The Right Treatment for Abdominal Aortic Aneurysms

Machine Learning Based Predictions of Postoperative Endograft Apposition Using 4D Statistical Shape Models Including Thrombus

Daan Roos

Master Thesis Technical Medicine

UNIVERSITY OF TWENTE.

umcg

The Right Treatment for Abdominal Aortic Aneurysms

Machine Learning Based Predictions of Postoperative Endograft Apposition Using 4D Statistical Shape Models Including Thrombus

Master thesis by:
Daan Roos

31-07-2024

Graduation Board:

Chair	prof. dr. R.H. Geelkerken
Clinical supervisor	prof. dr. J.P.P.M. de Vries
Technological supervisor	dr. J.M. Wolterink
Technical physician	dr. R.C.L. Schuurmann
Process supervisor	R.J. Lambers MSc.
External member	dr. J.A. Simmering

ABSTRACT

Determination of anatomic suitability is the most important factor in successful endovascular aneurysm repair (EVAR). As the function of the endograft is to prevent blood flow into the aneurysm sac from the healthy segment of the aorta, adequate seal is needed between proximal part of the graft and the aorta. Postoperative Shortest Apposition Length (SAL) less than 10mm following EVAR is associated with the development of a type 1a endoleak. Currently, procedures are planned by careful measurement of certain anatomic descriptors. The three dimensional morphology of the aorta is complex and might be oversimplified if only described by these anatomical parameters. In recent literature, a statistical shape model (SSM) is used to describe the morphology of the aortic neck. Also, a machine learning approach towards the preoperative prediction of SAL has been suggested. These recent approaches to predict postoperative SAL based on an SSM of the preoperative aortic neck used segmentations of the lumen of the aorta. It is hypothesized that for the prediction of postoperative SAL, also information about preoperative thrombus location, degree of circumference and thickness is needed. The aim of this study is to incorporate information of intramural thrombus into 4D SSMs and to provide insight in the possibility for continuous predictions of endograft apposition in the pararenal aorta.

A dataset is composed of patients that received EVAR and patients that received FEVAR. Manually obtained wall segmentations of the aorta are used in the parameterization process before SSM creation. Resulting thrombus thickness between the lumen and wall segmentations is incorporated in the SSM as the fourth dimension. The SSM is validated on compactness, generalization ability and specificity. Ensemble learning, in the form of Random Forest classifiers, is used to predict postoperative SAL labels < 10 mm or ≥ 10 mm, by the means of leave-one-out cross validation. Predictions of classification are compared to previously published methods, using the DeLong's test. Secondly, the possibility for continuous outcome of postoperative SAL is investigated in the pararenal aorta and the aortic neck specifically, by the use of regression algorithms. The root mean squared error (RMSE) is calculated for the continuous predictions and Bland-Altman plots are published to show potential bias.

The 4D SSMs are based on 216 patients with a mean SAL of 14.33 mm. Generalization, specificity and compactness performances confirmed that the novel SSMs are able to capture thrombus morphology changes in the population. Random Forest predictions of the binary classes show improved accuracy for the use of the 4D vs lumen-only SSMs (84% vs 72%, $p - value =< 0.001$). RMSE-values for the continuous prediction are 10.82 mm for a pararenal segment of the aorta extending 30 mm both proximal and distal. RMSE of 6.85 mm is shown for the aortic neck. The Bland-Altman plots show proportional bias for both algorithms, indicating that the current workflow has difficulty predicting SAL towards the limits of the data.

The first study shows the possibility to implement information on thrombus location and morphology in four dimensional statistical shape models of the pararenal aorta. The newly described framework using 4D SSMs outperforms recently published methods, predicting a clinically relevant measure for SAL in the aortic neck. Regression models have provided a first insight in the possibility to predict continuous outcome of SAL based on statistical shape models of the aorta. SSMs should be evaluated to determine if they adequately describe the aortic morphology needed for the prediction of SAL in that specific region. Current errors in predictions suggests that more effort is needed before the incorporation of continuous predictions in the surgical planning of endografts. The incorporation of intramural thrombus in the morphological description of the aorta used for predictions means a positive gain in accuracy. Implementation of the models as an automated warning system (when low SAL is predicted) during conventional planning seems feasible. Future research should focus on the increase of accuracy for (continuous) prediction and on the implementation of software. Both studies described in this thesis expand on previous work to help the progression of a *virtual stenting tool*.

PREFACE

Met het indienen van deze thesis sluit ik een periode af die voor altijd mijn studententijd zal heten. Daarbij kan ik terug kijken op vele mooie momenten en bovenal een erg leerzame tijd. Ik ben de verschillende stagebegeleiders dankbaar voor de vele mogelijkheden die er voor mij waren om in de Masterfase ervaring op te doen met Technische Geneeskunde in de praktijk. De afwisselende stages en variatie daarin hebben de afgelopen twee jaar tot ontzettend leuke jaren gemaakt. Graag bedank ik de Vaatchirurgen, Interventieradiologen, OK-personeel, vaatlaboranten en arts-assistenten van het UMCG en OZG voor het afgelopen jaar waarin ik me overal welkom heb gevoeld om ervaringen op te doen in het kader van mijn Klinische Specialisatiestage.

Richte en Jean-Paul wil ik daarbij extra bedanken voor hun begeleiding in het afgelopen jaar. Jullie hebben mij enthousiast gemaakt voor het onderwerp van deze thesis en ik waardeer hoe ik binnen de onderzoeksgroep ervaring op heb mogen doen met het uitvoeren en presenteren van medisch-technische studies. De adviezen hierin neem ik mee voor de toekomst, van jullie weet ik zeker dat ze van de juiste kennis en ervaring komen!

Jelmer, het was erg fijn om geregeld een aanvullende kijk op het werk te krijgen tijdens onze digitale gesprekken. Bedankt voor jouw waardevolle input, mocht het nog eens nodig zijn dat ik de tweeling in slaap "teams", dan hoor ik dat graag!

Mijn procesbegeleider Rianne wil ik graag bedanken voor de afgelopen twee jaar. Ondanks dat de intervisiebijeenkomsten vaak een onpraktische reis betekende, heb ik het altijd als leuke momenten ervaren. Deze momenten hebben er voor gezorgd dat ik meer uit mijn persoonlijke ontwikkeling heb kunnen halen. Ook mijn intervisiegenoten bedank ik voor het feit dat het altijd fijn was om samen te komen!

Cas, Daniëlle en alle andere onderzoekers van de Y3: het was heel fijn om met jullie te kunnen sparren over het onderzoek! Ondanks de soms drukke agenda's heb ik altijd het gevoel gehad aan te kunnen kloppen. Een congres/symposium met jullie was naast leerzaam ook altijd erg gezellig!

Ik ben ontzettend dankbaar voor de vele mooie momenten die ik tijdens mijn studie heb mogen beleven met familie en vrienden. Mijn huisgenoten van Bommelstein, vrienden van AES, Koh-i-Noor en de studie: onwijs bedankt! Mijn lieve ouders en Marijn & Teun verdienen extra aandacht, bedankt voor de onvoorwaardelijke steun en het vertrouwen. Ik kom nog altijd erg graag naar Zevenaar, waar altijd een liefdevol ontvangst is. Lieve Daniek, bedankt voor onze gesprekken en jouw steun. Jij betekent heel veel voor mij, ik heb ontzettend veel zin in wat de komende jaren voor ons in petto hebben!

Ik wens iedereen veel plezier met het lezen van deze thesis.

Daan Roos

Contents

1 Background

1.1 Abdominal Aortic Aneurysm

An abdominal aortic aneurysm (AAA) is an abnormal dilation of the abdominal aorta. AAAs are relatively common and rupture of an aneurysm is associated with high mortality rates [1]. Most AAAs are asymptomatic [2], these patients come to medical attention by targeted screening of patients at risk or as the result of abdominal imaging studies targeting other pathologies. When symptoms occur, patients commonly present with abdominal, back, or flank pain, but thromboembolism can also be present, leading to symptoms of limb ischemia [3].

An aneurysm is defined as the dilation of a blood vessel 50 percent greater than a healthy diameter [4]. The abdominal aorta is the most common anatomic site of arterial aneurysm [5]. AAAs are described relative to the involvement of the renal or visceral vessels, shown in Figure 1.

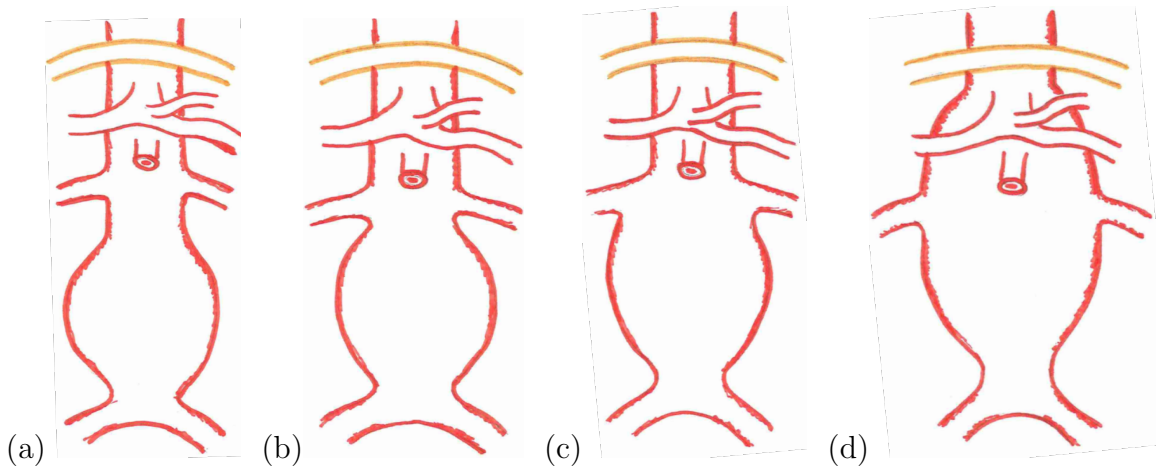


Figure 1: **Infrarenal AAA (a)**: The aneurysm originates distal to the renal arteries. There is a segment of nonaneurysmatic aorta extends distal to the origins of the renal arteries, called the aortic neck. **Juxtarenal AAA (b)**: the aorta at the level of the renal arteries is not aneurysmatic, but there is no segment of nonaneurysmatic aorta distal to the renal arteries.**Pararenal AAA (c)**: The aneurysmatic aorta involves the origins of the renal arteries, but the origin of the superior mesenteric artery (SMA) is not aneurysmatic. **Suprarenal AAA (d)**: involves the origins of one or more visceral arteries, but remains below the diaphragm.

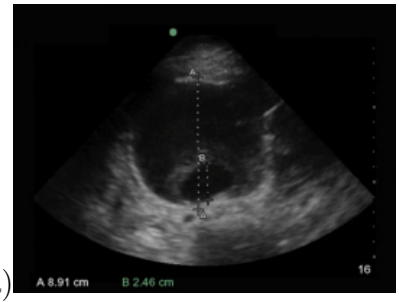
Most AAAs occur distally of the renal arteries and 15% of aneurysms can be defined as juxtarenal [6].

Clinical presentation

Risk factors for the development of an AAA include [7]:

- Smoking
- Male Sex
- Age
- Atherosclerosis
- Family history of AAA
- Present arterial aneurysms
- Connective tissue disorder
- History of aortic dissection

To find asymptomatic AAAs, screening of the population above a certain age can help identify patients. In the Netherlands no screening programs exist. In some cases, patients at risk are identified and referred by their general practitioner [8]. The presence of a pulsatile abdominal mass could be a predictor of AAAs. Approximately 30 percent of asymptomatic AAAs may be found when a pulsatile abdominal mass is palpated on routine physical examination [9]. AAA is often detected as an incidental finding on imaging studies performed for other purposes [10]. As atherosclerosis is a risk factor for the development of AAA, patients may also be identified through the examination of associated arterial disease. Symptomatic patients might be found when presenting with either abdominal, back, flank or pelvic pain. As symptomatic AAA has a high risk of aneurysm rupture, patients with acute abdominal pain should be examined for the possible diagnosis of an AAA.

(a) 

(b) 

Figure 2: (a) Transverse ultrasound image demonstrating a wall-wall aorta diameter measurement and aortic lumen diameter [11]. (b) Computed tomography angiography showing the aortic lumen in light grey (arrow), surrounded by thrombus located in the aneurysm.

Abdominal ultrasound and computed tomography (CT) are used for the diagnosis of AAA. Ultrasound is noninvasive, inexpensive, and has a sensitivity of 98 percent and specificity of 99 percent for the diagnosis of AAA [12] and can be used as a cost-effective screening tool for identifying aneurysms in patients with risk factors [13]. Intraobserver variability of this technique is reported as high as 4mm [14]. Computed tomography angiography (CTA) using intravenous contrast agents has the advantage of evaluating the aortic geometry in more detail. Acquisition of a CTA enables treatment planning. Both imaging techniques are shown in Figure 2.

1.2 AAA management

The treatment of patients diagnosed with AAA is focused on the prevention of rupture. Abdominal aortic aneurysms are classified according to the diameter and the presence of symptoms. For symptomatic patients, it is important to establish if the aneurysm has ruptured, using diagnostic techniques previously described. Asymptomatic patients can be monitored on an outpatient basis, where a diameter of the aneurysm above 55 mm for men [15] and above 50 mm for women [16] is indicative for aneurysm repair. Repair of the aneurysm can be accomplished by open surgery or endovascular treatment. Endovascular procedures are associated with a lower risk of perioperative morbidity and mortality compared to open surgery, however long-term mortality is not significantly different between the techniques [17] [18]. Patients with lower operative risks might benefit from open surgery. Open surgery of the abdominal aorta can be performed using different approaches [19] and involves replacement of the pathological segment with a tube or bifurcated prosthetic graft.

Endovascular aneurysm repair

Endovascular aneurysm repair (EVAR) involves the placement of modular graft components, mainly delivered via the femoral arteries, to exclude the aneurysm.The goal of EVAR is the exclusion of the aneurysm sac from circulation and thus lowering the risk of rupture. The type of endovascular graft needed for exclusion depends on the extent of the aneurysm. With all endovascular grafts, seal is important for exclusion of the aneurysm sac, with inadequate seal leading to endoleaks, or migration of the graft.

For an infrarenal aneurysm, seal can be achieved within a segment of healthy aorta between the aneurysm and renal arteries.This segment of the aorta is called the landing zone of the endograft. As the endograft is placed below of the renal arteries the kidneys remain part of the regular circulation of blood. For the other types of abdominal aneurysms extending more cranially, the healthy segment of aorta available as landing zone will also be found more cranially. Fenestrated EVAR (FEVAR) devices are available if the endograft would also exclude the renal and visceral arteries (i.e. superior mesenteric artery and celiac trunc). In this configuration, these arteries will remain part of blood circulation via bridging stents, sealed distally in the arteries and placed through fenestrations in the main part of the endograft. Figure 3 shows both EVAR (a) and FEVAR (b) configurations.

After placement of a (F)EVAR device, the patient needs lifelong follow-up, by ultrasound or CTA, to check for previously mentioned complications such as device migration and endoleaks, but also component separation, stent fracture or endograft infection. As failure of the bridging stents is possible, FEVAR results in more reinterventions than EVAR [20]. Reintervention rates for FEVAR and EVAR are 39% [21] and 16% [22] respectively. Overall, following endovascular procedures, re-intervention is needed more often compared to open surgery [20], indicating the importance of selecting the right treatment for each patient.

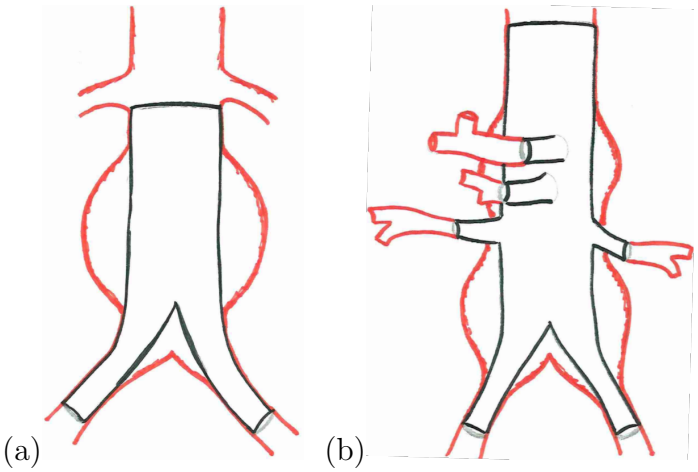


Figure 3: (a) Endovascular aneurysm repair (EVAR) placed below the renal arteries. (b) Fenestrated endovascular aneurysm repair (FEVAR) with fenestrations of the celiac trunk, superior mesenteric artery and renal arteries. The outline of the endograft is shown in black.

Endograft seal

Determination of anatomic suitability is the most important factor in successful EVAR. As the function of the endograft is to prevent blood flow into the aneurysm sac from the healthy segment of the aorta, adequate seal is needed with the graft both proximally at the aortic neck and distally in both the illiac arteries. An adequate seal depends on sufficient circumferential apposition of the graft with a healthy part of the aorta. To quantify the amount of seal, Geraedts et al. [23] define the shortest apposition length (SAL) as the shortest distance between the circumference of the proximal endograft fabric and the location where circumferential apposition of the endograft with the aortic wall is lost. Postoperative SAL less than 10mm following EVAR is associated with the development of an endoleak at the proximal part of the endograft, defined as a type 1a endoleak. This includes both SAL at first postoperative CTA [23], and diminishing apposition during subsequent follow-up [24]. Figure 4 shows the definition of SAL.

For infrarenal aneurysms, the morphology of the infrarenal aortic neck determines the SAL that can be achieved postoperatively. Specific values for diameter, length, supra- and infrarenal angulation [25] and aortic curvature [26] of the aortic neck are associated with stent migration and endoleaks. These factors are taken into account when planning endovascular treatment and deciding if the aortic neck is suitable as landing zone to provide sufficient apposition. Semiautomatic sizing software is available for the planning of endovascular therapies [27], quantifying these factors and objectively describing the morphology of the aorta.

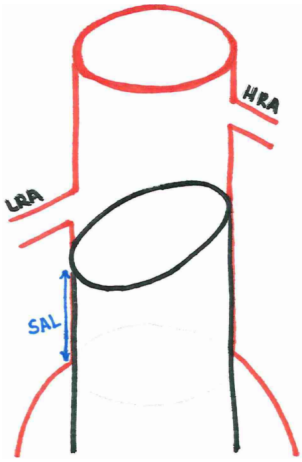


Figure 4: EVAR in situ displayed in black, showing the Highest Renal Artery (HRA) and the Lowest Renal Artery (LRA). The Shortest Apposition Lenght (SAL) is shown in blue. SAL is defined as the shortest distance between the circumference of the proximal endograft fabric and the location where circumferential apposition of the endograft with the aortic wall is lost [23].

The three dimensional (3D) morphology of the aorta, and in particular the aortic neck, is complex and might be oversimplified if only described by previously mentioned anatomical parameters. Oversimplification might lead to less real achieved apposition than anticipated, leading to enhanced risk for type 1a endoleaks [24]. Little consensus exists on an exact definition of a hostile aortic neck morphology and how these characteristics combined influence endograft outcome [28]. In recent literature, a statistical shape model (SSM) is used to describe the 3D morphology of the aortic neck [29]. A mathematical technique is used to model the shape variation of an anatomy of interest in a population returning a description of the variation of the shape within a population. This variation of individual aortas is relative to the mean shape of the anatomy in the dataset.

1.3 Statistical Shape Models

To form an SSM, principal components analysis (PCA) is used to create linearly independent components describing the variation of aortic morphology within a population [30]. The aortic neck morphology of a single patient can be described by a combination of scores for the found principal components (PCs). This makes the technique an objective way to describe aortic neck morphology.

The first step in performing statistical shape modeling is to convert the geometric information of the relevant part of the aorta to a discrete representation [31]. For the whole dataset this is done in a standardized manner. Van Veldhuizen et al. [29], developed a workflow to create a SSM of the aortic neck. A common anatomical landmark is used in the form of the origin of the lowest renal artery (LRA). All aortic necks in the dataset can be seen relative to the projection of the LRA on the center lumen line (CLL). Standardized parameterization along the CLL is used to create a set of points on the contour of the aortic neck segmentation.

Shape descriptors

New possibilities of morphological analysis are made possible by the increasing amount of data from medical imaging and computational abilities to process this data [31] Coefficients representing shapes could serve as statistical descriptors, able to capture characteristic changes in shape between individual cases of the dataset. Unsupervised learning is able to find structures and connections in the shape data, without the need for interpretation of clinical values. Clustering of PC-scores could identify subgroups within data, improving shape-based risk assessment and enabling treatment planning [32]. In a supervised framework where principal components are combined with labels indicating disease states, classifiers can be trained to enable diagnosis based on anatomical morphology [31] [33].

Supervised learning for EVAR

After first focusing on creating and validating a workflow to construct SSMs of the aortic neck, van Veldhuizen et al. [34] used the standard deviations for a set of PCs describing aortic neck morphology to train machine learning (ML) based prediction models. The model is trained using shape information from preoperative CT of EVAR-patients, who were labeled with a binary value for the shortest apposition length ($SAL \geq or < 10mm$). In literature this value is proposed as the threshold for SAL to pose risk for type IA endoleaks [35]. The model predicts the binary outcome based on PCs that describe a new morphology of an aortic neck. This provides a possibly valuable tool in the preoperative phase to identify patients who are less likely to achieve sufficient apposition after EVAR.

1.4 Technical Medical Challenges

Recent approaches to predict postoperative SAL based on a SSM of the preoperative aortic neck used segmentations of the lumen of the aorta. These segmentations resulted from semi-automated segmentation of CTA in 3mensio Vascular Workstation (version 9.1 SP2, Pie Medical Imaging BV, Maastricht, the Netherlands). The aortic lumen shows the part of the aorta where blood flow is possible, but fails to provide information on other factors that influence a possible landing zone for an endograft. Aneurysmatic aortas often show thrombus near the aortic wall (Figure 2 (b)). The location of thrombus would mean an unfavorable site for endograft placement as it compromises the amount of circumferential apposition that can be achieved by the graft with healthy parts of the aorta, needed for seal.

It is hypothesized that for the prediction of postoperative SAL, also information about preoperative thrombus location, degree of circumference and thickness is needed. Efforts have been made to classify the presence of thrombus in the aorta [36], but comparison between classifications of thrombus distinguishing a group of complicated EVARs and a control group fails to show significant difference [37].

The fact that classification of thrombus does not provide useful input for predictive models might be explained by the complex nature of thrombus morphology in the aortic neck. Here, the degree of circumference and thickness of thrombus could prove more unfavorable than the length of thrombus. This makes it difficult to objectively describe cutoff points where apposition of an endograft can be achieved and where it will not have sufficient apposition. SSMs could be expanded to comprise thrombus morphology in the aortic neck to better describe the overall anatomy influencing postoperative SAL. Segmentations of the aortic wall have have been used for the assessment of aneurysm progression [36] and include thrombus. Literature on fully automated software for the acquisition of such segmentations has recently been published [38] enabling the creation of large datasets with segmentations containing information on thrombus morphology.

Also the composition of datasets could be important, both for the creation of a relevant SSM as for the training of predictive algorithms. Current datasets include only patients preoperatively selected for EVAR. It is to be expected that non-relevant thrombus is present in these aortic necks and therefore the influence of thrombus on predictions based on these datasets would be small. If the dataset is expanded with patients treated with FEVAR, also juxta- and pararenal aneurysms will be represented, which can increase the clinical relevance of the predictive model. This model could provide valuable information for the task clinicians are faced with: to select the right endovascular device for their patients. *Is the aortic neck suitable for infrarenal EVAR, or does its morphology and/or presence of thrombus make it necessary to implant a FEVAR-device?* And also, *How many fenestrations are needed before sufficient proximal fixation of the FEVAR device is achieved?* are relevant questions here.

Accurate preoperative prediction of postoverative SAL could help in answering the first question. Models providing a prediction on the exact location where the endograft will lose circumferential apposition with the aortic wall, could answer question two. Any type of abdominal aneurysm would be presented to an SSM and the values for the principal components are used to predict to location relative to the lowest renal artery where circumferential apposition is to be achieved. Expanding on previous work described in the background for this thesis should help the development of a *virtual stenting tool*.

1.4.1 Hypothesis

Described technical-medical challenges form the basis for new hypotheses that are researched in this work:

Hypothesis 1: Including thrombus in statistical shape models (SSMs) improves the predictive accuracy of postoperative stent graft apposition length (SAL) in the aortic neck compared to models using only lumen segmentations.

Context: Investigating the impact of including thrombus in statistical shape models (SSMs) on the predictive accuracy of postoperative stent graft apposition length (SAL) in the aortic neck.

H0: *Prediction models for postoperative SAL in the aortic neck based on SSMs that include thrombus do not have significantly higher predictive accuracy compared to models that include only lumen segmentations.*

H1: *Prediction models for postoperative SAL in the aortic neck based on SSMs that include thrombus have significantly higher predictive accuracy compared to models that include only lumen segmentations.*

Hypothesis 2: The use of Statistical Shape Models (SSMs) enables accurate prediction of continuous outcomes for the end of circumferential apposition between the aortic prosthesis and the aortic wall in the juxtarenal aortic neck.

Context: Investigating whether statistical shape models (SSMs) enable the prediction of continuous outcomes for the end of circumferential apposition between an aortic prosthesis and the aortic wall in the juxtarenal aortic neck?.

H0: *SSMs do not enable the accurate prediction of continuous outcomes for the end of circumferential apposition between the aortic prosthesis and the aortic wall in the juxtarenal aortic neck.*

H1: *SSMs enable the accurate prediction of continuous outcomes for the end of circumferential apposition between the aortic prosthesis and the aortic wall in the juxtarenal aortic neck.*

1.4.2 Thesis outline

In the introduction of this thesis, background information is provided on the clinical relevance of research towards the preoperative prediction of postoperative apposition of endografts. The current state of technical development within this field is described resulting in relevant technical medical questions. For this thesis, the following objectives are described:

- The implementation of thrombus in a statistical shape model, by using outer wall segmentations of the aorta.
- Creating a dataset representing patients treated with EVAR and FEVAR, including morphological variation of abdominal aneurysms.
- Train prediction models with this new dataset, answering the question if adequate postoperative seal is to be achieved by EVAR (without fenestrations) in the infrarenal aortic neck.
- Expand the segmentations to include more possible continuous outcomes of pararenal SAL, within the combined EVAR/FEVAR-dataset.
- Predict the location where circumferential apposition of the proximal endograft with the aorta will end for the combined EVAR/FEVAR-dataset.

The second chapter of this thesis describes a study on the creation of a SSM, where thrombus is included using the outer wall segmentations of aortas. The SSM is based on a dataset combining patients receiving either EVAR or FEVAR, with the goal to capture a wide variety of morphological variations. The SSM is validated and predictions based on these SSMs are compared to conventional strategies (i.e. outcome of recent developments).

The third chapter describes a study towards clinical implementation of prediction models, first expanding the regions used for SSMs to include more locations of observed outcomes of postoperative SAL by the proximal endograft. Prediction models are adapted to provide continuous outcome, effectively predicting the location where apposition of the endograft with a newly presented aortic morphology will end. Subgroups are created to allow predictions of the juxtarenal aorta and the pararenal aorta separately.

2 Implementation of Intramural Thrombus in Statistical Shape Models Describing Aortic Morphology for Preoperative Prediction of Shortest Apposition Length in the Aortic Neck

2.1 Introduction

For adequate endovascular repair of abdominal aortic aneurysms (AAA) by endovascular aneurysm repair (EVAR), sufficient proximal seal is needed to prevent a type 1a endoleak (T1aEL) [25]. Seal is achieved by circumferential apposition of the endograft to the aortic wall, in the infrarenal neck. To determine if sufficient seal is feasible in the aortic neck of a particular patient, certain characteristics are quantified. Hostile aortic neck characteristics including length < 10 mm, diameter > 30 mm, high infra- and supra-renal angulation, a conic shape, and the presence of circumferential thrombus have been associated with an increased risk to develop a type Ia endoleak (T1aEL) [39] [28] [40]. In previous studies it was hypothesized that quantification of these measurements alone might lead to simplification of the complex morphology of the aortic neck [29]. In this earlier work, statistical shape models (SSMs) have been created using segmentations of the aortic neck lumen [29] [37], providing an alternative strategy for the quantification of aortic neck morphology.

In recent literature, descriptors of aortic morphology provided by statistical shape models have been used to train classification models [34]. These models provide a prediction of the shortest apposition length (SAL), defined as the shortest distance between the proximal circumference of the fabric and the first slice perpendicular to the centerline where circumferential apposition of the fabric with the aortic neck is lost [41]. Patients with SAL < 10 mm might be at higher risk of developing a T1aEL [23]. Classification models are able to predict if SAL will be achieved $\geq or < 10mm$ with 78% accuracy [34]. This model could be a valuable tool to preoperatively identify patients that will not achieve sufficient apposition after standard EVAR. These patients could then be evaluated for alternative AAA repair by open surgery or fenestrated EVAR (FEVAR).

The SSMs that are used to generate the input for these classification models are based on segmentations of the aortic lumen and the dataset only contains patients that received standard EVAR. Consensus is that the apposition of an endograft with the aortic wall is greatly influenced by the presence of thrombus in the aorta [28]. Currently no representation or quantification of thrombus location or morphology is provided for the classification model. A pararenal AAA, with significant thrombus in the aortic neck, could still be wrongfully presented as a slim lumen by automated prediction. The classification models could, in this case, predict SAL > 10 mm suggesting the patient is suitable for repair by standard EVAR, while the presence of thrombus would make this impossible.

SSMs however, can be expanded to include extra dimensions of data in addition to 3D morphology of a structure [42] [43]. Current development of automatic outer wall-segmentations of the aorta [38] [44] enables generation of large datasets containing information of thrombus morphology in the aorta. The goals of this study are to implement information on thrombus location and morphology in statistical shape models. The dataset used for the SSM includes juxta- and pararenal aortic aneurysms. Classification models using the newly generated SSMs are compared to previously used methods.

2.2 Method

A retrospective observational study is conducted towards the prediction of infrarenal apposition by EVAR. The goal of prediction models is to predict whether a patient had a SAL of either $\geq or < 10mm$ on the first operative CTA. This value is chosen as patients with a SAL <10 mm may have an increased risk of developing a T1aEL [23].

Population

Two previously used datasets [37] [45] are combined to represent patients with infra, juxta, para and suprarenal abdominal aneurysms. The first dataset includes patients that received technically successful primary EVAR and had the availability of a CTA scan within 6 months pre-EVAR and is previously described by van Veldhuizen et al. [37]. A subset is created for patients developing a T1aEL, detected after the first postoperative CTA. Exclusion criteria are treatment of a symptomatic or ruptured AAA, adjunct proximal fixation such as additional cuffs or endoanchors, computed tomography (CT) scans with insufficient contrast for assessment in a vascular work- station, and intentional low (>5 mm) positioning of the endoprosthesis relative to the lowest renal artery.

The second dataset included patients receiving a Zenith (Cook Medical Inc., Bloomington, IN, USA) fenestrated graft. Patients were consecutively treated between January 2012 and December 2019 in the University Medical Center Groningen (UMCG) [45]. Exclusion criteria are treatments for aortic dissections, symptomatic AAA or rupture, use of additional proximal fixation (e.g., endo- anchors or thoracic extension cuff), and endovascular treatment of the thoracic aorta in the past with overlap of the fenestrated graft.

Within the second group, an additional exclusion criterion for this study are FEVARs with a main body length shorter than the shortest apposition length on the first postoperative CTA. If circumferential apposition of the FEVAR-device with the aorta is achieved closer to the baseline (i.e. LRA) than the point where the main body of the device tapers, apposition might be lost by this tapering. These situations are deemed unsuitable as input for the prediction of postoperative apposition of an endograft as it might be a false representation of the apposition length that could have been achieved in the aorta without tapering.

Thrombus segmentation

Morphology of the aortic lumen, the center lumen line and the representation of the lowest renal artery on the center lumen line were semi-automatically obtained, following a measurement protocol using 3Mensio Vascular Workstation (version 9.1 SP2, Pie Medical Imaging B.V., Maastricht, the Netherlands). Outer wall segmentations of the aorta were obtained manually using 3D Slicer [46], following a predetermined protocol, by researchers WAV, NK and DR. The aorta, including the aortic outer wall, was segmented on axial slices with a maximum of 3mm intervals between the aortic bifurcation and cranially of the celiac trunk. These segmented slices were subsequently interpolated and smoothed using median smoothing.

Apposition evaluation

The calculation of postoperative circumferential endograft apposition with the aortic wall is previously described [41] and includes the use of Vascular Imaging Analysis (VIA) prototype software (Endovascular Diagnostics B.V., Bussum, The Netherlands). The CLL, aortic lumen segmentation and coordinates of the lowest renal artery are determined with previously used methods [23] in 3Mensio Vascular Workstation.

Preprocessing of aortic morphology

Before principal component analysis (PCA) can be performed, the aortic morphology was pre-processed by parameterization of the aorta, using a protocol developed by a UMCG research group [29]. The segment of the aorta included for PCA is determined by distance on the CLL relative to the representation of the lowest renal artery, a segment of 5cm proximal of the lowest renal artery and 2 cm distal of the artery. Following the protocol previously used by the research group, this segment is divided in eight equidistant contour rings with 1 cm intervals. 360-degrees parameterization of the aortic wall with steps of 10 degrees along these intervals resulted in 288 contour points that are used for point-to-point correspondence of all patients in the study.

This parameterization technique can be applied for the three dimensional representation of the aortic lumen and aortic wall using similar strategies. In this study, the previously used protocol is expanded, by the introduction of the aortic wall as a morphological descriptor and the implementation of a fourth dimension in the representation of the aorta.

Parameterization is based on vectors originating from the intervals on the CLL. 36 vectors are used perpendicular to the CLL with 10-degree intervals, resulting in 36 points for each contour interval where these vectors intersect with the segmentation of the aorta. Intersections can be found between the vectors and both the aortic lumen and aortic wall, but as the CLL is equal for both, the resulting contour points will correspond. The thrombus thickness for each contour point in the aneurysm is described by the correspondence between the contour points resulting from the two different segmentations. Calculating the Euclidean distance using equation 1, results in the thrombus thickness (δ).

$$\delta_n = \sqrt{(W_{x_n} - L_{x_n})^2 + (W_{y_n} - L_{y_n})^2 + (W_{z_n} - L_{z_n})^2} \tag{1}$$

Here, the thrombus thickness for each contour point (δ_n) is calculated taking the Euclidean distance between the three dimensional coordinates of each contour point on the aortic lumen and the aortic wall, $(L_{x_n}, L_{y_n}, L_{z_n})$ & $(W_{x_n}, W_{y_n}, W_{z_n})$ respectively. A representation of the lumen- and wall-contour points resulting from parameterization, including a vector representation of the thrombus and CLL is shown in Figure 5.

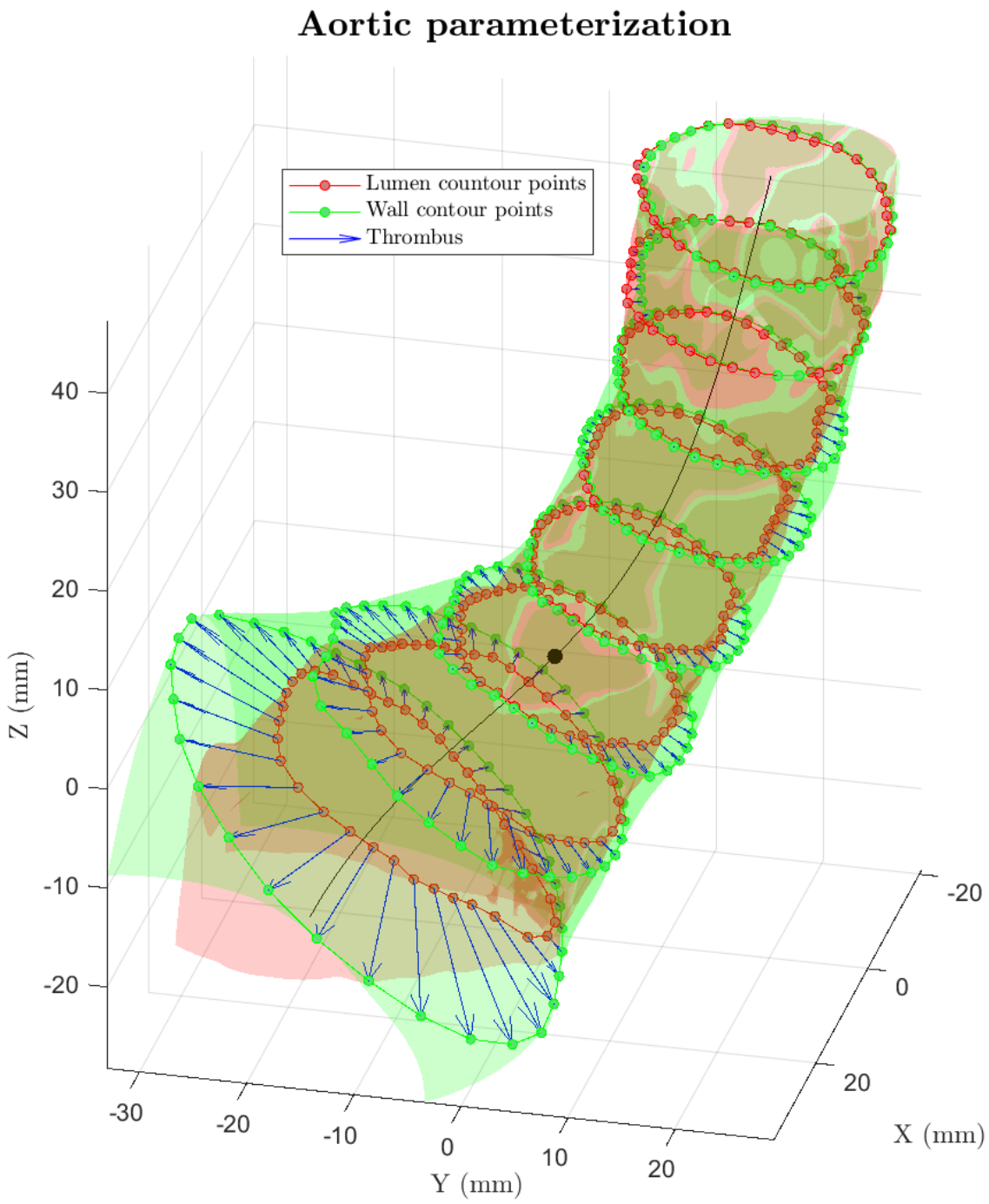


Figure 5: Example of aortic parameterization, showing the aortic lumen and wall segmentation in red and green, respectively. The center lumen line is shown in black with the addition of the representation of the lowest renal artery shown as a black dot. A blue vector representing the thrombus thickness is shown for each contour point.

In the case of strong angulation of the CLL, the perpendicular planes can cross before intersection with the segmentation. This would result in overlapping contour lines and is unfavorable for PCA and the creation of the SSM as these points will not follow the same distribution in all patients. To prevent this, within the original framework, the eight parallel contour points from each contour line are rearranged at equal distances along the longitudinal axis using interpolation. As in this study a comparison is to be made between previously used methods and the addition of the thrombus thickness in the fourth dimension, this interpolation is implemented as described before [37]. Interpolation is based on 3D lumen segmentation coordinates for the lumen and $4D_{lumen}$ models and on the 3D wall segmentation coordinates for the wall and $4D_{wall}$ models. In this way, both the addition of the fourth dimension and the options for the 3D-mesh (lumen or wall) is researched.

Four Dimensional Statistical Shape Model

Parameterization of the aortas in relation to the lowest renal artery registers the three and four dimensional data points the similar relative positions. The value of thrombus thickness for each contour points was directly integrated in the data matrix used for PCA [42] [43] to produce a model of shape variation correlated with variations in thrombus thickness. The four dimensional data matrix (M) contains the concatenated point coordinates of each contour and the corresponding thickness of thrombus, similar to conventional SSMs [47] as shown in equation 2. Different matrices are created with contour coordinates resulting from either lumen or wall segmentations.

$$M = \begin{bmatrix} x_1 & y_1 & z_1 & \delta_1 \\ x_2 & y_2 & z_2 & \delta_1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & \delta_n \end{bmatrix} \tag{2}$$

In this data matrix, x_i, y_i, z_i are the three-dimensional coordinates of contour coordinate i and δ_i is the corresponding thrombus thickness value. PCA is applied to this data matrix, without scaling to integrate size variations in the aortic morphology. Translation is performed by the parameterization process, for all patients the location of the LRA is used as the origin of the 3D structure. The morphology of an individual aorta can be described using scores of the principal components (PCs). PCs represent the deviation from the mean shape of aortas present in the dataset and are expressed in standard deviations (SDs). Unique shape variations are found by the SSM and objectively describe the individual aorta and its shape variation compared to other aortas in the dataset.

SSM Evaluation

The evaluation of statistical shape models has been described by Davies et al.[48]. Generally accepted evaluation of SSMs follows this principle and includes:

Compactness: Describes how well the model is represented by a specific amount of parameters. Compactness is used to evaluate how efficiently the variation in the dataset can be represented by the model.

Generalization ability: Depicts the ability of the model to represent new data samples, for a given number of modes of variation. Using a Leave-One-Out (LOO-)method, the model is constructed on all but one aorta. Next, the left-out aorta is fitted to the model and the Euclidean distance is measured between the fitted and the actual aorta. To evaluate the ability of the model to generalize the thrombus thickness, the value of δ can be used as a fourth dimension in the Euclidean distance calculations [49]. The generalization is calculated using equation 3 [50], where the LOO-process is performed for each example x_j. The Euclidean distance is calculated between the actual sample x_j and the fitted aorta $\tilde{x}_j$, for each number of samples n_s and for varying amount of modes (i.e. PCs) n_m.

$$G(n_m) = \frac{1}{n_s} \sum_{j=1}^{n_s} \left\| x_j - \tilde{x}_j(n_m) \right\| \tag{3}$$

Specificity: The specificity is used to quantify how similar new cases generated by the SSM are to cases included in the training set. An SSM should only generate new models that are representative of the described anatomy [42]. Specificity is calculated by creating a large amount (M) of artificial instances (y_A), and comparing these to real models included in the training set (x_j). In this study $M = 2,000$ artificial instances are created. This is done using a varying amount of modes (n_m). For each artificial instance, the Euclidean distance to the nearest element (x_j) in the dataset is found, following equation 4 [50].

$$S(n_m) = \frac{1}{M} \sum_{A=1}^{M} \min_j \left\| x_j - y_A(n_m) \right\| \tag{4}$$

For the 4D SSMs, specificity is shown separately for the first three dimensions, showing either lumen of wall contours and the possible variations in thrombus thickness. More specific models show smaller distances.

Classification

Supervised learning is used for the preoperative prediction of SAL. Scikit-learn [51] (version 1.5) is used in Python (version 3.11) to develop software for this model. Scikit enables the use of different classifiers for predictions, for each type of prediction different classifiers are preferred.

Ensemble learning is a process where multiple estimators are generated and combined to solve a particular classification problem. Ensemble learning is primarily used to improve the classification performance of a model, or reduce the likelihood of an unfortunate selection of a single poor estimator. Random Forest classifiers use ensemble learning to improve predictions in classification [52]. The base estimators in random forest are decision trees. A Random Forest randomly selects a set of features which are used to decide the best split at each node of the decision tree and combines the outcomes of all estimators for the final prediction. After Random Forest training, the influence of features on the predictions can be extracted, showing the importance of the features. In this study, four random forest classifiers are trained to provide binary classification outcome and to show feature importance based on the four different SSMs: Lumen, Wall and $4D_{lumen}$ and $4D_{wall}$.

Validation of the model is performed by leave-one-out cross validation. Subsequently, each patient is individually subtracted from the dataset. The four SSMs are constructed based on the reduced dataset and the left-out patient is fitted to the models. The Random Forest model is trained on PCs and binary SAL ($\geq or < 10mm$) of the reduced dataset and validation of the model is performed on the left out patient. A prediction is made based on the PCs of the fitted patient and compared to the true binary value for the SAL. Similar to previous machine learning based prediction-models used in the research group [34], true positive was defined as a patient with <10 mm predicted apposition and <10 mm actual apposition, a true negative as ≥10 mm predicted apposition and ≥10 mm actual apposition, a false positive is defined as <10 mm predicted apposition, but ≥10 mm actual apposition and a false negative as ≥10 mm predicted apposition, but <10 mm actual apposition.

Statistical analysis

Using the compactness curve, only the first PCs describing 95% of variation are used for evaluation. The Mann-Whitney U test is used to find significant differences in PCs between the groups with or without < 10 mm SAL. These significant PCs are found for the four strategies used for the creation of SSMs.

In previous studies, classification models were trained with only significantly different features by feature selection using the Mann-Whitney U test [34]. In this study we chose to incorporate all features, together describing 95% of shape variation. This is possible because the Random Forest classifier selects the most important features during training. Selecting features this way could make the implementation of an arbitrary p-value following statistical analysis as cutoff obsolete. Comparing feature importance resulting from the random forest classification to the results by the Mann-Whitney U test could validate that the random forest classifier uses the features that are significantly different.

The DeLong's test [53] is used as an evaluation metric for the area under the curve receiver operating characteristics curve (AUROC) resulting from the predicted probabilities by the three classification models. The test shows significant differences between the best performing model and the others, with p-values < 0.05 considered significant. Bootstrapping with 1,000 samples is performed to calculate 95% confidence intervals (CI) for evaluation of the classification models using Statkit [54]. Metrics that will be evaluated are: AUROC, accuracy, sensitivity and specificity.

2.3 Results

Study population

The first dataset included is described in detail by van Veldhuizen et al. [37] and consisted of 147 patients that received EVAR. Within this dataset, 54 patients developed a T1aEL and 93 patients are included in the uncomplicated group. The second dataset consisted of 74 patients that received FEVAR with at least one fenestration. Five (6.8%) of these patients where excluded from this study as the main body length of the fenestrated graft was shorter than the apposition length. In total 216 patients are included with a median SAL of 14.33 mm (24.26 mm - 0.73 mm).

SSM evaluation

The compactness curve for the combined dataset ($n = 216$) is shown in Figure 6. 95% of variation in morphology is described by 10, 12 and 17 PCs for the lumen, wall and both 4D SSMs respectively.

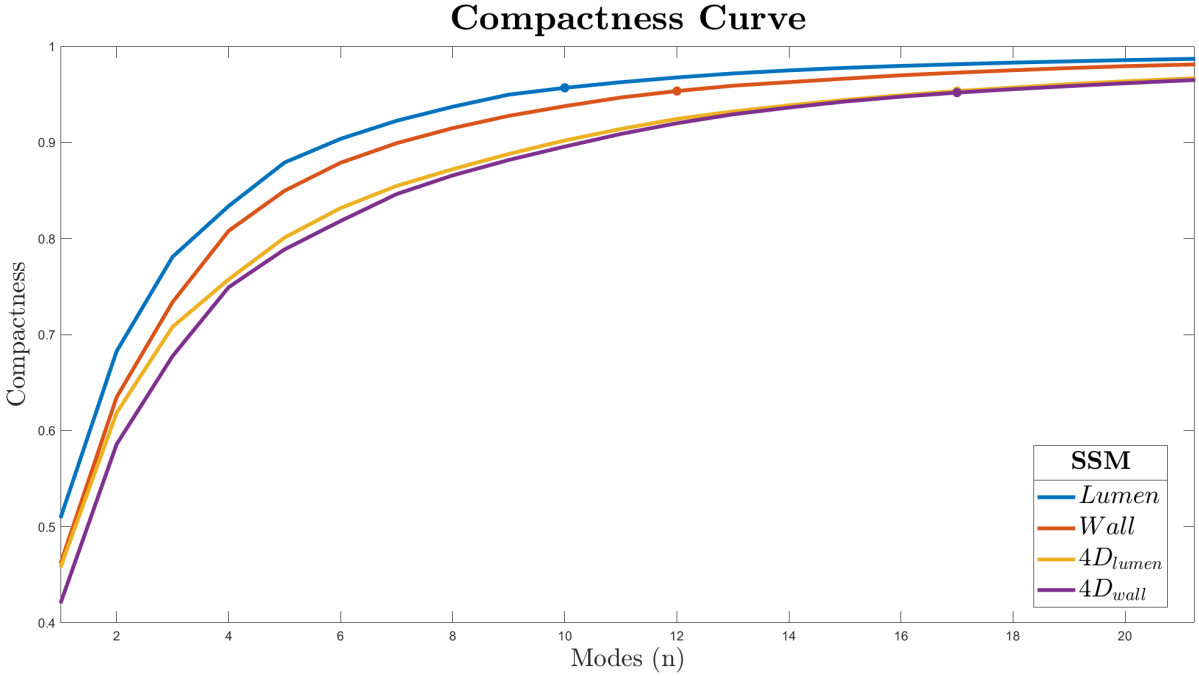


Figure 6: Compactness curve of the statistical shape models show cumulative morphological variation that is described by the principal components (PCs). The dots indicate the amount of PCs needed to describe 95% of morphological variation.

Mean generalization, or the ability of the SSM to reconstruct a given new patient-specific shape is shown in Figure 7. For the number of PCs when 95% of variation in morphology is described, the mean accuracy is 3.27 (95% CI 3.0–3.5) mm for the lumen, 3.8 (95% CI 3.4–4.1) mm for the wall, 4.2 (95% CI 3.8–4.5) mm for the $4D_{lumen}$ and 4.7 (95% CI 4.2–5.3) SSMs.

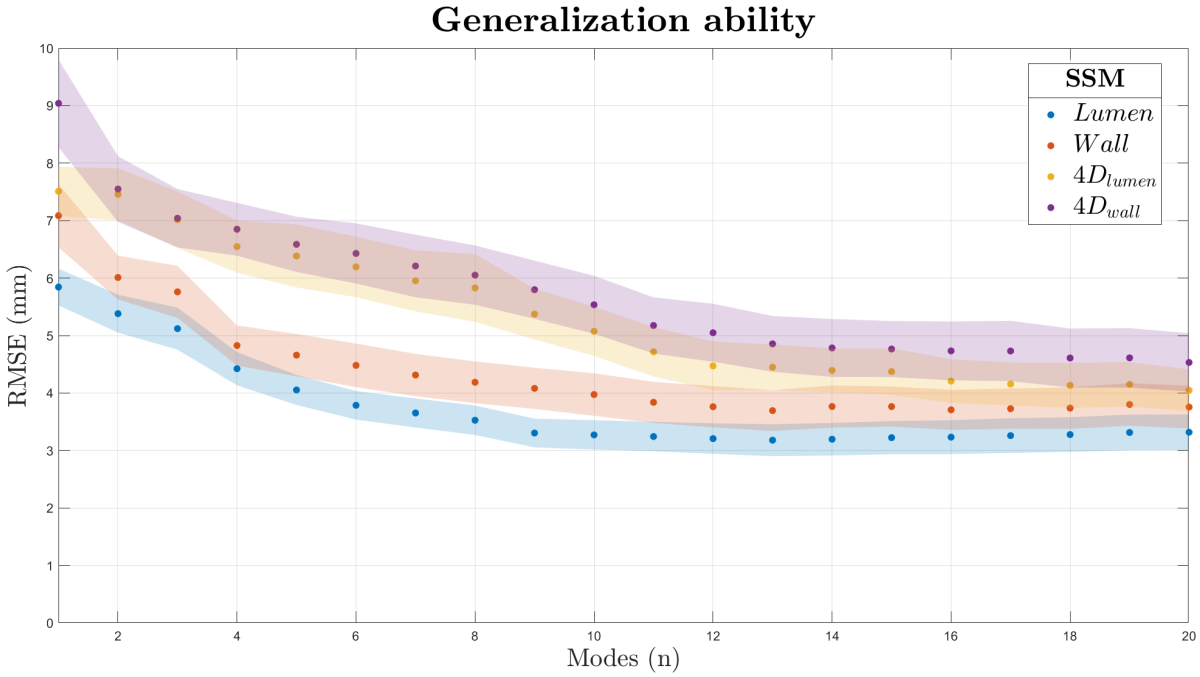


Figure 7: Generalization is shown as the root mean squared error (RMSE) between reconstructed samples and the actual sample, for all patients by leave-one-out (LOO). Mean values are shown as dots for statistical shape models (SSMs) with increasing n modes and are surrounded by the 95% CI.

The ability of the model to generate valid instances, also included in the training data is shown as the specificity in Figure 8, showing similar results for the 4D SSMs compared to the corresponding 3D SSM. Figure 9 shows the ability of the 4D SSMs to generate valid thrombus thickness, showing similar results for both methods.

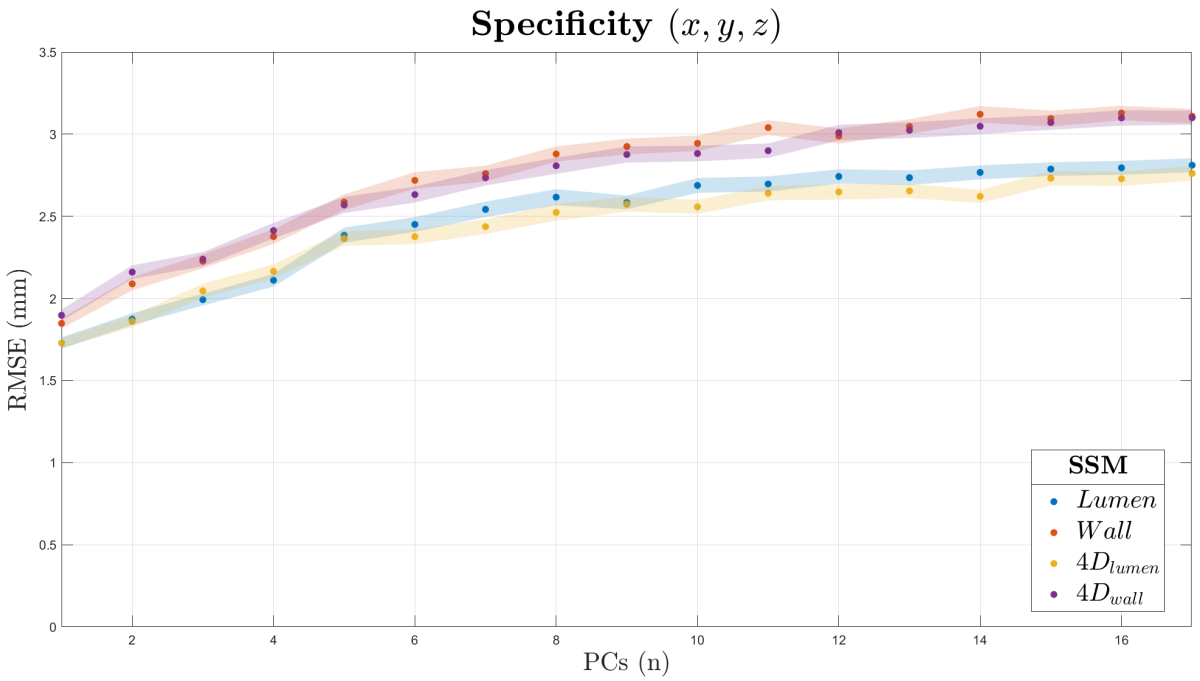


Figure 8: Specificity is shown as the lowest root mean squared error (RMSE) between 2000 random generated samples and the actual dataset. Mean values are shown as dots and are surrounded by the 95% CI.

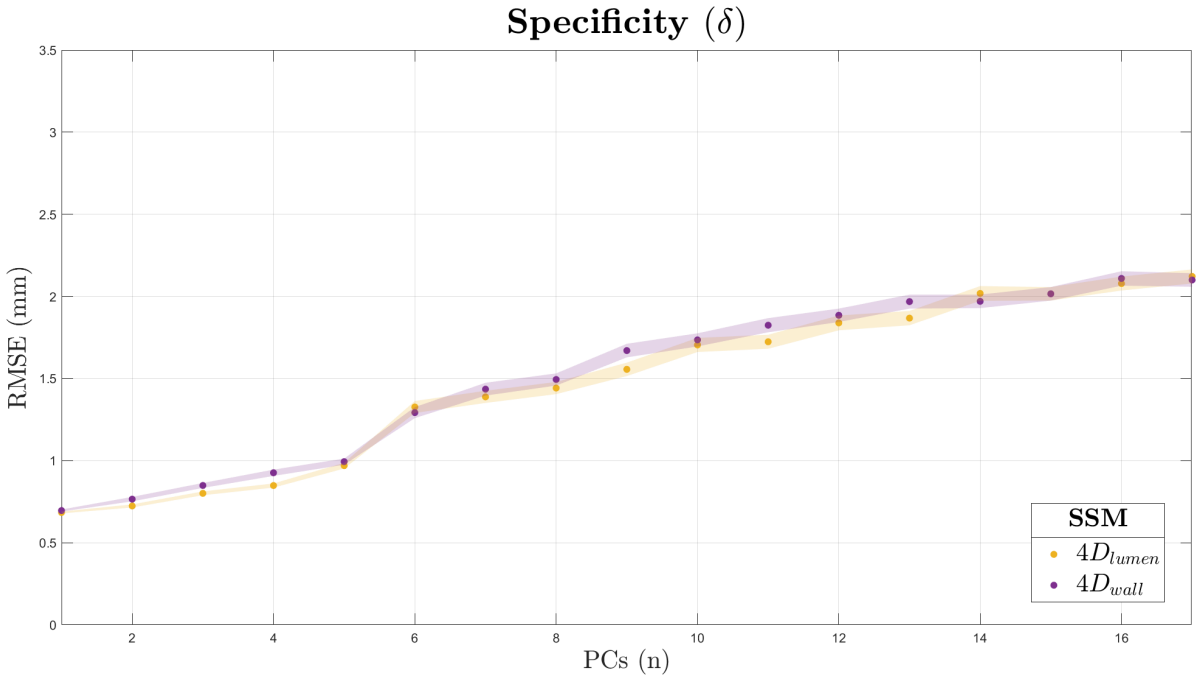


Figure 9: Specificity is shown as the lowest root mean squared errors (RMSE) between 2000 random generated samples and the actual dataset. Mean values are shown as dots and are surrounded by the 95% CI.

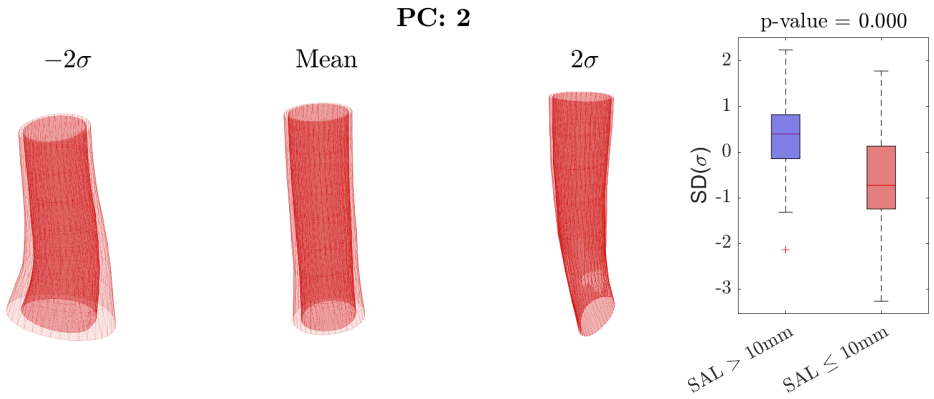


Figure 10: Visualization of PC 2 from the $4D_{wall}$-SSM, including the distribution of SDs between the two SAL-groups shown in the boxplot with the outcome of the MWU-test.

Statistical analysis

An exemplary visualization of a PC is provided in figure 10. Variations in morphology depicted by SD-values of PCs are not directly linked to traditional measurements, such as aortic neck length and diameter. The PCs, however, do contain information on these conventional measures, but are able to show new variety and combinations in varieties within the dataset. For the 4D SSMs variation in thrombus thickness between the SDs can be seen. An overview of all p-values following the Mann-Whitney U test are shown in Table 1.

Table 1: P-values following the Mann-Whitney U test comparing < 10 mm and SAL ≥ 10 mm groups shown for all PCs. Significant p-values are shown bold.

# Mode	Lumen	Wall	$4D_{lumen}$	$4D_{wall}$
1	0.0616	**0.0296**	**0.0477**	**0.0186**
2	**0.0001**	**0.0001**	**0.0001**	**0.0001**
3	0.2341	0.1815	0.2341	0.1441
4	**0.0001**	**0.0001**	**0.0001**	**0.0001**
5	**0.1251**	0.9593	**0.0019**	0.8088
6	**0.0288**	**0.0039**	**0.0001**	0.0532
7	0.3593	0.1779	**0.0001**	**0.0162**
8	**0.0086**	0.6732	**0.0178**	0.1160
9	0.7291	0.2359	**0.0001**	0.3342
10	**0.0202**	0.2767	0.1032	0.6112
11	-	0.2333	0.2777	0.0691
12	-	**0.0158**	0.2493	**0.0036**
13	-	-	0.2623	0.9982
14	-	-	0.4102	0.9593
15	-	-	**0.1554**	0.0681
16	-	-	0.3478	**0.0475**
17	-	-	**0.0110**	0.6993

Classification

After classification based on the PCs for 216 patients, the ROC-curves from the Random Forest models using four different SSMs are shown in Figure 11. The highest AUC value is 0.90 and is achieved by the model trained with an SSM of the $4D_{wall}$. Results of the classification models showing AUC, accuracy, sensitivity and specificity are listed in Table 2

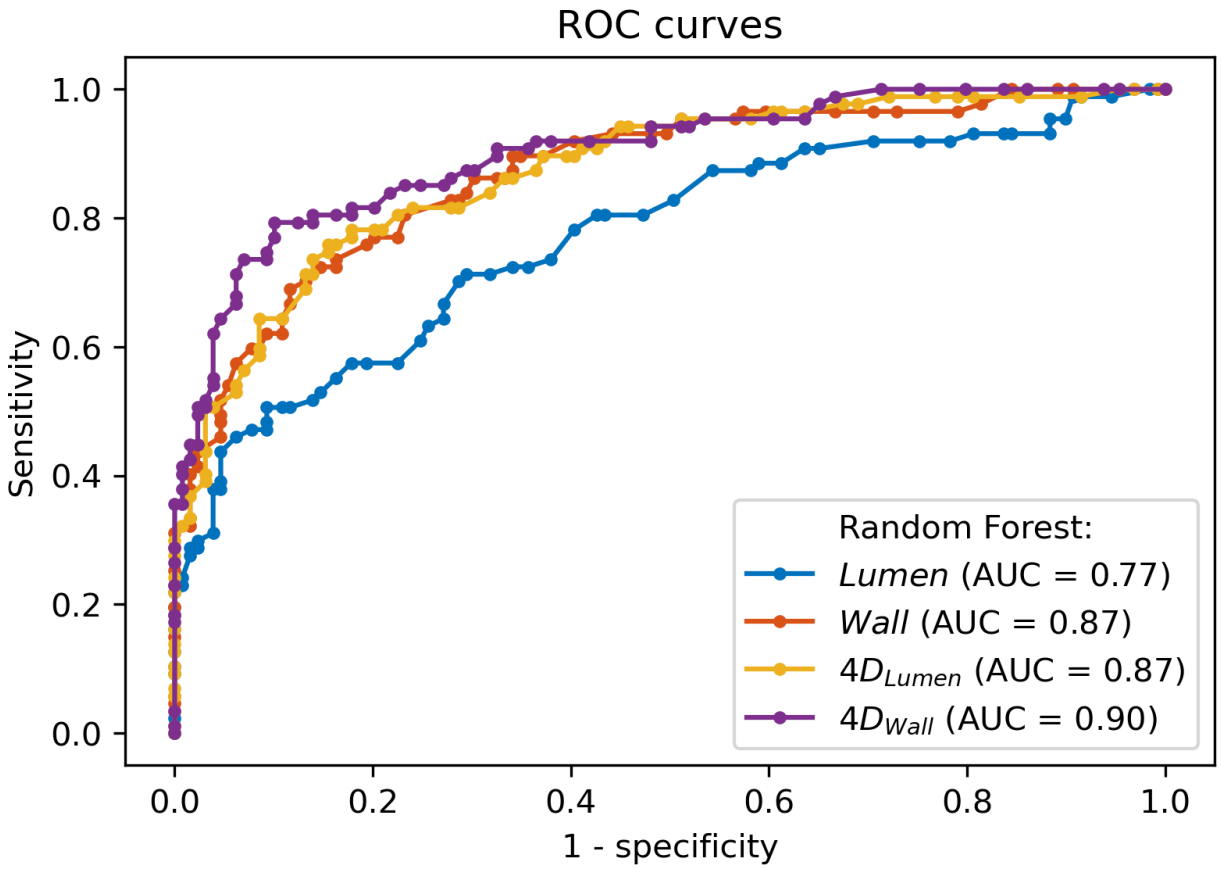


Figure 11: Receiver Operating Characteristics (ROC) curves and area under the curve (AUC) values of the four random forest prediction models.

Table 2: Results of the three classifications by random forest classification. 95% CI are shown for all metrics after bootstrapping by 1000 iterations [54]. The p-value is calculated with the use of the DeLong's test, comparing the Best Performing Model (BPM) to the other three.

	AUC	**Accuracy**	**Sensitivity**	**Specificity**	**p-value**
Lumen	0.77 (0.70 - 0.83)	0.72 (0.66 - 0.78)	0.52 (0.41 - 0.62)	0.86 (0.80 - 0.92)	*0.0001*
Wall	0.87 (0.82 - 0.92)	0.81 (0.75 - 0.86)	0.69 (0.59 - 0.79)	0.88 (0.83 - 0.94)	*0.0386*
4D$_{lumen}$	0.87 (0.82 - 0.92)	0.81 (0.75 - 0.86)	0.71 (0.62 - 0.80)	0.86 (0.81 - 0.92)	0.1929
4D$_{wall}$	0.90 (0.86 - 0.94)	0.84 (0.80 - 0.89)	0.71 (0.60 - 0.81)	0.94 (0.89 - 0.98)	BPM

Feature Importance

The random forest classifiers are able to show the importance of each feature for improvement of classification. An overview of principal component importance for classification is given in Figure 12, showing PC 2 as the most important feature.

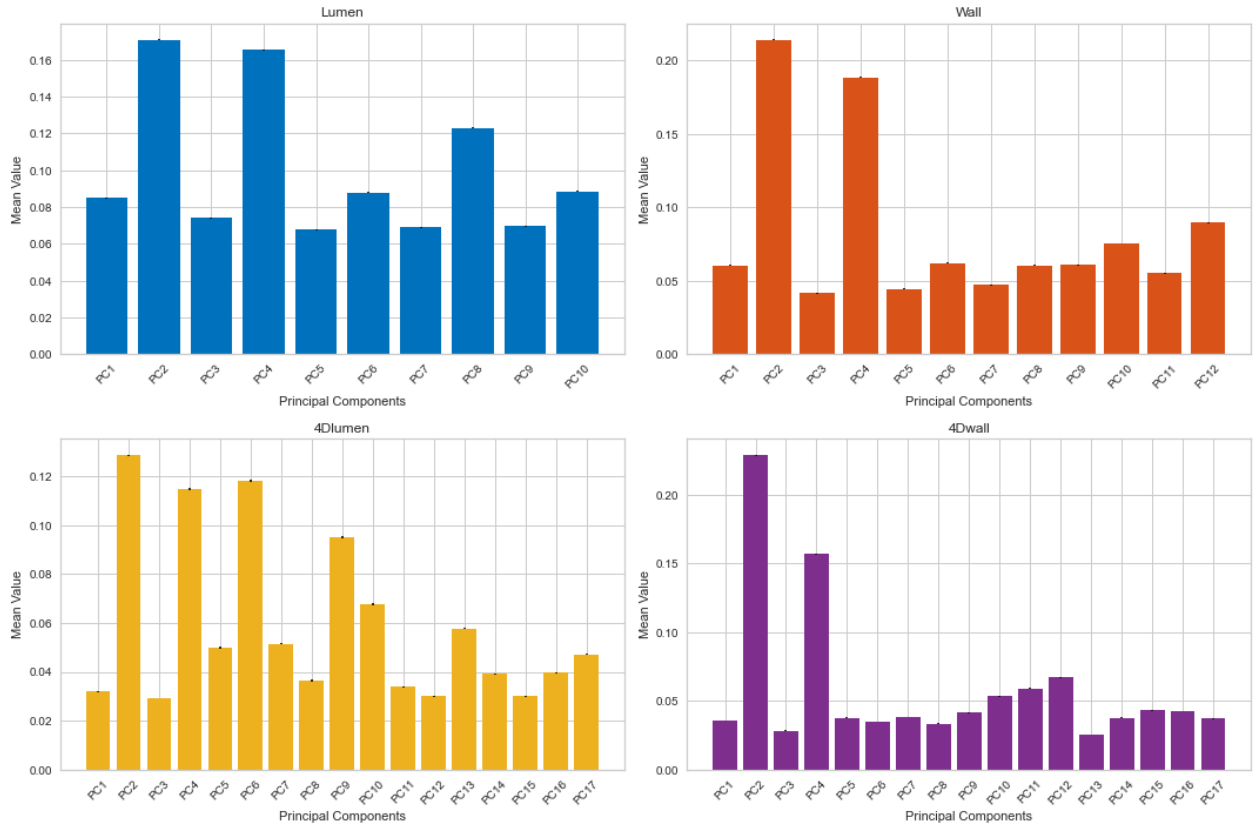


Figure 12: The impact of each individual feature on the classification outcome using the random forest classifier is shown for all classification models.

Comparing the feature importance shown in Figure 12 with the results from the Mann-Whitney U test shown in Table 1 provides an interesting insight in the construction of the Random Forest model. Low P-values following the MWU-test for a PC results in high feature importance of that PC in the classification model.

Visualizations of PC's with a feature importance ≥ 0.10 are shown in Appendix A.

2.4 Discussion

In this study, we presented a new, four dimensional, method to incorporate intramural thrombus into SSMs of abdominal aortic aneurysms. A previously used dataset within our research group [37] is expanded with patients that received FEVAR [45] to represent all types of abdominal aortic aneurysms, with varying thrombus morphology. The new SSMs are validated, using generally accepted quantifications of compactness, generalization and specificity. Random Forest classifiers trained with the novel 4D SSMs, outperform previously used classifiers [34] when predicting SAL < 10 mm and SAL ≥ 10 mm labels for the presented aortic morphology.

Interpretation of results

SSM evaluation shows that SSMs including thrombus information for PCA are more complex in comparison to 3D SSMs (i.e. need more PCs to describe 95% in shape variation, as seen in Figure 6). Both 4D models are less compact than the 3D models. This indicates that the dataset used for input includes more variation in shape than the other two. This increase in complexity is explained by the addition of the dimension containing thrombus thickness.

The dataset based on wall segmentations is more complex than the lumen SSM, suggesting that more anatomical variation is captured with the wall segmentations. This would support the part of the hypothesis, stating that SSMs based on only lumen segmentations might be an oversimplification of the true morphology of the (pararenal) aorta. Within the dataset, description of the true anatomy might be lost if only described by the lumen. The lumen might show as a straight tubular structure, but will in reality be surrounded by thrombus, providing anatomical variation to the dataset containing wall segmentations.

Over-fitted SSMs can suffer from an in ability to adequately describe the morphology of unseen cases [50]. The generalization ability curve (Figure 7) shows increasing RMSE-values for more complex SSMs. This can be explained by the fact that the SSMs describing most anatomical variation need more modes to do so and these relatively complex modes might not perfectly describe the unseen case. All SSMs reach a plateau with relatively similar RMSE, showing that also more complex SSMs and also a 4D SSM can be used to describe the aortic neck morphology of an unseen patient. Within the size of our dataset (N = 216) an equilibrium of generalization is achieved for the most complex models, suggesting this size of dataset is adequate for the creation of a SSM.

The 4D SSMs are able to generate 3D morphology of the lumen and wall similar to the 3D models, as shown by the specificity curve in Figure 8. Additionally these models are able to represent thrombus thickness present for all contour points included. After evaluation, the incorporation of thrombus thickness in a SSM of the aorta, to accurately describe an unseen instance is feasible.

Statistical analysis of the distribution of PC scores between the SAL < 10 mm and SAL $\geq$ 10 mm groups shows which PCs are important for the training of classifiers. In previous methods, the outcome of statistical analysis has been used to determine the input features for classifiers [34]. Using ensemble learning could make this step obsolete, allowing all PCs as input and relevant features are selected by the classification model. Comparison between the outcome of the Mann-Whitney U test and the feature importance following random forest (Figure 12) shows that important features found by the classification models also show a large variety in PC-scores between the SAL-groups. This implies that feature selection by the random forest could replace selection by the Mann-Whitney U test.

An added benefit of this is seen in the application of Leave One Out Cross Validation. Before, the selection features would be made on an SSM including all aortas in the dataset. As in this study, effectively 216 SSMs of 215 aortas are created to fit individual aortas to an SSM that is not trained on that instance, significant PCs could actually vary depending on the left out instance. This could be solved by feature selection using Random Forest.

An increased performance is seen in Figure 11 between all models that incorporated thrombus information as supposed to the classification based on an SSM of the lumen only. The ROC curve shows the superiority of the $4D_{wall}$ model, achieving an AUC of 0.90, confirmed by the DeLong's test. The increase in performance of the $4D_{wall}$ classification compared to the $4D_{lumen}$ is not significant. Improved accuracy of the models provide increased valuable tools for the risk assessment by preoperative prediction of apposition lengths < 10 mm.

In this assessment of risk, clinical relevance is highest in the prediction of low apposition lengths as opposed to the ability of a model to accurately predict longer apposition lengths. Clinical relevance is higher here, as a faulty prediction of larger SAL could influence the decision to implant a standard EVAR. This could result in an achieved SAL $< 10mm$ and increased chances of developing a T1aEL. The capability of the models to provide this clinically relevant prediction is quantified by the sensitivity of the classification models. It is interesting that the 4D SSMs allow prediction models to outperform the 3D wall model when comparing the sensitivity metric. With clinical relevance in mind, and the fact that the AUC and accuracy of the wall and 4D model show now significant difference, the use of 4D classification models could be preferred.

A relevant comparison can be made between this study and the previous study by our research group [34]. Slightly worse outcome is achieved for the classification models based on only lumen information in this study in comparison to the previous study (0.73 VS 0.78 accuracy), despite an increase in training dataset size. The difference in outcome could be explained by the fact that more complex aneurysms are added in the dataset with the introduction of FEVAR-patients. These aneurysms, with para- and supra-renal thrombus often still show a straight suprarenal segment of the aortic lumen. This lumen however, is circumferentially coated in thrombus, supporting the idea that these segmentations are less accurate representations of the clinical situation and have a negative influence on the predictive outcome of classification models.

Limitations

The influence of oversizing on the postoperative apposition length is clear [55] and it can therefore be stated that prediction models should include the degree of oversizing. Including a measure for oversizing for the FEVAR-patients however, is a challenge as correspondence between the actual achieved landing zone and the anticipated landing zone is difficult. Determining a value for oversizing in the FEVAR-situation would mean that a shared baseline need to be incorporated to calculate the relative oversizing to, as the proximal landing site varies between patients and is unsuitable for the description of the location where apposition is lost. As in this study the focus is put on the implementation of thrombus information in 4D SSMs, it is chosen not to implements values for oversizing in the classification models as implementations without careful validation might negatively influence the comparison between the different strategies for SSM creation under review.

A second limitation of this study lies in the acquisition of outer wall segmentations for the use of thrombus morphology description. Segmentations were obtained by a a predetermined protocol, but only by a single observer. The quality of these segmentations should be assessed by comparison of these segmentations between the observers. In recent literature, methods are described to automatically obtain outer wall segmentations [38] [56]. These methods could lower intra-observer variability, The meshes following automated segmentation should be validated for the use of SSM creation as described in this study.

2.5 Conclusion

This study shows the possibility to implement information on thrombus location and morphology in four dimensional statistical shape models of the pararenal aorta. These models accurately describe unseen instances. Resulting principal component scores after fitting of a new aortic morphology can be used to train ensemble learning classifiers. Implementing the described framework, binary classification of postoperative shortest apposition length (< 10 mm or ≥ 10 mm) achieved by EVAR can be based on preoperative CTA with an accuracy of 0.84%, outperforming previous methods. The new models could better help clinicians with the planning of endovascular treatment of abdominal aortic aneurysms.

3 Continuous prediction of the postoperative endograft apposition location in the pararenal aorta for AAA after EVAR

3.1 Introduction

When disease of the aorta is more extensive and involves branch vessels (see Figure 1), the complexity of the repair and risk of complications increases. An endovascular approach has become the preferred treatment for complex AAA in many high volume centers [20]. Chapter 2 describes the anatomic suitability of an infrarenal aneurysm for (standard) EVAR, as the predicted labels are chosen to describe this. If endovascular repair is preferred above open surgery and the aorta is not suitable for standard EVAR, fenestrated devices are available for repair.

Technical success rates for FEVAR are high, but after FEVAR reinterventions have been reported in up to 39% of patients [21]. BECS-associated complications included significant stenosis, occlusion, type 3 endoleak, or stent fracture [57]. Reports suggest that more complex FEVAR increases complication rates (4% vs. 18%) compared with standard FEVAR [58]. Excessively long sealing zones will result in larger coverage of the healthy aorta, bringing increased risk of spinal cord ischemia (SCI). In the latest clinical practice guidelines, it is therefore suggested that consideration should be given to limiting the aortic coverage to reduce the risk of SCI, however without compromising the proximal sealing zone [16]. Sufficient proximal seal above the renal arteries is suggested to be 20mm [59].

Planning where sufficient circumferential apposition of the endograft can be achieved in the complex AAA remains difficult, possibly influencing a trend of increased use of complex FEVAR [58]. In Chapter 2 we expanded SSMs to include all types of abdominal aneurysms and showed improved accuracy of predictive models for the outcome of EVAR, compared to previously used methods [34]. To enable planning of FEVAR based on predictive models, outcome should represent the exact location where circumferential apposition between the endograft and the aorta is lost. This location could be evaluated relative to the renal and visceral arteries, possibly in a 3D visualization, to provide insight in the complexity needed for FEVAR to achieve enough seal.

In Chapter 2 we have classified whether SAL will be $\geq 10mm$ or $< 10mm$. To aid the planning of FEVAR devices and to provide more information on the predicted apposition in the aortic neck, in this study, we aim to go beyond this binary outcome and predict the exact location where circumferential apposition is lost. The goal of the presented prediction models is to provide continuous outcome on the location where circumferential apposition by an endograft device and the pararenal aorta is lost. Preoperative prediction of this location could aid in the planning of (F)EVAR.

To provide the binary outcome previously described in this thesis, Random Forest classifiers were used. In this study the best pipeline using a regression algorithm is investigated to produce continuous outcome in a retrospective observational study.

3.2 Method

Dataset

The study population previously described in Chapter 2.2 is used for this study. Segmentations and the available contour lines are currently limited to 5 contour lines both proximal and distal from the LRA (i.e. baseline) with 10 mm intervals along the CLL Including the contour line at the baseline this results in a total of 11 contour lines. Postoperatively achieved SAL for the available dataset is shown in Figure 13.

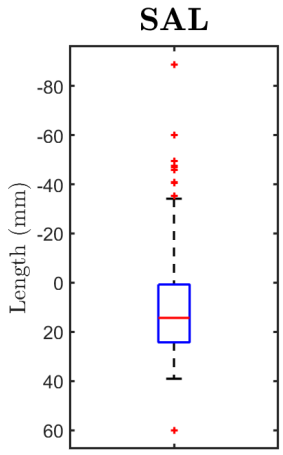


Figure 13: Boxplot showing postoperative SAL for all patients in the dataset used for Chapter 2. Median is shown as the red line with the top and bottom of the box showing the 25th and 75th percentiles, the whiskers extend to the most extreme data not considered outliers. Suprarenal loss of apposition is shown as negative values, as previously defined.

SAL included in the dataset extends beyond the region described by the contour coordinates. A region proximal of the location where apposition is lost should be incorporated in the SSM, to describe the morphology that results in apposition loss. The sealing of a fenestrated endograft is suggested to be determined by a proximal zone of 20 mm [58]. This would suggest that two proximal contour lines are needed for the evaluation of apposition loss at a certain point, implying that only SAL values $\geq$ -30 mm can currently be described by this dataset. Within this study, five proximal contour lines are used for the creation of the SSM and patients with SAL values $<$ -30 mm are excluded.

SSM extension

In previous studies [37] [34] and in the study described in Chapter 2 of this thesis, 2 distal contour lines from the baseline are used for the creation of the SSM. A problem within the previously described parameterization process is the possibility of overlapping contour lines with an increasing angle of the CLL. More distally from the baseline the change of this happening increases as the CLL enters the aneurysm. A visualization of increasing contour lines following the parameterization process is shown in Figure 14. In the dataset (n = 216), 48 parameterizations (22%) contain crossing contour lines.

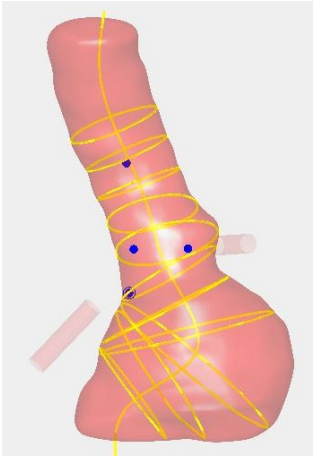


Figure 14: Wall segmentation of an abdominal aneurysm including the CLL and crossing contour lines after parameterization. Origins of the visceral arteries are shown as blue dots.

The interpolation step to solve the crossing of contour lines is previously described in Chapter 2.2. As seen in Figure 14, infrarenal angulation of the CLL can cause the contour lines to enter the aneurysm more vertically. This results in contour points that are unfavorable for the use of SSM creation as they greatly influence the mean shape and variation of morphology. A method to filter contour points that are parameterized further away from the CLL than other points in that contour line is applied in this study and described in Appendix B.

The described filtering method is able to still represent the shape of the contour point. If the diameter of a contour line increases, but filtering is needed because the contour line also vertically enters the aneurysm, the increase in diameter is maintained. Resulting contour points, however, might still negatively influence the SSM creation by the introduction of outliers.

To increase the region for SAL prediction, we wish to extend the SSM more distally. Here, the trade-off between an increased region and the increased incorporation of invalid contour points becomes relevant. It should also be noted that the clinical relevance of continuous prediction of SAL is lower as the SAL increases. With high SAL, it is clear that standard EVAR is possible and the difference between 20 or 30 mm SAL is clinically less relevant than the difference between 5 and 15 mm SAL. In this study, the SSM is extended distally with one contour line and patients with SAL $>$ 30 mm are excluded as those locations are not incorporated in the SSM. This results in 9 contour lines used for SSM.

Continuous prediction

Supervised learning is used for the preoperative prediction of SAL, with the goal to provide continuous outcome. To provide continuous outcome, an ensemble regressor is used in the form of the Random Forest Regressor available in Scikit-learn [51] (version 1.5) implemented in Python (version 3.11). In Random Forest Regression, the continuous prediction is an average of the predictions produced by the decision tree used to build the forest.

To determine the best decision forest for our predictions, hyperparameter tuning is performed using Tree-Based Pipeline Optimization (TPOT) [60]. Such automated machine learning algorithms can be used to optimize (hyper)parameters, by building a multitude of machine learning pipelines and selecting the best one based on the desired validation function [61]. In the case of a random forest regressor, for example the optimal number of estimators and maximum amount of features to use can be found as well as restrictions on the size of the individual decision trees.

To evaluate regression results for the whole included dataset and the aortic neck specifically, two separate SSMs are used for predictions. The first SSM includes all postoperative SAL between -30mm (suprarenal) and 30mm (infrarenal), the second SSM includes SAL between 0mm (baseline/LRA) and 30mm (infrarenal).

Oversizing

Oversizing of all included endografts is included as an feature to the features provided by the SSM. The oversizing is calculated using the diameter of the mainbody and the aortic diameter at the LRA. Using this calculation, negative oversizing is possible as in the case of FEVAR, the size of the endograft could be fitted to a part of the suprarenal aorta, where the diameter of the aorta is smaller than at the location of the LRA. The feature importance following random forest regression shows the relevance of implementing oversizing as a feature. Also the Mann-Whitney U test is used to compare oversizing between the $SAL \geq 10mm$ and $SAL < 10mm$ classification groups.

Evaluation

Continuous outcome of predictors is evaluated using the Root Mean Squared Error (RMSE) and is calculated using equation 5, where f_i are the predicted values, o_i are the observed (i.e. actual) values, for the total sample size (n).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2} \tag{5}$$

Here, a smaller value for the RMSE indicates good predictive capability of a model. The RMSE is measured in the same units as the target variable, and can also be described as the standard deviation of prediction errors (i.e. residuals).

The continuous predictions following the two regression algorithms are compared to the actual values in error plots. Bland-Altman plots [62] are used to display the relationship of the predicted and actual paired variables. In the Bland-Altman plots the differences between the predicted and actual variables are plotted against their means, visualizing the degree of agreement between the two. It shows (systematic) bias and includes the 95% CI of the bias (i.e. the limits of agreement).

Finally, continuous outcome after regression is transformed into binary outcome relative to 10 mm SAL and compared to outcome an algorithm based on classification. This is done to evaluate the clinical implications of the use of regression algorithms on the prediction of the possibility for EVAR placement as described in Chapter 2. Confusion matrices are produced for comparison of outcome and the accuracy of predictions are calculated for both methods. To test if the regression algorithm produces similar results to the classification algorithm, the McNemar test is performed.

3.3 Results

Study Population

Out of the 216 patients included for the study described in Chapter 2, 13 patients were excluded because the achieved $SAL < -30mm$ and 38 patients were excluded because the achieved $SAL > 30mm$. This left 165 patients included in this study with a median SAL of 12.56 mm (19.86 mm - 0.73 mm). The subset only including SAL in the aortic neck (i.e. positive SAL) consisted of 131 patients with a median SAL of 14.72 mm (30.94 mm - 7.82 mm).

Hyperparameters

TPOT hyperparameter tuning resulted in the hyperparameters shown in Table 3. For both subgroups, the criterion (i.e. loss function) 'squared error' is used, calculating the mean squared error for fitting of the model. Differences between in the models for two subgroups are seen in the *n_estimators* and the *min_sample_split*. The smaller $0 \geq SAL \geq 30$-group was tuned to include more decision trees (higher *n_estimators*) and required more samples to split an internal node (higher *min_sample_split*).

Table 3: Hyperparameters after tuning using the TPOT-algorithm

Hyperparameter	**$-30 \geq$ SAL ≥ 30**	**$0 \geq$ SAL ≥ 30**
criterion	Squared error	Squared error
n_estimators	100	300
max_features	4	4
max_depth	None	None
min_leaf_nodes	2	2
max_sample	None	None
min_sample_split	2	10

Regression models

Error plots comparing the predicted SAL to the actual SAL for both subgroups are shown in Figure 15. $RMSE$ for the $-30 \geq SAL \geq 30$ subgroup is 10.82mm, $RMSE$ for the $0 \geq SAL \geq 30$ subgroup is 6.85mm. Results for the $-30 \geq SAL \geq 30$ subgroup show lower error for positive SAL (i.e. loss of circumferential apposition in the aortic neck). The lower $RMSE$ for the $0 \geq SAL \geq 30$ subgroup supports this.

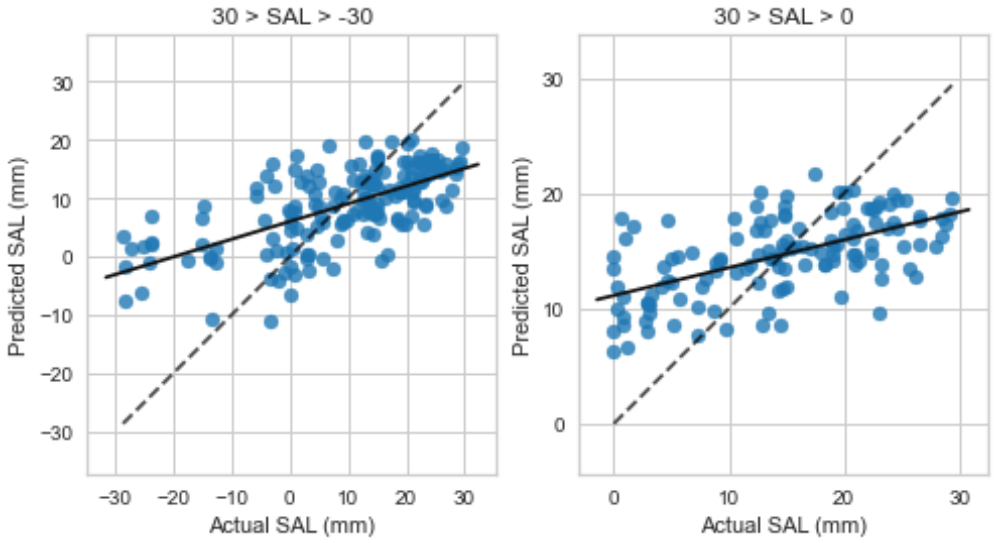


Figure 15: Error plots for the two subgroups, with actual SAL on the x-axis and predicted SAL on the y-axis in mm's. The identity (i.e. $x = y$) is shown as an interrupted line. The best fitted model is shown as an uninterrupted line.

Bland-Altman plots after regression are shown in Figure 16. Bias for the $-30 \geq SAL \geq 30$ subgroup is 0.15 (CI: -21 - 22) and bias for the $0 \geq SAL \geq 30$ subgroup is -0.17 (CI: -14 - 13). Proportional bias is seen for both subgroups, showing a trend of high predicted values for low actual SAL and low predicted values for higher SAL. Lowest differences in predicted and actual SAL is seen between 10 - 20 mm SAL, corresponding with the intersection of the identity and best fit seen in Figure 15. Mean bias is low for both subgroups with larger limits of agreement for the $-30 \geq SAL \geq 30$ subgroup compared to the $0 \geq SAL \geq 30$ subgroup.

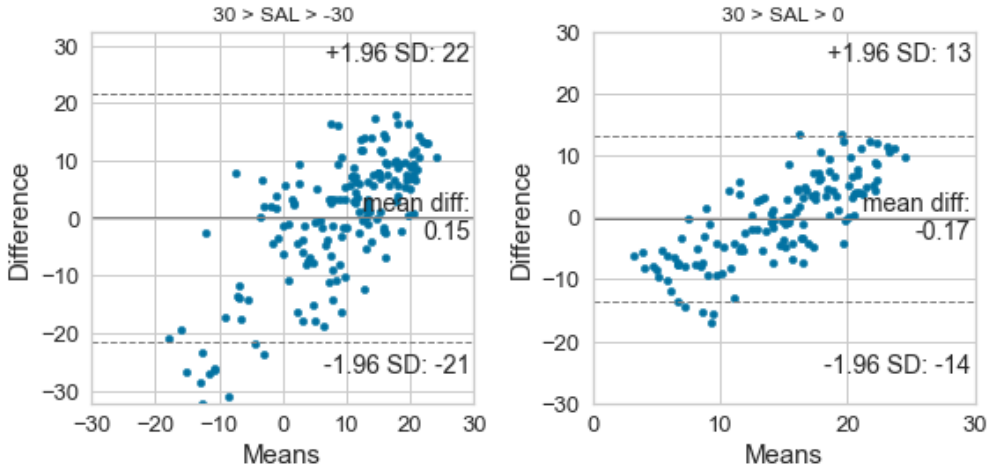


Figure 16: Bland-Altman plots for the two subgroups. The x-axis shows the mean difference between actual SAL and predicted SAL. The y-axis shows the difference following: $difference = y_{true} - y_{pred}$. The limits of agreement (i.e. 95%CI) is shown for both bias

Feature importance and oversizing

Oversizing is shown for both the $SAL \geq 10mm$ and $SAL < 10mm$ classification groups in Figure 17, showing a median oversizing of 17.15 (8.81 - 25.45) for the $SAL \geq 10mm$ group and median oversizing of 11.33 (5.02 - 24.03) for the $SAL < 10mm$ group. The Mann-Whitney U test shows no significant difference between the two groups with $p = 0.211$. As previously shown in Chapter 2, features that are not significantly different between the $SAL \geq 10mm$ and $SAL < 10mm$ classification groups show less feature importance and little influence to the predictions.

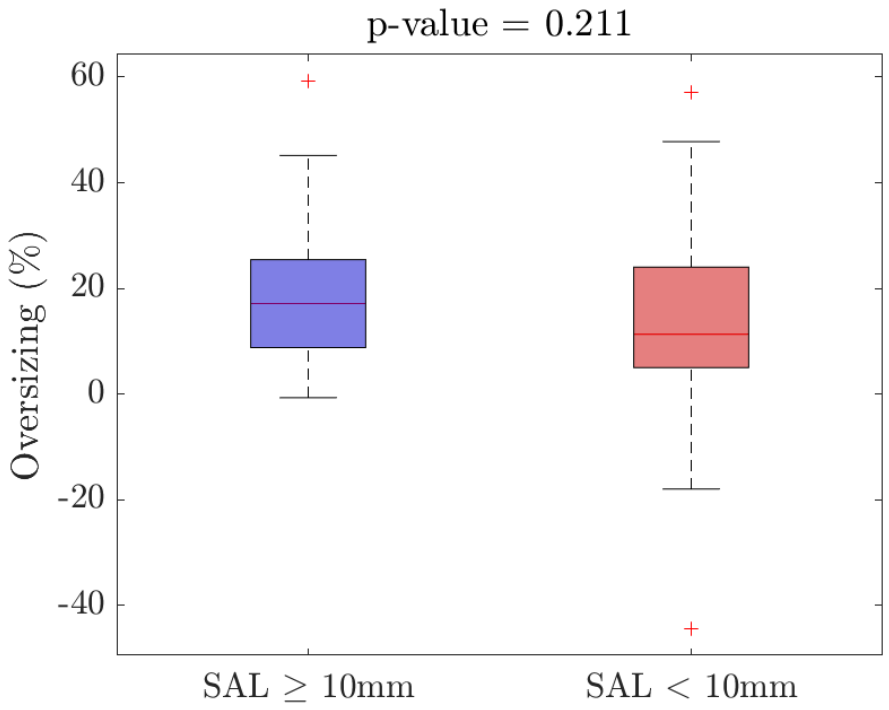


Figure 17: Boxplots showing the distribution of values of oversizing for the $SAL \geq 10mm$ and $SAL < 10mm$ classification labels. Median SAL is shown as a red line inside the box. The box itself represents the 25 and 75 percentiles.

Feature importance is shown for the largest subgroup in Figure 18, showing small importance of oversizing in the prediction of SAL. The new model depends mostly on feature "PC 2", whereas the $4D_{wall}$ model described in Chapter 2 (Figure 12 d), also shows high importance of feature "PC 4".

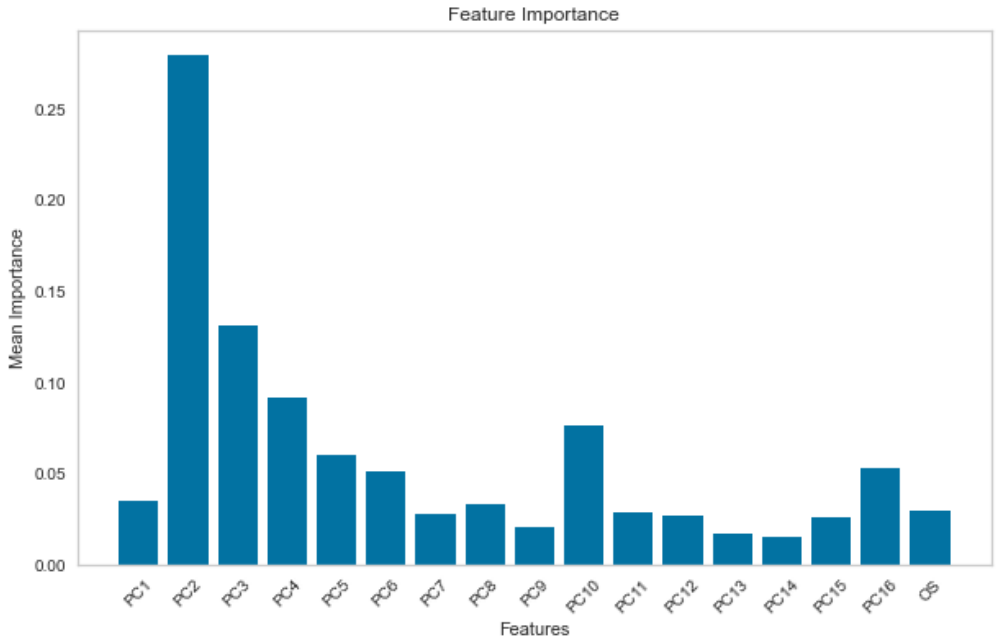


Figure 18: The impact of each individual feature on the classification outcome using the random forest regressor is shown for the $-30 \geq SAL \geq 30$ subgroup. Mean values are the result of LOOCV.

Classification ability

Random forest classifiers were trained on the larger subgroup ($-30 \geq SAL \geq 30$) and the results from the corresponding random forest regressor were binarized to represent the $SAL \geq 10mm$ and $SAL < 10mm$ classification labels. Results from both classifications are shown in confusion matrices displayed in Figure 19. Accuracy of the pipeline using regression for classification resulted in an accuracy of 73% and classification using the random forest classifier resulted in an accuracy of 77%. The McNemar test results in a p-value of 0.344, indicating no significant difference in model outcome. The contingency table used for the McNemar test is provided in Table 4.

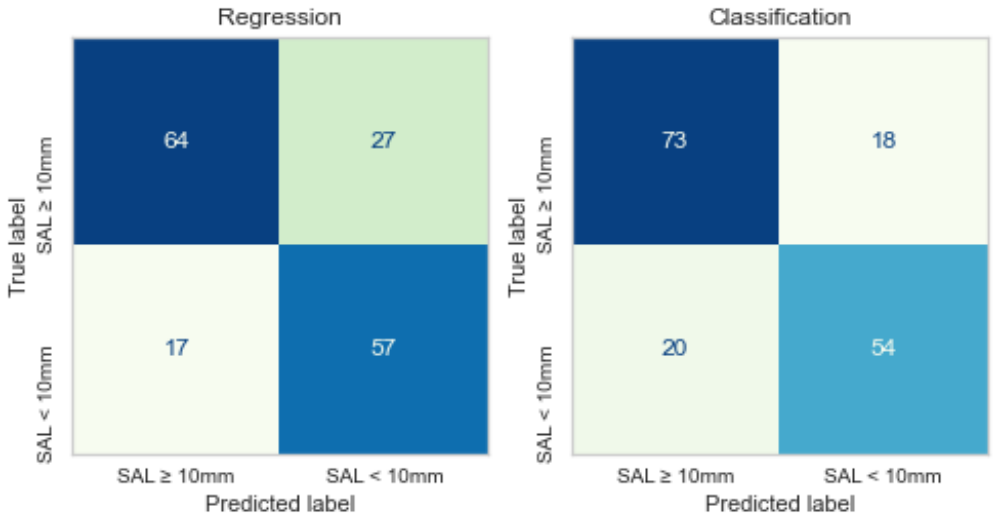


Figure 19: Confusion Matrices, showing the predicted and true labels for the $SAL \geq 10mm$ and $SAL < 10mm$ labels.

Table 4: Contingency table for the predictions shown in Figure 19. A and B represent the regression and the classifier algorithms respectively. 0 and 1 indicate false and correct classification respectively.

	B0	**B1**
A0	27	17
A1	11	110

3.4 Discussion

In this study, insight is provided on the capability of random forest regression algorithms to predict the location where circumferential apposition between an endograft and the aorta is lost. Hyperparameter tuning is an important step that is incorporated in this study, to train the best performing random forest regressors. For the creation of this pipeline, an automated machine learning tool is used in the form of the TPOT algorithm.

The random forest regression algorithm can be used to predict clinically relevant labels with similar accuracy to a random forest classification algorithm used for the same task. To support this, the error plots (Figure 15) show low predictive error for predictions around the threshold of 10 mm SAL. The best fits, however, show bias from the identity (i.e. *actual* = *predicted*). Bland-Altman plots (Figure 16) provide more detail on this bias. Mean bias is low for both subgroups, but clear proportional bias is seen for both subgroups. The model tends to predict values around the median value, and has difficulty predicting both higher and lower values. Such bias is unfavorable for the predictive models, resulting in large limits of agreement.

Continuous prediction in the aortic neck is better than the predictions based on the SSM that includes suprarenal loss of apposition. This might be influenced by the fact that the region where predictions are possible is larger for the bigger region. Towards the limits of the values within the training set, prediction becomes more difficult for the trained algorithms. Here, a larger region results in larger errors.

This bias might be influenced by the loss function used in the random forest regressor. The mean squared error criterion is used for both subgroups, also known as L2 loss. The squared error loss tends to pull the predictions toward the mean of the distribution of the target variable. If the distribution of target variables has a wide range, resulting models can exhibit proportional bias [63]. Weighted loss functions can be used that vary inversely with the magnitude of the observed values to counteract the influence of values towards the bounds of training data.

The root mean squared error that is observed is a representation of the mean error a prediction by the algorithm will have. Clinically workable error margins should be significantly smaller than the margins where clinical decisions are based upon. Standard FEVAR is possible when 20mm seal is achieved below the (scalloped) SMA. The point where apposition is lost can be predicted with a RMSE of 10.82mm. If this error is workable depends largely on the distance between the renal arteries and the SMA, which can differ largely in a population. The relatively large RMSE however, would be only usable for patients with relatively large distances between the SMA and the renal arteries. The RMSE of the aortic neck subgroup is smaller with a value of 6.85mm. Within the aortic neck, however margins for predictions are also smaller.

PC2 for the $4D_{wall}$ SMM is shown in Figure 23 and is dominant in both the classification and regression algorithms. This PC however, only describes the distal part of the region that is incorporated in the SSM. This could result in the large bias in the predictions of the $-30 \geq SAL \geq 30$ subgroup. Here, information on the suprarenal aortic morphology is needed, but the random forest regressor is mostly trained on features describing infrarenal morphology. In the future, SSMs could be dedicated to the region where prediction is needed. To determine this region a classification model could precede the regressor, labeling the aorta in several regions. With correct classification of the region of interest, an SSM could be created for a regression algorithm resulting in the exact location of loss of circumferential apposition.

The random forest regressor was expanded with information on oversizing of the endograft mainbody relative to the diameter of the aorta at the baseline (i.e. at the LRA). The Mann-Whitney U test shows little variance in the oversizing values within the dataset and no significant difference between $SAL \geq 10mm$ and $SAL < 10mm$ groups. Feature importance after training of the algorithms shows low importance of oversizing as an added feature within this dataset.

Limitations

For this study subgroups are selected for evaluation. The $-30 \geq SAL \geq 30$ subgroup is created to show the feasibility of continuous predictions on a larger region of the aorta. A smaller, $0 \geq SAL \geq 30$ subgroup is created to see if better predictions are possible in the aortic neck. The study shows that for the created SSMs, lower RMSE is achieved in the aortic neck. For a proper comparison between the two regions, also a seperate subgroup of the suprarenal aorta should be created using a dedicated SSM. To create a valid SSM, more para- and supra-renal aneurysms need to be incorporated in the dataset.

Currently, the TPOT algorithm is trained on a 80/20 train-test split of the data. Incorporating the TPOT algorithm in the LOOCV would make the cross validation computationally heavy. As we are dealing with a relatively small dataset for regression, it might be possible that the pipeline fitted on the training set might not be the best model used in the LOOCV. The TPOT algorithm could be used in this way, with better results, if the dataset is larger or the model is used for a consecutive dataset. The TPOT algorithm might also be expanded to search for different regression models, as using only a random forest regressor is currently limiting this study. The optimal pipeline might include another ensemble regressor, such as grandient boosting machines or even neural networks.

3.5 Conclusion

Regression models have provided a first insight in the possibility to predict continuous outcome of SAL based on statistical shape models of the aorta. SSMs should be evaluated to determine if they adequately describe the aortic morphology needed for the prediction of SAL in that specific region. In this study, the location where circumferential apposition between an endograft and the aorta is lost can be predicted with a RMSE of 6.85mm in the aortic neck and 10.82mm for a pararenal segment extending 30mm proximal and distal of the LRA. These distances should be observed relative to the margins where clinical decisions are based upon in the corresponding regions. This suggests that more effort is needed to produce smaller errors for the incorporation of continuous predictions in the surgical planning of endografts.

4 General Discussion

In this chapter, a general discussion is provided on the topics addressed in this thesis. First, views on the future perspective are provided, followed by suggestions for adaptation of the currently researched workflow supported by literature.

Developments in automated (F)EVAR planning

The studies described in this thesis fall in a category of recent developments in automated (F)EVAR planning. Other research is conducted towards the automated segmentation and measuring of aortic aneurysms [38]. Such models enable automated evaluation of patients in both preoperative/planning and postoperative/surveillance settings. Also the automatically generated aortic wall meshes might prove useful for the purpose of generating 4D SSMs The quality of automatically generated wall meshes should be investigated.

Also, FEVAR planning is automatized in recent literature, describing models that are capable of automated sizing [64] and alignment of fenestrations [65]. Preoperative predictions on the real achieved sealing zone based on these digitally proposed endografts could mean the next step in virtual stenting both in an elective and emergency setting.

Implementation of predictive models

It would be very interesting to research the implementation of the predictive models described in this thesis. The representation of the outcome of models could be made visual and presented to a team of physicians planning (F)EVAR. The margins of error by such models should also be incorporated. A first step in the implementation might not be to provide continuous outcome of postoperative SAL, however. Personalized predictions could prove useful as a warning mechanism, for example when patients that are predicted to be unsuitable for standard EVAR are highlighted by an automated warning system. The current model as described in Chapter 2 of this thesis shows promising accuracy for such applications of personalized medicine.

Preoperative representation of postoperative outcome could also prove useful in shared decision making, when options for treatments are discussed with patients. Prevalence of inadequate health literacy and low disease knowledge in patients with AAA is high [66], which is known to result in difficulties for shared decision making [67]. Clear visual representation of the treatment options might positively influence a patients understanding of the disease and the available options for treatment.

Parameterization

Chapter 1.3 describes the process of parameterization currently in use, where the centerline defines points on the aortic mesh. Difficulties arising from this workflow in the form of crossing contours are described in Chapter 3.2. The solution using median filtering, proposed in the second study, offers a solution for many cases. The shape of contours is still represented by the filtered contour points, offering valid input for the SSM. With large angles of the CLL, however, the filtered contour points would still negatively influence the creation of an SSM, currently still influencing the choice to not include distal contours #4 and #5. The problem with crossing contour lines could however also be solved if parameterization is performed in other ways.

Literature is available on the correspondence between tubular surfaces [68]. Other methods for parameterization of the aorta proposed use cylindrical b-spline deformations of the surface mesh. Correspondences following this method result in improved descriptors for statistical shape models of cylindrical anatomy such as the trachea and aorta [69].

In other work, Riemannian frameworks have shown improved capability to capture the non-linearity in shape variation for SSMs compared to use of Euclidean structures [33]. Here, the principal components following reconstruction of the SSM with differential coordinates, allowed better distinction between disease labels coupled to the dataset. In this thesis a similar goal is pursued, therefore the outcome of predictive models could potentially also benefit from the use of different parameterization techniques in the future.

Regression for continuous outcome

Principal components after SSM creation have been carefully evaluated in Chapter 2 of this thesis. PCs that are important in the classification algorithms for the prediction of SAL $\geq or < 10mm$ are found to have an important role in the regression algorithms too. After regression, categorization remains possible with similar accuracy to classification algorithms. Difficulty is found to predict continuous outcome with clinically acceptable error.

It is concluded in Chapter 3.5 that the right PCs should be available for the regressor to base predictions upon. This implicates the need for dedidcated SSMs of the suprarenal aorta, if continuous prediction of circumferential loss of apposition of a FEVAR device with the aortic wall is needed. The datasets that are currently available in the research group should be expanded to contain more patients that received FEVAR, for this purpose.

It would be interesting to research further improvement of the models described in this thesis, with the goal to minimize the error of continuous predictions. After this thesis it is hypothesised that the biggest improvement could be achieved by dedicated SSMs of the ROI, with sufficient sizes of datasets for each region.

With the optimization of regression algorithms and the application of dedicated SSMs for optimal continuous prediction, the remaining error should be evaluated to be useful for clinical applications. Better predictions might be found with the use of different model types. Here the use of neural networks might provide interesting improvement. Neural networks can capture complex, non-linear relationships within data [70]. This allows them to better handle cases where the relationship between the input features and the target variable is not straightforward, possibly improving regression outcome.

Digital Twin of the aorta

Especially using neural networks, previously unknown links in parameters might prove useful for the prediction of postoperative (F)EVAR success. Models can be presented with additional parameters in addition to the shape of the aorta. Data that could be useful include clinical parameters such as blood pressure and microbiology, but also demographic information such as age and sex.

A comprehensive digital representation of the aorta, also called the Digital Twin, could enable further improvement of predictive models. Parameterization of the aorta might be obsolete if the shape of the different aortas in the dataset is uniformly described by the mesh. Here the distance from the lumen to the outer wall can still be incorporated for each face or vertex in the mesh and improve predictive outcome compared to only using the lumen mesh.

It would be interesting to incorporate more information on the thrombus composition in the Digital Twin. Different compositions of thrombus might influence the achieved apposition of an endograft at that location. Recent developments in imaging techniques using dual energy computed tomography enables differentiation of different types of thrombus composition [71]. Information on thrombus composition could be added to the Digital Twin of a patients aorta to further improve (F)EVAR planning. Digital Twins are currently being developed for the purpose of coronary stenting [72] and mapping of stenosis in the ascending aorta [73], but also for the detection of AAA [74].

4.1 Conclusion

This first study in this thesis aimed to improve the preoperative prediction of postoperative SAL, to answer the question if a particular aortic neck is suitable for standard EVAR. To this extend four dimensional statistical shape models were created to include thrombus and prediction models based on these SSMs are compared to the use of a lumen SSM. Prediction models for postoperative SAL in the aortic neck based on SSMs that include thrombus have significantly higher predictive accuracy compared to models that include only lumen segmentations, allowing us to reject the first null hypothesis formulated in 1.4.1.

The second study describes a study towards clinical implementation of prediction models. Models are adapted to provide continuous outcome, effectively predicting the location where apposition of the endograft with a newly presented aortic morphology will end. Continuous outcome is possible using these algorithms, but further improvement is needed for safe implementation of these predictions, Suggestions for this improvement are provided in the thesis and are to be researched in the future. For now the second null hypothesis is not to be rejected.

Other suggestions for further research and the possible implications of related research are provided in the general discussion of the thesis, aimed at further improving predictive algorithms. Both studies described in this thesis expand on previous work to help the progression of a *virtual stenting tool*.

References

[1] P. W. Stather, D. A. Sidloff, I. A. Rhema, *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 (2014).

[2] S. Aggarwal, A. Qamar, V. Sharma, *et al.*, "Abdominal aortic aneurysm: A comprehensive review," (2011).

[3] B. T. Baxter, G. S. McGee, W. R. Flinn, *et al.*, "Distal embolization as a presenting symptom of aortic aneurysms," *The American Journal of Surgery* **160**(2), 197–201 (1990).

[4] K. W. Johnston, R. B. Rutherford, M. D. Tilson, *et al.*, "Suggested standards for reporting on arterial aneurysms," *Journal of Vascular Surgery* **13**(3), 452–458 (1991).

[5] E. L. Chaikof, M. F. Fillinger, J. S. Matsumura, *et al.*, "Identifying and grading factors that modify the outcome of endovascular aortic aneurysm repair," *Journal of Vascular Surgery* **35**(5), 1061–1066 (2002).

[6] V. Jongkind, K. K. Yeung, G. J. Akkersdijk, *et al.*, "Juxtarenal aortic aneurysm repair," *Journal of Vascular Surgery* **52**, 760–767 (2010).

[7] E. Altobelli, L. Rapacchietta, V. F. Profeta, *et al.*, "Risk factors for abdominal aortic aneurysm in population-based studies: A systematic review and meta-analysis," (2018).

[8] Gezondheidsraad, "Bevolkingsonderzoek naar aneurysma van de abdominale aorta (AAA)," Tech. Rep. 10 (2019).

[9] H. A. Fink, F. A. Lederle, C. S. Roth, *et al.*, "The accuracy of physical examination to detect abdominal aortic aneurysm," *Archives of Internal Medicine* **160**, 833–836 (2002).

[10] C. Van Walraven, J. Wong, K. Morant, *et al.*, "Incidence, follow-up, and outcomes of incidental abdominal aortic aneurysms," *Journal of Vascular Surgery* **52**(2) (2010).

[11] J. D. B. Watson, S. M. Gifford, and D. F. Bandyk, "Aortic aneurysm screening using duplex ultrasound: Choosing wisely who to examine," *Seminars in Vascular Surgery* **33**, 54–59 (2020).

[12] L. L. LaRoy, P. J. Cormier, T. A. Matalon, *et al.*, "Imaging of abdominal aortic aneurysms," *American Journal of Roentgenology* **152**, 785–792 (1989).

[13] T. Y. Lee, P. Korn, J. A. Heller, *et al.*, "The cost-effectiveness of a "quick-screen" program for abdominal aortic aneurysms," *Surgery* **132**, 399–407 (2002).

[14] K. Singh, K. H. Bonaa, S. Solberg, *et al.*, "Intra- and interobserver variability in ultrasound measurements of abdominal aortic diameter. The Tromsø study," *European Journal of Vascular and Endovascular Surgery* **15**, 497–504 (1998).

[15] C. Oliver-Williams, M. J. Sweeting, J. Jacomelli, *et al.*, "Safety of Men with Small and Medium Abdominal Aortic Aneurysms under Surveillance in the NAAASP," *Circulation* **139**, 1371–1380 (2019).

[16] A. Wanhainen, I. Van Herzeele, F. Bastos Goncalves, *et al.*, "Editor's Choice – European Society for Vascular Surgery (ESVS) 2024 Clinical Practice Guidelines on the Management of Abdominal Aorto-Iliac Artery Aneurysms," *European Journal of Vascular and Endovascular Surgery* **67**, 192–331 (2024).

[17] F. A. Lederle, J. A. Freischlag, T. C. Kyriakides, *et al.*, "Outcomes following endovascular vs open repair of abdominal aortic aneurysm: A randomized trial," *JAMA* **302**, 1535–1542 (2009).

[18] M. Prinssen, E. L. Verhoeven, J. Buth, *et al.*, "A Randomized Trial Comparing Conventional and Endovascular Repair of Abdominal Aortic Aneurysms," *New England Journal of Medicine* **351**, 1607–1618 (2004).

[19] N. Kawaharada, K. Morishita, J. Fukada, *et al.*, "Minilaparotomy abdominal aortic aneurysm repair versus the retroperitoneal approach and standard open surgery," *Surgery Today* **34**(10), 837–841 (2004).

[20] S. R. Patel, D. C. Ormesher, R. Griffin, *et al.*, "Comparison of Open, Standard, and Complex Endovascular Aortic Repair Treatments for Juxtarenal/Short Neck Aneurysms: A Systematic Review and Network Meta-Analysis," *European Journal of Vascular and Endovascular Surgery* **63**(5), 696–706 (2022).

[21] M. Sveinsson, B. Sonesson, T. Kristmundsson, *et al.*, "Long-term outcomes after fenestrated endovascular aortic repair for juxtarenal aortic aneurysms," *Journal of Vascular Surgery* **75**, 1164–1170 (2022).

[22] J. P. Becquemin, J. C. Pillet, F. Lescalie, *et al.*, "A randomized controlled trial of endovascular aneurysm repair versus open surgery for abdominal aortic aneurysms in low- to moderate-risk patients," *Journal of Vascular Surgery* **53**, 1167–1173 (2011).

[23] A. C. Geraedts, R. Zuidema, R. C. Schuurmann, *et al.*, "Shortest Apposition Length at the First Postoperative Computed Tomography Angiography Identifies Patients at Risk for Developing a Late Type Ia Endoleak After Endovascular Aneurysm Repair," *Journal of Endovascular Therapy* **31**, 274–281 (2024).

[24] R. Zuidema, A. C. Geraedts, W. A. van Veldhuizen, *et al.*, "Diminishing Endograft Apposition during Follow-Up Is an Important Indicator of Late Type 1a Endoleak after Endovascular Aneurysm Repair," *Journal of Clinical Medicine* **12** (2023).

[25] R. Zuidema, C. van der Riet, M. El Moumni, *et al.*, "Pre-operative Aortic Neck Characteristics and Post-operative Sealing Zone as Predictors of Type 1a Endoleak and Migration After Endovascular Aneurysm Repair: A Systematic Review and Meta-Analysis," (2022).

[26] R. C. Schuurmann, K. Ouriel, B. E. Muhs, *et al.*, "Aortic curvature as a predictor of intraoperative type Ia endoleak," *Journal of Vascular Surgery* **63**, 596–602 (2016).

[27] J. J. Reimerink, H. A. Marquering, A. Vahl, *et al.*, "Semiautomatic sizing software in emergency endovascular aneurysm repair for ruptured abdominal aortic aneurysms," *CardioVascular and Interventional Radiology* **37**(3), 623–630 (2014).

[28] E. M. Marone, A. Freyrie, C. Ruotolo, *et al.*, "Expert Opinion on Hostile Neck Definition in Endovascular Treatment of Abdominal Aortic Aneurysms (a Delphi Consensus)," *Annals of Vascular Surgery* **62**, 173–182 (2020).

[29] W. A. van Veldhuizen, R. C. Schuurmann, F. F. Ijpma, *et al.*, "A Statistical Shape Model of the Morphological Variation of the Infrarenal Abdominal Aortic Aneurysm Neck," *Journal of Clinical Medicine* **11**(6), 5–11 (2022).

[30] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," (2016).

[31] F. Ambellan, H. Lamecker, C. von Tycowicz, *et al.*, "Statistical Shape Models: Understanding and Mastering Variation in Anatomy," *Advances in Experimental Medicine and Biology* **1156**, 67–84 (2019).

[32] J. L. Bruse, M. A. Zuluaga, A. Khushnood, *et al.*, "Detecting Clinically Meaningful Shape Clusters in Medical Image Data: Metrics Analysis for Hierarchical Clustering Applied to Healthy and Pathological Aortic Arches," *IEEE Transactions on Biomedical Engineering* **64**(10), 2373–2383 (2017).

[33] C. von Tycowicz, F. Ambellan, A. Mukhopadhyay, *et al.*, "An efficient Riemannian statistical shape model using differential coordinates: With application to the classification of data from the Osteoarthritis Initiative," *Medical Image Analysis* **43**, 1–9 (2018).

[34] W. A. V. Veldhuizen, J.-p. P. M. D. Vries, A. Tuinstra, *et al.*, "Machine learning-based prediction of postoperative infrarenal endograft apposition for abdominal aortic aneurysms." (2024).

[35] F. B. Gonçalves, K. M. Van De Luijtgaarden, S. E. Hoeks, *et al.*, "Adequate seal and no endoleak on the first postoperative computed tomography angiography as criteria for no additional imaging up to 5 years after endovascular aneurysm repair," *Journal of Vascular Surgery* **57**(6), 1503–1511 (2013).

[36] M. Ribeiro, G. S. Oderich, T. Macedo, *et al.*, "Assessment of aortic wall thrombus predicts outcomes of endovascular repair of complex aortic aneurysms using fenestrated and branched endografts," *Journal of Vascular Surgery* **66**(5), 1321–1333 (2017).

[37] W. A. van Veldhuizen, R. C. Schuurmann, R. Zuidema, *et al.*, "A Statistical Shape Model of Infrarenal Aortic Necks in Patients With and Without Late Type Ia Endoleak After Endovascular Aneurysm Repair," *Journal of Endovascular Therapy* , 1–10 (2023).

[38] C. Caradu, B. Spampinato, A. M. Vrancianu, *et al.*, "Fully automatic volume segmentation of infrarenal abdominal aortic aneurysm computed tomography images with deep learning approaches versus physician controlled manual segmentation," *Journal of Vascular Surgery* **74**(1), 246–256 (2021).

[39] N. F. Oliveira, F. B. Gonçalves, K. Ultee, *et al.*, "Patients with large neck diameter have a higher risk of type IA endoleaks and aneurysm rupture after standard endovascular aneurysm repair," *Journal of Vascular Surgery* **69**(3), 783–791 (2019).

[40] R. C. Schuurmann, R. Kropman, K. Ouriel, *et al.*, "Remodeling of Abdominal Aortic Angulation and Curvature After Endovascular Aneurysm Repair in Patients With vs Without Late Type Ia Endoleak or Endograft Migration," *Journal of Endovascular Therapy* **28**(2), 342–351 (2021).

[41] R. C. Schuurmann, S. P. Overeem, K. van Noort, *et al.*, "Validation of a New Methodology to Determine 3-Dimensional Endograft Apposition, Position, and Expansion in the Aortic Neck After Endovascular Aneurysm Repair," *Journal of Endovascular Therapy* **25**, 358–365 (2018).

[42] J. Pitocchi, R. Wirix-Speetjens, G. H. van Lenthe, *et al.*, "Integration of cortical thickness data in a statistical shape model of the scapula," *Computer Methods in Biomechanics and Biomedical Engineering* **23**(10), 642–648 (2020).

[43] J. Zhang, J. Hislop-Jambrich, and T. F. Besier, "Predictive statistical models of baseline variations in 3-D femoral cortex morphology," *Medical Engineering and Physics* **38**, 450–457 (2016).

[44] F. Brutti, A. Fantazzini, A. Finotello, *et al.*, "Deep Learning to Automatically Segment and Analyze Abdominal Aortic Aneurysm from Computed Tomography Angiography," *Cardiovascular Engineering and Technology* **13**(4), 535–547 (2022).

[45] C. van der Riet, R. C. Schuurmann, S. N. Schuurman, *et al.*, "Mid-term proximal sealing zone evaluation after fenestrated endovascular aortic aneurysm repair," *Journal of Cardiovascular Surgery* **64**(2), 142–149 (2023).

[46] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic Resonance Imaging* **30**, 1323–1341 (2012).

[47] T. F. Cootes, "Statistical Models of Appearance," *Direct* (2004).

[48] R. Davies, C. Twining, and C. Taylor, "Establishing Correspondence," in *Statistical Models of Shape*, 1–17, Springer London (2008).

[49] J. Tabak, *Geometry: The Language of Space and Form, Facts on File math library*, Facts On File (2014).

[50] R. Davies, C. Twining, and C. Taylor, "Objective Functions," in *Statistical Models of Shape*, 1–28, Springer London (2008).

[51] F. Pedregosa, V. Michel, O. Grise, *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research* **12**(85), 2825–2830 (2011).

[52] A. Parmar, R. Katariya, and V. Patel, "A Review on Random Forest: An Ensemble Classifier," *Lecture Notes on Data Engineering and Communications Technologies* **26**, 758–763 (2019).

[53] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics* **44**, 837 (1988).

[54] H. Donker, "Why you should report confidence intervals on your test set. — Towards Data Science."

[55] J. van Prehn, F. J. Schlösser, B. E. Muhs, *et al.*, "Oversizing of Aortic Stent Grafts for Abdominal Aneurysm Repair: A Systematic Review of the Benefits and Risks," *European Journal of Vascular and Endovascular Surgery* **38**, 42–53 (2009).

[56] A. Abdolmanafi, A. Forneris, R. D. Moore, *et al.*, "Deep-learning method for fully automatic segmentation of the abdominal aortic aneurysm from computed tomography imaging," *Frontiers in Cardiovascular Medicine* **9**, 1040053 (2023).

[57] C. van der Riet, R. C. Schuurmann, E. L. Verhoeven, *et al.*, "Outcomes of Advanta V12 Covered Stents After Fenestrated Endovascular Aneurysm Repair," *Journal of Endovascular Therapy* **28**, 700–706 (2021).

[58] O. Yazar, G. Pilz da Cunha, M. W. de Haan, *et al.*, "Impact of stent-graft complexity on mid-term results in fenestrated endovascular aortic repair of juxtarenal and suprarenal abdominal aortic aneurysms," *Journal of Cardiovascular Surgery* **64**, 268–278 (2023).

[59] A. Katsargyris, K. Oikonomou, G. Kouvelos, *et al.*, "From the Society for Vascular Surgery Comparison of outcomes for double fenestrated endovascular aneurysm repair versus triple or quadruple fenestrated endovascular aneurysm repair in the treatment of complex abdominal aortic aneurysms,"

[60] R. S. Olson, R. J. Urbanowicz, P. C. Andrews, *et al.*, "Automating biomedical data science through tree-based pipeline optimization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, **9597**, 123–137, Springer Verlag (2016).

[61] T. T. Le, W. Fu, and J. H. Moore, "Scaling tree-based automated machine learning to biomedical big data with a feature set selector," *Bioinformatics* **36**, 250–256 (2020).

[62] J. Martin Bland and D. G. Altman, "STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT," *The Lancet* **327**, 307–310 (1986).

[63] T. e. a. Hastie, *Springer Series in Statistics The Elements of Statistical Learning*, vol. 27 (2009).

[64] T. M. Dillon, P. Liang, M. L. Schermerhorn, *et al.*, "A computational program for automated surgical planning of fenestrated endovascular repair," *Communications Engineering* **2**(1), 1–11 (2023).

[65] L. Derycke, J. Sénémaud, D. Perrin, *et al.*, "Patient Specific Computer Modelling for Automated Sizing of Fenestrated Stent Grafts," *European Journal of Vascular & Endovascular Surgery* **59**, 237–246 (2020).

[66] G. L. Van Leeuwen, M.-A. Kooijman, R. C. L. Schuurmann, *et al.*, "SYSTEMATIC REVIEW Health Literacy and Disease Knowledge of Patients With Peripheral Arterial Disease or Abdominal Aortic Aneurysm: A Scoping Review EJVES Open Access," (2024).

[67] D. M. Muscat, H. L. Shepherd, D. Nutbeam, *et al.*, "Health Literacy and Shared Decision-making: Exploring the Relationship to Enable Meaningful Patient Engagement in Healthcare," *Journal of General Internal Medicine* **36**, 521–524 (2021).

[68] J. Huysmans, T; Sijbers, "Parameterization and Correspondence for Improved Modeling, Analysis, and Visualization of Tubular Surfaces," 190 (2010).

[69] T. Huysmans, J. Sijbers, and V. Brigitte, "Automatic construction of correspondences for tubular surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**, 636–651 (2010).

[70] M. Rocha, P. Cortez, and J. Neves, "Evolution of neural networks for classification and regression," *Neurocomputing* **70**, 2809–2816 (2007).

[71] Y. H. Ding, M. Abbasi, G. Michalak, *et al.*, "Characterization of thrombus composition with multimodality CT based imaging: an in-vitro study," *Journal of neurointerventional surgery* **13**, 738 (2021).

[72] G. Poletti, L. Antonini, L. Mandelli, *et al.*, "Towards a Digital Twin of Coronary Stenting: A Suitable and Validated Image-Based Approach for Mimicking Patient-Specific Coronary Arteries," *Electronics 2022, Vol. 11, Page 502* **11**, 502 (2022).

[73] A. Rouhollahi, J. N. Willi, S. Haltmeier, *et al.*, "CardioVision: A fully automated deep learning package for medical image segmentation and reconstruction generating digital twins for patients with aortic stenosis," *Computerized Medical Imaging and Graphics* **109**, 102289 (2023).

[74] N. Kavan Chakshu, I. Sazonov, and P. Nithiarasu, "Towards enabling a cardiovascular digital twin for human systemic circulation using inverse analysis," **20**, 449–465 (2021).

[75] C. Leys, C. Ley, O. Klein, *et al.*, "Journal of Experimental Social Psychology Detecting outliers : Do not use standard deviation around the mean, use absolute deviation around the median," *Experimental Social Psychology* , 4–6 (2013).

Appendices

A Principal Component Visualization

Principal component variation of the important features (feature importance ≥ 0.10) following random forest classification are shown with the distribution of PC-scores for both < 10 mm and SAL ≥ 10 mm labels, including the results from the Mann-Whitney U test. Feature importance is shown in Figure 12.

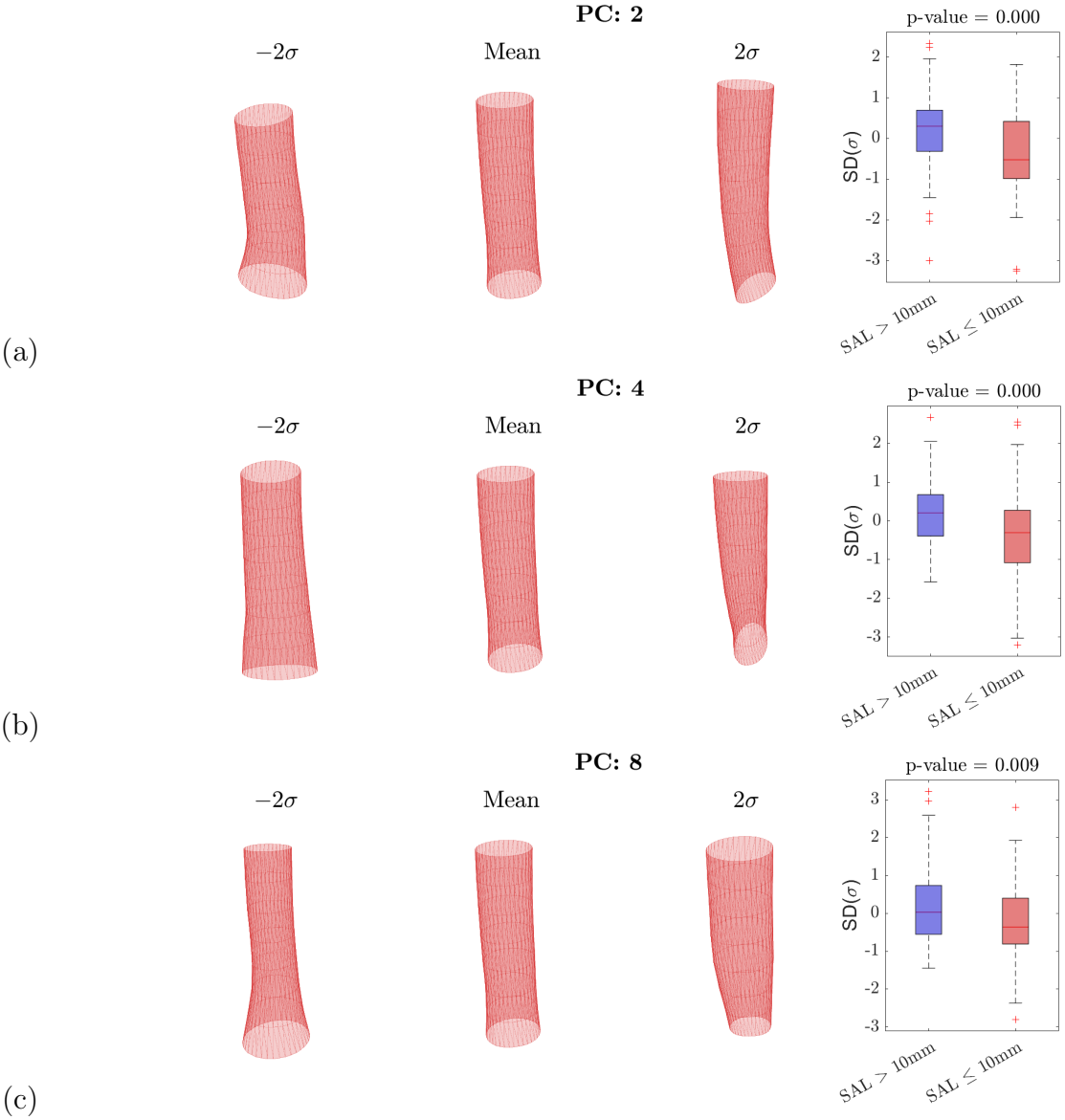


Figure 20: Visualization of PCs with feature importance ≥ 0.10 after training of the Random Forest classifier. Based on the lumen SSM, PC 2 (a), 4 (b) and 8 (c) are shown.

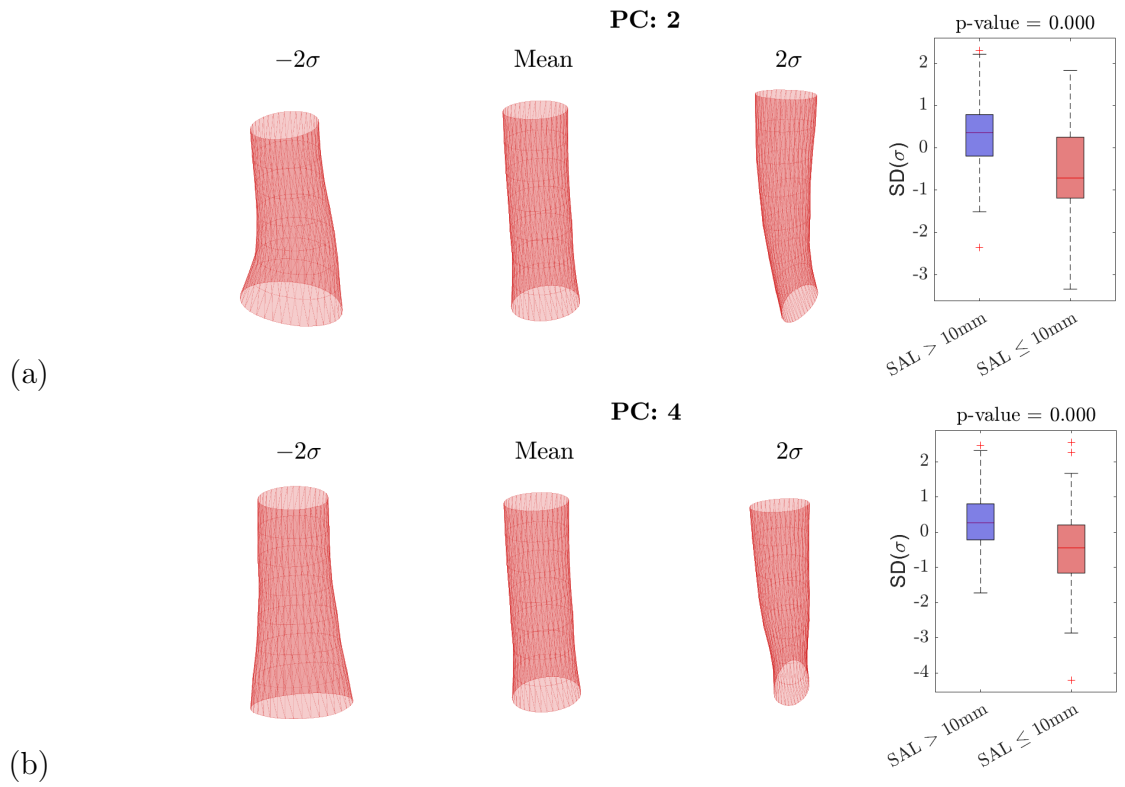


Figure 21: Visualization of PCs with feature importance ≥ 0.10 after training of the Random Forest classifier. Based on the wall SSM, PC 2 (a) and 4 (b) are shown.

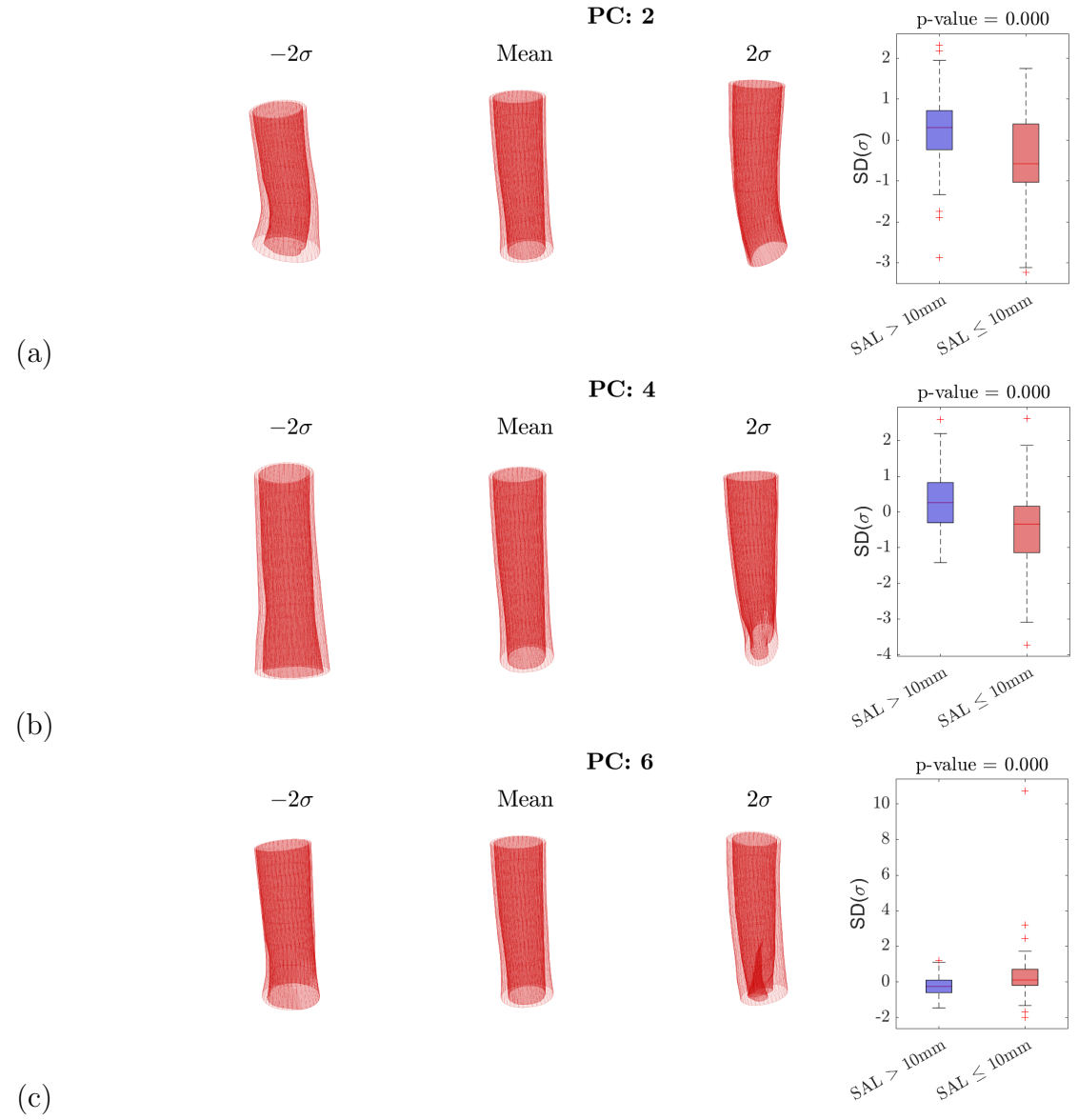


Figure 22: Visualization of PCs with feature importance ≥ 0.10 after training of the Random Forest classifier. Based on the $4D_{lumen}$ SSM, PC 2 (a), 4 (b) and 6 (c) are shown.

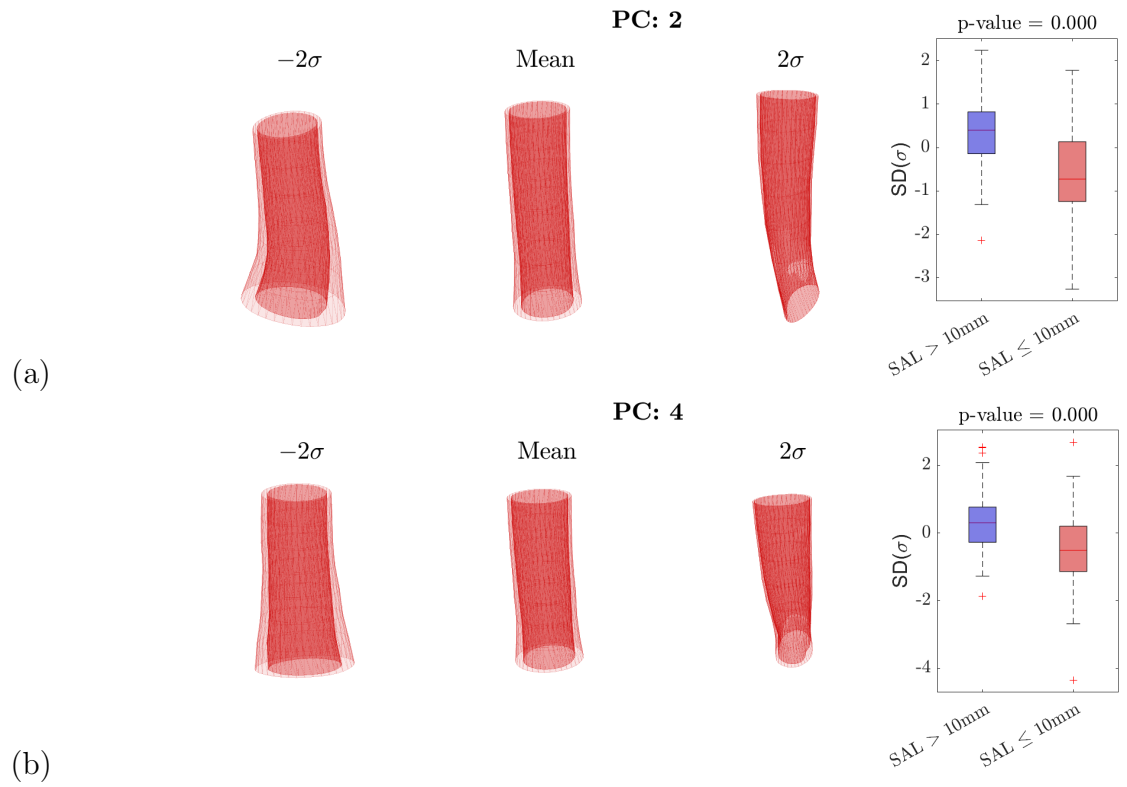


Figure 23: Visualization of PCs with feature importance ≥ 0.10 after training of the Random Forest classifier. Based on the $4D_{wall}$ SSM, PC 2 (a) and 4 (b) are shown.

B Contour Filtering Method

If a contour line enters an aneurysm and is influenced by a large angulation of the CLL, some contour points can be placed in unfavorable positions for SSM. A method to filter contour points that are parameterized further away from the CLL than other points in that contour line is introduced in this thesis and described in this section.

Vector length from the center point of the contour line to the 36 wall points is determined to find outliers. describing contour points that are placed further away from the center points than the other points in the contour, because they are placed in an aneurysm. An upper limit U for the distance from the CLL is identified using the Median Absolute Deviation (MAD) [75], with the *isoutlier*-function in Matlab (The Math Works, Inc. *MATLAB*. Version 2023a).

A filter magnitude of $K = 3$ is chosen for this purpose. Vector lengths exceeding the limit are filtered with this factor using equation 6. Here, the difference from the unfiltered distance x_i to the upper limit U is devided by factor K. This value is added to the upper limit and results in the new vector distance $\tilde{x}_i$. Contour filtering is shown in Figure 24.

$$\tilde{x}_i = U + \frac{x_i - U}{K} \tag{6}$$

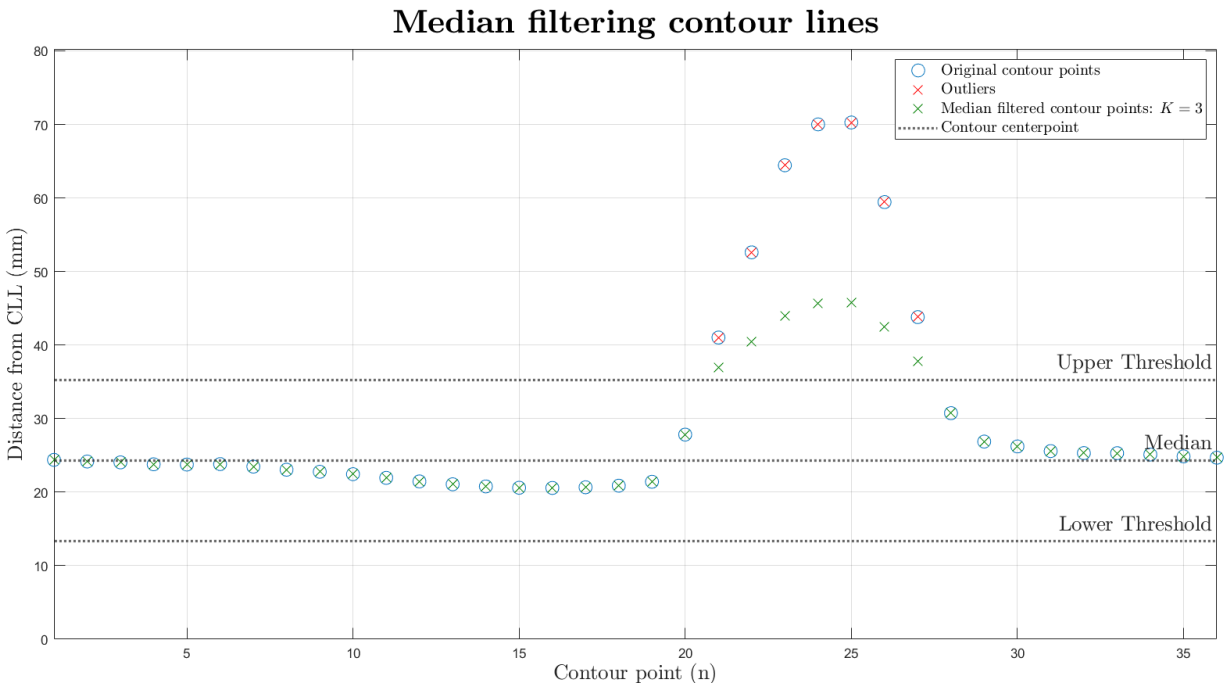


Figure 24: Distance from the CLL is shown for all contour point in one contour line as blue circles. The median value including the upper and lower threshold for outliers is shown as dotted lines. Outliers and the filtered distances are shown as red and green crosses respectively.

Using the filtering method described above, variation described by that contour line is maintained after filtering. The filtered distances are used to find new contour points in the same direction as the unfiltered contour points, using the new distances as vector length. An example of this is shown in Figure 25, where the contour line vertically enters the aneurysm resulting in contour points with large distances from the contour centerpoint.

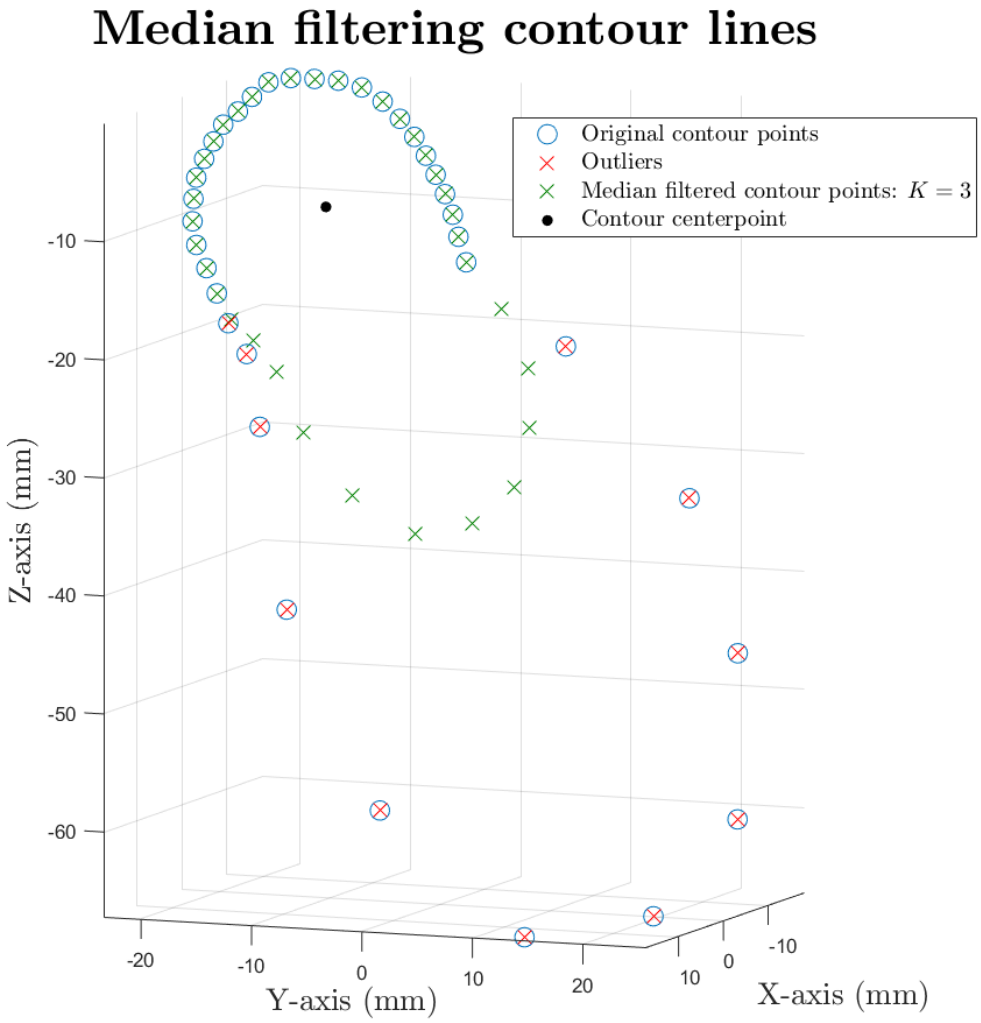


Figure 25: An example of the filtering method is shown. The original contour points are shown as blue circles. Outliers and the filtered distances are shown as red and green crosses respectively. The contour centerpoint is shown as a black dot.

The outliers shown would negatively influence the creation of an SSM. Interpolation of contour points based on these filtered points partly solves this issue. Difficulties might still arise for large CLL angulation between subsequent contour lines as the changing directions are unfavorable for interpolation even after filtering.