

Towards Automatic Detection Of Vessel Stenoses in Cerebral 3D DSA Volumes

F. Mualla, M. Pruemmer, D. Hahn, and J. Hornegger

Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg,
Germany

E-mail: firmualla@informatik.uni-erlangen.de

Abstract. Vessel diseases are very common reason of permanent organ damage, disability, and death. This fact necessitates further research for extracting meaningful and reliable medical information from the 3D DSA volumes. Murray's Law states that at each branch point of a lumen-based system the sum of the minor branch diameters each raised to the power x is equal to the main branch diameter raised to the power x . The principle of minimum work and other factors like the vessel type impose typical values for the junction exponent x . Therefore, deviations from these typical values may signal pathological cases. In this paper, we state the necessary and the sufficient conditions for the existence and the uniqueness of the solution for x . The second contribution is a scale- and orientation- independent set of features for stenosis classification. A support vector machine classifier was trained in the space of these features. Only one branch was misclassified in a cross validation on 23 branches. The two contributions fit into a pipeline for the automatic detection of the cerebral vessel stenoses.

PACS numbers: 42.30.Tz,07.05.Mh,87.57.-s

Submitted to: *Phys. Med. Biol.*

1. Introduction

Vessel diseases are a major reason of permanent organ damage, disability, and death. In this paper, we are interested in those diseases which affect the morphological properties of the vascular wall, and hence can be detected using computer vision methods, and then learned using machine learning techniques. Aneurysms, stenoses, and what we call long abnormalities exemplify this kind of diseases.

A ruptured cerebral aneurysm can cause permanent brain damage, disability, or death. Unfortunately, 80%-90% of aneurysms are not diagnosed until after they have ruptured [Olendorf et al., 2002].

Both stenoses and aneurysms can be described as "local" and "rapid" changes in the vessel's wall geometry. However, in some cases, the vessel may have a narrowing

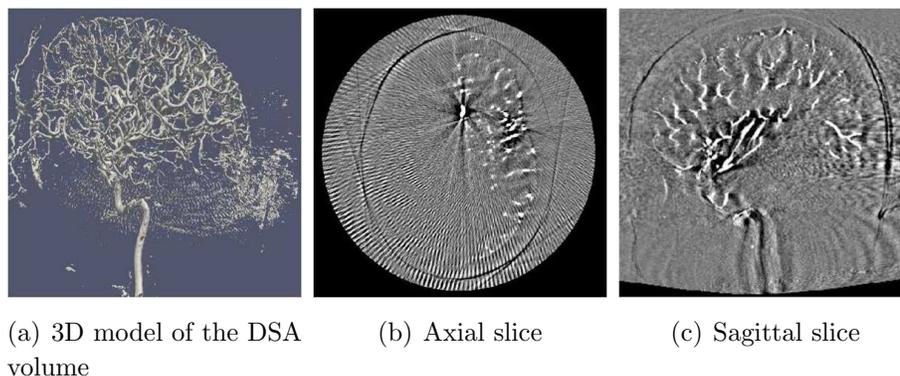


Figure 1. An Example of the input

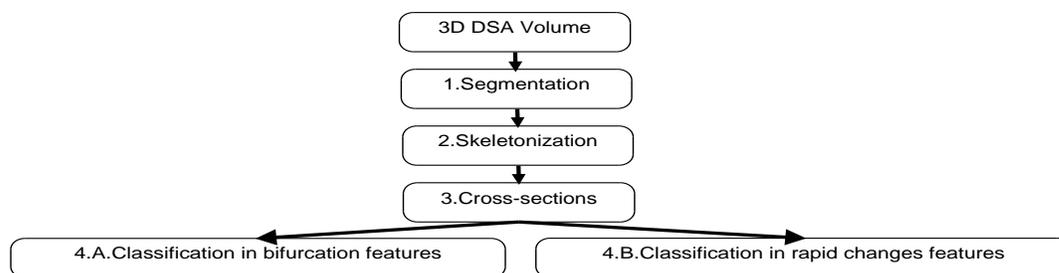


Figure 2. A system for the automatic detection of the vascular diseases

(or a widening) on a long distance, which can be in extreme cases as long as the whole branch. Such narrowings or widenings are termed "long abnormalities" in this paper.

Some aspects of the automatic disease detection are modality-dependent; for instance, the difficulty, and hence the quality of the segmentation. Therefore, it is important to mention the image modality under which our results were obtained. Our input images are 3D Digital Subtraction Angiography volumes computed from CT scans of the human head. Figure 1 shows an example of the input depicted from different perspectives.

An automatic system for the vascular diseases detection (figure 2), in the context of this paper, can be roughly described as a software which takes a 3D DSA volume as input, and gives a pointer to the suspected points (e.g. stenoses) in the whole vascular tree. To achieve this goal, the system must segment the vascular tree, extract the medial axis, and cut cross-sections perpendicular to this medial axis. Then two types of analysis are performed. First, a law in physics called Murray's law is utilized to compute features at the bifurcations of the vascular tree. These bifurcation features are meant to detect the "suspected" branch points. This kind of analysis was suggested in [Endress et al., 2011] to detect the long abnormalities. Second, another set of features is used to detect rapid pathological changes like stenoses and aneurysms at the non-bifurcational medial axis points. This set of features was used and validated in this paper on stenoses. Therefore, it is referred to by "stenoses features" through this text.

To our knowledge, there is no currently available system which can achieve this

task reliably and automatically on the cerebral 3D CT DSA volumes. This paper does not present such a system. Nevertheless, it adds two novel contributions to it. The first contribution fits in box number 4.A of figure 2 and described in section 2.3. The second fits in box number 4.B of the same figure and described in section 2.4.1.

The organization of the paper is as follows. In section 2.1 the vessels segmentation from the 3D DSA volumes is discussed. Section 2.2 discusses getting the medial axis of the segmented mask. In section 2.3 the junction-exponent of Murray’s law is studied and utilized. Section 2.4 addresses the stenosis classification problem. This includes a review of the used machine learning techniques, a description of our training data, and a novel set of features. The results are presented in section 3 and discussed with a summary of conclusions in section 4.

2. Materials and methods

2.1. Segmentation

Aiming for a generic, flawless segmentation framework for 3D vessels is probably illusory [Lesage et al., 2009]. Vesselness measures which depend on the eigenvalues of the Hessian of the intensity profile were used in the literature [Lorenz et al., 1997],[Frangi et al., 1998],[Aylward et al., 1996] for vessels segmentation. Unfortunately, and depending on the noise level, the information we obtained from these measures were not always sufficient. One may think of combining it with the intensity information and the homogeneity constraints in an EM-MRF scheme like [Zhang et al., 2002]. However, this is computationally expensive because of the multi-scale Hessian computation, the eigenvalue analysis, and the iterative scheme.

Level set methods became very popular tool in computer vision since they were introduced by Sethian and Osher in the 1980s [Osher and Sethian, 1988]. Antiga [Antiga, 2002] used the level set segmentation for segmenting 3D vessels. The vascular wall smoothness was controlled by an average second order term. The level set was attracted to the vascular wall using an attraction term to image regions of high gradient magnitude, which correspond to the vessel wall. We used this approach, as it gave sufficiently good results in reasonable time.

2.2. Skeletonization

The skeleton is a classical notion of geometry [Blum et al., 1967]. There exists several algorithms of skeletonization [Lantuéjoul and Maisonneuve, 1984],[Heijmans, 1994]. The first approach by mathematical morphology is in [Lantuéjoul, 1978]. It is called the skeleton of Lantuéjoul. In [Serra, 1988], a link was established with approaches based on the distance transform. In fact, the geodesic skeleton by influence zone (SKIZ) is the basic notion used in the watershed [Beucher and Lantuéjoul, 1979].

With morphological thinning approaches, erosion is applied iteratively on the object at hand until no further erosion can be carried out without breaking some

rules. These rules seek to guarantee the connectedness of the remaining voxels and to assure that almost the same amount of erosion was applied from all directions [Lee et al., 1994],[Lam et al., 1992].

Distance transform methods use the information of the distance transform [Borgefors, 1986]. Informally speaking, the ridge of this transform is the sought skeleton. The method we use [Antiga, 2002] is conceptually distance-transform based. The medial axis is defined as the curve C which minimizes the following energy functional:

$$\int_{0=C^{-1}(p_0)}^{L=C^{-1}(p_1)} F(C(s))ds$$

where C is the centerline curve, p_0 , p_1 are the start and end points, and $F(X) = \frac{1}{R(X)}$. R is the field of the maximum inscribed spheres inside the vessel. The domain of F is restricted to the Voronoi diagram of the boundary. Instead of following a variational approach, the wave propagation theory and the fast marching methods were used in the aforementioned reference to find a global optimum. We found the resulting medial axis satisfying in terms of centeredness, thinness, and continuity.

This algorithm requires the vascular wall to be represented as a polygonal surface. This makes its accuracy dependent on the local mesh resolution instead of the image resolution [Antiga, 2002]. Therefore, the segmentation result is converted to a surface mesh using Marching Cubes [Lorensen and Cline, 1987]. The resulting surface is smoothed using [Taubin, 1995] in order to remove the small segmentation artifacts. As a post-processing step, the resulting skeleton is slightly smoothed using a moving average filter.

2.3. The junction exponent

A concept will be borrowed from fluid dynamics and utilized to detect the unhealthy cases at the bifurcation points of the vascular tree. This concept is called Murray's law:

$$\sum_{i=1}^{m-1} a_i^x = a_m^x$$

where a_m is the main branch diameter, while the a_i s are the diameters of the minor branches, and m is the total number of branches at the considered bifurcation, so it is at least 3. Classical literature of Murray's law points out that a value of 3 for the junction exponent x is the "typical" value. Typical in the sense that best circulation efficiency will be achieved by the system whose branches fulfill Murray's law with this exponent. More information is in [Sherman, 1981].

Kaimovitz et al [Kaimovitz et al., 2008] showed in a study on porcine coronary arterial trees, that the typical exponent is a function of the vessel's order, and it is significantly far from Murray's optimum 3 in many cases. For example, it is close to 3 and 3.2 for order 1, close to 2.4 for order 3 and minimized for order 5 and 6 with a value close to 1.4. So, nature does not strictly follow Murray's mathematical conclusion.

Nevertheless, the junction exponent values are not arbitrary. [Endress et al., 2011] suggested using the law to detect the "long abnormalities" in the human vascular trees. We think that two requirements are needed to achieve this goal. First, a theoretical understanding of the mathematical properties of the junction exponent is necessary. Second, a decision boundary between the healthy and the pathological bifurcations must be learned in the bifurcation features. As mentioned in the introduction, these bifurcation features are derived from the junction exponent. In this paper, we address the first requirement. We also extract bifurcation features. The classification in the space of these features is planned as further work.

2.3.1. Theoretical analysis We are interested in solving Murray's equation:

$$h(x) = \left(\sum_{i=1}^{m-1} a_i^x \right) - a_m^x = 0, x > 0 \quad (1)$$

Dividing both sides by a_m^x yields:

$$g(x) = \left(\sum_{i=1}^{m-1} \left(\frac{a_i}{a_m} \right)^x \right) - 1 = 0, x > 0 \quad (2)$$

This division is important because $h(x)$ is not convex in general, whereas $g(x)$ is.

The equation is not always solvable, e.g. the configuration ($m = 3, a_3 = a_2$) yields an exponential which is never zero. We'll see later that there is a class of cases for which there is no solution of the equation. So, the problem is better formulated as a minimization problem:

$$\widehat{x} = \arg \min_x f(x) = \arg \min_x \left(\left(\sum_{i=1}^{m-1} \left(\frac{a_i}{a_m} \right)^x \right) - 1 \right)^2 = \arg \min_x (g(x))^2, x > 0 \quad (3)$$

Adding Lagrange Multiplier $\lambda \geq 0$ yields the following Lagrangian of the problem:

$$L(x, \lambda) = \left(\left(\sum_{i=1}^{m-1} \left(\frac{a_i}{a_m} \right)^x \right) - 1 \right)^2 - \lambda x \quad (4)$$

The necessary conditions for the optimum are:

$$f'(x) = 2g(x)g'(x) = \lambda$$

$$\lambda x = 0$$

where

$$g'(x) = \sum_{i=1}^{m-1} \ln\left(\frac{a_i}{a_m}\right) \left(\frac{a_i}{a_m}\right)^x \quad (5)$$

In case the constraint is inactive, $x > 0$ implies $\lambda = 0$ and the necessary condition turns out to be:

$$f'(x) = 2g(x)g'(x) = 0$$

In the following, we characterize the properties of the possible solution of the optimization problem (3) including its relationship with the possible solution of the equation (2). A set of statements will be claimed and proved in order to investigate these properties, starting with this definition:

Definition 1. We'll call the case when the condition $\exists k \in 1 \dots m - 1 : a_k \geq a_m$ holds a **degenerate configuration**.

Claim 1. $g(x)$ has at most one root in \mathbb{R} under a **non-degenerate** configuration, and it has no root in the feasible range under a **degenerate** configuration.

Proof:

- (i) If $a_m > a_i, i = 1..m - 1$ (i.e. non-degenerate case), then all the exponentials in g are strictly decreasing over the whole real axis \mathbb{R} . This implies that their sum, and hence g is strictly decreasing over \mathbb{R} , which leads to the fact that g has at most one root because it is strictly monotonic and continuous.
- (ii) If the condition in the previous case does not hold, then there is at least one of the exponentials in g increasing over \mathbb{R} ‡, which its minimum value for the feasible range approaches 1 as x approaches 0. The other exponentials in g are decreasing but strictly positive, which means that the sum $\sum_{i=1}^{m-1} (\frac{a_i}{a_m})^x$ is always strictly more than 1 in the feasible range, and hence $g(x) = 0$ is not solvable.

Claim 2. $g(x)$ has at least one root in the feasible range under a **non-degenerate** configuration.

Proof: In the non-degenerate case we have $\lim_{x \rightarrow \infty} g(x) = -1$ and $\lim_{x \rightarrow 0} g(x) = m - 2$. The latter is ≥ 1 because $m \geq 3$. $g(x)$ is continuous. The previous facts imply that there is at least one root in $]0, +\infty[$.

Claim 3. $g'(x)$ has at most one root in \mathbb{R} under **any** configuration, and it has no root in \mathbb{R} under a **non-degenerate** configuration.

Proof:

- (i) Having $g'(x)$ with more than one root contradicts the fact that g is a non-constant convex function.
- (ii) If $a_m > a_i, i = 1..m - 1$ (non-degenerate configuration), then it is obvious from equation (5) that a sum of exponentials with the same sign (negative) cannot be zero.

Claim 4. $g'(x)$ has at least one root in the feasible range under a **degenerate** configuration if and only if $\beta = \sum_{i=1}^{m-1} \ln(\frac{a_i}{a_m}) < 0$.

Proof:

- (i) Let us consider $\lim_{x \rightarrow \infty} g'(x)$ in the degenerate case, which is one of these two cases:

‡ even in the special case when the base is 1, the minimum value is 1

(a) There is at least k such that $a_k > a_m$:

$$\lim_{x \rightarrow \infty} g'(x) = \lim_{x \rightarrow \infty} \sum_{i=1}^{m-1} \ln\left(\frac{a_i}{a_m}\right) \left(\frac{a_i}{a_m}\right)^x = +\infty$$

because there is at least one increasing exponential due to the degenerate case.

(b) No $a_i, i : 1..m - 1$ is strictly larger than a_m , but there is at least k such that $a_k = a_m$ which leads to:

$$\lim_{x \rightarrow \infty} g'(x) = \lim_{x \rightarrow \infty} \sum_{i=1}^{m-1} \ln\left(\frac{a_i}{a_m}\right) \left(\frac{a_i}{a_m}\right)^x = C > 0$$

where C is a non-zero positive number.

(ii) Let us consider the limit at the other side of the feasible range:

$$\lim_{x \rightarrow 0} g'(x) = \lim_{x \rightarrow 0} \sum_{i=1}^{m-1} \ln\left(\frac{a_i}{a_m}\right) \left(\frac{a_i}{a_m}\right)^x = \sum_{i=1}^{m-1} \ln\left(\frac{a_i}{a_m}\right)$$

(iii) The second derivative $g''(x) = \sum_{i=1}^{m-1} \ln^2\left(\frac{a_i}{a_m}\right) \left(\frac{a_i}{a_m}\right)^x$ is always positive, which means that $g'(x)$ is increasing over \mathbb{R} , and hence over the feasible range.

We conclude that the continuous function $g'(x)$ has a root in the feasible range *iff* the limit of g' is negative when x approaches 0 .

Claim 3 leads to the following interesting result:

Result 1. *Solving the equation (2) and the optimization problem (3) are equivalent under a **non-degenerate** configuration!*

Claims 1 and 3 lead together to:

Result 2. *$f(x)$, though not convex in general, but it has one stationary point at most.*

The proof of claim 4 yields that under a degenerate configuration, there is no solution if and only if $\beta = \sum_{i=1}^{m-1} \ln\left(\frac{a_i}{a_m}\right) \geq 0$ and this corresponds with strictly positive g' over the feasible range. On the other hand, g is strictly positive over the feasible range in the degenerate case as well (see proof of claim 1). So $f'(x) = 2g(x) * g'(x)$ is strictly positive over the feasible range when there is no stationary point of f , which leads to the following result:

Result 3. *When $f(x)$ has no stationary point, it is minimized (with its value zero) when x approaches zero from the right hand side.*

Of course, this case is not interesting as it cancels the effect of the individual diameters a_i by raising to power 0.

2.3.2. *Interpretation of the condition $\beta < 0$* It turns out that $\beta < 0$ has an interesting interpretation:

$$\begin{aligned} \sum_{i=1}^{m-1} \ln\left(\frac{a_i}{a_m}\right) < 0 &\Leftrightarrow \exp\left(\sum_{i=1}^{m-1} \ln\left(\frac{a_i}{a_m}\right)\right) < 1 \Leftrightarrow \prod_{i=1}^{m-1} \exp\left(\ln\left(\frac{a_i}{a_m}\right)\right) < 1 \Leftrightarrow \\ &\prod_{i=1}^{m-1} \frac{a_i}{a_m} < 1 \Leftrightarrow \frac{\prod_{i=1}^{m-1} a_i}{a_m^{m-1}} < 1 \Leftrightarrow \prod_{i=1}^{m-1} a_i < a_m^{m-1} \end{aligned}$$

Or:

$$\sqrt[m-1]{\prod_{i=1}^{m-1} a_i} < a_m \quad (6)$$

This tells simply that $\beta < 0$ is equivalent to the geometric mean of the minor branch diameters being less than the main branch diameter. Cases which fulfill this condition are the non-degenerate cases plus a subset of the degenerate cases.

2.3.3. *Finding the junction exponent in practice* Newton's method is a well-known root-finding and optimization technique with a quadratic convergence rate. However, it does not take into account the problem constraints. The theoretical results of section 2.3.1 will be utilized to use Newton's method properly, such that the convergence is guaranteed and the constraint is fulfilled. In the following, a detailed discussion of using Newton's method to minimize our objective function f (optimization problem (3)) is presented and justified:

The degenerate condition given in definition 1 is checked:

- (i) Non-degenerate case: The root of $g(x)$ is found, Newton for finding roots is guaranteed to converge to the root because $g(x)$ is convex, and there is one and only one root according to claims 1 and 2. As this root is in the feasible range (claim 2), the constraint will be fulfilled automatically, even when the iterative procedure hits values outside the feasible range during the iterations. There is no need to check the root of $g'(x)$ according to claim 3.
- (ii) Degenerate case: The condition $\beta < 0$ pointed out in claim 4 is checked:
 - (a) If fulfilled, the stationary point of $g(x)$ is found. Because of the convexity, Newton's method is guaranteed to converge to one and only one solution due to claims 3 and 4. This solution is also automatically inside the feasible range due to claim 4 even if Newton's method hits the negative side of the x-axis during the iterations.
 - (b) If the condition is not fulfilled, then there is no solution due to claim 4.

2.4. *Stenosis classification*

2.4.1. Stenosis features A literature review of the stenosis classification reveals the fact that not so many automatic stenosis detection approaches were suggested. Among them, Zeng [Zeng, 2011] detected the cardiac stenoses in the 2D angiograms by measuring the ratio between the minimum diameter and the most frequent diameter in the vessel. TeBmann et al. [TeBmann et al., 2008] used the intensity, the intensity gradient, and the Haar-like features in order to train an Ada-Boost classifier to detect the coronary stenotic lesions in the CTA images. Some other works are completely different from our approach. Either because the imaging modality is inherently different [Goodenday et al., 1997], or because the visual pattern is different [Makram-Ebeid, 1999],[Christodoulou et al., 2003],[Mittal et al., 2010], or even the signal is acoustic [Pavlopoulos et al., 2004]. The method suggested in this paper measures the morphological changes of the vascular wall as opposed to the intensity-dependent features in [TeBmann et al., 2008] and the perfusion pattern detection in [Goodenday et al., 1997]. Therefore, the closest approach to ours is the one followed in [Zeng, 2011].

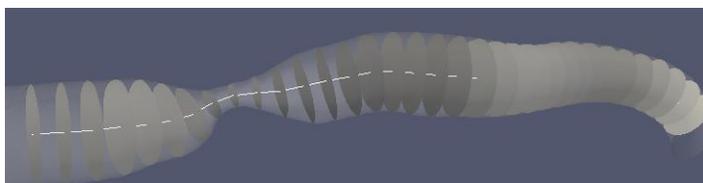


Figure 3. A 3D model of a branch. The white line is the medial axis. Perpendicular to it, one cross-section was cut every 1 mm. In the training data, a smaller interval was taken in order to use all available information.

In our approach, we extract a feature vector at each point of the medial axis of the vessel. The intended outcome of each feature vector is to give a judge whether the corresponding medial axis point belongs to a stenotic or a healthy vascular segment. The features must be independent of the size and the orientation of the vessel. In the following, we propose five features to serve this goal.

At each medial axis point, we cut a cross-section perpendicular to the medial axis. This is shown in figure 3. For each cross-section, we compute the maximum diameter, the minimum diameter, the area, and the circumference. Obviously, these geometric measures cannot be used directly as an input to a machine learning algorithm because they are dependent on the vessel size. However, their derivatives w.r.t. the distance or the abscissa along the medial axis contain important and discriminative information.

Assume a function $t(s)$ which increases before $s = s_0$ and then decreases after it (or vice versa). This behavior of t at s_0 can be well captured by its curvature, or its second derivative. If t is the maximum diameter of the vessel, and s is the spatial dimension along the medial axis of the vessel (abscissa), then the first case may correspond to an aneurysm, and the second may correspond to a stenosis. t can be also any other geometric measure like the minimum diameter, the area, and the circumference. The second derivative can capture the rapid change, but it is still dependent on the vessel

size. For instance, a linear scaling of the vessel diameter by a factor a will scale the 2nd derivatives of the maximum diameter, the minimum diameter, and the circumference by a . And scale the 2nd derivative of the area by a^2 . Therefore, the derivative must be normalized.

The discretization of the curvature is given by the well-known finite differences formula:

$$c(t, s, d) = \frac{t(s+d) - 2t(s) + t(s-d)}{d^2} \quad (7)$$

where t is the geometric measure as a function of the abscissa s along the medial axis, and d is the scale.

To normalize it, we can divide by some measure M of the vessel size like the maximum diameter in the whole branch, or the average diameter in the whole branch:

$$c_n(t, s, d) = \frac{1}{M} \frac{t(s+d) - 2t(s) + t(s-d)}{d^2} \quad (8)$$

A possibly better normalization option than the average in the whole branch is the simple average at the extraction point:

$$c_n(t, s, d) = \frac{t(s+d) - 2t(s) + t(s-d)}{\frac{t(s+d)+t(s)+t(s-d)}{3} d^2} \quad (9)$$

We think that the second is better because no global information about the whole branch is needed. In some applications, e.g. in 3D navigation, the segmentation of the vascular tree into branches might be unknown. Therefore, a value like the average diameter in the branch is unavailable.

Substituting t in (9) with the maximum diameter, the minimum diameter, the area, and the circumference of the cross sections generates four features.

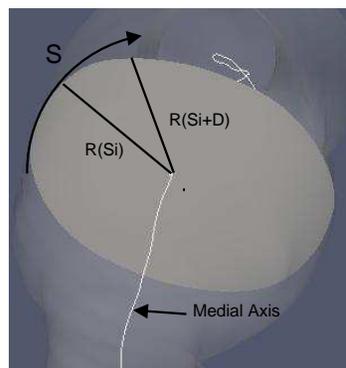


Figure 4. Illustration of the feature SectionCurvature

A fifth feature uses also the second derivative, but it has a different purpose. Rather than quantifying the changes along the medial axis like the previous four features, it quantifies the change along the enclosing curve of the cross-section. The feature is

defined by the following formula:

$$C_n(s) = \frac{1}{P(s)} \sum_{i=1}^{P(s)} \left| \frac{R(S_i + D(s)) - 2R(S_i) + R(S_i - D(s))}{D(s)} \right| \quad (10)$$

where R is the radius of the cross-section as a function of the abscissa S along the enclosing curve. s is the abscissa along the medial axis. $P(s)$ is the points number of the cross-section's enclosing curve at s . $D(s)$ is the scale defined as a specified ratio of the enclosing curve circumference at s . E.g. $D = 10\%$.

There are two differences between d in (9) and D in (10). First, D is a function, whereas d is a constant. Second, D is not squared. These two changes were made to make the feature independent of the vessel size. Any linear scaling of the vessel diameter will cause the same scaling of the circumference, and hence of D .

This feature measures how "weird" or uncircular the cross-section is. It is zero for the circular sections in spite of their radii.

Table 1 summarizes the five features. We have shown how to make the features size-independent. And because they are extracted from cross-sections which are perpendicular to the medial axis, they are also independent of the vessel orientation. The features capture the vascular wall changes in spite of the change's type. Therefore, they can be -in principle- used for aneurysms and stenoses. In this paper, we use them for stenoses.

Feature name	Formula
MaxDiameterCurvature	Substitute t with the maximum diameter in (9)
MinDiameterCurvature	Substitute t with the minimum diameter in (9)
AreaCurvature	Substitute t with the area in (9)
CircumferenceCurvature	Substitute t with the circumference in (9)
SectionCurvature	Use (10)

Table 1. Stenosis features

2.4.2. Support vector machines and kernel methods The development in the statistical learning theory led to the invention of the support vector machine (SVM) classifier [Boser et al., 1992]. The SVM training algorithm tries to find the linear decision boundary which maximizes the margin between two classes. A linear decision boundary can be expressed by the equation $\alpha x + \beta = 0$ where α is an n -dimensional normal vector, n is the dimensionality of the feature space, and β is the intercept. Maximizing the margin is achieved by minimizing the L2-norm of the normal vector α [Scholkopf and Smola, 2001].

The corresponding minimization problem is:

$$\begin{aligned} &\text{find } \alpha \text{ and } \beta \text{ such that :} \\ &g(\alpha) = \frac{1}{2} \|\alpha\|^2 + C \sum_i \xi_i \text{ is minimized under the conditions} \\ &\{y_i(\alpha^T x_i + \beta) \geq 1 - \xi_i, \xi_i \geq 0 \text{ for all patterns } x_i\} \end{aligned} \quad (11)$$

where x_i is the i th training pattern . y_i is its label which is either $+1$ or -1 . ξ_i s are the slack variables. They are responsible for the soft-margin. C is the SVM cost parameter.

Allowing the classifier to make some training errors is desirable as simpler decision surfaces tend to generalize better [Hamel, 2009]. The parameter C controls this trade-off between the over-fitting and the under-fitting.

The optimization problem is commonly solved in its dual form which yields a quadratic with a negative semi-definite Hessian. And hence, a concave objective function:

$$g(\lambda) = -\frac{1}{2}\lambda_{(1,p)}^T A_{(p,p)} \lambda_{(p,1)} + 1^T \lambda \rightarrow \max$$

$$\text{subject to } \sum_{i=1}^p \lambda_i y_i = 0, C \geq \lambda_i \geq 0$$
(12)

where p is the number of the training patterns, λ is a vector of Lagrange multipliers, and A is a matrix whose elements are the inner products of the training patterns.

The SVM discussed so far finds a linear decision boundary. If the classes are not linearly separable, then kernel methods are commonly used. The matrix A in (12) is a matrix of inner products, therefore it is positive semi-definite. Using the kernel trick [Scholkopf and Smola, 2001], one can replace A with other positive semi-definite matrices. This is equivalent to mapping the feature vectors into a higher dimensional feature space, where they can be linearly separable.

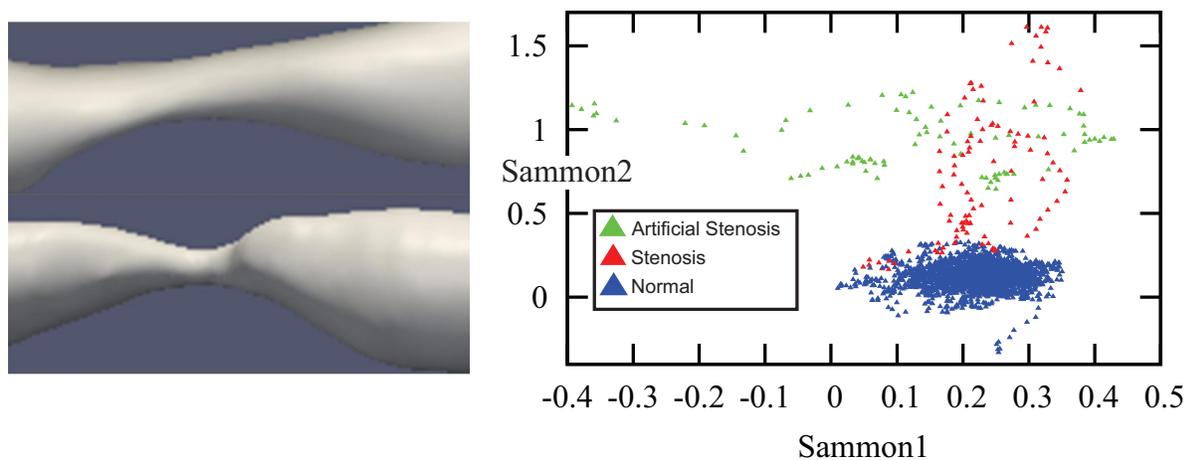
The simplest kernel is the linear kernel $A_{i,j} = k(x_i, x_j) = \langle x_i, x_j \rangle$ which is the one used in the previous formulation. Another possible kernel is the Gaussian Radial Basis Function RBF given by $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, where γ is the RBF kernel parameter. Other common kernels include the polynomial, sigmoid, and others (see [Abe, 2010]). The Gaussian RBF kernel is a reasonable first choice, and it has fewer numerical difficulties and hyperparameters than the polynomial kernel [Hsu et al., 2003]. In [Lin and Lin, 2003], it was concluded that the sigmoid kernel is in general not better than the RBF. Furthermore, if a complete model selection has been conducted using the RBF kernel, then there is no need to consider the linear kernel [Keerthi and Lin, 2003]. Therefore, we started with the RBF. We tested also the linear kernel in order to investigate the linear separability.

One major advantage of the SVM over the artificial neural networks (ANNs) is the convexity of the optimization problem which leads to a global optimum, as opposed to a local optimum in the ANNs. Another advantage is the immunity against outliers, which can be controlled by the cost parameter C . Third is the generalization ability because of the very nature of the SVM (see [Abe, 2010]).

Unlike the probabilistic classifiers, the support vector machine gives a hard decision. We use the SVM for stenosis classification and we need a probabilistic result which expresses the severity of the stenosis. For this purpose, we use the method of Platt [Platt, 1999]. He fit a sigmoid function after the SVM in order to get a probability measure from the distances to the decision boundary.

One important aspect of our classification problem is the imbalance of the training data, as the number of healthy feature vectors is much larger than the number of the stenotic feature vectors. Veropoulos et al [Veropoulos et al., 1999] suggested using two cost parameters C^+ and C^- to compensate for the imbalance. The ratio by which the positive misclassifications outweigh the negative misclassifications is called the cost factor $F = \frac{C^+}{C^-}$. This model is called Control-Sensitivity SVM or CS SVM. [Lee et al., 2001] reported considerable improvement using the CS SVM compared to the conventional SVM. As mentioned above, our data is imbalanced. Therefore, we use the CS SVM.

2.4.3. Training data Two classes are considered. "normal" for the healthy feature vectors and "stenosis" for the stenotic feature vectors. Luckily enough, data for the class "normal" is available because any healthy branch can be used to provide feature vectors for this class. However, data for the class "stenosis" is not sufficiently available. To circumvent this problem, we added artificial stenoses. These stenoses were generated from natural branches which contain surface deformations by over-deforming the surface. Figure 5(a) shows a natural and an artificial stenosis. A medical expert evaluated the artificial stenoses in order to assess their similarity to the natural stenoses. He gave an average grade of 65 %.



(a) An example of an artificial stenosis (top) and a natural one (bottom)

(b) Sammon map of the training data

Figure 5. Training data

The natural dataset is extracted from 6 DSA volumes. It contains 12 normal branches and 4 stenoses. The artificial dataset is 7 stenoses. Overall, we have 23 branches divided into 12 normal and 11 stenotic branches. Each branch gives a set of feature vectors, one feature vector per medial axis point. The final result is 3087 normal feature vectors and 189 stenotic feature vectors.

In order to get an insight into the five-dimensional feature space, we used the Sammon map visualization [Sammon, 1969]. It maps n dimensional feature vectors non-linearly into 2 or 3 dimensional feature space s.t. the "structure" is approximately preserved. Figure 5(b) shows the Sammon map of our five-dimensional feature space.

Normalization is a common pre-processing step of the classification pipeline. Simple normalization techniques like max-min and z-score are not without pitfalls (see [Jain et al., 2005]). The stenosis features have similar dynamic ranges. Therefore, we chose to train the classifiers without normalization.

3. Results

3.1. Junction exponent

The results of the junction exponent analysis are theoretical (section 2.3). What we show here is, more or less, a proof of concept.

We solve the optimization problem (3) using the procedure described in section 2.3.3. Obviously, the diameters a_i s must be known. Each diameter a_i is estimated by the formula $\hat{a}_i = \sum_{j=1}^{L_i} \text{MD}(i, j)$, where L_i is the number of the cross-sections in branch i , $\text{MD}(i, j)$ is the maximum diameter of the j th cross-section in branch i . A second choice is to estimate a_i using the minimum diameter instead of the maximum. A third is to use the mean of the maximum and the minimum diameters. These three ways of estimating the a_i s give different solutions of (3) and provide more information. In addition, the residual defined by equation (2) provides another kind of information.

Figure 6 exemplifies the process. First, cross-sections are assigned to branches in 6(a). Second, an average diameter of each branch is computed 6(b). In this example, the maximum diameters were averaged. Third, the optimization problem (3) is solved. Figure 6(c) shows the optimum. While figure 6(d) shows the residual.

In figure 6, there is a widening in one of the minor branches. Figure 6(c) shows that at the bifurcation related to the branch which contains the widening, we got a high junction exponent value compared to the other bifurcation.

Figure 7 shows another example. Here, each diameter a_i is estimated using the average of the mean diameters in branch i . Bifurcation A contains an aneurysm in its right minor branch. The branches related to bifurcation B do not contain rapid surface changes like stenoses or aneurysms, but the main branch is narrower than one of the minor branches (degenerate case). Both the junction exponent and the residual are distinguishable at A and B .

One of our theoretical results was that the residual is not zero if and only if the case is degenerate. This can be seen in bifurcation A of figure 6 and bifurcations A and B of figure 7.

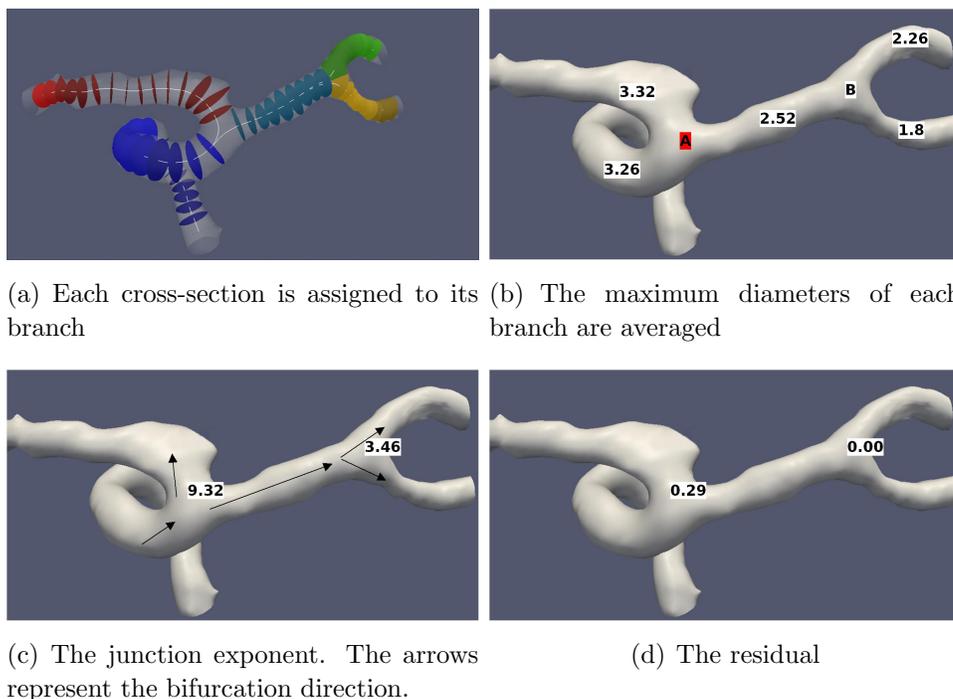


Figure 6. Computing the junction exponent

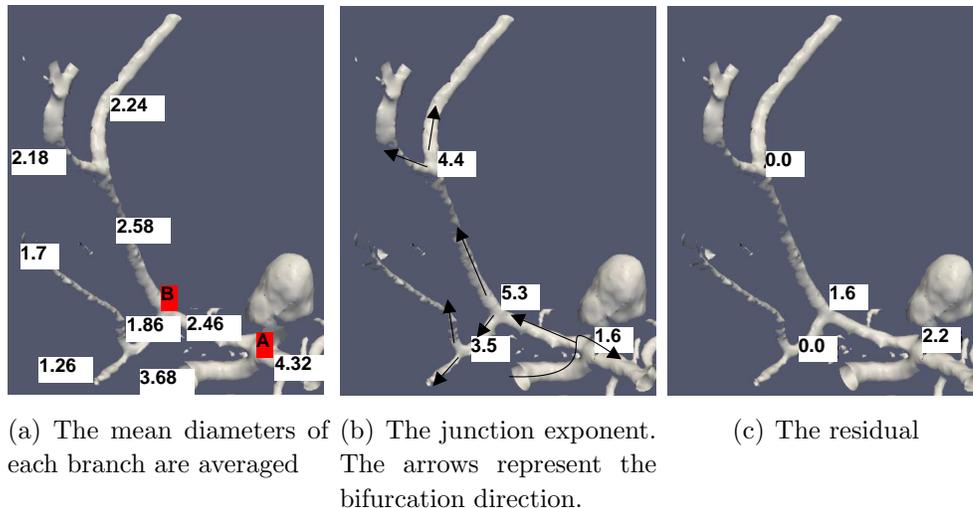


Figure 7. The junction exponent related information with an aneurysm

3.2. Stenosis features

SectionCurvature is meant to capture the weirdness of the cross-section. Figure 8(a) shows this feature extracted along a branch. The cross-section at which the feature is maximized in this branch is plotted. As can be seen, the enclosing curve of this cross-section suffers from severe uncircular deviations along its perimeter.

Figure 8(b) shows CircumferenceCurvature along a branch. The stenosis here is a decrease of the diameter followed by an increase. But what if there is a sever

decrease which is not followed directly by an increase? Figures 8(c) and 8(d) show that CircumferenceCurvature can be discriminative in this case as well.

Figures 8(a),8(b), and 8(c) are real stenoses. In Figure 8(d), we over-smoothed the surface to make an ideal simulation of the case when the change from one side is rapid and from the other side is gradual.

SectionCurvature should express the weirdness of the cross-section in a size-independent way. We chose the feature vectors for which the ratio between the minimum diameter and the maximum diameter is more than 0.9. These cross-sections are nearly circular. Therefore, SectionCurvature must be close to zero. Figure 9(a) shows that SectionCurvature is very close to zero for those cross-sections, regardless of the cross-section’s size. The range of the SectionCurvature axis in the figure is the dynamic range of this feature after removing the extreme outliers.

CircumferenceCurvature should reflect the change in a size-independent way. Figure 9(b) shows a set of stenotic feature vectors. As can be seen, CircumferenceCurvature does not increase with the increasing circumference.

Figure 9(c) shows the expected positive correlation between CircumferenceCurvature and AreaCurvature in the whole training dataset. While figure 9(d) shows that this correlation is much less between MaxDiameterCurvature and MinDiameterCurvature because most cross-sections are not circular.

3.3. Stenosis classification

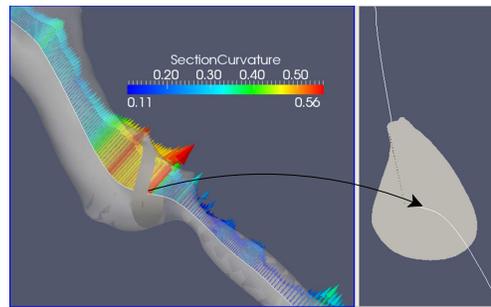
To evaluate the classification performance, we use the balanced error rate $BER = \frac{1-TPR+FPR}{2}$. Where TPR is the true positive rate or the sensitivity. FPR is the false positive rate or $1 - \text{specificity}$. The BER was chosen because it can adapt to the data imbalance.

Estimating the performance measure is done using K-fold cross validation. One may think of using leave-one-out cross validation LOOCV because it is an unbiased estimator of the test error [Smola and Bartlett, 2000] and it can be efficiently computed with the SVMs [Abe, 2010]. However, if the data is densely sampled, LOOCV may overfit. Therefore, we have decided for $K = 5$.

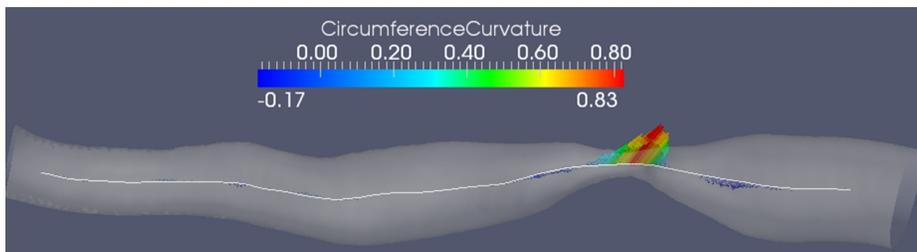
Parameter Search was performed to find the best RBF kernel in the space of γ , C , and F similar to [Hsu et al., 2003]. A rough estimation of the BER minimum was found with very coarse grid search under 3-fold CV. Then a finer grid search with 5-fold cross validation was performed near the previous optimum. As the data is imbalanced, stratified sampling was used in both of them. The same search strategy was applied with a linear kernel, where we have two parameters C and F . The results are shown in table 2. Figure 10 shows a ROC curve comparison of the corresponding models.

To evaluate each model, we applied CV on the 23 branches. The following error function was used:

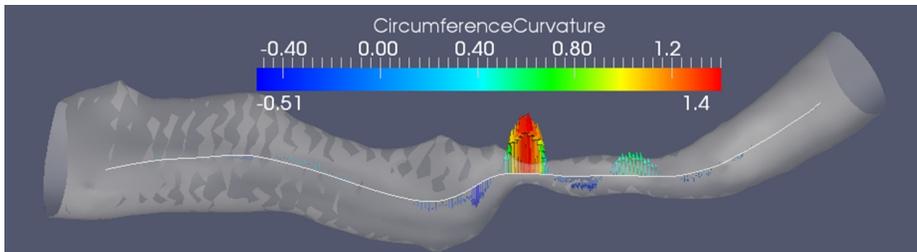
$$E(w_s, w_h) = \frac{w_s \sum_{i=1}^{n_s} \frac{mc_i}{n_i} + w_h \sum_{i=1}^{n_h} \frac{mc_i}{n_i}}{w_s n_s + w_h n_h}$$



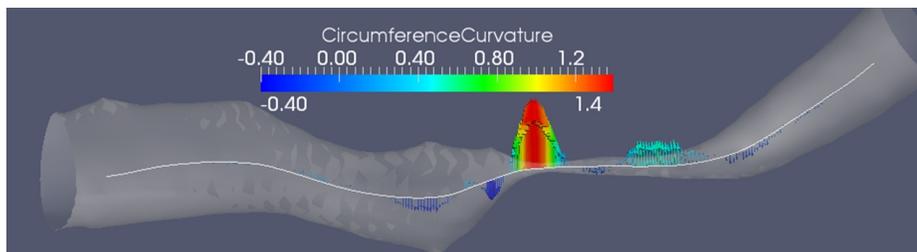
(a) Left: SectionCurvature with scale $D = 10\%$. Right: The cross-section at which the feature is maximized in this branch.



(b) CircumferenceCurvature extracted with scale $d = 1$ mm. The branch is approximately 35 mm long. The minimum of the minimum diameters is 0.92 mm.



(c) CircumferenceCurvature extracted with scale $d = 1$ mm. The branch is approximately 25 mm long. The minimum of the minimum diameters is 0.24 mm.



(d) CircumferenceCurvature on an over-smoothed version of the previous branch.

Figure 8. Stenosis features shown on stenotic branches

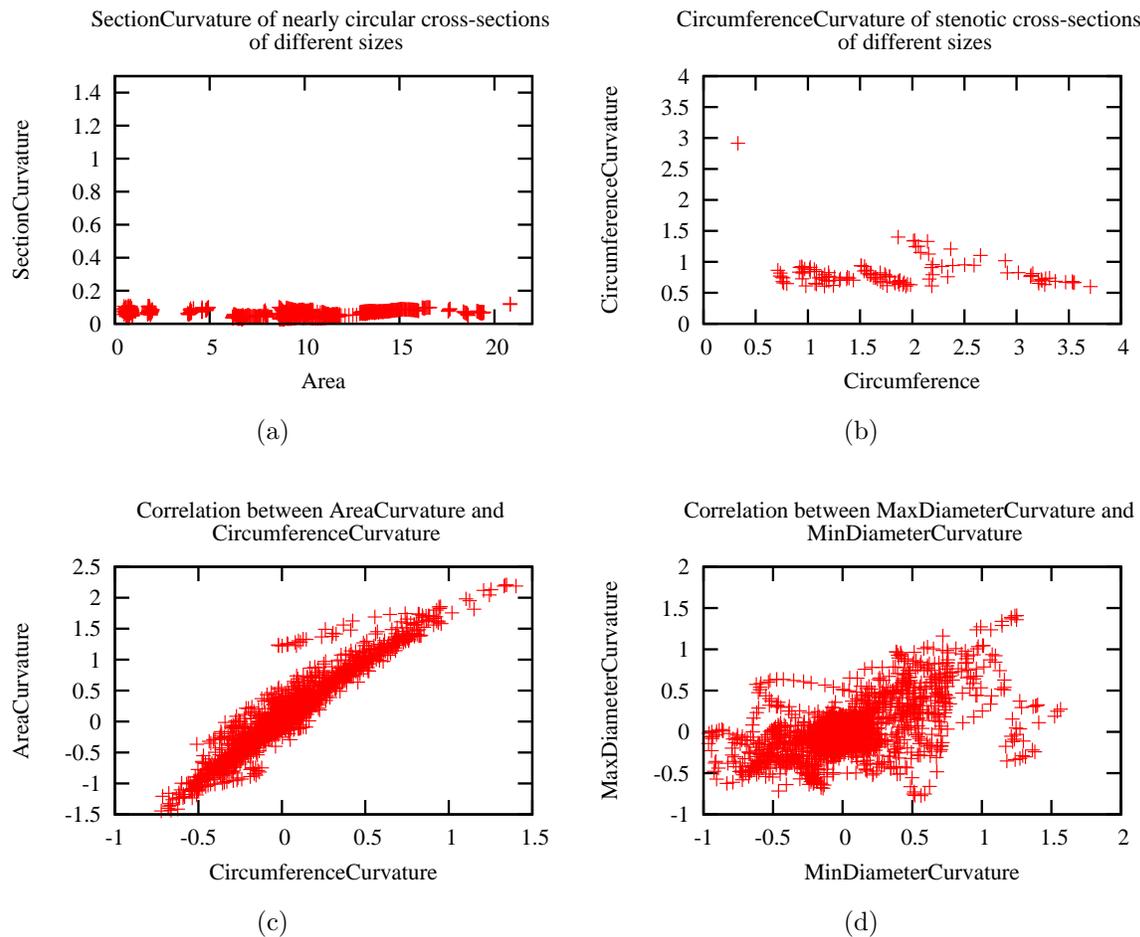


Figure 9. Some statistical properties of the stenosis features

Classifier	C	F	γ	BER
SVM with linear kernel	300	10	-	0.01
SVM with RBF kernel	250	0.1	25	0

Table 2. Results of the parameter search

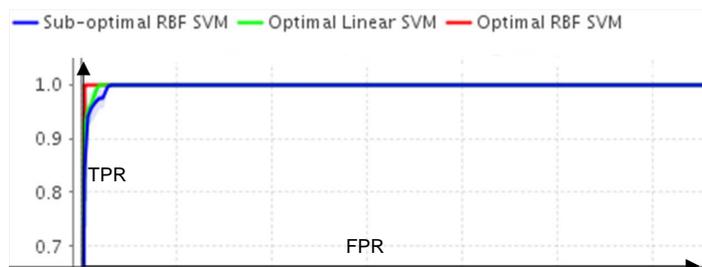


Figure 10. ROC curve comparison. The hard decision of each model was made probabilistic using [Platt, 1999]. The curves are computed with 5-fold CV and stratified sampling.

where n_s is the number of the stenotic branches in the training data. n_h is the number of the healthy or the normal branches. In our case, $n_s = 11$ and $n_h = 12$. n_i and mc_i are respectively the number of the feature vectors and the number of the misclassifications in the i th branch. w_s and w_h are the misclassification weights of the stenotic and the healthy branches, respectively. This error function makes the contribution of each branch in the error independent of the feature vectors number in this branch. Obviously, its range is $[0, 1]$. The results are shown in table 3.

Classifier	$E(1, 1)$	$E(2, 1)$	$E(10, 1)$
Sub-Optimal SVM with RBF kernel	0.192	0.265	0.352
Optimal SVM with linear kernel	0.064	0.084	0.111
Optimal SVM with RBF kernel	0.067	0.083	0.106

Table 3. Model Evaluation: Results of the branches CV

Figure 11 shows the distribution of the error among the branches. Branch number 19 is the only branch which was completely misclassified by both classifiers. By checking it, we noticed that its narrowing is not really serious. It was the only stenosis with light narrowing. Therefore, deleting it from the training data (by the CV) caused the classification error.

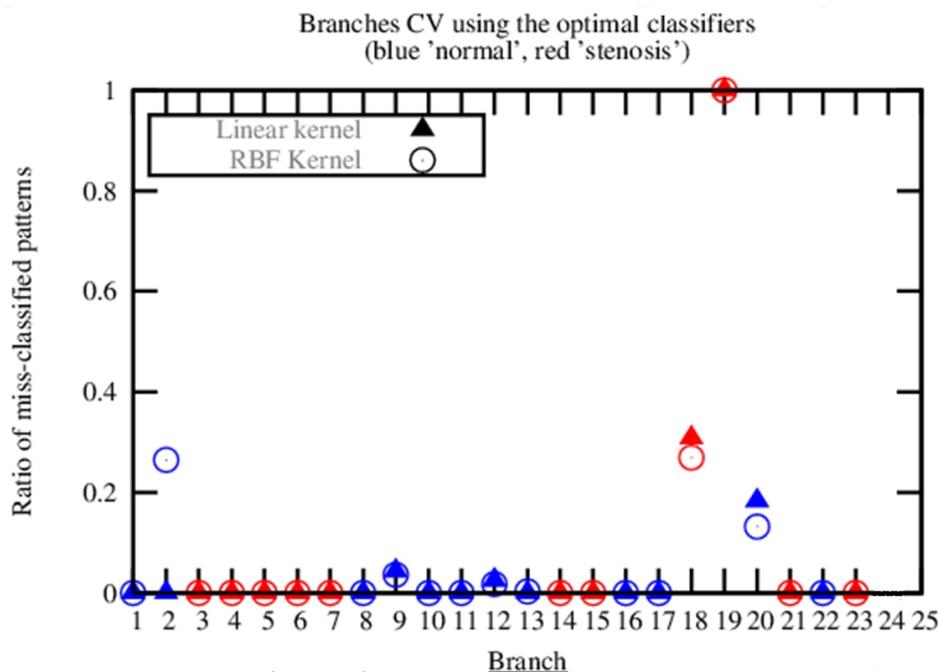


Figure 11.

Figure 12 shows the result of Platt probability fitting with the optimal RBF SVM. Of course, the test branch was removed from the training data. There are two stenoses in the branch, and they were detected with high probability.

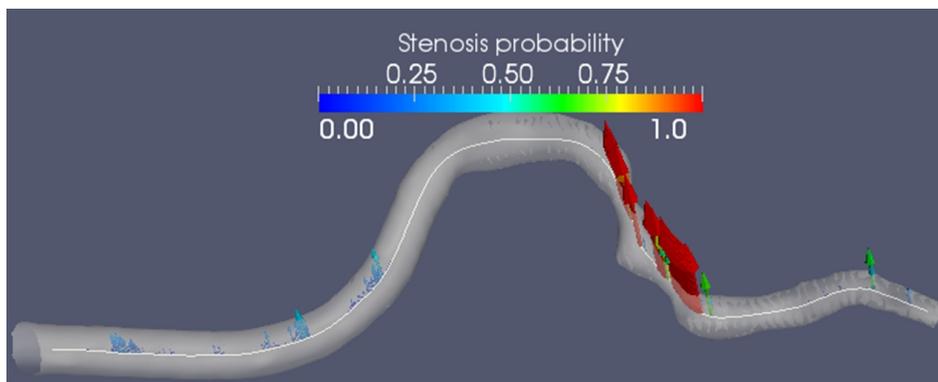


Figure 12. Testing the optimal RBF classifier with Platt probability

4. Conclusions and discussion

We presented novel features for stenosis detection. These features are independent of the vessel size and orientation. They are designed to capture the rapid changes in the vascular wall. Therefore, in principle, they can be used to detect other types of the vascular wall changes like aneurysms. In this paper, we investigated the features with stenoses only.

An SVM classifier was used to separate the healthy from the stenotic feature vectors. The Sammon map of the feature space and the high classification performance both showed that the stenotic and the healthy feature vectors are well separated.

Sometimes the narrowing of the vessel could be too long. In an extreme case, it could be as long as the whole branch. Such cases are not rapid surface changes. Therefore, they cannot be detected with the stenosis features. [Endress et al., 2011] suggested using the junction exponent of Murray’s law to detect such cases. We solved Murray’s equation for the junction exponent as an optimization problem. We proved that the optimization problem and the equation are equivalent only when the diameters configuration is not degenerate. By a degenerate diameters configuration we mean that at least one minor diameter is larger than the main diameter. This could be the case, for example, if the main branch contains a long narrowing, or if a minor branch contains an aneurysm.

We discussed the necessary and the sufficient conditions for the existence of an optimal junction exponent. Basically three cases were found and proved: First, if the case is not degenerate, then there is an existent and unique solution of the equation. Second, if the case is degenerate but the geometric mean of the minor diameters is still less than the main diameter. In this case, there is no solution of Murray’s equation, but there is an existent and unique solution of the optimization problem. Third, if the case is degenerate with the geometric mean of the minor branch diameters larger than the main branch diameter. In this case, there is no solution, i.e. no optimal junction exponent exists.

The stenosis features and the theoretical analysis of the junction exponent were

both developed with the following system in mind. A system which takes a cerebral 3D DSA volume, segments it, and gets the skeleton. Then two types of analysis are performed. First, it uses the junction exponent to identify the suspected bifurcation points as suggested in [Endress et al., 2011]. The junction exponent can be computed depending on the maximum, the minimum, or the mean diameters, and for each of them there is a residual. We called all these information bifurcation features. Second, another set of features are extracted at the non-bifurcational points and fed into an SVM classifier to detect rapid pathological changes like stenoses and aneurysms. Since the validation was performed on stenoses, we called this set of features stenosis features.

This paper added two contributions to the previous system. Further researches will concentrate on classification in the bifurcation features, investigating the automatic skeletonization, and performing the segmentation with higher degree of automation.

Acknowledgments

We highly appreciate the cooperation of Mr. Koch and Mr. Trini from Siemens AG Healthcare Sector, Imaging & Therapy Division, Forchheim, Germany. This work was partially financed by the ministry of higher education, Syria.

References

- [Abe, 2010] Abe, S. (2010). *Support Vector Machines for Patterns Classification*. Advances in Computer Vision and Pattern Recognition. Springer, second edition.
- [Antiga, 2002] Antiga, L. (2002). *Patient-Specific Modeling of Geometry and Blood-Flow in Large Arteries*. PhD thesis, POLITECNICO DI MILANO.
- [Aylward et al., 1996] Aylward, S., Bullitt, E., Pizer, S., and Eberly, D. (1996). Intensity ridge and widths for tubular object segmentation and description. In *Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, MMBIA '96, pages 131–, Washington, DC, USA. IEEE Computer Society.
- [Beucher and Lantuéjoul, 1979] Beucher, S. and Lantuéjoul, C. (1979). Use of watersheds in contour detection. In *International Workshop on image processing, real-time edge and motion detection-estimation*, pages 17–21.
- [Blum et al., 1967] Blum, H. et al. (1967). A transformation for extracting new descriptors of shape. *Models for the perception of speech and visual form*, 19(5):362–380.
- [Borgefors, 1986] Borgefors, G. (1986). Distance transformations in digital images. *Computer vision, graphics, and image processing*, 34(3):344–371.
- [Boser et al., 1992] Boser, B., Guyon, I., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM.
- [Christodoulou et al., 2003] Christodoulou, C., Pattichis, C., Pantziaris, M., and Nicolaides, A. (2003). Texture-based classification of atherosclerotic carotid plaques. *Medical Imaging, IEEE Transactions on*, 22(7):902–912.
- [Endress et al., 2011] Endress, A., Lautenschlaeger, S., Pickert, N., and Trini, M. (2011). Patent pending with the application number de102010042387: Auswertungsverfahren für ein gefäßsystem beschreibenden bilddatensatz.

- [Frangi et al., 1998] Frangi, R. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (1998). Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention*, pages 130–137.
- [Goodenday et al., 1997] Goodenday, L., Cios, K., and Shin, I. (1997). Identifying coronary stenosis using an image-recognition neural network. *Engineering in Medicine and Biology Magazine, IEEE*, 16(5):139–144.
- [Hamel, 2009] Hamel, L. (2009). *Knowledge Discovery with Support Vector Machines*. Wiley.
- [Heijmans, 1994] Heijmans, H. J. A. M. (1994). *Morphological Image Operators*. Academic Press, Boston.
- [Hsu et al., 2003] Hsu, C., Chang, C., Lin, C., et al. (2003). A practical guide to support vector classification.
- [Jain et al., 2005] Jain, A., Nandakumar, K., and Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern recognition*, 38(12):2270–2285.
- [Kaimovitz et al., 2008] Kaimovitz, B., Huo, Y., Lanir, Y., and Kassab, G. (2008). Diameter asymmetry of porcine coronary arterial trees: structural and functional implications. *American Journal of Physiology-Heart and Circulatory Physiology*, 294(2):H714–H723.
- [Keerthi and Lin, 2003] Keerthi, S. and Lin, C. (2003). Asymptotic behaviors of support vector machines with gaussian kernel. *Neural computation*, 15(7):1667–1689.
- [Lam et al., 1992] Lam, L., Lee, S., and Suen, C. (1992). Thinning methodologies-a comprehensive survey. *IEEE Transactions on pattern analysis and machine intelligence*, 14(9):869–885.
- [Lantuéjoul, 1978] Lantuéjoul, C. (1978). *La squelettisation et son application aux mesures topologiques de mosaïques polycristallines*. Dissertation, Centre de Morphologie Mathématique, Ecole des Mines, France, Paris.
- [Lantuéjoul and Maisonneuve, 1984] Lantuéjoul, C. and Maisonneuve, F. (1984). Geodesic methods in quantitative image analysis. *Pattern Recognition*, 17(2):177–187.
- [Lee et al., 2001] Lee, K., Gunn, S., Harris, C., and Reed, P. (2001). Classification of unbalanced data with transparent kernels. In *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, volume 4, pages 2410–2415. IEEE.
- [Lee et al., 1994] Lee, T., Kashyap, R., and Chu, C. (1994). Building skeleton models via 3-d medial surface/axis thinning algorithms. *CVGIP: Graphical Model and Image Processing*, 56(6):462–478.
- [Lesage et al., 2009] Lesage, D., Angelini, E., Bloch, I., and Funka-Lea, G. (2009). A review of 3d vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis*, 13(6):819–845.
- [Lin and Lin, 2003] Lin, H. and Lin, C. (2003). A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods. *submitted to Neural Computation*.
- [Lorensen and Cline, 1987] Lorensen, W. and Cline, H. (1987). Marching cubes: A high resolution 3d surface construction algorithm. In *ACM Siggraph Computer Graphics*, volume 21, pages 163–169.
- [Lorenz et al., 1997] Lorenz, C., Carlsen, I.-C., Buzug, T. M., Fassnacht, C., and Weese, J. (1997). Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2d and 3d medical images. In *CVRMed*, pages 233–242.
- [Makram-Ebeid, 1999] Makram-Ebeid, S. (1999). Digital image processing method for automatic detection of stenoses. US Patent 5,872,861.
- [Mittal et al., 2010] Mittal, S., Zheng, Y., Georgescu, B., Vega-Higuera, F., Zhou, S., Meer, P., and Comaniciu, D. (2010). Fast automatic detection of calcified coronary lesions in 3d cardiac ct images. *Machine Learning in Medical Imaging*, pages 1–9.
- [Olendorf et al., 2002] Olendorf, D., Jeryan, C., and Boyden, K., editors (2002). *The GALE Encyclopedia of Medicine*. Gale Group, 2 edition.
- [Osher and Sethian, 1988] Osher, S. and Sethian, J. (1988). Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49.
- [Pavlopoulos et al., 2004] Pavlopoulos, S., Stasis, A., and Loukis, E. (2004). A decision tree-based

- method for the differential diagnosis of aortic stenosis from mitral regurgitation using heart sounds. *Biomed Eng Online*, 3(1):21.
- [Platt, 1999] Platt, J. (1999). Probabilistic outputs for support vector machines. *Bartlett P. Schoelkopf B. Schurmans D. Smola, AJ, editor, Advances in Large Margin Classifiers*, pages 61–74.
- [Sammon, 1969] Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Trans. Comput.*, 18:401–409.
- [Scholkopf and Smola, 2001] Scholkopf, B. and Smola, A. J., editors (2001). *Learning With Kernels*. MIT Press.
- [Serra, 1988] Serra, J. (1988). *Image Analysis and Mathematical Morphology: Theoretical Advances*, volume 2. Academic Press, London.
- [Sherman, 1981] Sherman, T. F. (1981). On connecting large vessels to small. The meaning of Murray’s law. *The Journal of General Physiology*, 78(4):431–453.
- [Smola and Bartlett, 2000] Smola, A. J. and Bartlett, P. J., editors (2000). *Advances in Large Margin Classifiers*. MIT Press, Cambridge, MA, USA.
- [Taubin, 1995] Taubin, G. (1995). Curve and surface smoothing without shrinkage. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 852–857.
- [TeBmann et al., 2008] TeBmann, M., Vega-Higuera, F., Fritz, D., Scheuring, M., and Greiner, G. (2008). Learning-based detection of stenotic lesions in coronary ct data. *Vision, modeling, and visualization 2008: proceedings, October 8-10, 2008, Konstanz, Germany*, page 189.
- [Veropoulos et al., 1999] Veropoulos, K., Campbell, C., and Cristianini, N. (1999). Controlling the sensitivity of support vector machines. In *Proceedings of the international joint conference on AI*. Citeseer.
- [Zeng, 2011] Zeng, G. (2011). Apparatus for stenosis estimation. US Patent 20,110,081,057.
- [Zhang et al., 2002] Zhang, Y., Brady, M., and Smith, S. (2002). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *Medical Imaging, IEEE Transactions on*, 20(1):45–57.