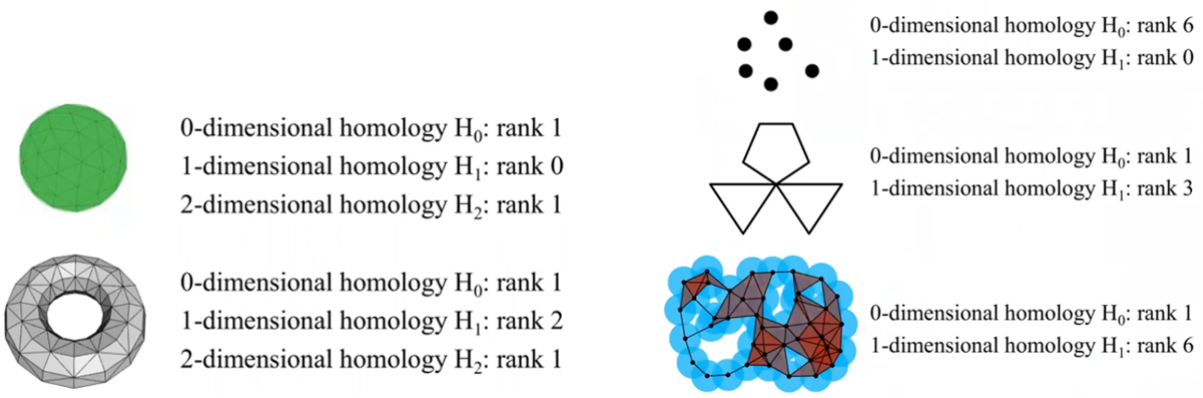


Figure 4.1: Here is possible to see the number of k-dimensional holes in different objects. The first figure shows an hollow sphere and an hollow torus. Images taken from [90].

Homology describes the shape of an object up to a large set of deformations. More precisely, homology portray an object up to *homotopy equivalence*.

Two topological spaces X and Y are homotopy equivalent if there are maps $f : X \to Y$ and $g : Y \to X$ such that $g \cdot f \simeq i_X$ and $f \cdot g \simeq i_Y$ [90, 93]. With $f \simeq g$ meaning that there is a continuous map $F : X \times I \to Y$ such that $F(x, 0) = h(x)$ and $F(x, 1) = k(x)$ for all $x \in X$. Then, F is called *homotopy* of h to k and h and k are said *homotopic*.

Summarizing, two objects are homotopy equivalent, e.g. are considered the same shape

if, starting from one, is possible to obtain the other using stretching and bending but not tearing [89].

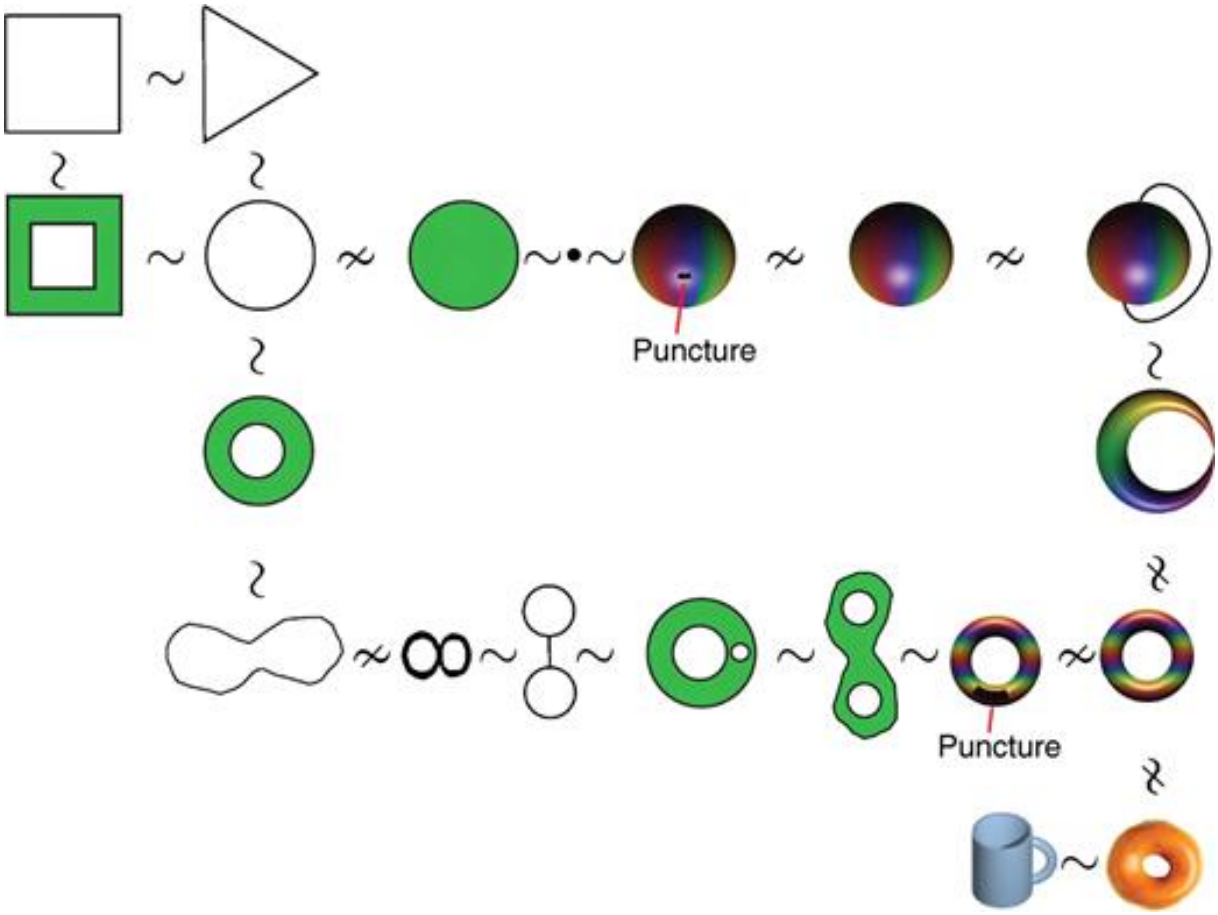


Figure 4.2: Example of homotopy and homology equivalent objects. Image from [94].

Persistent Homology describes the differences that occur in the homology of a shape when a parameter changes. To formalize the statement it is necessary to define the concept of filtration. Given a simplicial complex Σ, a filtration is a finite sequence of subcomplexes Σ^i, such that $\emptyset = \Sigma^0 \subseteq \Sigma^1 \subseteq \cdots \subseteq \Sigma^m \subseteq \Sigma$.

For $p, q \in \{0, \ldots, m\}$ such that $p \leq q$, the (p, q)-persistent k-homology group $H_k^{p,q}(\Sigma)$ consists of the k-cycles included from $C_k(\Sigma^p)$ into $C_k(\Sigma^q)$. Formally, $H_k^{p,q}(\Sigma) := Im(i_k^{p,q})$ with $i_k^{p,q}$ the linear map induced by the inclusion of complexes between the two subcomplexes $H_k(\Sigma^p)$ and $H_k(\Sigma^q)$. Thus, Persistent Homology provides additional information about a shape of an object compared to the classical homology, in fact the persistence of a cycle provides information about the importance of the cycle itself.

The concept of persistence pair is useful to interpret and visualize the topological features (cycles) detected by this method, representing their relevance. Given a filtration, $\{\Sigma^p \mid$

$0 \leq p \leq m\}$ of Σ, a persistence pair (q, p) is an element in $\{0, \cdots, m\} \times (\{0, \cdots, m\} \cup \infty)$ such that $p < q$ with $p, q \in \mathbb{R}$. In this definition, p represents the step of the filtration at which the homology class is born and q the step in which it *dies*, meaning that it merges in a class born earlier. So, the quantity $q - p$ can be used to discriminate between relevant and not classes, representing the lifespan of that class. If q is ∞ it means that there is no value in the filtration at which the homology class dies.

One of the most important qualities of the persistence pair, and therefore of the Persistent Homology, is the simplicity of the visualization. Among the many visualization methods, one of the most used is the Persistence Diagram. Given a dimension equal to k, the Persistence Diagram associated is the Cartesian plane (p, q) where the persistence pairs, of dimension k, are drawn as points. Since $q > p$, with p and $q > 0$, in Persistence Diagrams all the points lie in the first quadrant above the main diagonal. The persistence of a pair, $q - p$, is the distance of the associated point from the diagonal on the ∞-norm.

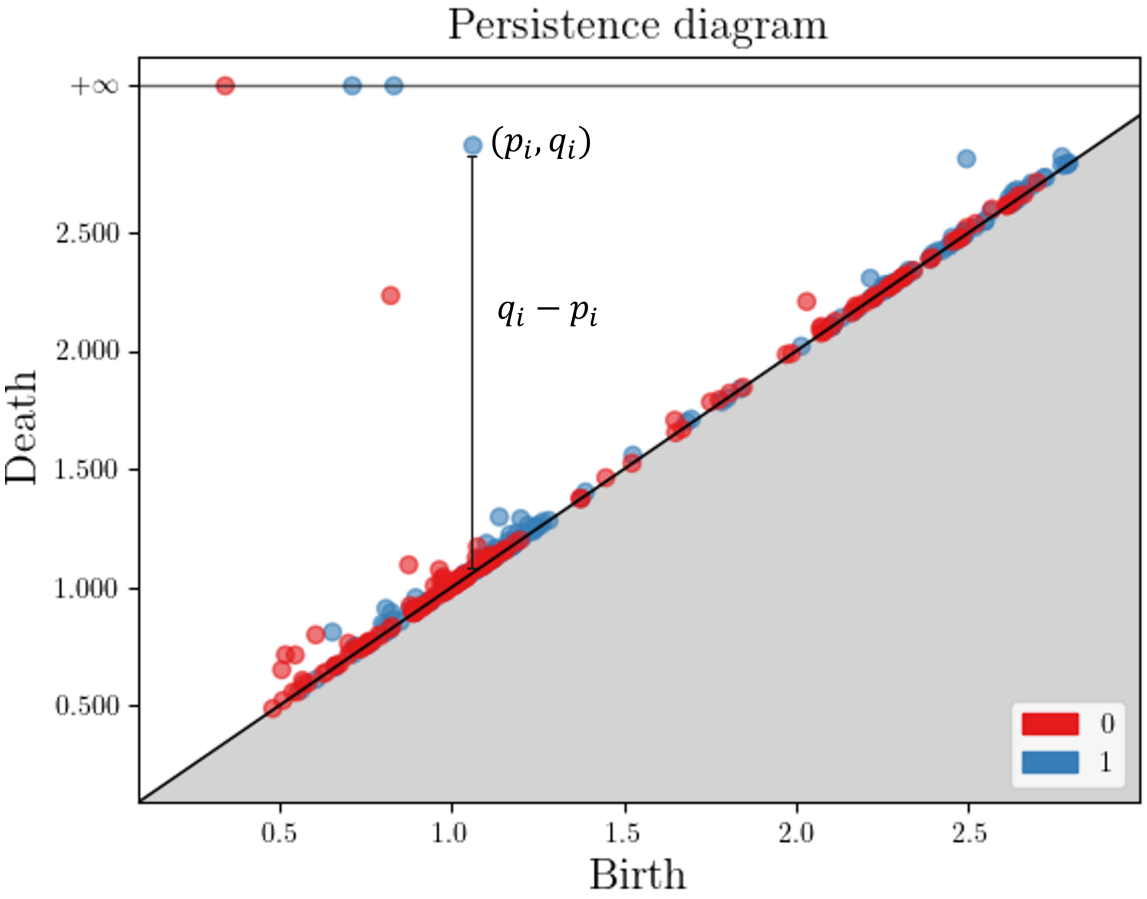


Figure 4.3: An example of Persistence Diagram, the points in red are 0-dimensional cycles and the blue ones are 1-dimensional cycles. The distance in the ∞-norm of each point to the diagonal is the persistence of the relative pair.

An important aspect of Persistence Diagram is that, using proper functions, it is possible to add the notion of distance and pseudo-distance between different ones with proved stability results. In this way it is possible to quantify the dissimilarity of two different diagrams and perform classification, clustering and dimensionality reduction techniques.

- Bottleneck Distance: It creates bijections between pairs in different diagrams. In order to allow comparisons between diagrams with different number of points, a match can be made with another pair or with a point on the diagonal. All matches have a value according to the ∞-norm distance. The cost of each matching is the maximum cost of any two pairs. Then the Bottleneck distance is the infimum of all costs as the matching varies among all the possible matches [95]:

$$W_\infty(X,Y) = \inf_{\eta:X\to Y} \sup_{x\in X} ||x-\eta(x)||_\infty, \text{ with } \eta \text{ a perfect matching.} \tag{4.2a}$$

  This distance has relevant stability results [96, 97].

- Wasserstein Distance:

$$W_q(X,Y) = \inf_{\eta:X\to Y} ||x-\eta(x)||_\infty^q, \text{ with } \eta \text{ a perfect matching.} \tag{4.3a}$$

  Again, as in the Bottleneck distance case, a match can be done between persistence pair and points on the diagonal [95]. Results for stability can also be obtained using this distance [98].

  Another possible distance is the Sliced Wasserstein Distance (SWD), an approximation of the Wasserstein distance. The Sliced Wasserstein Distance approximates the Wasserstein distance by computing the distances between the projections of the two distributions onto a set of random 1-dimensional slices. While the approximation may not be exact, it is computationally more efficient than computing the exact Wasserstein distance.

4.2. Filtration

Filtration is usually defined in two ways in topological data analysis.

- If the object is a set $\{X_1, X_2, \cdots, X_m\} \subseteq \mathbb{R}^n$, the filtration Σ is obtained starting from the set of points and increasing the radius of the balls centered in each point. A k-simplex generated by vertices $v_0, \cdots, v_k$ is created when every two balls centered in each of the vertices intersect. In this way there is no need for a pre-existing

simplicial complex to start the procedure. This filtration is usually called the *rips* filtration of $\{X_1, X_2, \cdots, X_m\}$.

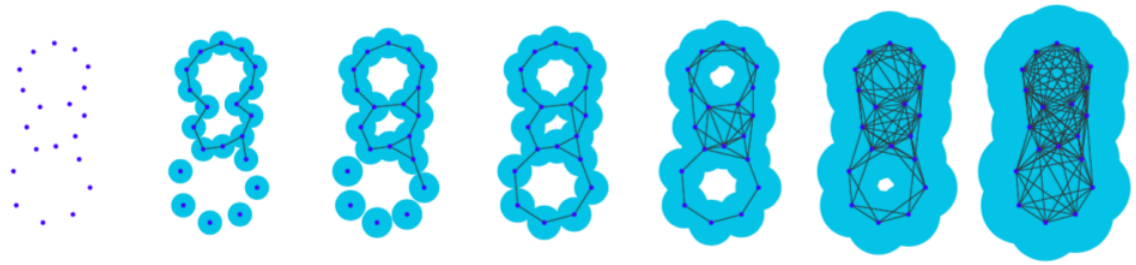


Figure 4.4: A visual representation of the process described above. Image from [99].

- Another way is to define a filtration over a pre-existing simplicial complex Σ and is especially suited for meshes. The method consists in assigning a value to each simplex of Σ, with $f : \Sigma \to \mathbb{R}$. Edges are added to the filtration at the moment in which the two points forming it are both in the filtration or later. The same is equal for facets that are added at the time in which the 3 points (or edges) creating it are added or later. To formalize, f assign a value to all the vertex, edges and faces of which Σ is made. For a vertex $[a, b]$, $f([a, b]) =\geq max\{f(a), f(b)\}$, while for the face $[a, b, c]$, $f([a, b, c]) =\geq max\{f(a), f(b), f(c)\}$

This second approach is the most suited to the purpose of this thesis, using the mesh itself as starting complex. It induces an ordering between the points and consequently, edges and faces of the mesh. To use it is necessary to define the set of value that is associated to each point. For this scope, the centerline was chosen. In fact, the filtration f is created by computing the Euclidean distances between each point of the mesh with each point of the centerline and keeping only the smallest one. In this way the values of $f : \Sigma \to \mathbb{R}$ are simply the distances of a set of points to a curve in space.

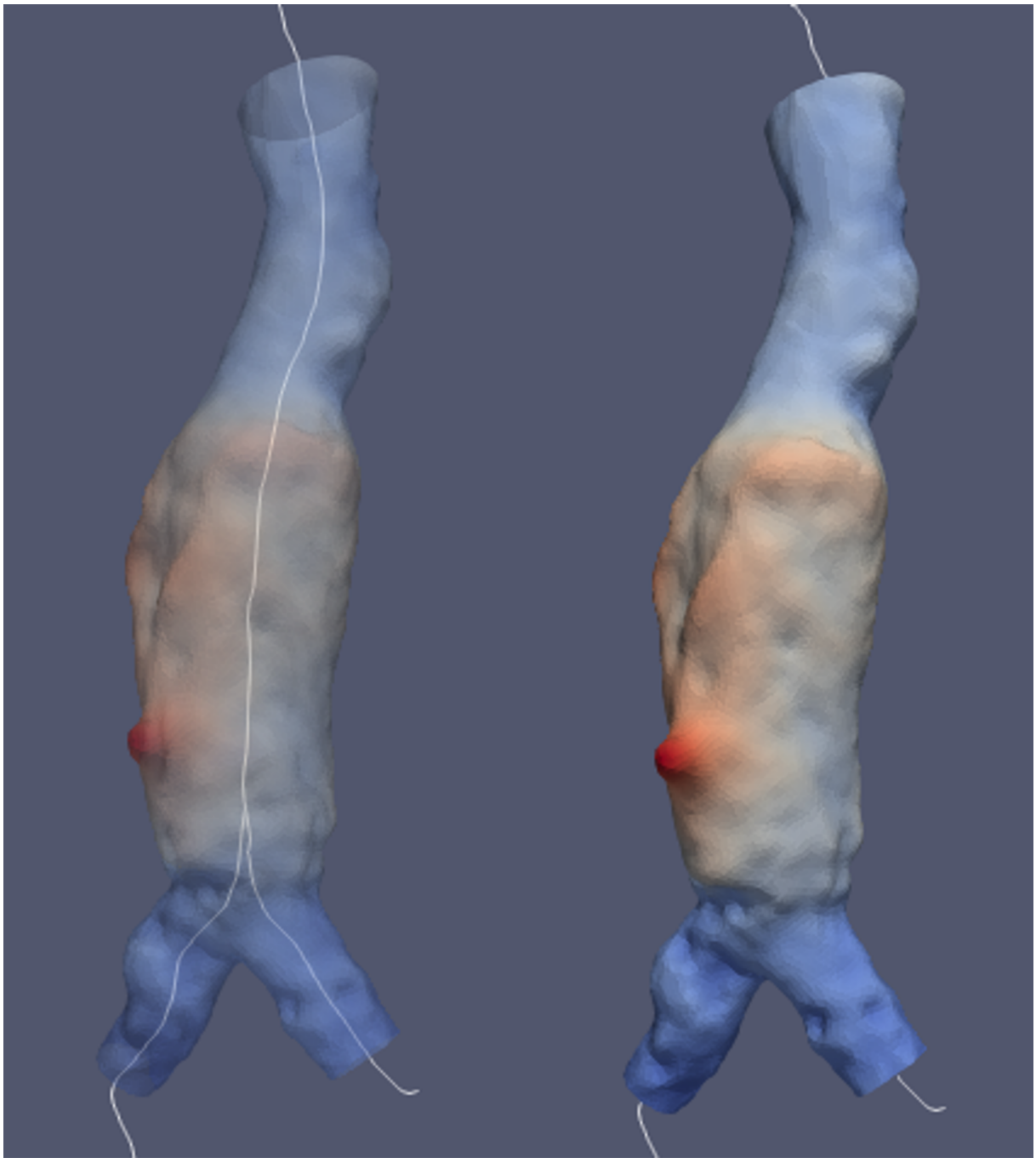


Figure 4.5: Here a mesh of the aorta of a patient with AAA. Points are coloured based on the minimum distance from the centerline. The more red a point is the more distant is from the centerline, and viceversa.

This seems to be the most natural way to study the irregularities in the lumen's surface, in the sense of indentations and bumps, made by stretches of the aortic wall, keeping track at the same time of the AAA's diameter. Irregularities can be also caused by the presence of calcifications and thrombus, that will be analysed in the future chapters.

A last step in the process, which leads to obtaining the Persistence Diagrams, is the normalization of the filtration. This is necessary in order to make possible the comparison and analysis of distinct Persistence Diagrams. Different patients can in fact have different aorta and iliacs' healthy dimensions making harder to compare their diagrams. The reason behind this decision are the same discussed in second chapter of this thesis. To perform the normalization of the filtration all the values associated with a simplex of Σ have been divided by the average neck radius.

Neck's average radius has been computed with the following steps, since even establish where the AAA's neck ends is not trivial:

1. Create a linear interpolation of the centerline's points.
2. For a given parameter $s = 0.5cm$, 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 [100, 101], 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:

$$\frac{|M_i - M_{i+j}|}{M_i} < 0.15 \quad \forall j = 1, \cdots, k \qquad \textit{with k=10} \tag{4.4a}$$

$$M_i < 3.5cm \tag{4.4b}$$

   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 approximation of the neck's radius.

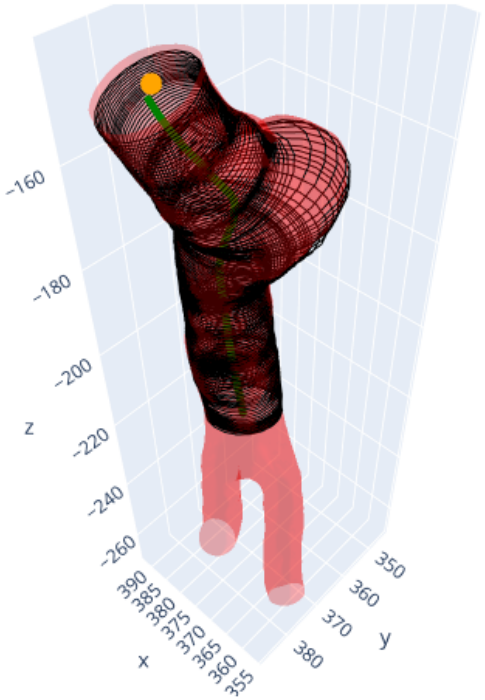


Figure 4.6: A visualization of the 3D mesh of an AAA with an example of two fitting ellipses. The green curve is the centerline while the orange point is the neck seed

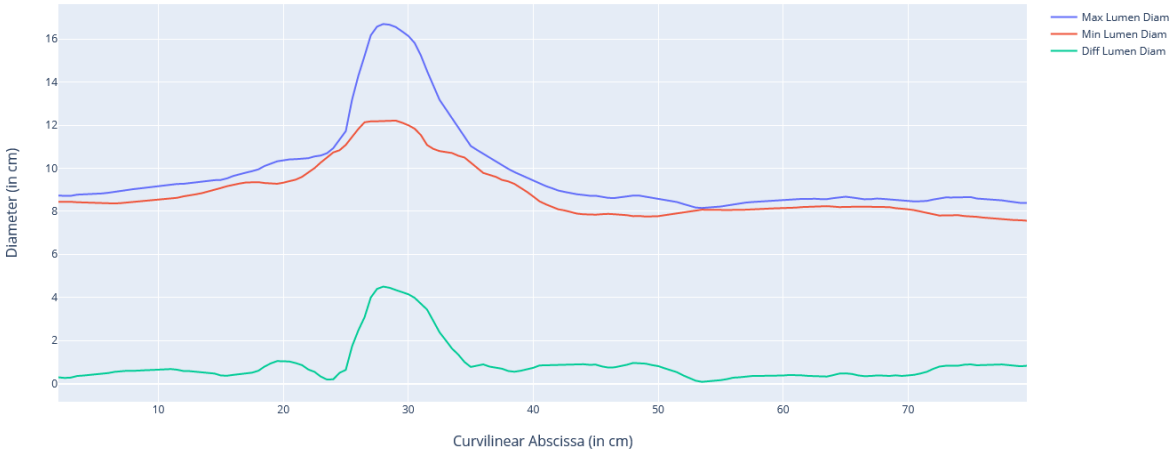


Figure 4.7: A plot of the semi-major axis (in red), the semi-minor axis (in blue) and their difference (in green). The x-axis represent the curvilinear abscissa s. Both axis are in mm.

In this way we build an algorithm in order to keep in check the rising value of the aorta's maximum diameter as we walk it down towards the AAA.

Moreover, while the ellipse approximation of aorta's slice can be inaccurate at the height of the AAA, the sane portion of the aorta, the neck, is regular enough to justify this method and the AAA's possibly wrong approximation is not considered. At the same time using fitting ellipse built in this way is more robust to errors can can happen at the moment of the segmentation. In this way, an estimate of the average neck radius is obtained.

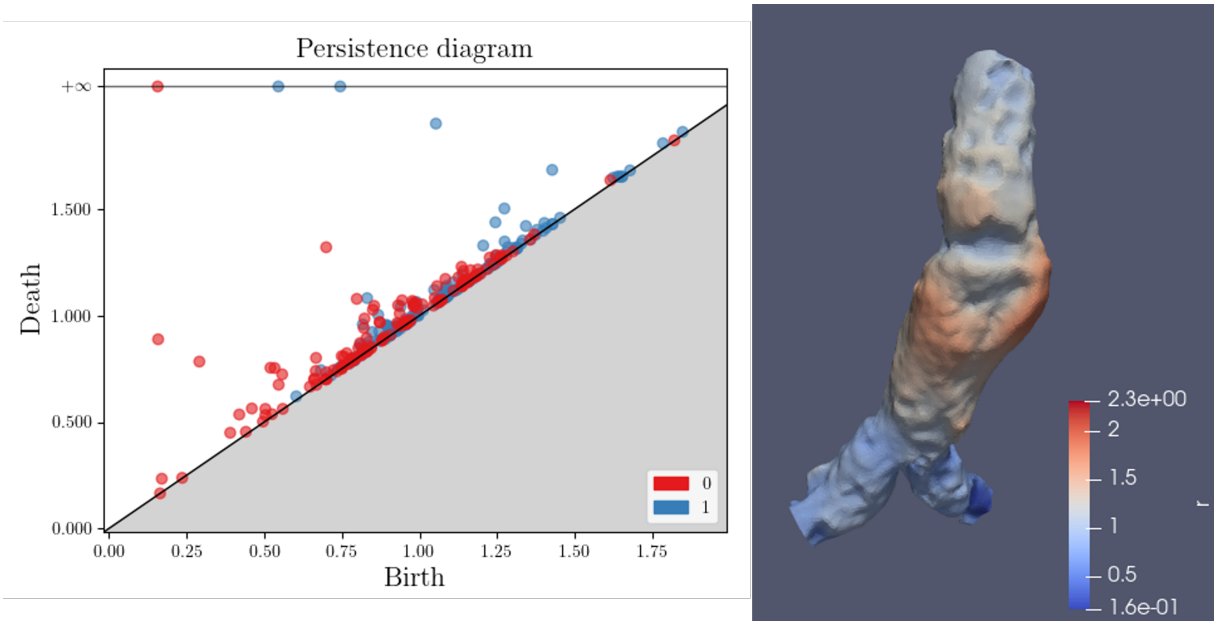


Figure 4.8: The starting mesh, coloured according to each points' filtration value and the resulting Persistence Diagram. Further details will be analysed in the next chapters.

4.3. Critical Points of the Pipeline

The major critical points of this procedure are two:

- Seed selection: The selection of seeds, the points on the CTA scan that define the region to segment are taken manually due to the difficulty in the implementation of a fully automatic segmentaion algorithm. However, points are manually chosen using the same criteria for all the patients, leading to consistent results.
- Number of Points: The number of points in each mesh and centerline is variable, due to the difference in volume and surface of each distinct aorta. Consequently, this is true for the number of edges and faces influencing also the centerline's algorithm.

So, the maintenance of the same mesh's resolution for each mesh on one hand assure a certain degree of consistency, while on the other hand makes the smaller aorta's meshes less refined respect to the bigger ones. Anyway, distances developed to compare Persistence Diagrams are specifically made to compare diagrams with different number of persistence pairs and the difference in mesh refinement is negligible.

The procedure has been conducted on *python 3.8* using the packages *ripser 0.6.4* [102] for computing the Persistence Diagram and *persim 0.3.1* [103] for the computation of the distances. In this way, all 22 meshes have been turned into Persistence Diagrams, ready to be visualized, analysed and studied.

5 | Results

Persistence diagrams are a tool used in topological data analysis TDA to study the shape and structure of complex data. A persistence diagram is a visual representation of the birth and death "times" of topological features, such as connected components, loops, and voids, that persist across different scales of a dataset.

Chapter 3 and 4 contain the long series of procedures and choices that led to the creation of the Persistence Diagrams, the final data of this study, on which the analysis that are the real core of this work are based. The ability of Persistence Diagrams to explain, visualize and summarize the topological features seems to be especially suited for the aim of this thesis.

The software used to visualize all the meshes, the CTA scans and produce the plots that will be showed in the next pages is *Paraview* version *5.11.0*.

5.1. H0 and H1 Cycles

Before looking at the Persistence Diagrams themselves is important to understand the morphology of an AAA, the effects of features such as thrombus or calcifications have on it and the physical meaning of element in H0 and H1 (zeroth and first homology groups).

After this step it will be shown in detail how each aspect of this pathology affects the topology of the AAA and consequently the associated Persistence Diagram.

First of all is important to see what is an element H0 or H1 on the mesh itself.

- H0: A 0-cycle is simply a path connected component. Note that any 1-cycle is also contained in a 0-cycle since is made by path connected simplexes. Roughly speaking, a 0-cycle is any set of connected points, from the single detached one to the full simplicial complex, e.g. the complete mesh.

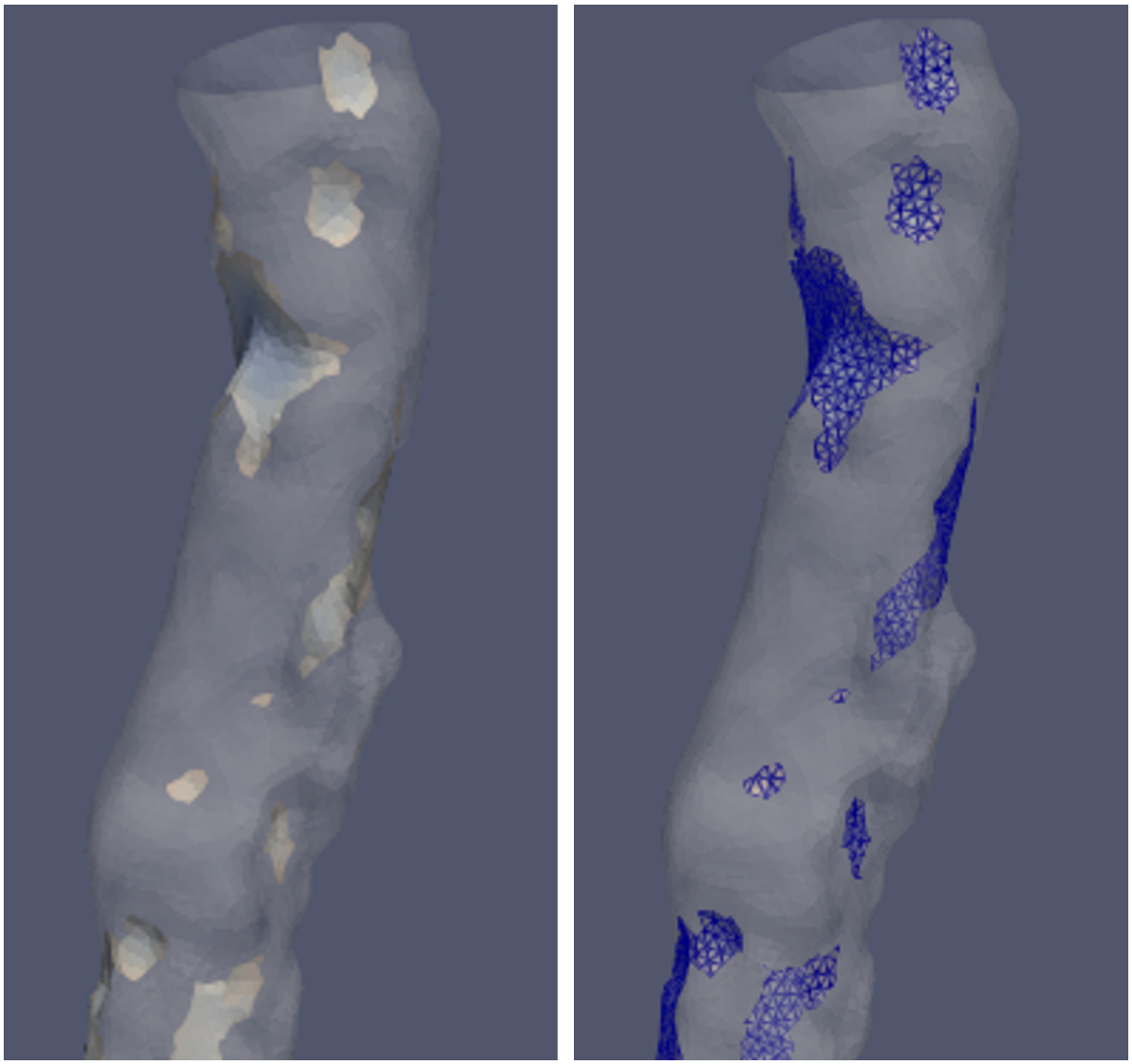


Figure 5.1: Example of many 0-cycles, with no H1 loops.

- H1 Loop: Closed paths surrounding an hole are 1-cycles and can be divided into two groups, as showed in Figure 5.2:

    1. The natural surface of the aorta is equivalent to a tube. If the points create a ring around the centerline, a 1-dimensional loop is made. These cycles are never filled up as the filtration value increases but can die when merged with another 1-cycle that is born earlier.

    2. The surface of the aorta can be irregular, presenting several outgrowths that create a temporary hole that will be filled up as the filtration value increases. All these cycles will eventually die by being filled up or merged.

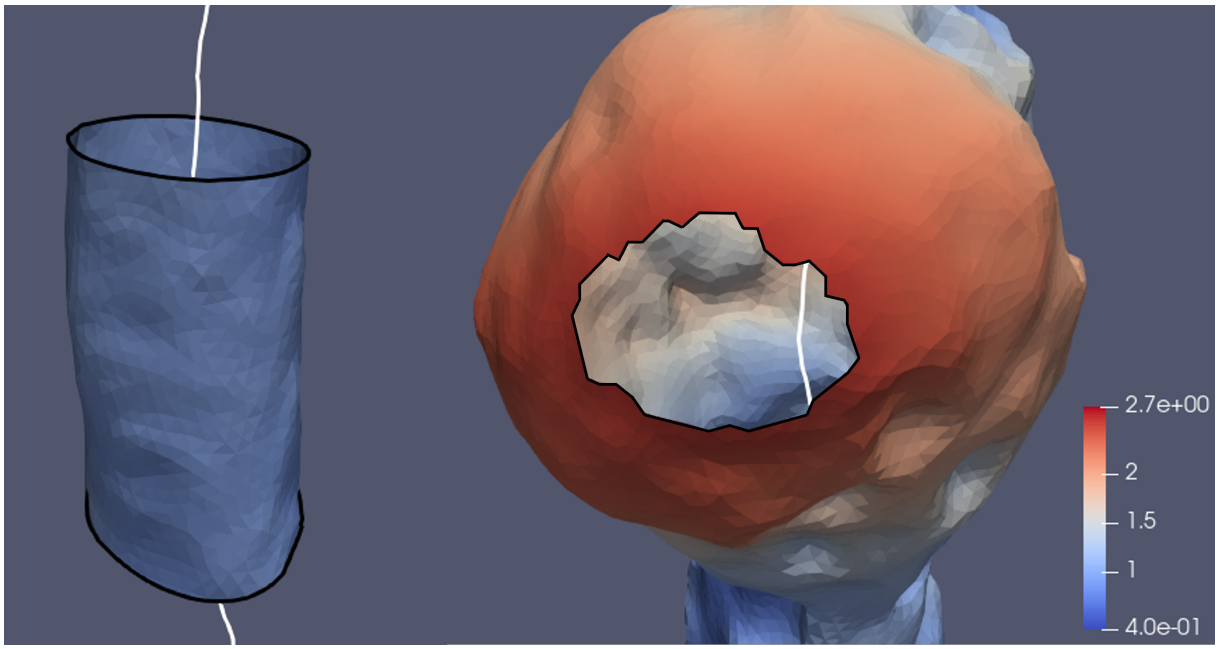


Figure 5.2: An example of the two types of 1-dimensional loop, with the centerline and the meshes' contours highlighted respectively in white and black. In the left figure it is possible to see the first type of 1-cycle, that surrounds the centerline due to the aortic vessel being in a tubular shape. Note that the loop on the top and the one on the bottom are equivalent: they represent the same element in H1. In the right one instead is shown the second type of H1 loop, in fact is caused by a bulge in the surface (the AAA in this case) and does not envelops the centerline.

5.2. The Role of Filtration

The role of the normalization of filtration can now be explained. Normalization is a mean to reveal at which distance from the centerline a persistence pair is born and dies relatively to the size of each patients' mean aortic neck radius. This choice is related to the fact that the aortic neck is the only portion of aorta assured to not be affected by calcification, thrombus or AAA. With reasons similar to those explained in Chapter 2 of this thesis, the normalization process is a way to compare different patients using the healthy part of the abdominal aorta as a baseline. In this way it is possible to focus on the relative growth of the AAA rather than its raw dimensions.

In fact, using this procedure it's assured that the value 1.0 on the x and y-axis of the Persistence Diagram is referred to a point that is born or dies at the same distance from the centerline as the aortic mean neck's radius.

Moreover, if the diameter of a sane Abdominal aorta is, in mean, around 2-3 cm, the one

of an healthy iliac artery is in the range of 1.0-2.0 cm [104], that is around 0.5 to 0.8 after standardization. These values shrink when looking at a point affected by calcification coming to a range of 0.2-0.4.

If an iliac aneurysm occurs, the diameter of iliac artery can increase up to 1.0 in the most concerning cases. In this way, anything that is collocated near the value 1.0 is located around the sane portion of the aorta or, in rarer cases on an aneurysmatic iliac. Points with a lower filtration value are instead below the aortic bifurcation. Then, points at AAA's height can instead have a more variable range, from 1.0 up to 7.0 and more in the most concerning case (See Figure 5.3 for a schematic visualization).

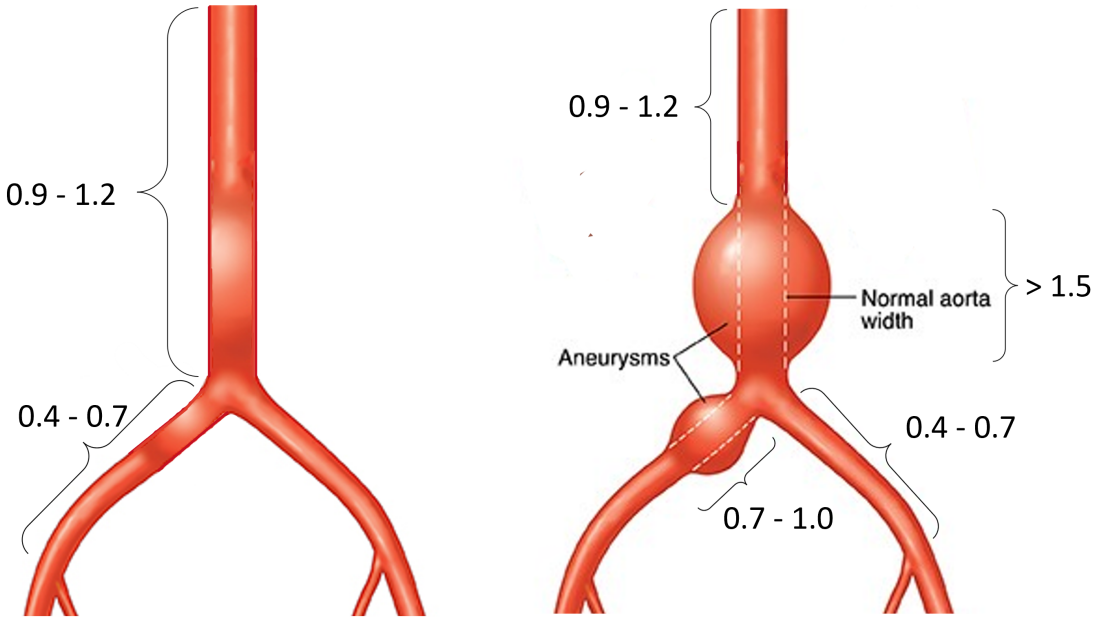


Figure 5.3: A visualization of an healthy and an aneurysmatic aortas' ranges of filtration.

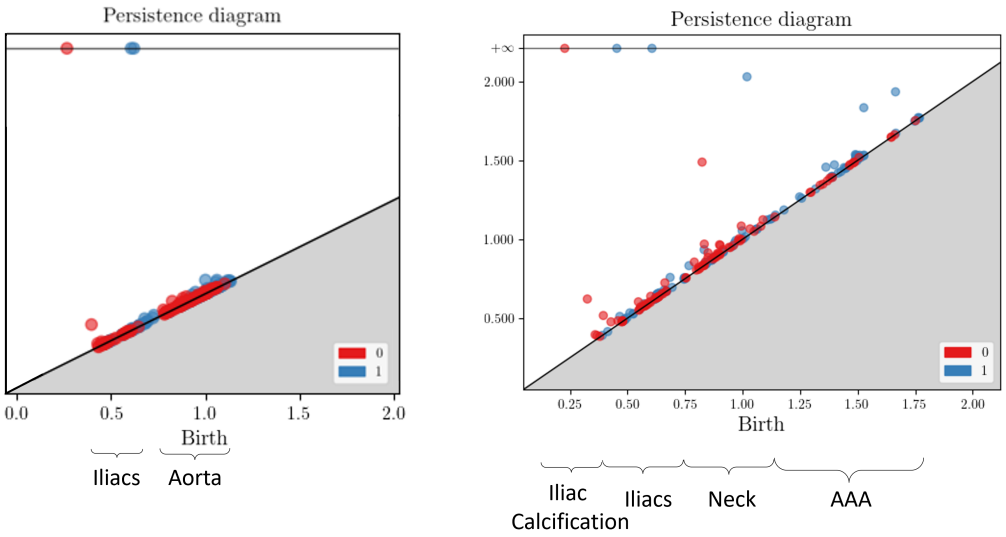


Figure 5.4: A translation of the previous image (Figure 5.3) on the Persistence Diagram. Here is also possible to see an initial comparison between the diagrams of a completely healthy aorta and an aneurysmatic one. It is worth to note that, on the healthy aorta's, H0 points (in red) that belong to an iliac artery and the ones on the main aortic body are clearly separated

These are the upstream reasons that brought to the choice of this particular filtration. In this way, it is possible to use an important indicator of the AAA, as the relative size (i.e., normalized distance), having at the same time an evidence of where the k-cycle is born or dies on the aorta.

5.3. Persistence

A persistence pair is characterized by three variables: type of cycle, birth and death, with the formers explained in the previous pages. Instead of considering a pair's death, its persistence can be equally considered ($Persistence = r_{Death} - r_{Birth}$). The more a pair persists, the more relevant is the topological features associated to it. So, the ones associated to a steep bulge or concavity are the further from the diagonal.

On the other hand, if a pair fades just after being born it means that the topological cycle associated to it is not relevant and is just a small group of points with a filtration value slightly smaller than the others surrounding it. These irregularities are something completely normal and physiological in a real, biological, structure. All these pairs lie near the main diagonal and could be removed to obtain a clearer diagram.

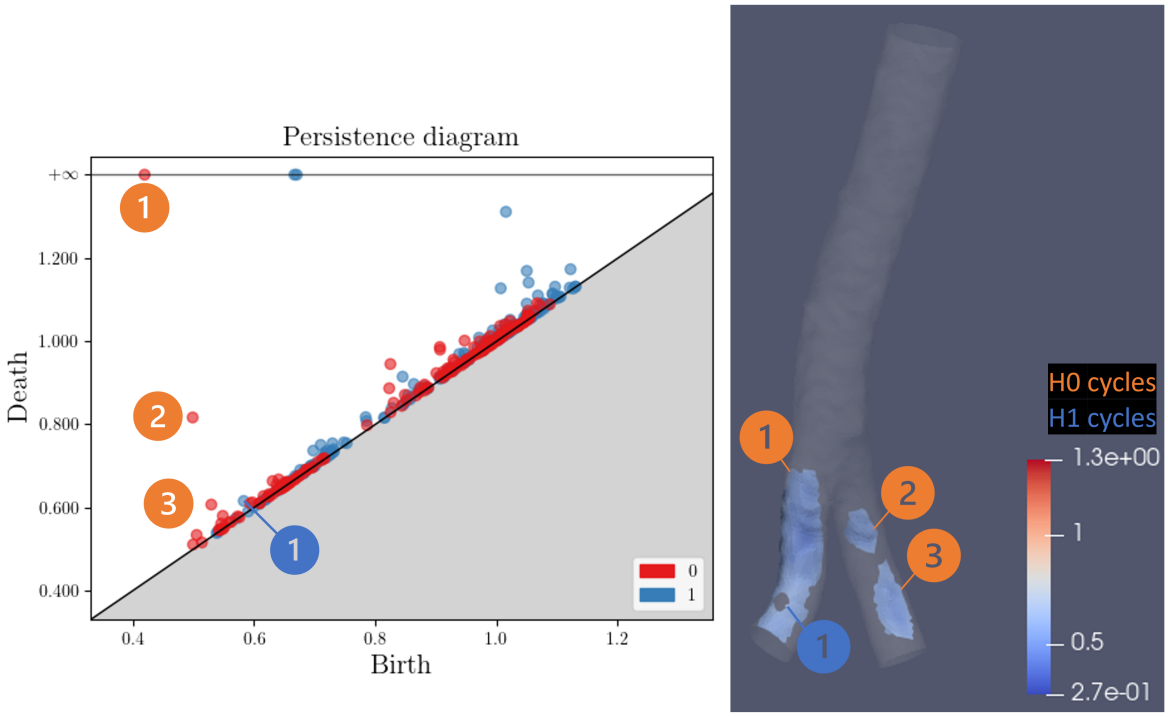


Figure 5.5: An example of a 1-dimensional cycles with very low persistence. It is not associated to any relevant feature of the aorta. The 0-dimensional cycles (1) and (2) represent instead relevant feature, being the two iliac arteries.

5.4. Application of Persistent Homology

The final simplicial complex Σ, obtained by adding all the simplexes to the filtration, is an object with a single 0-cycle (since there are no not-connected simplex) and two 1-cycles.

The existence of two 1-dimensional cycles and not three, as the number of the branches considered in the segmentation (aorta, left iliac and, right iliac artery) could suggest, is not trivially explained. It can be understood by using homological equivalence and visualizing the mesh like a tube (aorta and one of the two iliacs) with an hole on its wall (the other iliac). For this reason, all the Persistence Diagrams must have only one H0-cycle and two H1-loops with infinite persistence to be correct. This property is something worth noticing since it is an initial evidence on how strictly linked are the mesh and the associated persistence diagram.

These are the most relevant homological features, one being the whole set of points and

the other two representing the two iliac branches. These three points are also used as a proof of correctness of the Diagram. In fact, if a diagram has more of these pairs an error has been made during the Segmentation phase, such as the creation of points not connected to anything else or the presence of an hole in the segmentation.

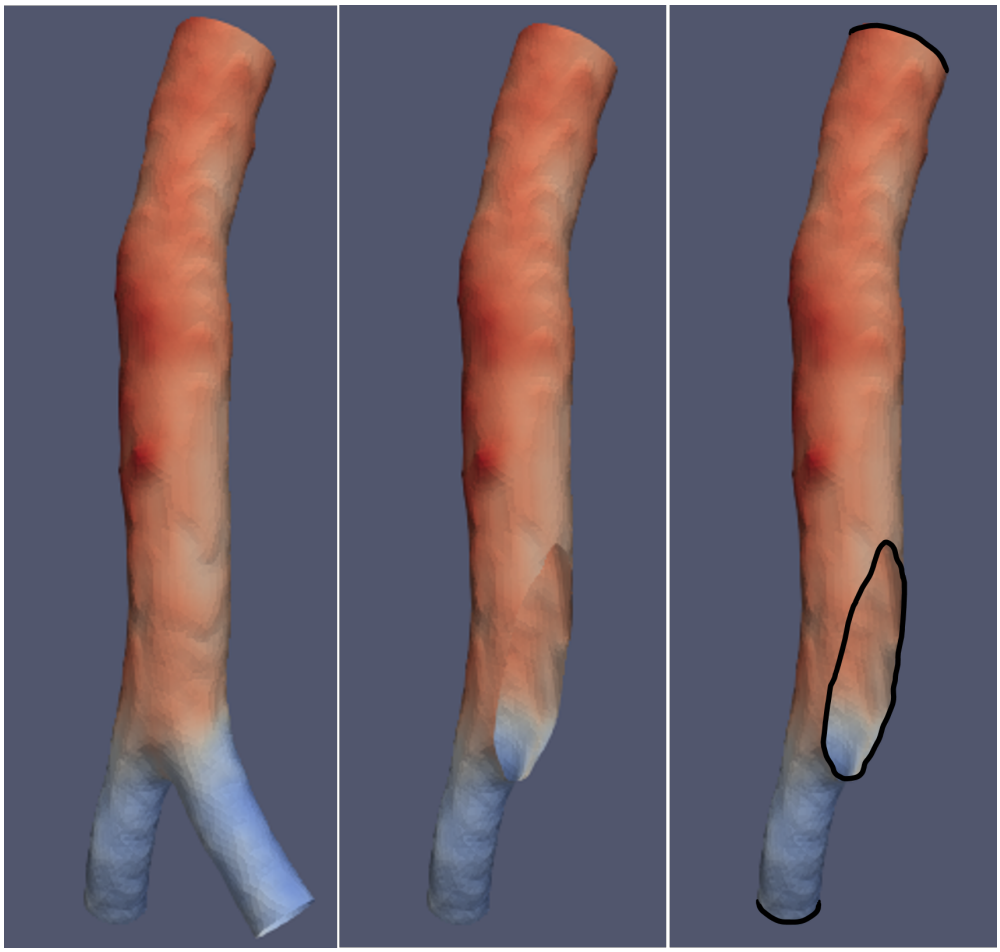


Figure 5.6: Example of a healthy aorta (on the left) and the homopogical equivalent tube with an hole (on the right), the same aorta without an iliac artery.

There are three other relevant points with high persistence:

1. The first vertex (point) that is added to the filtration is always from one of the two iliacs. That same point establishes the birth of the single 0-cycle with infinite persistence, as explained in the previous pages. The first point that belongs to the other iliac added to the filtration established the birth of a second early 0-cycle with high persistence that will be merged in the first one as the filtration value rises (see point "(2)" in Figure 5.07). All except the Persistence Diagram with an iliac aneurysm have this point. In fact, the creation of this 2 persistence pairs is linked to

the iliac arteries' smaller radius. If the iliacs have comparable diameters there are two clusters of points that are added at a similar filtration value bulding up these 0-cycles that merge together. If there is an iliac aneurysm, only the points of the healthy iliac are normally added merging together creating the 0-cycle with infinite persistence. On the contrary, since the aneurysmatic iliac has an higher diameter, and thus its points have an higher filtration value with respect to an healthy one, the points on the second iliac will be gradually added to the other one without creating a 0-cycle with high persistence.

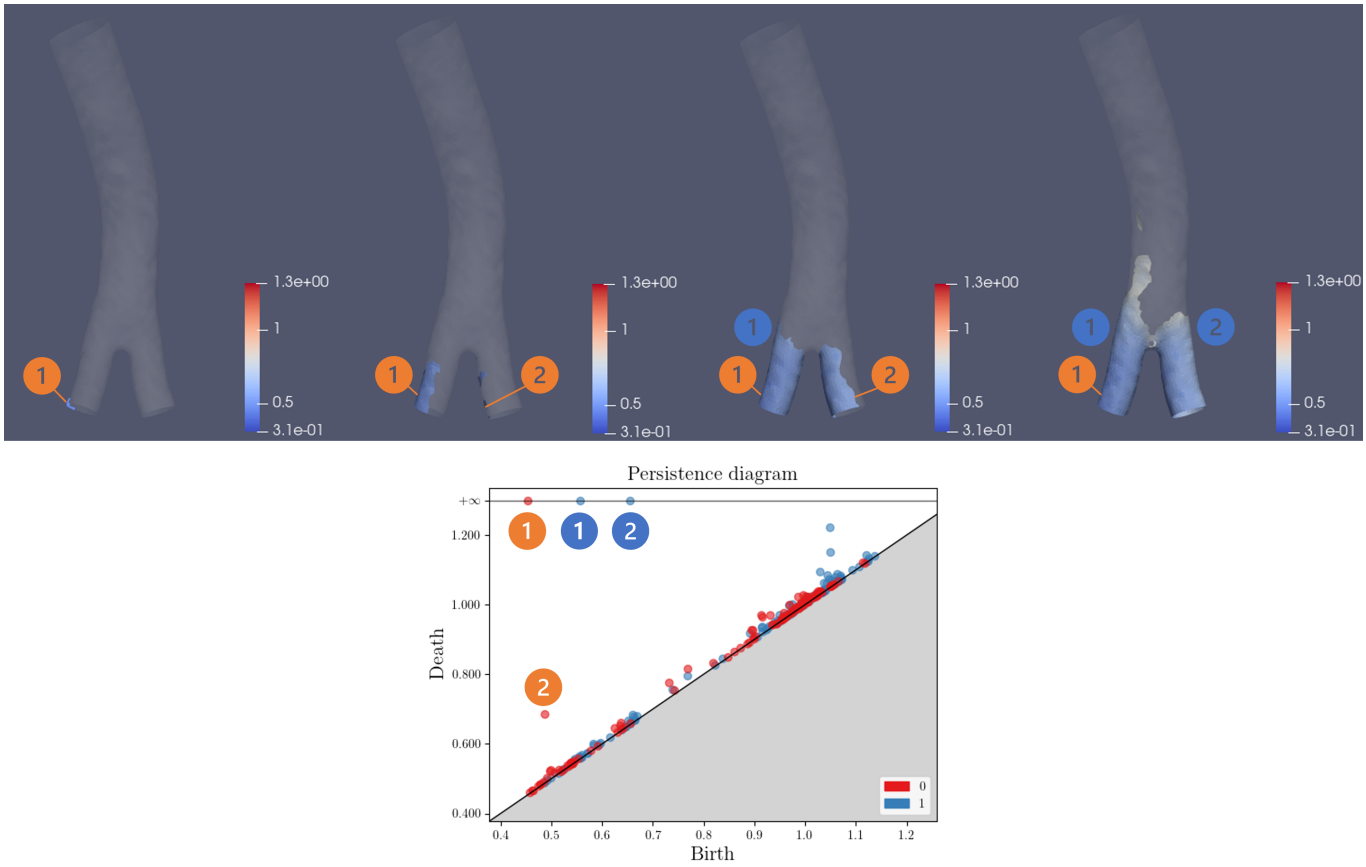


Figure 5.7: Here a visualization of the mesh of an healthy aorta at various filtration's values. In order. 0.45: the first points are added to the left iliac artery creating the first 0-cycle, the one with infinite persistence. 0.5: after a short span also points from the right iliac are added, creating the 0-cycle with the second highest persistence. 0.6: the left iliac artery (the smaller) is almost complete and the first 1-cycle with infinite persistence is created. 0.7: the two iliacs are completed, the 0-cycle mentioned before has been merged with the first one and the two 1-cycles representing the iliac arteries are finally created. Some points of the sane portion of the aorta are starting to join the simplicial complex.

2. A third notable 0-cycle with high persistence can be encountered in Persistence

Diagrams of aneurysmatic patients and is associated with the aortic neck. When an AAA occurs there is a "separation" between the mesh's points (see Figure 5.8). In fact, the AAA's points between the neck and the bifurcation have a filtration value noticeably higher than the rest and create a barrier, with respect to the filtration value, of point that separates the iliac arteries from the neck, as can be seen in Figure 5.8.

3. The final notable cycle created by the AAA is the cycle associated with the tubular structure of the neck. This is the last cycle to die and it is merged with the first 1-cycle associated to the first iliac artery when the very last point of the mesh is added to the filtration. This point is recognizable since its birth is around the value 1.0 and its death has the highest filtration value among all the diagram's points.

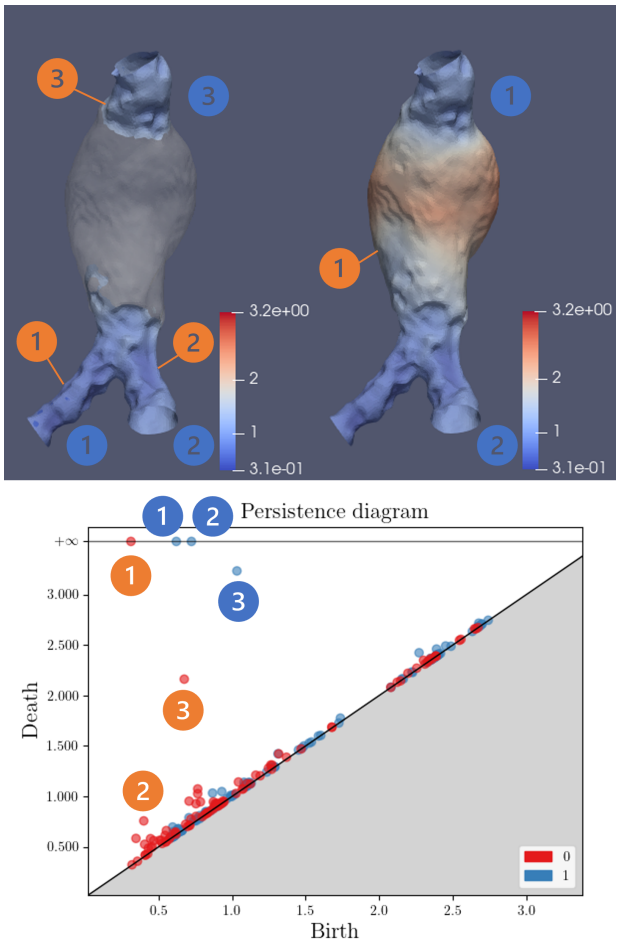


Figure 5.8: Here is possible to see the division between the filtration values caused by the AAA mentioned above. This is responsible for the creation of the two notable loops marked with "(3)".

To summarize, there are two main differences that make the persistence diagrams of patients with AAA easily distinguishable from the others:

1. The persistence pairs of AAA patients' can born (and die) at a bigger values. In fact, no healthy aorta has a point with a distance from the centerline higher than 1.4, while the points of an aneurysmatic aorta can exceed the value of 7.0.
2. The presence of at least one 0-cycle and one 1-cycle with high persistence, due to the presence of the AAA as explained above.

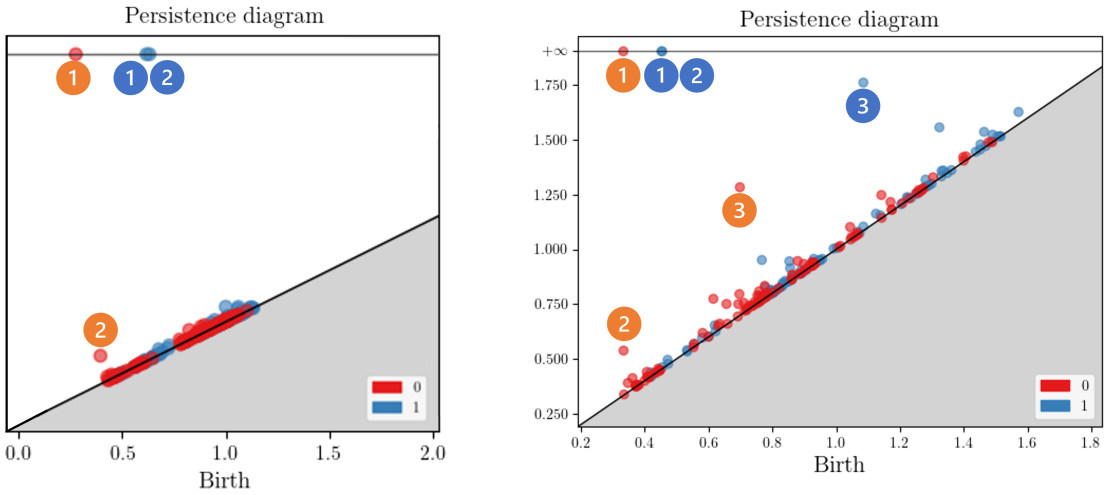


Figure 5.9: A comparison between the Persistence Diagrams of patients with healthy (on the left) and aneurysmatic (on the right) aorta. The fact that the points in the last diagrams are distributed on a bigger scale is generated by the presence of an AAA. The loops marked with the symbol "(3)", are associated with the aortic neck.

Another phenomenon to note is that the 0-cycles are mainly distributed in the first part of the diagram, in opposition to the 1-cycles, that are on the second half. This is due to the fact that H1 Loops need more points to be created.

Is now possible to see in detail some wall's characteristics often observed in AAA.

- Iliac aneurysm: The aneurysm can also affect one or both the iliac arteries in the same way it affects the abdominal aorta, resulting in a stretch of the wall. Iliac aneurysm can be also associated with calcification and thrombus, in the same way as the AAA.

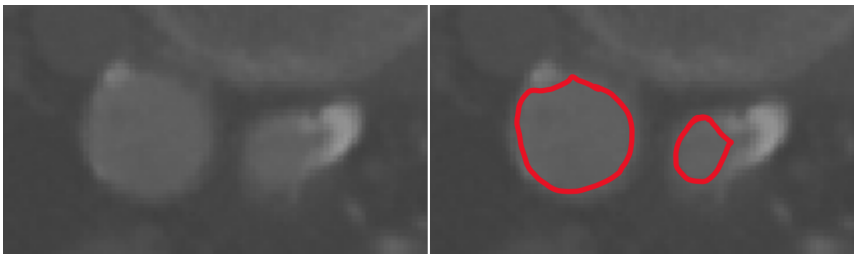


Figure 5.10: Here a CTA scan slice of a patient affected by iliac aneurysm on the right iliac artery. The difference in dimension is accentuated by the presence of calcification on the left one.

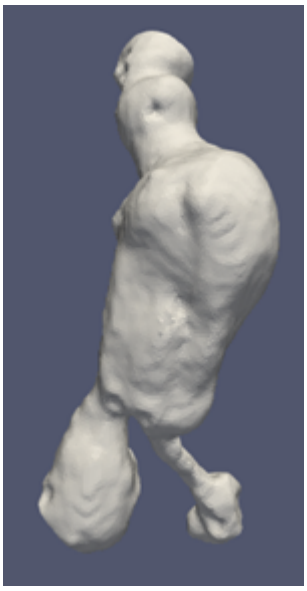


Figure 5.11: Here the segmentation of the same CTA scan showed in Figure 5.10, allowing a 3D visualization of the condition of the patient's aorta.

The presence of an iliac aneurysm prevent the formation of the second notable 0-cycle (see Figure 5.12). The points are firstly added to the healthy/calcified iliac, and, only after it is completed, the points on the aneurysmatic one are slowly added to the pre-existing 0-cycle that represent the iliac rather than create two independent 0-cycles (one per iliac) that are merged subsequently.

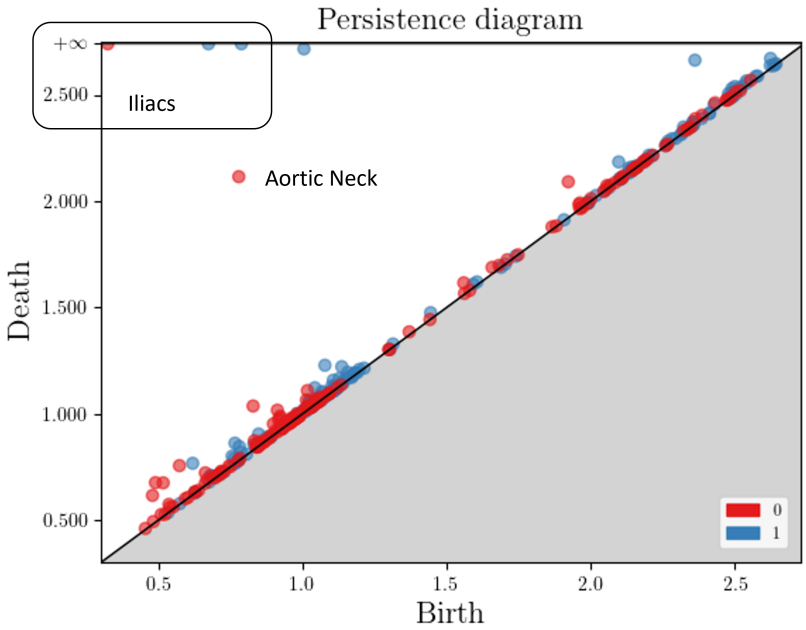


Figure 5.12: The Persistence diagram of a patient with an aortic aneurysm and iliac aneurysm. The lack of an early 0-cycle with high persistence is visible.

- Calcification: calcifications are always attached to the aortic wall, on the internal side. Being a calcium formation, the calcification creates a local thickening of the wall, resulting in the formation of indentations in the lumen, when in contact with the calcified wall.

  Calcifications are usually located on the iliacs and on the AAA's body, rarely on the sane portion of the aorta. When calcifications form near each other can create the opposite effect, resulting in an outgrowth of the lumen (see Figure 5.13, 5.14 and 5.15).

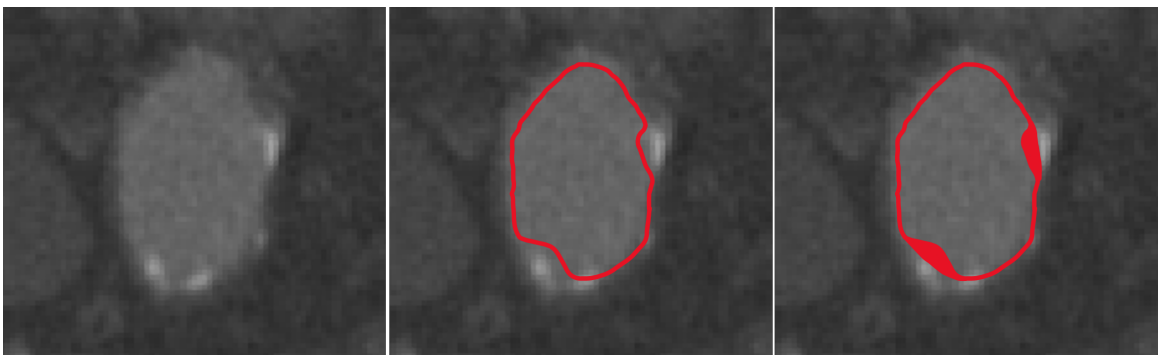


Figure 5.13: CTA slice with separated calcifications

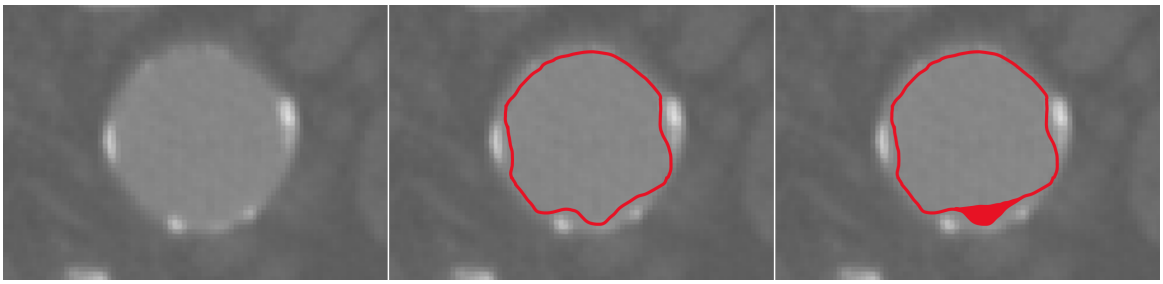


Figure 5.14: CTA slice with multiple calcifications

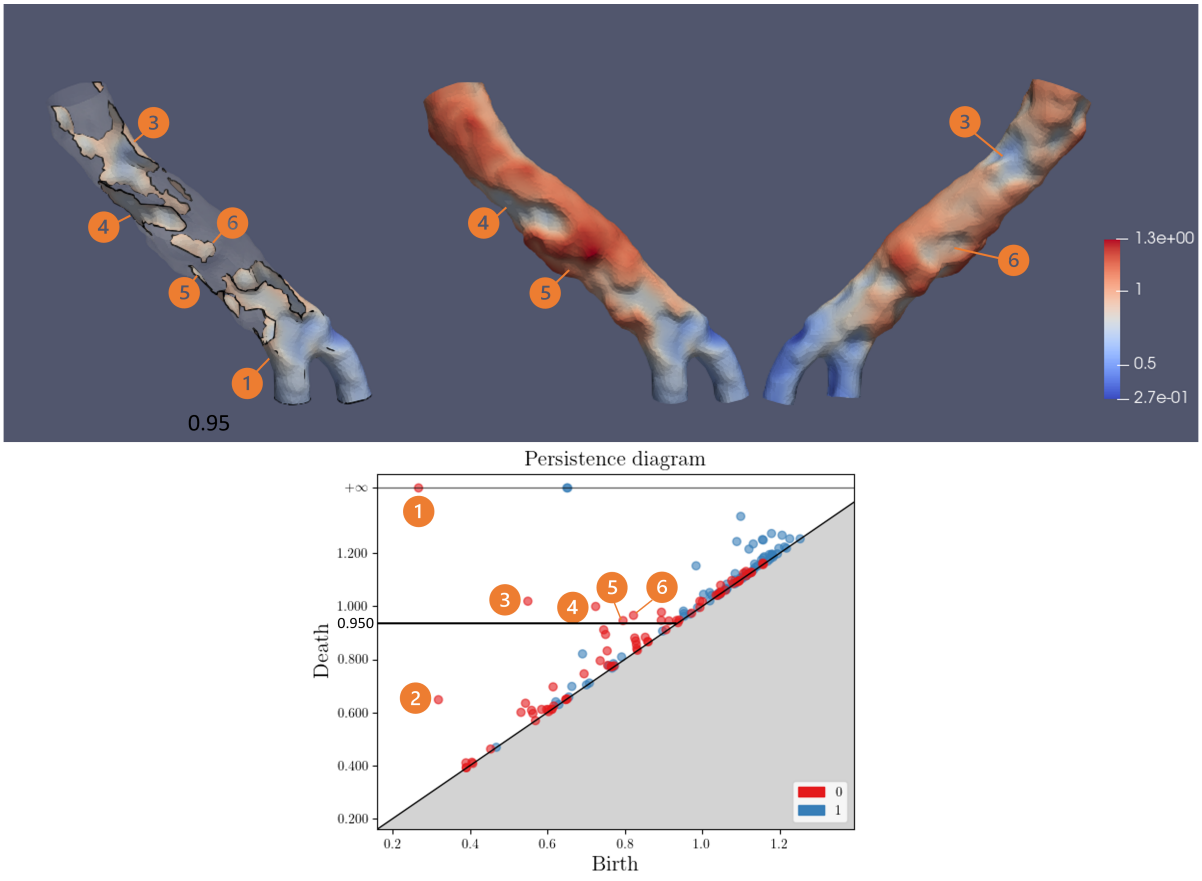


Figure 5.15: In this example a non aneurysmatic aorta with calcifications can be see. Calcifications can be easily recognised in the Persistence Diagram. In the left image, the filtration value is fixed at 0.95, just under the diameter of the healthy parts of the aorta. The 0-cycles numbered as "(1)" and "(2)" are the mandatory H0 cycles of the two iliacs, while "(3)", "(4)", "(5)" and "(6)" are calcifications. The more the calcification is deep and steep the bigger the associated pair's persistence is.

- Thrombus: Thrombus is often a non-homogeneous substance, which can flake off with the flow of blood creating wide ledges in the lumen (see Figure 5.16, 5.17 and

5.18). Thrombus is most likely to be located in the area most affected by AAA, due to the concentration of metabolites associated with inflammation.

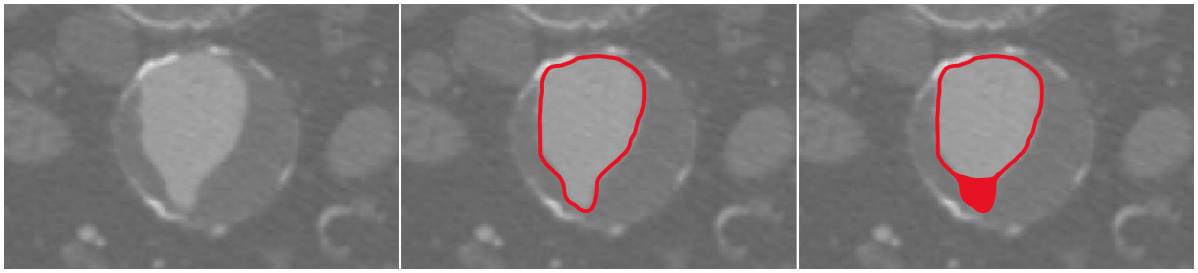


Figure 5.16: CTA slice of an AAA with thrombus

5.5. Detailed Description of an Aneurysmatic Aorta through Persistent Homology

All the results explained in the previous pages have been manually verified by simultaneously looking at the Persistence Diagram and at the simplicial complexes created by the filtration. It is also of particular interest to analyse some of these comparisons, in order to understand the power of this tool in describing this particular pathology. In fact, as will be shown in this section, is possible to retrieve with stunning accuracy each of the topological features creating each point of the persistence diagram and locate it on the aorta's mesh.

The focus is on a severe case of AAA, with thrombus and iliac calcification, to better understand how the methodology works and analyze how the topology changes as the filtration value, r, rises.

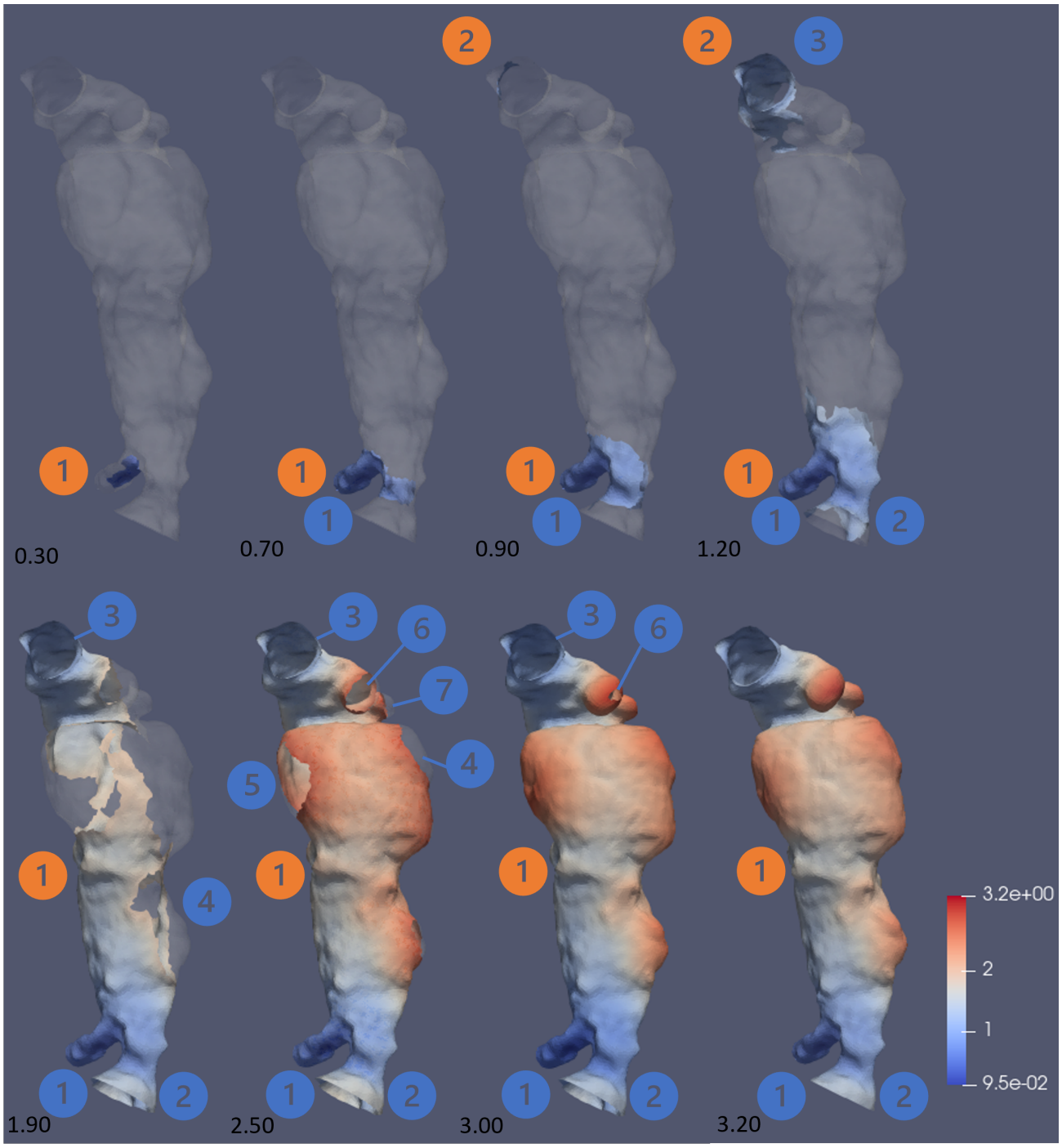


Figure 5.17

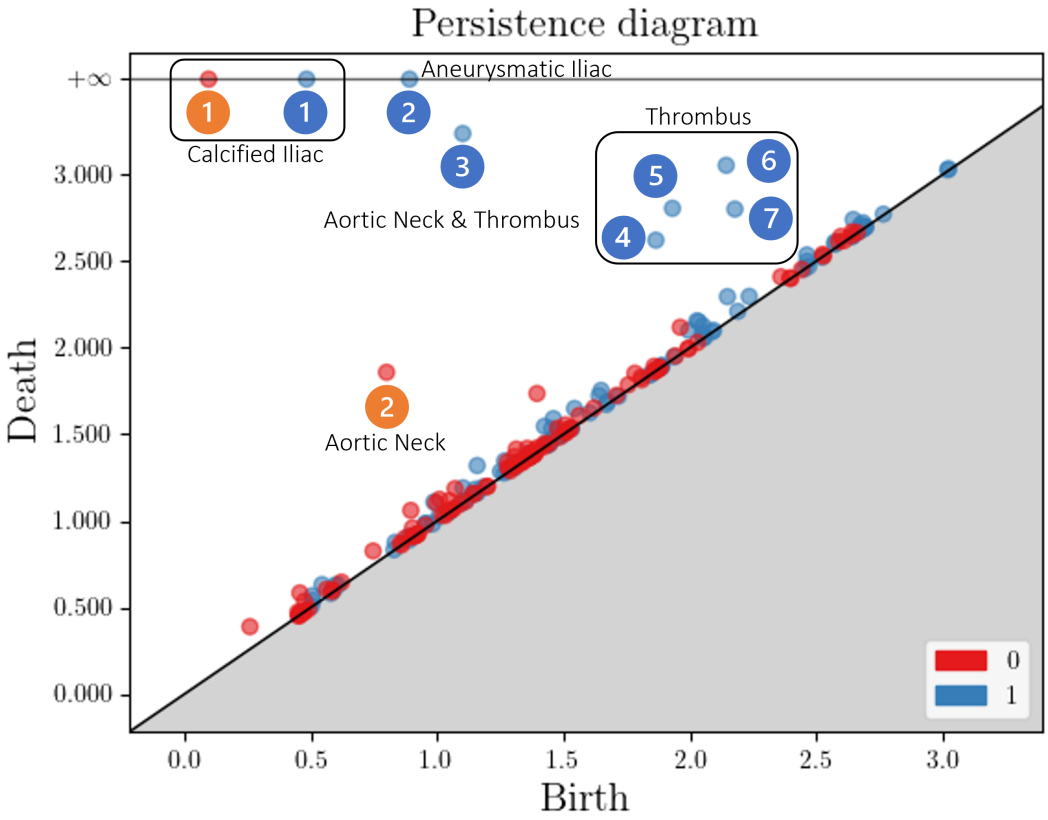


Figure 5.18: The Persistence Diagram associated with the mesh shown in Figure 5.17.

In the first image of Figure 5.17 ($r = 0.30$), the first, small, cluster of points has been already added to the filtration. In fact, because of the iliac calcification, the minimum value of r is 0.0912476.

In the second and third image ($r = 0.70$, $r = 0.90$) is visible the previously described effect of the iliac aneurysm: while all the points of the calcified iliac are added to the filtration, with even the first 1-cycle created, only a part of the aneurysmatic iliac has been recruited to the simplicial complex. This makes the appearance of the second iliac's 0-cycle not possible. Moreover, some points of the neck are starting to appear in the third image.

In the fourth image ($r = 1.20$) it is possible to see the appearance of the second 1-cycle with infinite persistence, the one of the aneurysmatic iliac, born at a filtration value over 0.90. This value is significantly higher the value found in non-aneurysmatic iliacs.

At $r = 1.90$ the two connected components of the iliacs "(1)" and aortic neck "(2)" are now merged together, so that the simplicial complex consists of a single connected

component. Moreover, the thrombus' effects are starting to be visible with the H1 loop "(4)".

Increasing r up to 2.50, the 1-cycles created by the presence of deep inlets in the trombus became more evident.

The second-last image shows the final step before the completion of the mesh. The majority of relevant 1-cycle that could die has died with the exception of "(6)", related to a recess in the thrombus. The hole representing the end of the H1 Loop "(1)" is not visible since is covered by the mesh but is existent and is as well an indent in thrombus, prolonging the life of the H1 Loop "(3)" well over the filtration value 3.00.

Finally, the last image shows the complete mesh, with only the three canonical cycles with infinite persistence.

5.6. Distances and MDS

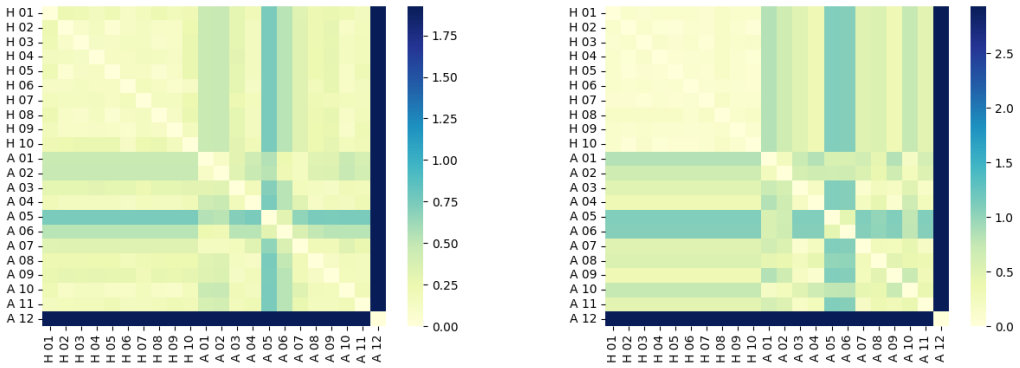
The previous section was dedicated to understanding how Persistent Homology translates the aortic features in persistence pairs, mostly in a qualitative way.

Given the excellent results, the applications of Persistent Homology to this clinical field could extend to classification and clustering, to stratify patients based on risk of rupture, morphological features, or AAA's evolution. To do that, it is necessary to add a definition of distance, already available thanks to the results of the studies in Persistent Homology [95].

The two notions of distance used, Bottleneck and Wasserstein, introduced in Chapter 4 can be used now to quantify the differences between each AAA.

Using these distances, both the similarity in the sense of 0-cycles and 1-cycles has been quantified obtaining four heatmaps. In all the heatmaps the first ten patients have healthy aortas, while the twelve remaining are affected by AAA.

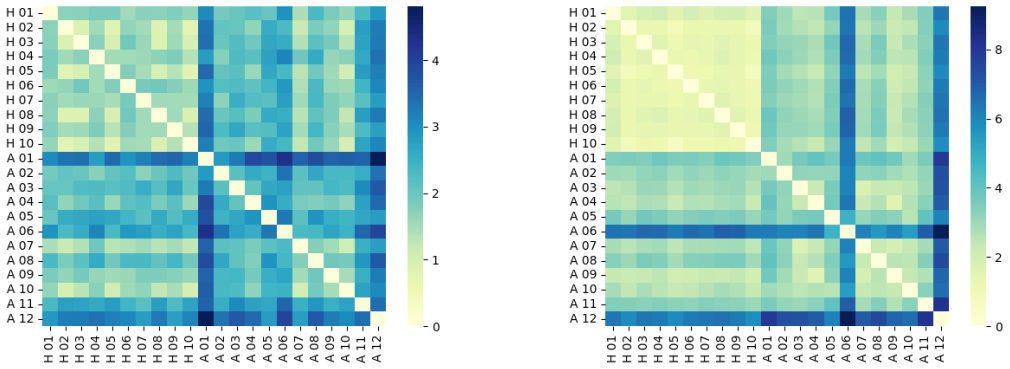
All the distances are computed using the python library *persim*, while the heatmaps are plotted using *seaborne*.



(a) (b)

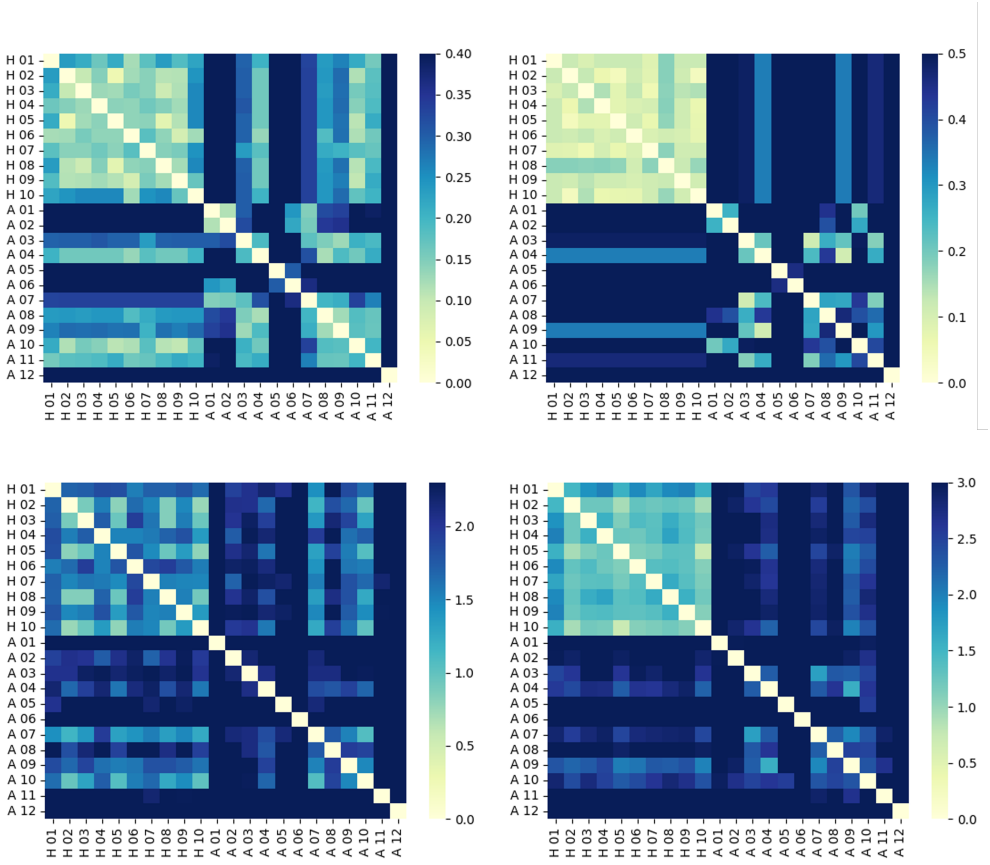
Figure 5.19: Heatmaps based on the Bottleneck distance matrices comparing each couple of persistence diagrams, the one on the left considers only H0 cycles while the one on the right the H1 loops.

The results obtained using Bottleneck distance are coherent with the definition despite having different couples of diagrams associated with the same distance value. The reason behind this is that this distance is the maximum weight in the solution to the minimum weight perfect matching problem [95]. Since the points in the diagonal are included to make a solution always existent, the matching occurs between the "furthest" pair of one of the two diagrams (the one that drives the same results) and one point on the diagonal of the other.



(a) (b)

Figure 5.20: Heatmaps based on the Wasserstein distance matrices comparing each couple of persistence diagrams, the one on the left considers only H0 cycles while the one on the right the H1 loops.



(a)

Figure 5.21: The same heatmaps showed before (same order) with the color range capped, to highlight the closeness of healthy patients to each other.

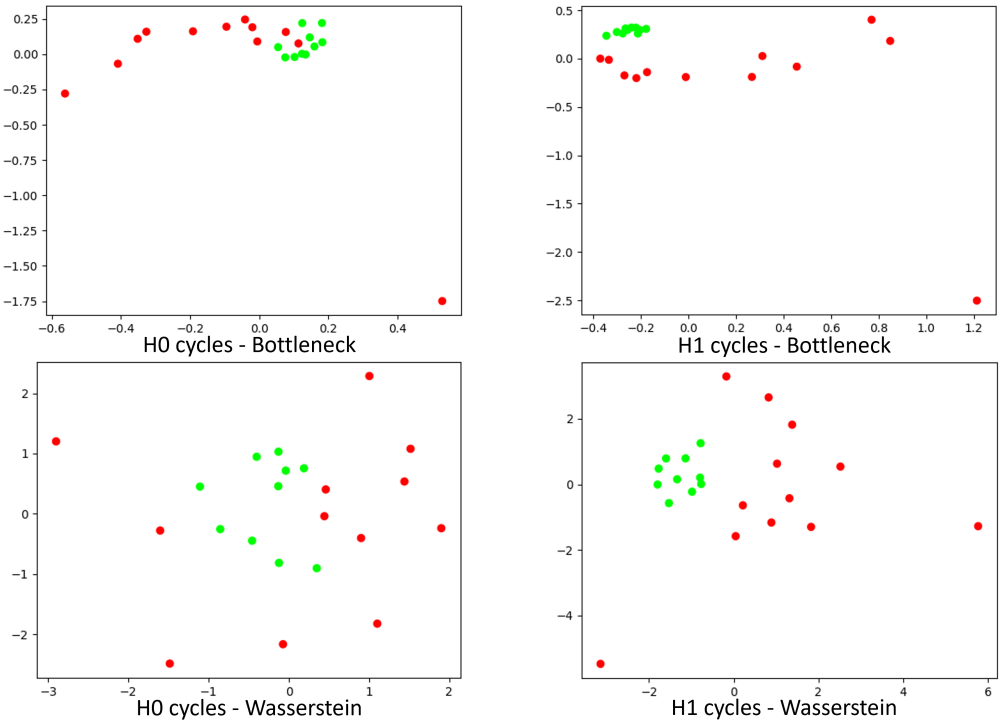
The fact that healthy patients are closer to each other is expected. In fact, healthy aortas have a contained variability, that is even reduced by the filtration's normalization. On the other hand aneurysmatic aortas can, in their pathological state, be so vastly different in terms of morphology that even the normalization is not enough.

Something that is worth to point out is that this distinction between healthy and aneurysmatic aortas is trivial. In fact, at the light of above, is enough to remember that only an aorta affected by AAA can have persistence pairs whose filtration value is well higher than 1.0.

The distance matrices computed contain the dissimilarities between all pairs of objects. To have a better comprehension of the results could be worth to have another visual representation. One of the best tools for this purpose is Multidimensional Scaling (MDS) [105]. The goal of MDS is to represent these pairwise distances (or dissimilarities) as points in an Euclidean space with given dimension while preserving their original relationships

as much as possible [105].

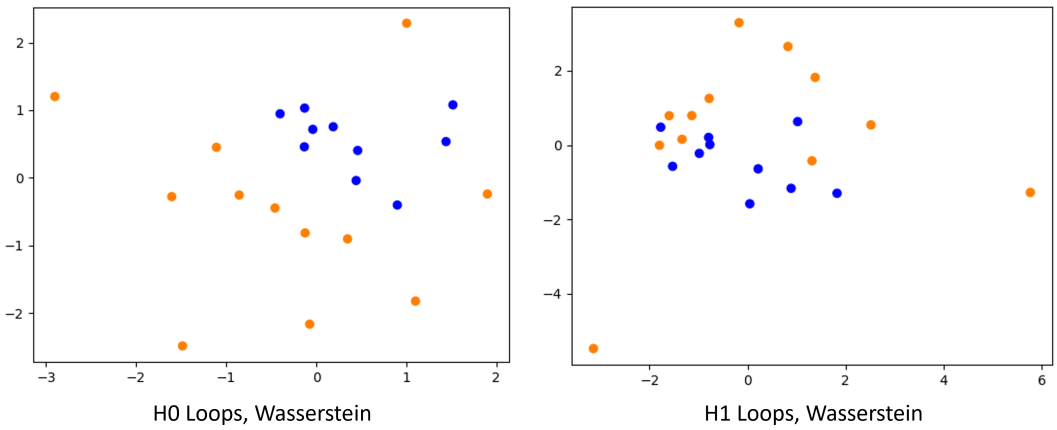
This can help to identify clusters or patterns in the data, which may provide insights into underlying structures or relationships between the objects. In this study MDS will be used both to check and to understand the results obtained through persistent homology.



(a)

Figure 5.22: The four plots resulting from MDS using two dimensions.

Looking at Figure 5.22, the green points are associated to an aorta without an aneurysm, while the red ones are associated to aneurysmatic aortas. As expected, by looking at the heatmaps in Figure 5.22 healthy patients are closer to each other and are distributed around a concise region, in a way that the aneurysmatic aortas could be considered outliers. This is especially evident by looking at the H1 graphs (on the right). This phenomenon seems to point the fact that an aneurysm is better described by 1-cycles rather than 0-cycles. In fact, the AAA is mainly associated with thrombus that, as explained in the previous sections, produces high-persistence H1 loops.



(a)

Figure 5.23: Here, MDS plots using Wasserstein distance are showed.

In Figure 5.23 the blue points are associated with aortas, both aneurysmatic and healthy, without calcification. The points in yellow are instead associated with calcified aortas. Something worth noticing is that, in opposition with the division ofhealthy/aneurysmatic aortas, calcification seems to be mainly represented by 0-cycles rather than 1-cycles. This is due to the fact that calcification tends to create 0-cycles with medium-low persistence and rarely affects 1-cycles.

6 | Conclusions

6.1. Future Developments and Limitations

Possible future developments are many and can be aimed both at enhancing the analysis, by solving the major limitations encountered, or further expand the work done.

The main limits of this study can be categorized into four main issues, as the developments aimed to settle them:

- Difficulty of extraction of the Persistence Diagram from the CTAs.
- Scarcity of data.
- Lack of Thrombus
- Minimization of the human intervention.

The most important one is the evident difficulty of extracting the Persistence Diagram from the CTA scan. In fact, the passage from raw image to diagram needs various steps such as the extraction of the lumen's surface and its centerline (i.e., the segmentation), the computation of distances between the mesh's points and the centerline, the computation of neck's mean diameter and, finally, the persistence diagram's calculation. In fact, this long procedure introduces the possibility to have errors at each step of the process, requiring more time for validation.

This leads to the second type of limit: future improvements should focus on the replication of the previous findings in larger cohorts, considering also epidemiological and clinical risk factors related to AAA development and growth.

Another limitation is related to the lack of thrombus, given the difficulties explained in Chapter 3 to obtain it. In this way, it will be possible to have a complete representation of the whole AAA. This, if combined with hemodynamical studies, could open the way to define a new model for quantification of AAA's rupture risk.

A final aspect that could be improved is the minimization of the human intervention during

segmentation, building an algorithm capable of recognizing autonomously the location to segment, enhancing the mesh's comparability.

A further development could regard a sensitivity study. In fact, given the strict connection between segmentation and final diagram, it could be worth performing it on the high number of parameters used, such as the thickness of the mesh's resolution. In this way, it could be possible to find the best range of parameters in order to maintain the accuracy observed while minimizing the computational costs.

Focusing on the statistical core of the work, the use of different distances as well combination of H0 and H1 distances, the exploitation of other models to classify or cluster patients or make regression by adding clinical features as covariates could also represent another interesting exploration.

6.2. Conclusions

Given the lack of similar studies on AAA, this thesis must be intended with pioneering purposes, with two different statistical analysis about AAA's morphologies that have been realized using different types of data.

An initial part of this work (Chapter 2) consists of a cluster analysis that uses aortic descriptors such as AAA's dimensions (volume, maximum diameter and, tortuosity) and angles to obtain five groups of patients characterized by different levels of expected AAA's severity. These results, judged interesting in the eyes of clinicians, provide an initial overview on the different aortic deformations caused by an AAA. Despite that, the initial dataset has been put aside, as well as the initial aim, given the lack of local features to better describe the aortic morphology. In fact, it was chosen to opt for another type of data: instead of global descriptors of the aorta a new dataset has been built combining CTA scans and Persistent Homology.

CTA scans are the primal source of data regarding the aortic shape allowing an accurate representation. Then, the CTA have been processed through the process called segmentation, creating a dataset of discretized aortas. These data have then been transformed into Persistence Diagrams, a fundamental tool of Topological Data Analysis, obtaining an accurate summary of its topological features based on the distance of the lumen's surface points from the center of the aortic vessel.

Then, an in deep comparison between the original mesh and the diagram obtained has been done for every patient, linking the most important aortic features associated to the AAA to the diagram's points. The capacity to enclose and summarize the topological

information has been then tested using particular distances obtaining promising results.

This whole work can be considered a first important step for the application of Persistent Homology on this pathology. In fact, many studies suggest correlations between the aortic morphology and AAA's evolution and rupture, that seems to be strictly related to haemodynamics that take place into the vessel [53–55]. The acquisition of data regarding this aspect of the lumen is something that seems to be out of reach for traditional statistics. In fact, it fails the task of collecting local features in a robust way, as showed in the first chapter of this work and confirmed by the lack of papers of interest published on this topic, especially when compared with the number of haemodynamical and mechanical studies.

The choice of the filtration function provides a natural way to describe the vessel's shape associating a clinical descriptor widely used in literature [41] such as the local diameter to each surface point. Normalization of distances assures consistency with different patients and enhances diagram's readability clarifying at which distance from the center of the vessel a change in homology occurs. Moreover, wall's features such as calcifications and blisters on the lumen's surface leave a visible footprint of the diagram in the form of persistence pairs that can be easily recognised.

On the mathematical side, results of Persistent Homology make possible to immerse the diagrams in a mathematical space embedded with a definition of distance.

In conclusion, Topological Data Analysis allows to expand the use of statistics to this clinical problem in a mathematically rigorous way, retrieving a low-dimensional representation of the aorta that is effective for research purposes.

This brings to an accurate summary of the aortic morphology approved by clinician experts in this pathology, making Persistent Homology a promising mathematical tool that in the future could be adopted by scholars in the always growing awareness of AAA and other vascular pathologies.

Acronyms

AAA Abdominal Aortic Aneurysm. i–iv, 6–11, 13–19, 32–35, 41, 46, 52, 53, 63–66, 69, 71–73, 77, 78, 80, 82, 85, 87, 88, 91–93, 107–111

ANOVA Analysis of Variance. 36, 37, 108

CTA Computed Tomography Angiography. i, 8, 16, 18–22, 41, 43–49, 51, 53, 54, 69, 79–82, 91, 92, 107, 109, 111

DBSCAN Density-Based Spatial Clustering. 24, 41

DICOM Digital Imaging and Communications in Medicine. 46

EVAR Endovascular Aneurysm Repair. 11–13, 15, 17, 33, 107

ILT intraluminal thrombus. 7, 11, 15, 16

ITK Insight Segmentation and Registration ToolKit. 51

KNN K-Nearest Neighbors. 16

MDS Multidimensional Scaling. 87–89, 112

OSI Oscillatory Stress Index. 15

PACS Picture Archiving and Communication System. 46, 54

RPI Rupture Potential Index. 15

SVM Support Vector Machine. 16

SWD Sliced Wasserstein Distance. 61

TDA Topological Data Analysis. i, 55, 57, 69

Tetgen Quality Tetrahedral Mesh Generator. 51

VMTK Vascular Modelling Toolkit. 51, 52

VTK Visualization ToolKit. 51

WSS Wall Sheer Stress. 15, 16

Bibliography

[1] Classification & structure of blood vessels. https://training.seer.cancer.gov/anatomy/cardiovascular/blood/classification.html. Accessed: 2022-11-20.

[2] Anatomy, arteries. https://www.ncbi.nlm.nih.gov/books/NBK547743/. Accessed: 2022-11-20.

[3] How does the blood circulatory system work? https://www.informedhealth.org/how-does-the-blood-circulatory-system-work.html. Accessed: 2022-11-20.

[4] K.C. Kent. Abdominal aortic aneurysms. *The New England Journal of Medicine*, 11 2014.

[5] Tunica intima. https://en.wikipedia.org/wiki/Tunica_intima. Accessed: 2022-11-20.

[6] Cardiovascular. http://microanatomy.net/cardiovascular/cardiovascular_system.htm. Accessed: 2022-11-20.

[7] Screening for abdominal aortic aneurysm. https://www.uptodate.com/contents/screening-for-abdominal-aortic-aneurysm. Accessed: 2022-11-20.

[8] Aneurysm. https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/aneurysm. Accessed: 2022-11-20.

[9] Aortic aneurysm. https://vijaythakore.com/aortic-aneurysm/. Accessed: 2022-11-20.

[10] K. Singh, K. H. Bønaa, B.K. Jacobsen, L. Bjørk, and S. Solberg. Prevalence of and risk factors for abdominal aortic aneurysms in a population-based study. *American Journal of Epidemiology*, 154(3):236–244, 8 2001.

[11] Ateosclerosi. https://it.wikipedia.org/wiki/Aterosclerosi. Accessed: 2022-11-20.

[12] Atherosclerosis. https://www.hopkinsmedicine.org/health/conditions-and-diseases/atherosclerosis. Accessed: 2022-11-20.

[13] B. Breznik, H. Motaln, and T.L. Turnšek. Proteases and cytokines as mediators of interactions between cancer and stromal cells in tumours. *Biological Chemistry*, 398(7):709–719, 6 2017.

[14] P.K. Shah. Inflammation, metalloproteinases, and increased proteolysis. *Circulation*, (10):2115–2117, 96 1997.

[15] Davies M.J. Aortic aneurysm formation. *Circulation*, 98(3):193–195, 7 1998.

[16] Thrombus. `https://en.wikipedia.org/wiki/Thrombus`. Accessed: 2022-11-20.

[17] M.J. Thubrikar, F. Robicsek, M. Labrosse, V. Chervenkoff, and B.L. Fowler. Effect of thrombus on abdominal aortic aneurysm wall dilation and stress. *The Journal of Cardiovascular Surgery*, 44(1):67–77, 2 2003.

[18] S. Rahmani, M. Alagheband, A. Karimi, M. Alizadeh, and M. Navidbakhsh. Wall stress in media layer of stented three-layered aortic aneurysm at different intraluminal thrombus locations with pulsatile heart cycle. *Journal of Medical Engineering and Technology*, 39(4):239–245, 5 2015.

[19] F. Riveros, G. Martufi, T. C. Gasser, and J. F. Rodriguez-Matas. On the impact of intraluminal thrombus mechanical behavior in aaa passive mechanics. *Annals of Biomedical Engineering*, 43(9):2253–2264, 9 2015.

[20] E. Georgakarakos, C.V. Ioannou, S. Volanis, Y. Papaharilaou, J. Ekaterinaris, and A.N. Katsamouris. The influence of intraluminal thrombus on abdominal aortic aneurysm wall stress. *International Angiology*, 28(4):325–333, 8 2009.

[21] A.J. Boyd. Intraluminal thrombus: Innocent bystander or factor in abdominal aortic aneurysm pathogenesis? *JVS Vascular Science*, 4(2), 5 2021.

[22] Calcification. `https://en.wikipedia.org/wiki/Calcification`. Accessed: 2022-11-20.

[23] Coronary artery calcification. `https://my.clevelandclinic.org/health/diseases/22953-coronary-artery-calcification`. Accessed: 2022-11-20.

[24] Calcification of the arteries. `https://www.news-medical.net/health/Calcification-of-the-Arteries.aspx`. Accessed: 2022-11-20.

[25] R.V.C. Buijs, T.P. Willems, R.A. Tio, H. H. Boersma, I.F.J. Tielliu, R. H. J. A. Slart, and C. J. Zeebregts. Calcification as a risk factor for rupture of abdominal aortic aneurysm. *European Journal of Vascular and Endovascular Surgery*, 46(5):542–548, 11 2013.

[26] S.J. Haller, A. F. Azarbal, and S. Rugonyi. Predictors of abdominal aortic aneurysm risks. *Bioengineering*, 7(3), 7 2020.

[27] D.P. Nathan, C. Xu, A.M. Pouch, K.B. Chandran, B. Desjardins, J.H. 3rd Gorman, R.M. Fairman, R.C. Gorman, and B.M. Jackson. Increased wall stress of saccular versus fusiform aneurysms of the descending thoracic aorta. *Annals of Vascular Surgery*, 25(8):1129–1137, 11 2011.

[28] Abdominal aortic aneurysm. `https://www.stanfordchildrens.org/en/topic/default?id=abdominalaorticaneurysm-85-P08247`. Accessed: 2022-11-20.

[29] Abdominal aortic aneurysm (aaa). `https://redbacteria.wordpress.com/tag/aaa/`. Accessed: 2022-11-20.

[30] F. Lanas, P. Serón, and A. Lasan. Cardiovascular disease in latin america: The growing epidemic. *Progress in Cardiovascular Diseases*, 57(3):262–267, 11-12 2014.

[31] European cardiovascular disease statistics 2017. `https://ehnheart.org/cvd-statistics.html`. Accessed: 2022-11-20.

[32] Stent graft. `https://www.sciencedirect.com/topics/medicine-and-dentistry/stent-graft`. Accessed: 2022-11-20.

[33] Endovascular aneurysm repair. `https://cacvi.org/services/peripheral-procedures/endovascular-aneurysm-repair/`. Accessed: 2022-11-20.

[34] Abdominal stent-graft system - p190015. `https://www.fda.gov/medical-devices/recently-approved-devices/treor-abdominal-stent-graft-system-p190015`. Accessed: 2022-11-20.

[35] Endoleak. `https://radiopaedia.org/articles/endoleak`. Accessed: 2022-11-20.

[36] Endoleak. `https://it.wikipedia.org/wiki/Endoleak`. Accessed: 2022-11-20.

[37] Endoleak classification system. `https://www.researchgate.net/figure/Endoleak-classification-system-9_fig1_315921086`. Accessed: 2022-11-20.

[38] M. Cury and et al. Aortic disease in the young: Genetic aneurysm syndromes, connective tissue disorders, and familial aortic aneurysms and dissections. *International Journal of Vascular Medicine*, 1 2013.

[39] R.M. Sandford and et al. The genetic basis of abdominal aortic aneurysms: A review. *European Journal of Vascular and Endovascular Surgery*, 33(4):381–390, 4 2007.

[40] K.C. Kent and et al. Analysis of risk factors for abdominal aortic aneurysm in a cohort of more than 3 million individuals. *Journal of Vascular Surgery*, 53(1):263–264, 7 2010.

[41] T. Lanne and et al. Diameter and compliance in the male human abdominal aorta: Influence of age and aortic aneurysm. *European Journal of Vascular Surgery*, 6(2):178–184, 3 1992.

[42] F.A. Lederle and et al. Smokers' relative risk for aortic aneurysm compared with other smoking-related diseases: a systematic review. *Journal of Vascular Surgery*, 38(2):329–334, 7 2003.

[43] R.V.C. Buijs and et al. Calcification as a risk factor for rupture of abdominal aortic aneurysm. *European Journal of Endovascular Surgery*, 46(5):542–548, 11 2013.

[44] G.W.H. Schurink and et al. Thrombus within an aortic aneurysm does not reduce pressure on the aneurysmal wall. *Journal of Vascular Surgery*, 31(3):501–506, 5 2000.

[45] L. Speelman and et al. The mechanical role of thrombus on the growth rate of an abdominal aortic aneurysm. *Journal of Vascular Surgery*, 51(1):19–26, 1 2010.

[46] J.P. Carpenter and et al. Impact of exclusion criteria on patient selection for endovascular abdominal aortic aneurysm repair. *Journal of Vascular Surgery*, 34(6):1050–1054, 12 2001.

[47] Greenhalgh R.M. Comparison of endovascular aneurysm repair with open repair in patients with abdominal aortic aneurysm (evar trial 1), 30-day operative mortality results: randomised controlled trial. *Lancet (London, England)*, pages 843–848, 9 2004.

[48] S.C.V. Paravastu and et al. A systematic review of open versus endovascular repair of inflammatory abdominal aortic aneurysm. *European Journal of Vascular and Endovascular Surgery*, 38(3):291–297, 9 2009.

[49] G.W.H. Schurink and et al. Endoleak after stent-graft treatment of abdominal aortic aneurysm: a meta-analysis of clinical studies. *The British Journal of Surgery*, 86(5):581–587, 5 1999.

[50] R.A. Baum and et al. Endoleaks after endovascular repair of abdominal aortic aneurysms. *Journal of Vascular and Interventional Radiology*, pages 1111–1117, 9 2003.

[51] Stather P.W. and et al. A review of current reporting of abdominal aortic aneurysm mortality and prevalence in the literature. *European Journal of Vascular and Endovascular Surgery*, 47(3):240–242, 3 2014.

[52] A. Karthikesalingam and et al. Mortality from ruptured abdominal aortic aneurysms: clinical lessons from a comparison of outcomes in england and the usa. *Lancet (London, England)*, 3 2014.

[53] C.M. Colciago and et al. Analysis of morphological and hemodynamical indexes in abdominal aortic aneurysms as preliminary indicators of intraluminal thrombus deposition. *Biomechanics and Modeling in Mechanobiology*, 19(3):1035–1053, 7 2020.

[54] M. Piccinelli and et al. Impact of hemodynamics on lumen boundary displacements in abdominal aortic aneurysms by means of dynamic computed tomography and computational fluid dynamics. *Biomechanics and Modeling in Mechanobiology*, 12(6):1263–1276, 11 2013.

[55] B.J. Doyle and et al. Finite element analysis rupture index (feari) as an additional tool for abdominal aortic aneurysm rupture prediction. *Vascular Disease Prevention*, 6:114–121, 2009.

[56] J.P. Vande Geest and et al. A biomechanics-based rupture potential index for abdominal aortic aneurysm risk assessment. *Annals of the New York Academy of Science*, 1085(1):11–21, 11 2006.

[57] A. Sheidaei and et al. Simulation of abdominal aortic aneurysm growth with updating hemodynamic loads using a realistic geometry. *Medical Engineering & Physics*, 33(1):80–88, 1 2011.

[58] G. Giannoglou and et al. Predicting the risk of rupture of abdominal aortic aneurysms by utilizing various geometrical parameters: Revisiting the diameter criterion. *Angiology*, 57(4):487–494, 8-9 2006.

[59] E.K. Kerut, P. To, K.L. Summers, C. Sheahan, and M. Sheahan. Statistical and machine learning methodology for abdominal aortic aneurysm prediction from ultrasound screenings. *Echocardiography (Mount Kisco, N.Y.)*, 36(11):1989–1996, 11 2019.

[60] F. Lareyre, C. Cédric, M. Adam, and J. Raffort. Prediction of abdominal aortic aneurysm growth and risk of rupture in the era of machine learning. *Angiology*, 71(8):767, 9 2020.

[61] B. Rengarajan, W. Wu, C. Wiedner, D. Ko, S.C. Muluk, M.K. Eskandari, P.G.

Menon, and E.A. Finol. A comparative classification analysis of abdominal aortic aneurysms by machine learning algorithms. *Annals of Biomedical Engineering*, 48(4):1419–1429, 1 2020.

[62] T.W. Wakefield, W.M. Jr Whitehouse, S. C. Wu, G.B. Zelenock, J.L. Cronenwett, E.E. Erlandson, R.O. Kraft, S.M. Lindenauer, and J.C. Stanley. Abdominal aortic aneurysm rupture: Statistical analysis of factors affecting outcome of surgical treatment. *Surgery*, 91(5):586–596, 5 1982.

[63] A. Monsalve-Torra, D. Ruiz-Fernandez, M. Marin-Alonso, A. Soriano-Payá, J. Camacho-Mackenzie, and M. Carreño-Jaimes. Using machine learning methods for predicting inhospital mortality in patients undergoing open repair of abdominal aortic aneurysm. *Journal of Biomedical Informatics*, 62:195–201, 7 2016.

[64] M. De Bruijne and et al. Model-based segmentation of abdominal aortic aneurysms in cta images. *Proceedings*, 5032, 5 2003.

[65] H.A. Hong and U.U. Sheikh. Automatic detection, segmentation and classification of abdominal aortic aneurysm using deep learning. *2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA)*, pages 242–246, 2016.

[66] A.K. Golla and et al. Automated screening for abdominal aortic aneurysm in ct scans under clinical conditions using deep learning. *Diagnostic (Basel, Switzerland)*, 11(11):2131, 11 2021.

[67] F. Kyriakou, W. Dempster, and D. Nash. A methodology to quantify the geometrical complexity of the abdominal aortic aneurysm. *Scientific Reports*, 9, 11 2019.

[68] H.F. Morton, O.J. Deters, F.F. Mark, C.B. Bargeron, and G.M. Hutchins. Arterial geometry affects hemodynamics. a potential risk factor for atherosclerosis. *Atherosclerosis*, 46(2):225–231, 2 1983.

[69] O. Tanweer, T.A. Wilson, E. Metaxa, H.A. Riina, and H. Meng. A comparative review of the hemodynamics and pathogenesis of cerebral and abdominal aortic aneurysms: Lessons to learn from each other. *Journal of Cerebrovascular and Endovascular Neurosurgery*, 16(4):335–349, 12 2014.

[70] Dbscan. https://en.wikipedia.org/wiki/DBSCAN.

[71] J. Gareth, D. Witten, T. Hastie, and R. Tibshirani. *Introduzione all'apprendimento statistico*. PICCIN, Padova, Italy, 1 edition, 2020.

[72] Dendrogram. https://en.wikipedia.org/wiki/Dendrogram. Accessed: 2022-11-20.

[73] V. Shweta, S. Shweta, and B. Neha. Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pages 568–573, 2019.

[74] Analysis of variance. https://en.wikipedia.org/wiki/Analysis_of_variance. Accessed: 2022-11-20.

[75] Permutational analysis of variance. https://en.wikipedia.org/wiki/Permutational_analysis_of_variance. Accessed: 2022-11-20.

[76] Permutation tests for univariate and multivariate - data alessia pini. Applied Statistics - Course Material.

[77] Tomograph. https://kirurgia.ge/en/teqnologia/Computed-tomograph-Toshiba-Aquilion-Lightning-SP/22. Accessed: 2022-11-20.

[78] Hounsfield scale diagram. https://radiopaedia.org/cases/hounsfield-scale-diagram. Accessed: 2022-11-20.

[79] Marco Aiello and et al. How does dicom support big data management? investigating its use in medical imaging community. *Insights into Imaging*, 12(164), 11 2021.

[80] V: Luboz, X. Wu, K. Krissian, C. Westin, R. Kikinis, S. Cotin, and S. Dawson. A segmentation and reconstruction technique for 3d vascular structures. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, pages 43–50, 2005.

[81] N. Senthilkumaran and S. Vaithegi. Image segmentation by using thresholding techniques for medical images. *Computer Science & Engineering: An International Journal (CSEIJ)*, 6(1):1485–1492, 2 2016.

[82] Vmtk. http://www.vmtk.org/. Accessed: 2022-11-20.

[83] Itk. https://itk.org. Accessed: 2022-11-20.

[84] Vtk. http://www.vtk.org. Accessed: 2022-11-20.

[85] Tetgen. https://wias-berlin.de/software/index.jsp?id=TetGen&lang=1. Accessed: 2022-11-20.

[86] Luca Antiga. *Patient-Specific Modeling of Geometry and Blood Flow in Large Arteries*. PhD thesis, Politecnico di Milano, 2002.

[87] Centerlines. `http://www.vmtk.org/tutorials/Centerlines.html`. Accessed: 2022-11-20.

[88] W. De Vos, J. Casselman, and G.R.J. Swennen. Cone-beam computerized tomography (cbct) imaging of the oral and maxillofacial region: A systematic review of the literature. *International Journal of Oral and Maxillofacial Surgery*, 38(6):609–625, 6 2009.

[89] An introduction to persistent homology. `https://www.youtube.com/watch?v=UxEH7WySO60`. Accessed: 2022-11-20.

[90] Ulderico Fugacci, Sara Scaramuccia, Federico Iuricich, and Leila De Floriani. Persistent Homology: a Step-by-step Introduction for Newcomers. In Giovanni Pintore and Filippo Stanco, editors, *Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference*. The Eurographics Association, 2016.

[91] Simplex. `https://en.wikipedia.org/wiki/Simplex`. Accessed: 2022-11-20.

[92] Munkres James R. *Elements of Algebraic Topology*. Addison-Wesley Publishing Company, Boston, United States, 1984.

[93] Munkres James R. *Elements of Algebraic Topology*. Addison-Wesley Publishing Company, Boston, United States, 1984.

[94] Gurjeet Singh, Facundo Memoli, Tigran Ishkhanov, Guillermo Sapiro, Gunnar Carlsson, and Dario L. Ringach. Topological analysis of population activity in visual cortex. *Journal of Vision*, 8(8), 8 2008.

[95] Persistence diagrams. `https://mtsch.github.io/PersistenceDiagrams.jl/v0.3/`. Accessed: 2022-11-20.

[96] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37:103–120, 1 2007.

[97] Alex McClearly and Amit Patel. Bottleneck stability for generalized persistence diagrams. *Proceedings of the AMS*, 2020.

[98] Primoz Skraba and Katharine Turner. Wasserstein stability for persistence diagrams. *arXiv*, 2020.

[99] P. Bubenik. Statistical topological data analysis using persistence landscape. *Journal of Machine Learning Research*, 16:77–102, 2015.

[100] Andrew Fitzgibbon, Maurizio Pilu, and Robert B. Fisher. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480, 5 1999.

[101] Radim Halir and Jan Flusser. Numerically stable direct least squares fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 5 2000.

[102] Ripser. `https://github.com/ripser/ripser`. Accessed: 2022-11-20.

[103] Persim. `https://github.com/scikit-tda/persim`. Accessed: 2022-11-20.

[104] Mohammad Wazzan, Ahmed Abduljabbar, n Amr Ajla, Rani Ahmad, Turki Alhazmi, Ayman Eskandar, Khalid Khashoggi, Farah Alasadi, Shahad Howladar, and Yousra Alshareef. Reference for normal diameters of the abdominal aorta and common iliac arteries in the saudi population. *Cureus*, 14(10), 10 2022.

[105] F. Trevor Cox and A. A. Michael Cox. *Multidimensional Scaling*. Chapman & Hall/CRC, United Kingdom, 2001.

List of Figures

Acknowledgements

Here you might want to acknowledge someone.