

Research



Cite this article: Zhao X, Gold N, Fang Y, Xu S, Zhang Y, Liu J, Gupta A, Huang H. 2018 Vertebral artery fusiform aneurysm geometry in predicting rupture risk. *R. Soc. open sci.* **5**: 180780. <http://dx.doi.org/10.1098/rsos.180780>

Received: 17 May 2018
Accepted: 9 October 2018

Subject Category:

Mathematics

Subject Areas:

biomedical engineering/biomathematics/medical computing

Keywords:

aneurysm geometry, rupture risk prediction, machine learning

Authors for correspondence:

Yibin Fang
e-mail: fangyibin@vip.163.com
Huaxiong Huang
e-mail: hhuang@fields.utoronto.ca

Vertebral artery fusiform aneurysm geometry in predicting rupture risk

Xiukun Zhao^{1,2}, Nathan Gold^{1,3}, Yibin Fang^{2,4},
Shixin Xu^{1,2}, Yongxin Zhang⁴, Jianmin Liu⁴,
Arvind Gupta^{1,2,5} and Huaxiong Huang^{1,2,3}

¹Centre for Quantitative Analysis and Modelling (CQAM), The Fields Institute, Toronto, Ontario M5T 3J1, Canada

²The Fields Institute for Research in Mathematical Sciences, Toronto, Ontario M5T 3J1, Canada

³Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada

⁴Department of Neurosurgery, Changhai Hospital, Second Military Medical University, Shanghai, China

⁵Department of Computer Science, University of Toronto, Toronto, Ontario M5T 3J1, Canada

XZ, 0000-0001-5624-4926; HH, 0000-0002-1266-9428

Cerebral aneurysms affect a significant portion of the adult population worldwide. Despite significant progress, the development of robust techniques to evaluate the risk of aneurysm rupture remains a critical challenge. We hypothesize that vertebral artery fusiform aneurysm (VAFA) morphology may be predictive of rupture risk and can serve as a deciding factor in clinical management. To investigate the VAFA morphology, we use a combination of image analysis and machine learning techniques to study a geometric feature set computed from a depository of 37 (12 ruptured and 25 un-ruptured) aneurysm images. Of the 571 unique features we compute, we distinguish five features for use by our machine learning classification algorithm by an analysis of statistical significance. These machine learning methods achieve state-of-the-art classification performance ($81.43 \pm 13.08\%$) for the VAFA morphology, and identify five features (cross-sectional area change of aneurysm, maximum diameter of nearby distal vessel, solidity of aneurysm, maximum curvature of nearby distal vessel, and ratio of curvature between aneurysm and its nearby proximal vessel) as effective predictors of VAFA rupture risk. These results suggest that the geometric features of VAFA morphology may serve as useful non-invasive indicators for the prediction of aneurysm rupture risk in surgical settings.

1. Introduction

Aneurysms, pathological dilation of blood vessel and weakening of the vessel wall, affect a significant portion of the adult population [1–3]. There exists an extensive literature with a large focus on possible correlations between the occurrence of aneurysms and hydrodynamic factors [4–6]. The findings are, however, inconsistent and sometimes conflicting. For example, high wall stress, low wall stress, low pressure, turbulence and flow instability have been identified as possible causes [7,8]. This is probably due to the fact that aneurysms are caused by multiple factors, with the exact cause not fully understood. Left untreated, some aneurysms may grow and rupture, causing uncontrollable haemorrhage. On the other hand, treatment also carries a certain risk, sometimes also causing undesirable consequences. Therefore, it is a life-and-death decision whether and what surgical intervention should be performed. To help surgical decision-making, it is desirable if certain features of aneurysms correlated to the likelihood of rupture can be identified.

There are two main types of aneurysms, saccular and fusiform, characterized by distinct morphologies. In saccular aneurysms, the contour/circumference of aorta remains intact, and is mostly uninvolved, with an eccentricity that involves only a part of the vessel wall contour. Fusiform aneurysms, conversely, result in complete distortion of the aortic contour, with a concentric formation along the vessel. Saccular aneurysms occur in the descending thoracic aorta and other locations. Such aneurysms are considered more dangerous since they tend to rupture well below the critical diameters of fusiform recommended for surgery [9]. On the other hand, an intracranial fusiform aneurysm is often associated with ischaemia, mass effect, or bleeding [10,11]. Rupture of fusiform aneurysms, especially those located in the posterior circulation, is often lethal, despite aggressive treatment. Endovascular treatment has been the primary method for vertebral artery fusiform aneurysms (VAFA); however, the risk of treatment cannot be ignored, especially when vital branches are involved [12]. As such, the evaluation of lesion rupture risk and personalized treatment plans for fusiform aneurysms represents an important clinical goal.

Haemodynamic and morphological studies of saccular aneurysms have demonstrated the value of haemodynamic and morphological evaluation in predicting the rupture of the aneurysms [13–15]. However, less work has been done on fusiform aneurysms, possibly due to their more complex morphology. It has been recognized that the morphological characteristics of fusiform aneurysms may play an important role in evaluating rupture tendency [16]. However, manual measurement and assessment on the basis of three-dimensional (3D) reconstruction are time-consuming and prone to error.

Recent advances in machine learning techniques have shown promise that these types of approaches can be highly effective in medical research [17]. Currently, the application of machine learning in the morphological study of cerebrovascular disease is limited. This study represents one of the first attempts to explore the utility of machine learning techniques to evaluate VAFA morphology for the prediction of rupture risk, as well as to assist in decision-making for clinical management of this disease.

2. Methods

2.1. Patients and image data

A total of 37 patients (23 males, 14 females, mean age = 52.43 years, s.d. = 10.12) with 12 ruptured and 25 un-ruptured aneurysms were included in the study.

All 3D vertebral artery aneurysm images are stored in stereolithography (STL) format files. Each file consists of a triangulated surface of the three-dimensional blood vessel, from which vertex information of the triangles was obtained.

2.2. Geometric characteristics

Many observational studies have been conducted to assess the predictors of aneurysm rupture and to guide physicians in decision-making [18–21]. Generally, predictors are categorized into geometric, haemodynamic and clinical characteristics. In this paper, we focus on the relationship between the rupture risk of an aneurysm and geometric properties of an aneurysm and its corresponding parent artery.

The open source software MeshLab [22] was used for processing and editing initial 3D triangular meshes, including segmentation, noise removal and the surface smoothing process. After the initial segmentation process, every 3D image was simplified to a region containing the aneurysm and their

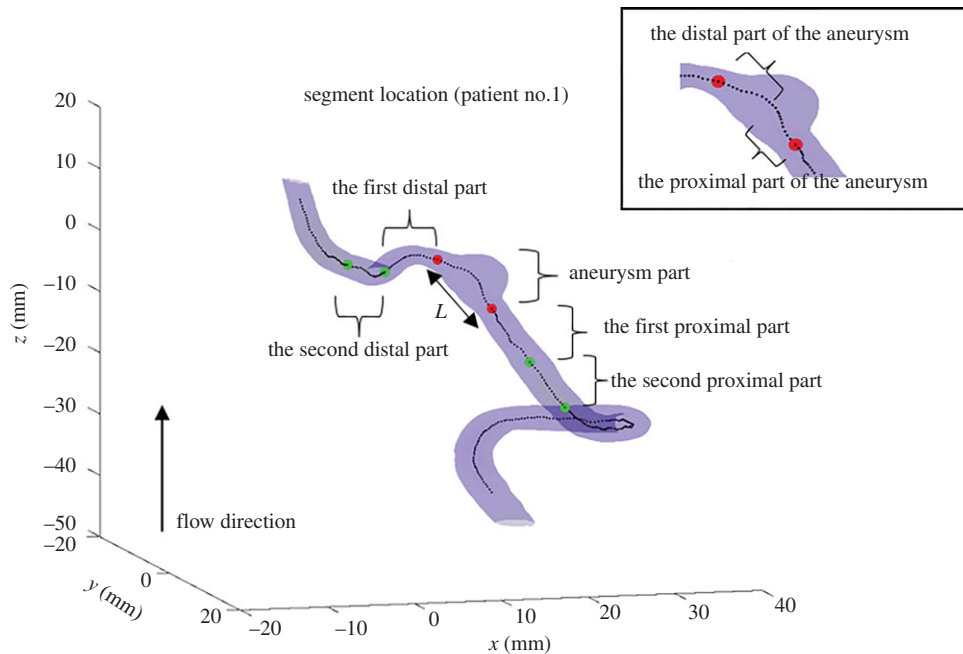


Figure 1. Segment location. Five segments, including the aneurysm part, the first proximal part, the second proximal part, the first distal part, and the second distal part, have the same length L (mm). The aneurysm part also consists of the proximal and distal segments, which is shown in the top right corner.

parent artery. The refined meshes with removing noise and smoothing surface were exported from MeshLab to be used for further analysis via Matlab [23].

To calculate the geometric properties of aneurysms, the 3D images were first voxelized. Based on the meshes forming the subject's surface, a 3D binary matrix was generated, where the voxels had 3D location information.

The centreline representation of the vessel is simpler than the volume or surface rendering while preserving the topology of the whole vessel trees [24,25]. Thus, we extracted the centrelines of blood vessels and calculated geometric indices based on the centrelines. The 3D medial surface thinning algorithm was used to find the skeleton of the 3D blood vessels [26,27]. In order to obtain a complete centreline, we queued the discrete points that formed the skeleton according to their locations, and then connected the data points by the 3D curve fitting algorithm.

By the definitions of the tangent unit vector, the normal unit vector and the binormal unit vector, together with the Frenet–Serret formulae, we calculated the curvature and the torsion of the centreline. In addition, the cross-sectional profiles and the eight corresponding geometric variables were obtained based on the central path. The geometric indices include maximum, minimum and equivalent diameters (the diameter of a circle with the same area as the region), cross-sectional area, area change rate, eccentricity, solidity (the proportion of the pixels in the convex hull) and extent (the ratio of pixels in the region to pixels in the total bounding box). A detailed explanation of these geometric indices is given in appendix A.

The geometric indices are ten sequences with respect to the centreline location. To analyse these geometric variables more conveniently, we calculated statistical features based on them and used these computed features for the statistical analysis and machine learning classification. Figure 1 shows an aneurysm and its parent artery of patient no. 1. We hypothesized that the aneurysmal shape and its nearby region affected the rupture risk significantly. Hence, we chose five segments with the equal length L (mm) as the feature extraction region shown in figure 1, where L is the length of an aneurysm, determined based on the cross-sectional area. Furthermore, in order to analyse the aneurysmal shape, the aneurysm part was further divided into two segments for the feature calculation. At these seven segments, the maximal value, mean, standard deviation, integral and variation of the ten geometric value sequences were calculated respectively. We also took the ratios between the values of two segments of the aneurysm part, and the ratios between the aneurysm part, the first distal part and the first proximal part.

Besides the previous features, some features related to the aneurysmal size and shape were considered to be important as well, such as the maximal diameter and the length of the aneurysm,

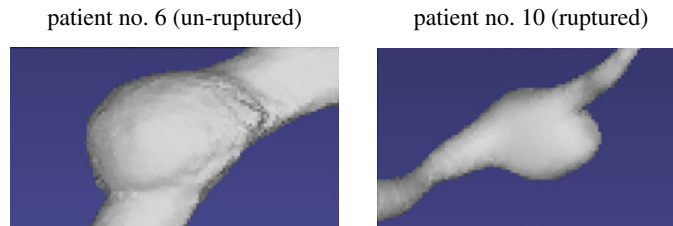


Figure 2. Comparison of aneurysmal shapes. Patient no. 6 is an un-ruptured case shown in the first picture, and the second picture includes a ruptured case. The proximal blood vessels are at the bottom of the pictures.

width–length ratio, the angle and the distance between two centreline tangent vectors at the necks of the bulge (refer to figure 6 in the appendix A), and the asymmetry factor. A complete feature list is presented in appendix A.

We computed 571 geometric features in total. The feature set is so large compared to the dataset that over-fitting problem will happen when classifying the aneurysm rupture risk. Therefore, the next step is to analyse and select features by statistical analysis.

2.3. Machine learning and statistical analysis

Statistical analysis and machine learning prediction were performed using Matlab. We applied *t*-tests for the geometric features, with a null hypothesis of no interaction between features when separating the classified groups. The *t*-statistics were compared for each feature as a measure of how effective it was at separating groups. Based on the *t*-test, 20 features with the largest *t*-statistics were selected. Feature selection based on their individual ranking may also contain redundant information, so not all features are required. We then checked the correlation of these features and removed 15 redundant features if their Pearson correlation coefficients were above a threshold of 0.5.

After the feature selection process, four typical machine learning algorithms: random forest (RF) [28], support vector machine (SVM) [29], *k*-nearest neighbours (KNN) [30] and subspace discriminant (SD) [31], were applied for classifying the ruptured and un-ruptured cases. A brief description of the machine learning algorithms is presented in appendix B.

Given our relatively small sample size, we used a procedure commonly referred to as a leave-one-out cross-validation to maximize the use of our dataset. This means that of our total sample of 37 patients, we set aside one case to use as the testing set and the remaining 36 cases to use as the training set. The procedure was repeated 37 times, each time using a different case as the test set. This procedure was used for assessing the performance of the four models and therefore choosing the best prediction model.

To validate the classification model, another resampling method train-test split was used for evaluating the performance of the chosen optimal model. Specifically, the whole dataset was randomly divided into a training set (80%) and a test set (20%). The chosen model was trained on the training set and applied to predict the data on the test set. To ensure our model's performance was not biased by a particular data partition, we randomly reassigned the cases into new training and test datasets and repeated the machine learning process 30 times.

3. Results

3.1. Geometric feature analysis

The highest ranked indices from the feature selection procedure are as follows: the ratio of cross-sectional areas between the proximal and distal parts of the aneurysm ($t_{37} = 2.6469, p = 0.0129$), the total maximum diameter at the nearby distal part ($t_{37} = 2.1634, p = 0.0384$), the total solidity of the aneurysm ($t_{37} = 2.0398, p = 0.0497$), the maximum centreline curvature at the nearby distal part ($t_{37} = 2.0191, p = 0.0519$), and the ratio of centreline curvature between the aneurysm and the nearby proximal parts ($t_{37} = 2.0107, p = 0.0531$). We set a *p*-value < 0.05 as the criterion for statistical significance.

Table 1 shows the relationship between the above five geometric characteristics and rupture in VAFA. The ratio of cross-sectional areas between the proximal and distal parts of an aneurysm represents the shape of an aneurysm. From this table, we find that most of the ruptured aneurysms have smaller ratio values than the un-ruptured cases, meaning that if the proximal part of an aneurysm is smaller than its

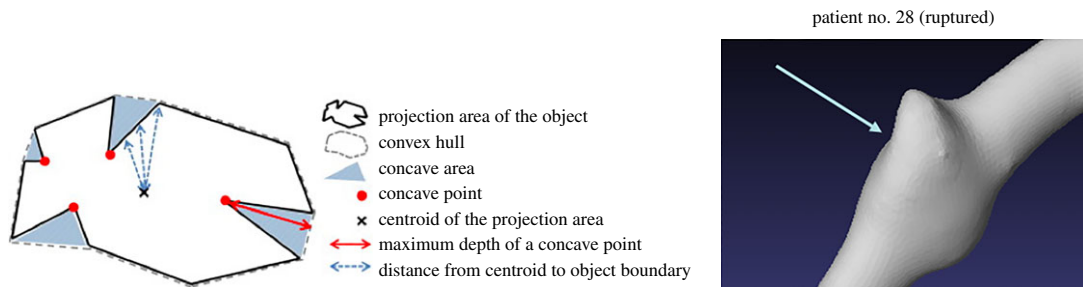


Figure 3. Solidity of an aneurysm. The blue area is the concave area on the left picture [32]. Patient no. 28 is a ruptured case shown on the right, which has a relatively large concave area. The arrow points out the concave region of the aneurysm.

Table 1. The Relationship between geometric characteristics and rupture in VAFA ($\bar{x} \pm$ s.d.).

geometric characteristic	rupture	un-rupture	<i>t</i> -stat	<i>p</i> -value
ratio of cross-sectional areas between the proximal and distal parts of the aneurysm (unitless)	0.95 ± 0.25	1.21 ± 0.34	2.6469	0.0129
total maximum diameter at the nearby distal part (mm ²)	27.55 ± 14.82	40.76 ± 21.78	2.1634	0.0384
total solidity of the aneurysm (mm)	10.35 ± 2.90	12.88 ± 4.58	2.0398	0.0497
maximum centreline curvature at the nearby distal part (mm ⁻¹)	0.29 ± 0.12	0.46 ± 0.38	2.0191	0.0519
ratio of centreline curvature between the aneurysm and the nearby proximal part (unitless)	-0.65 ± 0.91	0.72 ± 3.16	2.0107	0.0531

Table 2. Comparison of four models (leave-one-out).

machine learning algorithm	accuracy	<i>t</i> -statistic	<i>p</i> -value
SVM	81.08%		
RF	72.92%	0.8216	0.4141
KNN	75.68%	0.5583	0.5784
SD	72.97%	0.8216	0.4141

distal part, the aneurysm has a higher rupture risk. Two cases are taken as the examples to show the comparison of aneurysmal shapes in figure 2. The second geometric index shows that for most ruptured cases, the integral of the maximum diameter of the distal blood vessel near an aneurysm are usually smaller compared to un-ruptured cases. The solidity of an aneurysm is related to the regularity of vascular wall. When an aneurysm has a larger concave area (a smaller solidity value), as shown in figure 3, it is more likely to rupture. The last two features reflect the degree of curvature of the blood vessel near an aneurysm.

3.2. Classification

Based on the leave-one-out cross-validation method, the comparison results of four machine learning models are shown in table 2. The model with the lowest overall error estimate is SVM, with an average accuracy of 81.08%. To compare the models, we performed a hypothesis test on the performance difference of each model, using the SVM as the baseline. The results show that although SVM performed better than other models, the difference between model classification performance was not statistically significant. This is most likely due to the relatively small size of the data. Figure 4 depicts the receiver operating characteristics (ROC) curve produced by SVM. The area under the curve

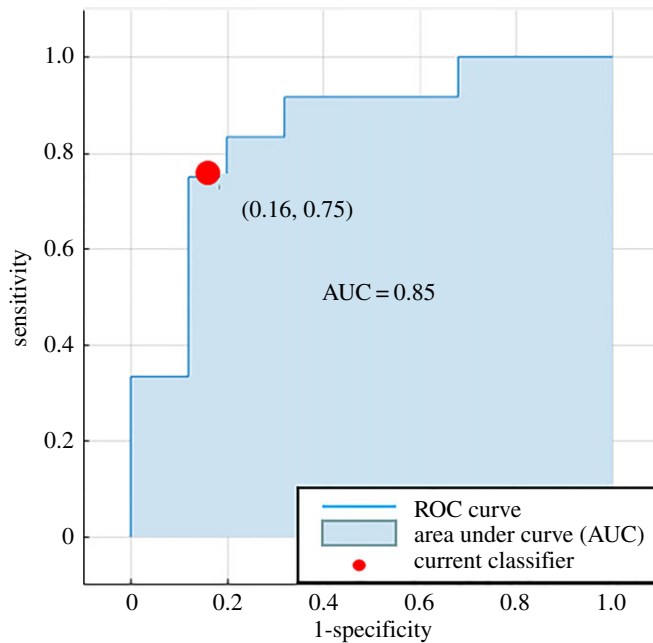


Figure 4. ROC curve produced by SVM. The blue area indicates the area under the curve (AUC). The optimal operating point on the ROC curve is denoted by the red point (0.16, 0.75).

Table 3. Comparison of four models (train-test data split).

machine learning algorithm	accuracy	<i>t</i> -statistic	<i>p</i> -value
SVM	81.43 (± 13.08) %		
RF	76.67 (± 13.78) %	1.3732	0.1750
KNN	79.05 (± 14.88) %	0.6583	0.5130
SD	79.52 (± 13.88) %	0.5471	0.5864

(AUC) is 0.85. The optimal operating point on the ROC curve shows the sensitivity of the classifier SVM is 75%, and the specificity achieves 84%.

We further evaluated the performance of the methods using the train-test data split. From table 3, we find that the performance difference between these methods was not statistically significant, but the SVM model still performed better than others. It achieved an average classification accuracy of 81.43% ($\pm 13.08\%$) on the test set.

4. Discussion

The objectives of our study were to use clinical data of the patients to evaluate a set of geometric characteristics describing the size and shape of an aneurysm and nearby blood vessel information and to apply a collection of machine learning models to discriminate ruptured and un-ruptured aneurysms. The rapid pace at which medical image data are being generated has resulted in a gap between the collection of data and analysis for decision-making. Machine learning combined with image processing techniques can help in identifying potentially useful patterns in image data and build models that are capable of predicting rupture risk of an aneurysm based on the known patient data.

The results of this study show that some geometric indices related to the size and shape of an aneurysm and nearby blood vessel were correlated with rupture risk. The most effective features for our current data include the cross-sectional area change of an aneurysm, the total maximum diameter of the nearby distal blood vessel, the total solidity of an aneurysm, the maximum curvature of the nearby distal blood vessel, and the ratio of curvature between an aneurysm and the nearby proximal blood vessel. These geometric variables are suitable for further study with more image data.

However, some geometric indices such as length, width and width–length ratio of the aneurysm, which are effective for distinguishing saccular aneurysms, we deemed to be not statistically significant for predicting the rupture risk of VAFA ($t_{37} = 1.7301$ ($p = 0.0943$), $t_{37} = 0.6647$ ($p = 0.5123$) and $t_{37} = 0.9036$ ($p = 0.3742$), respectively).

Due to the complexity of fusiform aneurysm morphology, manual measurement of geometric indices from 3D images is a difficult and time-consuming task. In our study, we extracted geometric information automatically by image processing techniques, which have the benefit of being much faster and with improved accuracy as compared to manual measurements. For the current dataset, the overall classification accuracy achieved 81%, much higher than chance (50%). The performance including sensitivity and specificity of the SVM classifier also proved that machine learning is an efficient method to differentiate ruptured and un-ruptured cases.

Although machine learning is transforming modern medicine and was proved to be useful for our study, we still need to consider its limitations. First, machine learning learns through historical data. The bigger the data and the longer it is exposed to these data, the better it will perform. Our current dataset is relatively small, so the geometric characteristics and the machine learning model need to be further validated on a larger dataset. Second, machine learning systems cannot always provide rational reasons for a particular prediction or decision. Thus, human collaboration is necessary to better evaluate the outputs of these systems.

5. Conclusion

In this paper, we have brought image analysis, statistical analysis and machine learning techniques together to study VAFA morphology, thereby predicting the rupture risk of VAFA. This method extracted multiple geometric characteristics automatically and accurately from the complex morphology of VAFA. Hypothesis testing and machine learning models further validated that some of these geometric indices were effective and meaningful predictors for the rupture risk of VAFA. Our results suggest the popular SVM machine learning classification technique can be used to classify ruptured and un-ruptured aneurysms successfully with high accuracy. Therefore, we conclude that studying the relationship between geometric characteristics and rupture in VAFA and thereby predicting the rupture risk via machine learning techniques is a promising research direction.

The application of machine learning algorithms to 3D medical image analysis is still in its infancy. However, machine learning combined with geometric feature extraction as a non-invasive method has great potential as an early predictor of the rupture risk, and ultimately provides a useful tool for the personalized treatment of VAFA. In addition, our approach can be extended further for studying 3D medical images of the other type of aneurysms and serve as a useful risk predictor for researchers and surgeons.

Ethics. Ethical approval for the retrospective study was obtained from the institutional review board of Changhai Hospital, Shanghai, China. The requirement for informed consent was waived by the review board due to the retrospective design of the study.

Data accessibility. Data supporting this paper are openly available at the GitHub page: https://github.com/ngold5/aneurysm_classification.

Authors' contributions. X.Z. carried out the image and statistical analysis, participated in the design of the study and drafted the manuscript; N.G. participated in statistical analysis and helped draft the manuscript; Y.F. collected image data, participated in image analysis and study design and helped draft the manuscript; S.X. participated in image analysis and commented on the manuscript; Y.Z. and J.L. participated in data collection; A.G. and H.H. designed the study, coordinated the study and commented on the manuscript. All authors gave final approval for publication.

Competing interests. The authors declare no competing interests.

Funding. This research is supported in part by the Fields Institute for Research in Mathematical Science (X.Z., S.X., H.H.), the Natural Sciences and Engineering Research Council of Canada (N.G., H.H.), China Scholarship Council No. 201603170357 (Y.F.), National Natural Science Foundation of China 81301004 (Y.F.), UBC (A.G.), the University of Toronto (A.G.), and the Centre for Quantitative Analysis and Modelling (X.Z., N.G., S.X., A.G., H.H.).

Appendix A. Definition of geometric features

By the definitions (A.1–A.3) of the tangent unit vector T , the normal unit vector N and the binormal unit vector B , together with the Frenet–Serret formulae (A.4), we have the curvature κ and the torsion τ of the

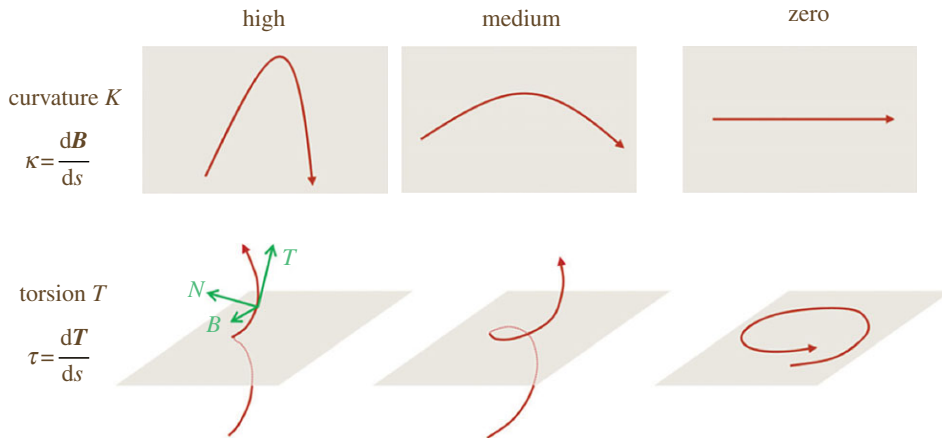


Figure 5. Curvature and torsion. The first row shows different curvature, and the second row shows different torsion [33].

centreline. To reduce the influence of the unsmooth centreline, the sliding window strategy was also involved to calculate the curvature and the torsion.

$$T = \frac{dr/ds}{\|dr/ds\|}, \quad (\text{A.1})$$

where r is a curve in Euclidean space, and s represents the arc length.

$$N = \frac{dT/ds}{\|dT/ds\|}, \quad (\text{A.2})$$

$$B = T \times N \quad (\text{A.3})$$

and

$$\begin{bmatrix} T' \\ N' \\ B' \end{bmatrix} = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{bmatrix} \begin{bmatrix} T \\ N \\ B \end{bmatrix}. \quad (\text{A.4})$$

Based on the curvature and torsion of the centreline, we found the central path and therefore calculated another eight geometric variables. The concrete explanation of these 10 geometric indices are shown as follows.

A.1. Centreline curvature κ (mm^{-1})

Curvature κ represents the rate of change of the angle through which the tangent to a curve turns in moving along the curve. The curvature κ is the solution of the Frenet–Serret formulae (A.4).

A.2. Centreline torsion τ (mm^{-1})

In three dimensions the torsion τ of a curve measures how sharply it is twisting out of the plane of curvature. The torsion τ is solved by the Frenet–Serret formulae (A.4). Figure 5 shows some examples of curvature and torsion.

A.3. Cross-sectional area A (mm^2)

The cross-sectional area A is the area of a two-dimensional shape that is obtained when a three-dimensional object. Along the centreline, a series of slices containing the cross-sectional profiles are obtained. The cross-sectional area A is calculated by the number of pixels in the region.

A.4. Rate of change of cross-sectional area ACR (mm)

The rate of change of cross-sectional area ACR is the change of the cross-sectional area with respect to the position change at the centreline r .

$$\text{ACR} = \frac{dA}{dr}.$$

A.5. Maximal diameter $MaxD$ (mm)

The maximal diameter $MaxD$ is the maximal distance between the centreline and the blood vessel surface.

A.6. Minimal diameter $MinD$ (mm)

The minimal diameter $MinD$ is the minimal distance between the centreline and the blood vessel surface.

A.7. Equivalent diameter $EquivD$ (mm)

Equivalent diameter $EquivD$ is the diameter of a circle with the same area as the region.

$$EquivD = \sqrt{\frac{4A}{\pi}}$$

A.8. Eccentricity e (unitless)

The eccentricity e is the ratio of the distance between the foci f of the ellipse and its major axis length $MaxD$. The value is between 0 and 1. An ellipse whose eccentricity is 0 is a circle, while an ellipse whose eccentricity is 1 is a line segment.

$$e = \frac{f}{MaxD}$$

A.9. Solidity SLD (unitless)

The solidity SLD is the proportion of the pixels in the convex hull.

$$SLD = \frac{A}{A_{convex}}$$

where A_{convex} is the area of the smallest convex polygon that can contain the region.

A.10. Extent EXT (unitless)

The extent EXT is the ratio of pixels in the region to pixels in the total bounding box.

$$EXT = \frac{A}{A_{bb}}$$

where A_{bb} is the area of the bounding box that can contain the region.

In addition to the above features, some geometric indices are important as well. The maximal diameter D_{max} (mm), the length L (mm), the proximal neck diameter $D_{neck,p}$ (mm), the distal neck diameter $D_{neck,d}$ (mm) of the aneurysm, and the ratios between these four factors and the normal diameter of the parent vessel D_{normal} (mm) were used to represent the aneurysmal size.

To describe the aneurysmal shape, we introduced an imaginary curve to represent the centreline of the normal vessel at the aneurysm region and compared it with the centreline of an aneurysm since there was no information on the changes of morphometric characteristics regarding an aneurysm. The hobby curve, a spline interpolation algorithm, was applied to connect the centrelines at the necks of the bulge [34].

Figure 6 depicts the real centreline of an aneurysm and the supposed centreline of the normal blood vessel. The distance between the real and supposed curves was calculated, and the maximum value, the accumulated value, the average, the standard deviation and the variation of the distance were taken as the geometric features. The angle and the distance between two centreline tangent vectors at the necks of the bulge, shown in figure 6, and the asymmetry factor $(1 - (d/D_{max}))$ are also the features for the further analysis.

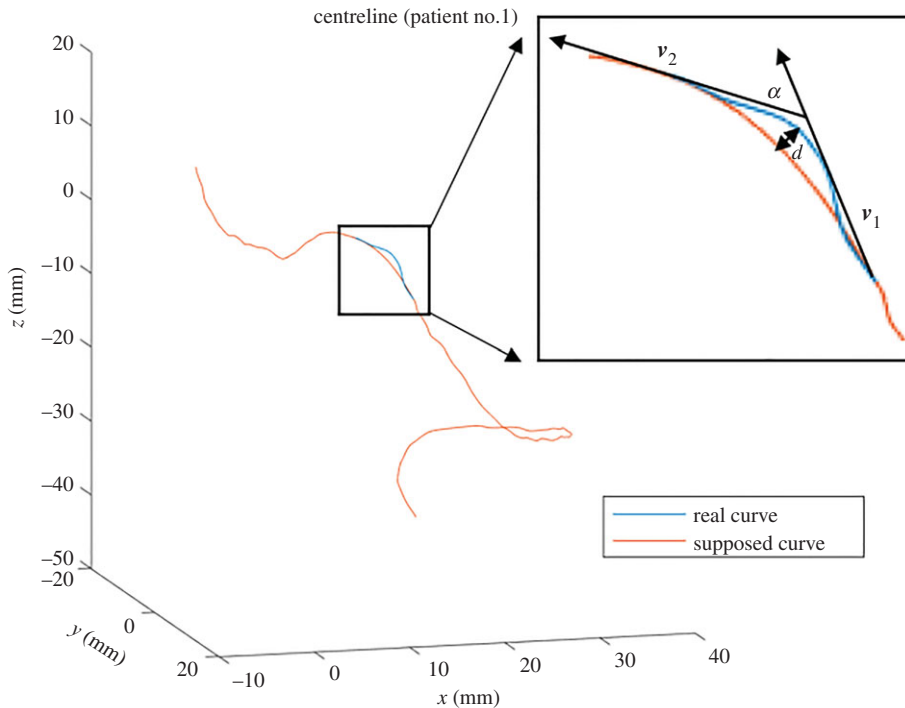


Figure 6. Centreline. Real curve is the centreline of an aneurysm, and the supposed curve is the centreline of the normal blood vessel. The aneurysm part is zoomed in and shown in the small black box. In the black box, the two rays are centreline tangent vectors at the necks of the bulge, and the maximum distance of two curves is denoted by d .

Appendix B. Brief description of machine learning algorithms

B.1. Random forest

Random forest (RF) [28] is an ensemble learning method for classification, that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees. Random forest corrects for decision trees' habit of over-fitting to their training set.

B.2. Support vector machine

Support vector machine (SVM) [29] classifies data by finding the best hyperplane that separates data points of one class from those of the other class. The best hyperplane for an SVM is defined as the hyperplane with the largest margin between the two classes; that is the maximal width of the slab parallel to the hyperplane that has no interior data points.

B.3. k -Nearest neighbours

The k -nearest neighbours algorithm (KNN) [30] is a non-parametric method used for classification. Given a set X of n points and a distance function, k -nearest neighbours search finds the k closest points in X to a query point or set of points. Various metrics can be used to determine the distance. KNN-based algorithms are widely used as benchmark machine learning rules.

B.4. Subspace discriminant

Subspace discriminant (SD) [31] is a high-quality ensemble model composed of a combination of multiple linear discriminant classifiers, where a random subset of the predictors is for each learner. The linear discriminant analysis assumes that different classes generate data based on different Gaussian distributions [31]. To train a classifier, the fitting function estimates the parameters of a Gaussian distribution for each class.

1. Etmnan N, Rinkel GJ. 2016 Unruptured intracranial aneurysms: development, rupture and preventive management. *Nat. Rev. Neurol.* **12**, 699–713. (doi:10.1038/nrneurol.2016.150)
2. Brown Jr RD, Broderick JP. 2014 Unruptured intracranial aneurysms: epidemiology, natural history, management options, and familial screening. *Lancet Neurol.* **13**, 393–404. (doi:10.1016/S1474-4422(14)70015-8)
3. Hao W, Gong S, Wu S, Xu J, Go M, Friedman A, Zhu D. 2017 A mathematical model of aortic aneurysm formation. *PLoS ONE* **12**, e0170807. (doi:10.1371/journal.pone.0170807)
4. Wang G, Liu L, Wen L, Cao Y, Pei Y, Zhang D. 2017 Morphological characteristics associated with rupture risk of multiple intracranial aneurysms. *Asian Pac. J. Trop. Med.* **10**, 1011–1014. (doi:10.1016/j.apjtm.2017.09.015)
5. Holman E, Peniston W. 1955 Hydrodynamic factors in the production of aneurysms. *Am. J. Surg.* **90**, 200–209.
6. Parshin DV, Kuianova IO, Yunoshev AS, Ovsyannikov KS, Dubovoy AV. 2017 On the mechanics of cerebral aneurysms: experimental research and numerical simulation. *J. Phys. Conf. Ser.* **894**, 012071. (doi:10.1088/1742-6596/894/1/012071)
7. Xu L, Liang F, Gu L, Liu H. 2018 Flow instability detected in ruptured versus unruptured cerebral aneurysms at the internal carotid artery. *J. Biomech.* **72**, 187–199. (doi:10.1016/j.jbiomech.2018.03.014)
8. Thiriet M. 2008 *Biology and mechanics of blood flows: Part II. Mechanics and medical aspects*. New York, NY: Springer.
9. Criado FJ. 2013 Aneurysm morphology matters: fusiform vs. saccular. *J. Endovasc. Ther.* **20**, 207–209. (doi:10.1583/1545-1550-20.2.207)
10. Little JR, St Louis P, Weinstein M, Dohn DF. 1981 Giant fusiform aneurysm of the cerebral arteries. *Stroke* **12**, 183–188.
11. Drake CG, Peerless SJ. 1997 Giant fusiform intracranial aneurysms: review of 120 patients treated surgically from 1965 to 1992. *J. Neurosurg.* **87**, 141–162. (doi:10.3171/jns.1997.87.2.0141)
12. Siddiqui AH, Abula AA, Kan P, Dumont TM, Jahshan S, Britz GW, Hopkins LN, Levy EI. 2012 Panacea or problem: flow diverters in the treatment of symptomatic large or giant fusiform vertebrobasilar aneurysms. *J. Neurosurg.* **116**, 1258–1266. (doi:10.3171/2012.2.JNS111942)
13. Ujiie H *et al.* 1999 Effects of size and shape (aspect ratio) on the hemodynamics of saccular aneurysms: a possible index for surgical treatment of intracranial aneurysms. *Neurosurgery* **45**, 119–129; discussion 129–130. (doi:10.1227/00006123-199907000-00028)
14. Dhar S, Tremmel M, Mocco J, Kim M, Yamamoto J, Siddiqui AH, Hopkins LN, Meng H. 2008 Morphology parameters for intracranial aneurysm rupture risk assessment. *Neurosurgery* **63**, 185–196. (doi:10.1227/01.NEU.0000316847.64140.81)
15. Xiang J, Tutino VM, Snyder KV, Meng H. 2014 CFD: computational fluid dynamics or confounding factor dissemination? The role of hemodynamics in intracranial aneurysm rupture risk assessment. *Am. J. Neuroradiol.* **35**, 1849–1857. (doi:10.3174/ajnr.A3710)
16. Backes D, Vergouwen MD, Velthuis BK, van der Schaaf IC, Bor AS, Algra A, Rinkel GJ. 2014 Difference in aneurysm characteristics between ruptured and unruptured aneurysms in patients with multiple intracranial aneurysms. *Stroke* **45**, 1299–1303. (doi:10.1161/STROKEAHA.113.004421)
17. Mazzaferri J, Larrivee B, Cakir B, Sapieha P, Costantino S. 2018 A machine learning approach for automated assessment of retinal vasculature in the oxygen induced retinopathy model. *Sci. Rep.* **8**, 3916. (doi:10.1038/s41598-018-22251-7)
18. Shum J, Martufi G, Di Martino E, Washington CB, Grisafi J, Muluk SC, Finol EA. 2011 Quantitative assessment of abdominal aortic aneurysm geometry. *Ann. Biomed. Eng.* **39**, 277–286. (doi:10.1007/s10439-010-0175-3)
19. Zanaty M, Chalouhi N, Tjoumakaris SI, Gonzalez LF, Rosenwasser RH, Jabbour PM. 2014 Aneurysm geometry in predicting the risk of rupture: a review of the literature. *Neurol. Res.* **36**, 308–313. (doi:10.1179/1743132814Y.0000000327)
20. Kontopodis N, Metaxa E, Papaharilaou Y, Georgakarakos E, Tsetis D. 2013 Changes in geometric configuration and biomechanical parameters of a rapidly growing abdominal aortic aneurysm may provide insight in aneurysms natural history and rupture risk. *Theor. Biol. Med. Model.* **10**, 67. (doi:10.1186/1742-4682-10-67)
21. Fang Y, Wu Y, Lv N, Chen Y, Huang Q, Liu J. 2015 Role of width-length ratio in evaluating the rupture risk of fusiform vertebro-basilar artery aneurysm. *Chin. J. Neurosurg. Dis. Res.* **14**, 15–17.
22. Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G. 2008 MeshLab: an open-source mesh processing tool. Sixth Eurographics Italian Chapter Conference, 129–136.
23. MATLAB and Statistics Toolbox Release 2016b, The MathWorks, Inc., Natick, Massachusetts, United States.
24. Cornea ND, Silver D, Min P. 2007 Curve-skeleton properties, applications, and algorithms. *IEEE Trans. Vis. Comput. Graph.* **13**, 530–548. (doi:10.1109/TVCG.2007.1002)
25. Wang Z, Chi Y, Huang W, Venkatesh SK, Tian Q, Oo T, Zhou J, Xiong W, Liu J. 2010 Comparisons of centerline extraction methods for liver blood vessels in imageJ and 3D slicer. In *APSPA ASC 2010 – Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Biopolis, Singapore, 14–17 December*, pp. 276–279.
26. Lee T, Kashyap RL, Chu C. 1994 Building skeleton models via 3-D medial surface/axis thinning algorithms. *Comput. Vis. Graph. Image Process.* **56**, 462–478. (doi:10.1006/cgip.1994.1042)
27. Kerschnitzki M, Kollmannsberger P, Burghammer M, Duda GN, Weinkamer R, Wagermaier W, Fratzl P. 2013 Architecture of the osteocyte network correlates with bone material quality. *J. Bone Miner. Res.* **28**, 1837–1845. (doi:10.1002/jbmr.1927)
28. Breiman L. 2001 Random forests. *Mach. Learn.* **45**, 5–32. (doi:10.1023/A:101093340)
29. Cortes C, Vapnik V. 1995 Support-vector networks. *Mach. Learn.* **20**, 273–295. (doi:10.1007/BF00994018)
30. Altman NS. 1992 An introduction to kernel and nearest-neighbour nonparametric regression. *Am. Stat.* **46**, 175–185. (doi:10.1080/00031305.1992.10476879)
31. Fisher RA. 1936 The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188.
32. Terdenge L, Kossuch JA, Schembecker G, Wohlgemuth K. 2017 Potential of gassing crystallization to control the agglomeration degree of crystalline products. *Powder Technol.* **320**, 386–396. (doi:10.1016/j.powtec.2017.07.044)
33. Pitt WR, Montalvao RW, Blundell TL. 2014 Polyphony: superposition independent methods for ensemble-based drug discovery. *BMC Bioinformatics* **15**, 324. (doi:10.1186/1471-2105-15-324)
34. Hobby JD. 1986 Smooth, easy to compute interpolating splines. *Discrete Comput. Geom.* **1**, 123–140.