



## Shape-Based Analysis of Intracranial Aneurysms

Juchler, Norman

**Abstract:** Intracranial aneurysms (IAs) are malformations of larger arteries in the brain that are associated with a structural weakening of the vessel wall. Unruptured IAs are prevalent in 2-5% of the population and are detected ever more frequently due to the increased availability of medical imaging. Albeit the majority of IAs develops asymptotically, the rare rupture of an IA causing a subarachnoid hemorrhage can have detrimental effects on the patient's health or even cause the patient's death. Therefore, clinicians are more often faced with difficult treatment decisions where they must weigh the costs of treatment against the risks of aneurysms to rupture. So far it is not possible to non-invasively determine the condition of the affected vessel wall region. Clinicians are therefore seeking for biomarkers that describe the structural stability of IAs. IA morphology, as seen in angiographic imaging, holds the potential for such a biomarker. Recent pathobiological studies suggest that structural wall instability is reflected in the geometry of the aneurysm lumen. This thesis project investigated the imaging-based morphological assessment of IAs. A first, data-driven approach, was based on a quantitative shape analysis derived on 3D surface geometries of 750 aneurysms. The author benchmarked established and novel morphometric parameters in terms of their predictive capacity for the disease status of the aneurysm, with the non-sphericity index and normalized Zernike energies performing best. He observed that shape is a stronger predictor for disease status than aneurysm size alone and confirmed the existing belief that IA morphology is associated with rupture. A second, psychometric approach, addressed the indistinct notion of morphological irregularity used by clinicians to characterize IA shape. Based on rating data from 13 clinical experts and 26 laypersons, the perceived irregularity of 134 aneurysms was quantified and used to identify the morphological constituents of overall irregularity. The author demonstrated that irregularity represents a continuous characteristic, with the risk of rupture increasing as the irregularity increases. Both approaches revealed a pronounced dependency of the shape on the anatomical location of the aneurysm. Combining shape and location substantially improved the accuracy of classification models for the IA rupture status. Other clinical aspects such as patient sex, age, smoking status or a history of blood hypertension did not play a significant role in the experiments. For future work, it is of great importance that the scientific community establishes a reference database to which new datasets can be related. In terms of morphology, the AneuX morphology database, which was developed in the context of this thesis project, could serve as such a reference. This thesis provides a refined, standardized taxonomy for morphological characteristics and offers a methodology to quantify subjective assessments of shape by humans. It contributes a software toolbox for morphometric analyses, and a new multicentric database comprising 750 aneurysms. Based on the comprehensive study of quantitative shape features, the author promotes the use of non-sphericity and an objective notion of irregularity for the clinical assessment IA shape.

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

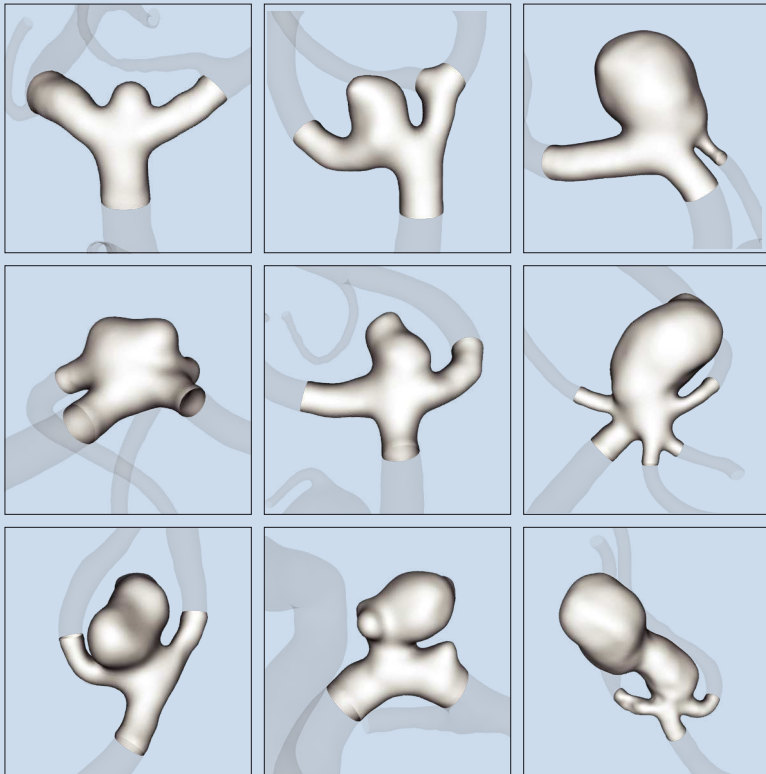
Juchler, Norman. Shape-Based Analysis of Intracranial Aneurysms. 2020, University of Zurich, Faculty of Science.

# Shape-Based Analysis of Intracranial Aneurysms

Norman Juchler  
Dissertation



University of  
Zurich <sup>UZH</sup>



# Shape-Based Analysis of Intracranial Aneurysms

---

**Dissertation**

zur

Erlangung der naturwissenschaftlichen Doktorwürde  
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

**Universität Zürich**

von

**Norman Juchler**

**Promotionskommission**

Prof. Dr. Vartan Kurtcuoglu

Prof. Dr. Sven Hirsch

Prof. Dr. Renato Pajarola

Zürich, 2020



# Acknowledgments

This thesis is the culmination of a collaboration between the Institute of Applied Simulation (IAS) at the Zurich University of Applied Sciences (ZHAW) in Wädenswil and The Interface Group at the University of Zurich (UZH). The project was funded mainly by the AneuX project (SystemsX.ch, Grant MRD 2014/261), with financial support from the ZHAW and the NCCR Kidney.CH consortium.

I would like to thank Prof. Dr. Sven Hirsch for his trust, continued support and the freedom he gave me. I am equally indebted to Prof. Dr. Vartan Kurtcuoglu for always offering constructive feedback, guidance and encouragement along the way. My thanks also go to Prof. Dr. Renato Pajarola for his active role on the thesis committee.

The thesis would not have been possible without the substantial support from our clinical collaborators. Prof. Dr. Philippe Bijlenga from Geneva University Hospitals (HUG) not only acted as data provider and principal investigator for the AneuX project, but actively shaped the leading scientific questions of this thesis. Prof. Dr. Daniel Rüfenacht, alongside Prof. Dr. Isabel Wanke and their colleagues at the Hirslanden Klinik in Zurich, assisted as medical advisors and patiently educated me about the intricate aspects of aneurysm research.

I am obliged to my working colleagues at the IAS for their support: Erich Zbinden, Dr. Georg Spinner, Dr. Susanne Suter, Dr. Martin Rerabek, Dr. Stefan Glüge, Dr. Ivo Kälin, Ueli Ebnöther, and Sofia Rey. I am also grateful for the support and guidance of Dr. Sabine Schilling, who helped shape the work presented in this dissertation and inspired me to pursue a new hobby. Dr. Victor Garcia was a frequent supporter professionally but above all he has been a good friend. Wishing you luck, buddy!

Thank you to The Interface Group for providing me with a convivial academic home. I wish the current team the best of luck in your scientific endeavors. Thank you, Dr. Diane de Zélicourt, Dr. Kartik Jain, and Dr. Stefano Buoso for fruitful discussions and inspiration. Thank you, Dr. Willy Kuo, for your notorious sarcasm. And of course, thank you J-floor people, for not letting me be the only loser on Friday nights.

I had the pleasure of interacting with a very open and supportive scientific community. The work by Prof. Dr. Juan Cebral and Dr. Félicitas Detmer from George Mason University heavily inspired my work. I would also like to thank Dr. Eliisa Netti, Dr. Suresh Raghavan, and Dr. Hitomi Anzai for their collaboration.

Rarely acknowledged but being of such tremendous importance to the scientific world, I would like to express my gratitude to the many digital platforms, open source communities, and knowledge databases that enabled this work. Python greatly stimulated my pleasure in coding. While my relationship with VTK is complicated, I am thankful to the community that helped me numerous times.

Last but not least, I would like to thank my family and friends, who patiently endured my absence in the last months and years.

Zurich, April 2020,  
Norman Juchler

# Contents

<b>Acknowledgments</b>	<b>iii</b>
<b>Contents</b>	<b>v</b>
<b>Abstract</b>	<b>ix</b>
<b>Zusammenfassung</b>	<b>xi</b>
<b>Abbreviations</b>	<b>xiii</b>
Anatomical locations	xiv
<b>1 Introduction</b>	<b>1</b>
1.1 Intracranial aneurysms	2
1.1.1 Epidemiology and risk factors	3
1.1.2 Cerebrovascular anatomy	6
1.1.3 Pathobiology	9
1.2 Management strategies for unruptured IAs	11
1.3 Tools for clinical decision making	14
1.3.1 Treatment guidelines and scoring models	14
1.3.2 Prediction models for aneurysm disease status	15
1.4 Quantitative assessment of aneurysms	18
1.5 Purpose and goals	20
<b>2 Morphometric analysis of aneurysm shape</b>	<b>21</b>
2.1 Abstract	22
2.2 Introduction	23
2.2.1 Related works	24
2.3 Material and methods	27
2.3.1 Imaging and patient data	27
2.3.2 Data processing	29
2.3.3 Morphometric description of the aneurysms	31
2.3.4 Analysis of the diagnostic capability of aneurysm shape	34
2.3.5 Software tools	35
2.4 Results	36
2.4.1 Univariate analysis	36
2.4.2 Multivariate analysis	41
2.4.3 Validation using external data	41
2.4.4 Dependency on the cut configuration	41



2.5	Discussion	43
2.5.1	Which features encode disease status?	43
2.5.2	How relevant is the cut configuration?	45
2.5.3	Do multiple shape predictors lead to a better model?	46
2.5.4	What is the effect of location as predictor?	47
2.5.5	External validation results	48
2.5.6	Comparison with the Detmer model	49
2.5.7	Limitations	51
2.6	Conclusions	52
2.A	Appendix: Geometry Indices	53
2.A.1	Indices measuring the aneurysm size	53
2.A.2	Indices measuring the aneurysm shape	53
2.B	Appendix: Distribution-derived metrics	54
2.B.1	Curvature-based features	54
2.B.2	Writhe-based features	55
2.C	Appendix: Zernike Moment Invariants	57
2.D	Appendix: Contributions	60
2.D.1	AneuX morphology database	60
2.D.2	The Geometric Modelling Toolkit	61
2.D.3	GMTK File Viewer	62
<b>3</b>	<b>Quantification of perceived irregularity</b>	<b>63</b>
3.1	Abstract	64
3.2	Introduction	64
3.3	Materials and methods	65
3.3.1	Imaging and patient data	65
3.3.2	Quantitative shape description	66
3.3.3	Qualitative shape assessment	67
3.3.4	Processing of the rating data	70
3.4	Results	74
3.4.1	Rating data	74
3.4.2	Multivariate quantitative model for perceived irregularity	76
3.5	Discussion	78
3.6	Conclusions	82
<b>4</b>	<b>Diagnostic value of irregularity</b>	<b>83</b>
4.1	Abstract	84
4.2	Introduction	85

4.3	Materials and methods	86
4.3.1	Imaging and patient data	86
4.3.2	Morphometric quantification of the IA lumen	87
4.3.3	Morphological assessment of IA lumen by human raters	88
4.3.4	Relationship between irregularity and morphological attributes	89
4.3.5	Relationship between morphology and clinical factors	91
4.4	Results	92
4.4.1	Rating data	92
4.4.2	Relationship between morphological attributes and irregularity	93
4.4.3	Relationship between morphology and clinical factors	96
4.5	Discussion	98
4.5.1	Limitations	100
4.6	Conclusions	101
4.7	Acknowledgments	102
<b>5</b>	<b>Discussion</b>	<b>103</b>
5.1	Morphological variability of aneurysm shape	106
5.2	The value and practicality of morphometric indices	109
5.2.1	Non-sphericity	109
5.2.2	Normalized Zernike energies $ZN_{surf}$	110
5.2.3	Other metrics	110
5.2.4	Morphometrics and qualitative characteristics	111
5.2.5	Hemodynamic indices	112
5.2.6	Robustness and cross-modal applicability	114
5.3	Human perception of aneurysm shape	115
5.4	IA morphology and anatomical location	118
5.5	Methodological considerations	122
<b>6</b>	<b>Conclusion and outlook</b>	<b>131</b>
<b>7</b>	<b>References</b>	<b>133</b>



# Abstract

Intracranial aneurysms (IAs) are malformations of larger arteries in the brain that are associated with a structural weakening of the vessel wall. Unruptured IAs are prevalent in 2-5% of the population and are detected ever more frequently due to the increased availability of medical imaging. Albeit the majority of IAs develops asymptotically, the rare rupture of an IA causing a subarachnoid hemorrhage can have detrimental effects on the patient's health or even cause the patient's death. Therefore, clinicians are more often faced with difficult treatment decisions where they must weigh the costs of treatment against the risks of aneurysms to rupture.

So far it is not possible to non-invasively determine the condition of the affected vessel wall region. Clinicians are therefore seeking for biomarkers that describe the structural stability of IAs. IA morphology, as seen in angiographic imaging, holds the potential for such a biomarker. Recent pathobiological studies suggest that structural wall instability is reflected in the geometry of the aneurysm lumen.

This thesis project investigated the imaging-based morphological assessment of IAs. A first, data-driven approach, was based on a quantitative shape analysis derived on 3D surface geometries of 750 aneurysms. The author benchmarked established and novel morphometric parameters in terms of their predictive capacity for the disease status of the aneurysm, with the non-sphericity index ( $NSI$ ) and normalized Zernike energies ( $Z_N^{\text{surf}}$ ) performing best. He observed that shape is a stronger predictor for disease status than aneurysm size alone and confirmed the existing belief that IA morphology is associated with rupture.

A second, psychometric approach, addressed the indistinct notion of morphological *irregularity* used by clinicians to characterize IA shape. Based on rating data from 13 clinical experts and 26 laypersons, the perceived irregularity of 134 aneurysms was quantified and used to identify the morphological constituents of overall irregularity. The author demonstrated that irregularity represents a continuous characteristic, with the risk of rupture increasing as the irregularity increases.

Both approaches revealed a pronounced dependency of the shape on the anatomical location of the aneurysm. Combining shape and location substantially improved the accuracy of classification models for the IA rupture status. Other clinical aspects such as patient sex, age, smoking status or a history of blood hypertension did not play a significant role in the experiments.

For future work, it is of great importance that the scientific community establishes a reference database to which new datasets can be related. In terms of morphology, the AneuX morphology database, which was developed in the context of this thesis project, could serve as such a reference.

This thesis provides a refined, standardized taxonomy for morphological characteristics and offers a methodology to quantify subjective assessments of shape by humans. It contributes a software toolbox for morphometric analyses, and a new multicentric database comprising 750 aneurysms. Based on the comprehensive study of quantitative shape features, the author promotes the use of non-sphericity and an objective notion of irregularity for the clinical assessment IA shape.

# Zusammenfassung

Intrakranielle Aneurysmen (IA) sind Missbildungen grösserer Hirnarterien, die mit einer strukturellen Schwächung der Gefässwand einhergehen. Unrupturierte IA treten bei 2-5% der Bevölkerung auf und werden aufgrund der zunehmenden Verfügbarkeit medizinischer Bildgebung immer häufiger entdeckt. IA bilden sich weitgehend asymptomatisch, können aber in seltenen Fällen rupturieren. Die daraus resultierende Subarachnoidalblutung ist bekannt für ihre gravierende Auswirkung auf die Gesundheit der Patient\_innen. Daher stehen Mediziner\_innen immer häufiger vor schwierigen Entscheidungen, bei denen sie die Kosten einer Behandlung gegen die Risiken einer Aneurysmaruptur abwägen müssen.

Bislang ist es nicht möglich, den Zustand der betroffenen Gefässwandregion nicht-invasiv zu bestimmen. Mediziner\_innen suchen daher nach Biomarkern, welche die strukturelle Stabilität von IA beschreiben. Die Morphologie von IA, wie sie mittels angiographischer Bildgebung zu sehen ist, birgt das Potenzial für einen solchen Biomarker. Neuere pathobiologische Studien deuten darauf hin, dass sich die strukturelle Wandinstabilität in der Lumen-Geometrie des Aneurysmas widerspiegelt.

Das vorliegende Dissertationsprojekt untersuchte die bildgebungs-basierte morphologische Beurteilung von IA. Ein erster, datengetriebener Ansatz verfolgte die quantitative Analyse von 3D-Geometrien für insgesamt 750 Aneurysmen. Der Autor verglich bekannte und neuartige morphometrische Parameter hinsichtlich ihrer Vorhersagefähigkeit des Aneurysmenzustands. Dabei schnitten Non-Sphericity-Index (*NSI*) und die normalisierten Zernike-Energien ( $Z_N^{\text{surf}}$ ) am besten ab. Er stellte fest, dass die Beurteilung der Aneurysmenform eine bessere Einschätzung des Krankheitsstatus' erlaubt als die im klinischen Kontext sehr verbreitete Vermessung der Aneurysmengrösse.

Ein zweiter, psychometrischer Ansatz befasste sich mit dem unscharfen Begriff der morphologischen Irregularität, der oft zur allgemeinen Charakterisierung der Aneurysmenform verwendet wird. Auf der Grundlage von Bewertungsdaten von 13 klinischen Fachleuten und 26 Laien wurde die wahrgenommene Irregularität (engl.: perceived irregularity) von 134 Aneurysmen quantifiziert und dazu verwendet, die morphologischen Elemente

der Irregularität zu bestimmen. Der Autor zeigte auf, dass die Irregularität ein kontinuierliches Merkmal von Aneurysmen darstellt, welche die Rupturwahrscheinlichkeit widerspiegeln.

Beide Ansätze offenbarten eine ausgeprägte Abhängigkeit der Form von der anatomischen Lage des Aneurysmas. Die Kombination von formbeschreibenden Prädiktoren mit der Lokalisation verbesserte die Genauigkeit der Klassifikationsmodelle für den IA-Rupturstatus erheblich. Andere klinische Aspekte wie Geschlecht, Alter, Raucherstatus oder eine Vorgeschichte von Bluthochdruck spielten in den Experimenten keine signifikante Rolle.

Für zukünftige Arbeiten ist es von grosser Bedeutung, dass die wissenschaftliche Gemeinschaft eine Referenzdatenbank erarbeitet, auf die neue Datensätze bezogen werden können. Die Morphologiedatenbank AneuX, die im Rahmen dieses Dissertationsprojekts entwickelt wurde, könnte in Bezug auf die Morphologie als solche Referenz dienen.

Diese Dissertation präsentiert eine verfeinerte, standardisierte Taxonomie für die morphologischen Merkmale von Aneurysmen und beschreibt eine Methodik zur Quantifizierung subjektiver Bewertungen der Aneurysmenform. Diese Arbeit umfasst eine Software-Toolbox für die morphometrische Beschreibung von Aneurysmen, sowie eine multizentrische Datenbank mit 750 Aneurysmen. Basierend auf der umfassenden Untersuchung quantitativer Formmerkmale empfiehlt der Autor die Verwendung von Non-Sphericity und einen objektiven Begriff der Formirregularität für die klinische Beurteilung von Aneurysmen.

## Abbreviations

3DRA	3D rotational angiography
AR	Aspect ratio
aSz	Aneurysm size
AUC	Area under the ROC curve
CFD	Computational fluid dynamics
CoW	Circle of Willis
CTA	Computed tomography angiography
CV	Cross-validation
EI	Ellipticity index
GI	Geometry index
GLN	Total Gaussian curvature, normalized
HUG	University Hospitals Geneva
IA	Intracranial aneurysm
IQR	Interquartile range
LASSO	Least absolute shrinkage and selection operator
LOWESS	Locally weighted scatterplot smoothing
LR	Logistic regression
MAD	Median absolute deviation
ML	Machine learning
MLN	Total mean curvature, normalized
MRA	Magnetic resonance angiography
NSI	Non-sphericity index
OLS	Ordinary least squares
PA	Prediction accuracy
PCA	Principal component analysis
RMSE	Root-mean-square error
ROC	Receiver operating characteristic
SAH	Subarachnoid hemorrhage
std	Standard deviation
SVM / SVR	Support vector machine / support vector regression
UI	Undulation index
ZM / ZMI	Zernike moment / Zernike moment invariants



## Anatomical locations

ACA	Anterior cerebral artery
AComA	Anterior communicating artery
BA tip	Tip of basilar artery
ICA bif	Internal carotid artery, terminal bifurcation
ICA cav	Internal carotid artery, intracavernous segment
ICA chor	Internal carotid artery, choroidal segment
ICA oph	Internal carotid artery, ophthalmic segment
MCA	Middle cerebral artery, M1 segment and upper trunk
MCA bif	Middle cerebral artery, bifurcation of the M1 and M2 segments
PCA	Posterior cerebral artery
PComA	Posterior communicating artery
VB other	Other locations in the vertebrobasilar trunk

See Figure 1.2 for a schematic overview of the locations.

# 1 Introduction

Intracranial aneurysms (IAs) are malformations of cerebral arteries that bear a risk of rupture. Unruptured aneurysms can be treated invasively. As the risks of treatment do not always outweigh the risks of rupture, clinicians must carefully assess each individual case before deciding to intervene.

Already the medical term *aneurysm*, which originates from the Greek *ἀνεύρυσμα* [a'nevrizma] meaning “dilation”, refers to the disease as a morphological anomaly. The shape indeed plays an important role in the context of diagnosing and treating unruptured IAs (uIAs). Shape is an aneurysm property that can be extracted from angiographic data. As imaging of blood vessels is the clinicians’ primary source of information when deciding on the best management strategy for IAs, the availability of shape information is thus very high. The shapes of IAs display great morphological variability that is related to the pathobiology of the aneurysm. Consequently, it is thought that shape can serve as a proxy for the disease status and thus is often used as a decisive factor in determining possible treatment options. The geometry of the aneurysm, its neck and the adjacent vasculature are taken into account by interventionists when planning the treatment.<sup>1</sup>

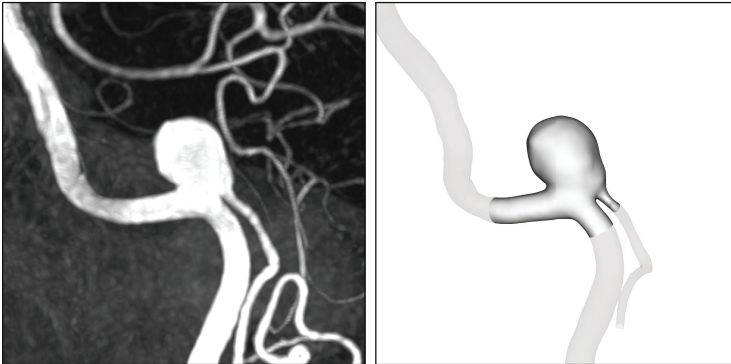
Despite the apparent clinical relevance of the aneurysm shape, its characteristics have yet been assessed by clinicians mostly qualitatively. While the measurement of the *size* of the aneurysm, its neck and the vascular caliber are common, little if any quantitative metrics are used to document the aneurysm *shape*.

The principal goals of this thesis were to evaluate the ability of quantitative morphology to predict the disease status of aneurysms and to make improvements on how the information of shape could be better used in the clinical context. This introductory chapter first presents the aneurysm disease and then motivates the clinical need for quantitative methods to support clinical decision making.

## 1.1 Intracranial aneurysms

Aneurysms in general are vascular diseases characterized by a structural weakening of the vessel wall leading to a widening or outward bulging of the affected vessel wall region. Aneurysms are most commonly found in the major cerebral arteries<sup>1,2</sup> and in the abdominal and thoracic aorta.<sup>3,4</sup> Although predominantly asymptomatic, aneurysms may lead to a catastrophic dissection or rupture of the vessel wall and pose a serious threat to the patients. Despite having several pathobiological aspects and risk factors in common,<sup>5</sup> aortic and cerebral aneurysms are treated as two different diseases because of differences in the vessel wall anatomy, mechanobiology and available treatment options.

*Intracranial aneurysms.* This thesis deals exclusively with *saccular intracranial aneurysms* that are characterized by a focal dilation of cerebral arteries (Figure 1.1). Intracranial aneurysms form a relatively wide range of diseases. Pathologists distinguish between saccular aneurysms, fusiform aneurysms (segmental ectasias), dissecting aneurysms, and other rarer types of aneurysms.<sup>6</sup> Saccular IAs are by far the most common type of IAs, accounting for 70-90% of all IAs.<sup>7</sup>



**Figure 1.1:** Exemplary saccular aneurysm at the bifurcation between the right vertebral artery (VA, large vessel) and the posterior inferior cerebellar artery (PICA, small vessel) from the AneuX morphology database (see Chapter 2). The aneurysm has a size of about 10mm. **Left:** The aneurysm as seen in medical imaging (3D rotational angiography). **Right:** 3D model of the aneurysm and the surrounding vasculature extracted from the same angiography.

### 1.1.1 Epidemiology and risk factors

*Clinical significance.* Saccular IAs usually form and grow silently without creating any symptoms. However, the aneurysmal lesion may rupture, causing blood to leak into the subarachnoid space. This subarachnoid hemorrhage (SAH), a type of stroke, is known for its devastating effects on the patient. The case fatality rate for aneurysmal SAH is 50%, and about 50% of the survivors suffer from long-term cognitive impairment.<sup>8-10</sup> Ruptured IAs account for 85% of (non-traumatic) SAH.<sup>9</sup>

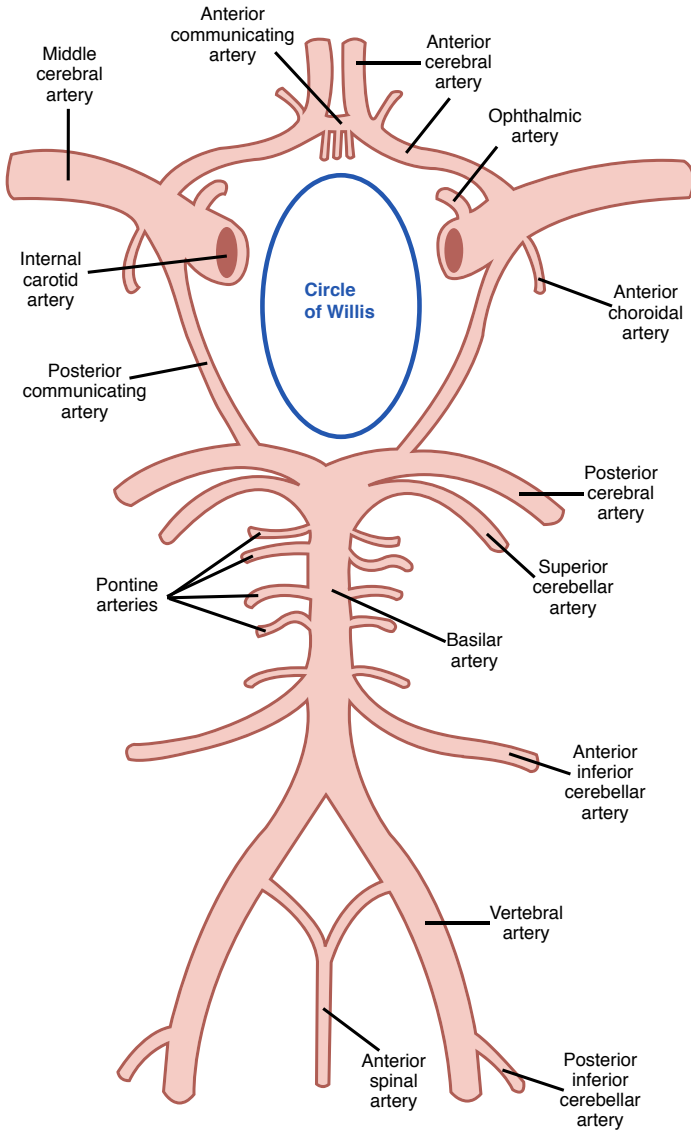
*Epidemiology.* The saccular IA is a common disease. An extensive meta-study by Vlák et al.<sup>11</sup> estimated the prevalence of unruptured IAs to be 3.2% (95% confidence interval, 1.9-5.2) in a population without specific comorbidities and adjusted for sex and age. The same study stated that unruptured IAs were more prevalent in females than in males, with a prevalence ratio of 1.61 (95% CI 1.02-2.54). Patients frequently develop multiple aneurysms. According to two different studies, about 30% of patients that present with a ruptured aneurysm harbored multiple aneurysms.<sup>12,13</sup>

*Risk factors for IA formation and growth.* Although IAs are more common in older people, aneurysms cannot be regarded simply as wear and tear of the arterial walls in a lifelong exposure to physical and biochemical stresses, as not all older patients are affected. Certain predispositions and risk factors seem to favor IA initiation and growth.<sup>1,11,14-17</sup> A familial history for ruptured IAs (prevalence ratio: 3.4) and female sex (prevalence ratio: 1.6, odds ratio: 2.0) increase the likelihood for harboring an unruptured aneurysm.<sup>11,18</sup> Hypertension (odds ratio: 2.9) and the current smoking status (odds ratio: 3.0) are the most frequently cited modifiable risk factors.<sup>1,17</sup> IAs are also more frequently observed in patients either autosomal dominant polycystic kidney disease, a brain tumor or atherosclerosis.<sup>11</sup> Regular physical exercise is associated with a lower prevalence of unruptured IAs.<sup>17</sup> Genetic risk factors affecting the cellular and pericellular components of the vessel wall have also been identified.<sup>1,19</sup> How and to what extent the genetic predisposition affects the development of aneurysms is subject to ongoing research. According to Vlák et al.<sup>11</sup>, no significant relation was found between patient age and the prevalence of unruptured aneurysms. The prevalence of aneurysm *growth* was estimated by Backes et al.<sup>16</sup> to be 9% in a pooled cohort of 4'972 unruptured aneurysms

followed up for a total of 13'987 patient-years. Another study by Backes et al.<sup>20</sup> observed growth in 14% of 1'909 aneurysms monitored over 5'782 patient-years. In a study by Chien et al.<sup>21</sup> involving 520 unruptured IAs followed up for 1'417 aneurysm-years, the average growth rate was measured as 0.085 mm/month for large aneurysms (> 7mm) and 0.030 mm/month for small aneurysms (< 3mm). Based on an epidemiological study, Koffijberg et al.<sup>22</sup> concluded that aneurysms most likely are not growing at a constant rate.

*Rupture risk.* Not all aneurysms eventually rupture. The annual incidence rate for IA rupture was estimated to be 0.8-1.9%.<sup>23-27</sup> A large meta-analysis based on data from 8'382 patients followed up for a total of 29'166 patient-years revealed an overall 1-year risk of aneurysm rupture of 1.4% (95% CI 1.1-1.6) and a 5-year risk of 3.4% (95% CI 2.9-4.0). These incidence rates for rupture, however, varied greatly with factors such as aneurysm size, aneurysm location, patient age and other factors.<sup>1,28</sup> In the same study, the mean age of patients suffering from aneurysmal SAH was 60.7 years.<sup>28</sup>

*Risk factors for aneurysm rupture.* Most risk factors that apply for aneurysm growth also apply to aneurysm rupture.<sup>16</sup> Consumption of strong alcohol or narcotics, smoking and hypertension favor IA rupture.<sup>1,28</sup> Furthermore, a positive familial history for SAH, the presence of multiple aneurysms, observation of growth in imaging series, and irregular morphology substantially increase the risk of aneurysm rupture.<sup>1,28-32</sup> Larger aneurysms are associated with a substantially larger risk of rupture.<sup>28</sup> Compared to small aneurysms (<5.0 mm of size), the hazard ratios (HR, ratio of rupture rates) were measured at 1.1 for the range of aneurysm sizes 5-7 mm, 2.3 for the size range 7-10 mm, 5.5 for the size range 10-20 mm and 20.8 for aneurysms larger than 20mm. The rupture rate also varies considerably for different anatomical locations, with higher hazard ratios in the posterior circulation.<sup>28,33</sup>



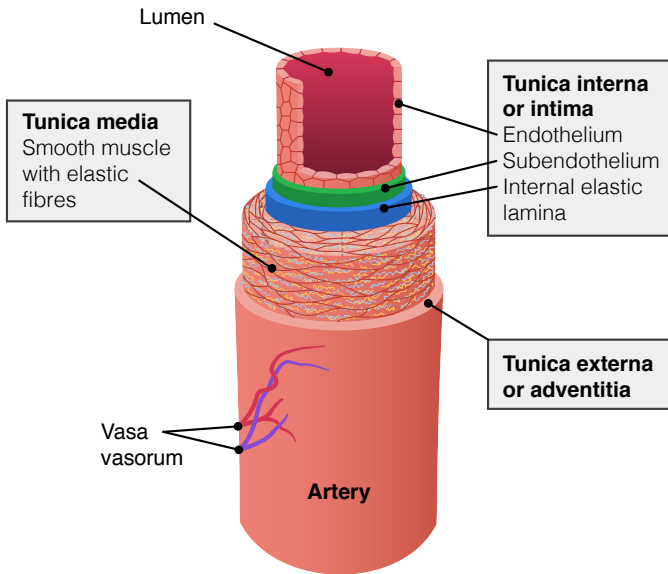
**Figure 1.2:** Major cerebral arteries and the Circle of Willis. (Source: Wikimedia Commons, modified, license: public domain content)

### 1.1.2 Cerebrovascular anatomy

*Circle of Willis.* Most IAs develop in the proximity of the Circle of Willis (CoW), a circular network formed by the major cerebral arteries (Figure 1.2). The CoW displays great anatomical variability.<sup>34–36</sup> Anatomical studies revealed that about 50% of the CoWs show anomalies such as hypoplasia, accessory vessels or missing vessels.<sup>34,36</sup> Not only the functional availability of vascular segments differs between individuals, but also the topological configuration of the CoW is subject to change.<sup>37</sup> This variability is believed to be an important factor in aneurysm formation, as it has a direct influence on local flow conditions and the stress acting on the vessel wall.<sup>1,38,39</sup>

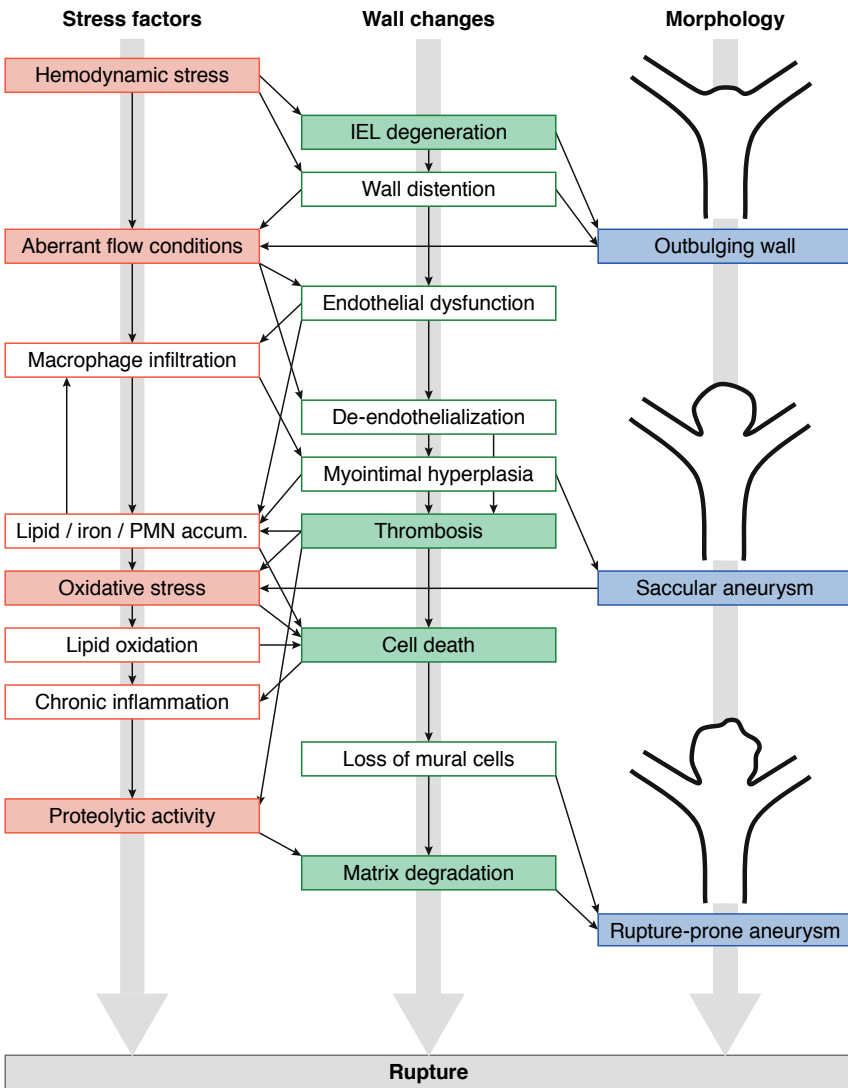
*Structure of the artery.* Arteries are organized in three layers: intima, media and adventitia (Figure 1.3). The innermost layer, the *intima*, consists of the endothelium and the internal elastic lamina, which are separated by a thin layer of connective tissue (subendothelium, also known as basement membrane). The endothelium is a monolayer of endothelial cells that forms a protective barrier between the blood flowing through the vessel lumen and the vascular tissue. It plays a central role in blood clotting (thrombosis), immune system signaling (inflammation) and the signaling cascade leading to vasoconstriction (or vasodilation). The internal elastic lamina is a layer of elastic tissue with elastin fibers. It is important for bearing the mechanical stress exerted on the vessel wall.<sup>6</sup> The *media* is composed of densely packed smooth muscle cells embedded in an extracellular matrix composed of elastin and collagen fibers.<sup>5</sup> The *adventitia* contains fibroblasts and white blood cells.<sup>1</sup> The average wall thickness of the major cerebral arteries varies between 0.45–0.65mm.<sup>40</sup>

*Intra- and extracranial differences.* In contrast to extracranial arteries, intracranial arteries lack a fully developed external elastic lamina between media and adventitia. They have a considerably thinner adventitia and a smaller wall thickness relative to the vessel caliber.<sup>1,5,6</sup> Furthermore, intracranial arteries are surrounded by cerebrospinal fluid of the subarachnoid space instead of connective tissue as in extracranial arteries. These structural differences are thought to be the primary reasons why saccular IAs have a very distinctive pathology and morphology compared to aneurysms in other parts of the body.<sup>1,6</sup>



**Figure 1.3:** Anatomy of a cerebral artery. (Source: Wikimedia Commons, modified, license: CC-BY-SA 3.0)





**Figure 1.4:** Pathobiological model for the disease progression in IAs and the relationship with morphology. Adapted by permission from Springer Nature: *Acta Neuropathologica*, Frösen, Tulamo et al.<sup>41</sup>, copyright 2015. Abbreviations: IEL – internal elastic lamina; PMN – polymorphonuclear leukocytes; accum. – accumulation.

### 1.1.3 Pathobiology

The pathobiology of intracranial aneurysms is complex. The wall-degenerating processes are non-sequential, interrelated and stochastic, depend on various time-varying internal and external factors, and affect the vessel wall both microscopically and macroscopically (Figure 1.4).

The overall development of an aneurysm can be divided into three stages: formation, growth and rupture.<sup>5</sup> It is thought that aberrant blood flow caused by anatomical variations or homeostatic imbalance, and the structural composition of the arterial vessel wall are the primary factors leading to aneurysm formation.<sup>1,5,41,42</sup> High mechanical load can result in alterations in the internal elastic lamina, leading to an initial outbulging of the vessel wall.<sup>41,43</sup> Conversely, wall deformation alters the hemodynamic flow, which may further intensify the mechanical stress exerted on the vessel wall.<sup>41,42</sup> Increased shear stress in exposed regions of the vessel wall can lead to focal degeneration of the endothelium. This impairs its protective function, enabling the influx of macrophages and triggering a multitude of inflammatory responses that have a detrimental effect on the structural constituents of the vessel wall.<sup>41,44</sup> Additionally, the oxidative stress (imbalance between production and destruction of free radicals) on the cellular components of the vessel wall increases, entailing a loss of mural cells through apoptosis.<sup>45</sup> Overall, the laminar structure of the vessel wall gradually deteriorates, further reducing its structural resilience to mechanical forces and allowing additional wall distension.<sup>41,43</sup> This development may be exacerbated by the adverse effects of luminal thrombosis, in which oxidative stress and degradation of the extracellular matrix are enhanced by neutrophils trapped in the fibrin networks of the thrombus.<sup>41</sup>

At the same time, inherent wall repair by means of smooth muscle cell proliferation and synthesis of new collagen matrix may partially compensate for the deteriorating structural wall properties. In addition, the protective, angiogenic and matrix metabolizing effects of luminal thrombosis can also have a beneficial effect on wall regeneration.<sup>41</sup> Through migration of smooth muscle cells and fibroblast, the thrombus may organize and become an integral part of the aneurysmal wall.<sup>44</sup>

Aneurysm growth comes to a halt when regenerative and degenerative processes within the wall reach a balance. Episodes of temporary growth can alternate with periods of stability. Prolonged exposure to risk factors such as smoking, or hypertension can disrupt the temporary equilibrium and revive aneurysm growth. In an advanced state, chronic inflammation, neovascularization in response to hypoxic conditions, progressive infiltration of leukocytes and matrix degeneration through increased proteolytic activity can further deteriorate the mechanical properties of the vessel wall – until the tissue can no longer withstand the mechanical forces acting on it and eventually ruptures.<sup>1,41,44</sup>

Aneurysmal pathogenic processes in part resemble those of atherosclerosis, a chronic inflammatory disease of the arterial wall that exposes patients to the risks of an ischemic stroke. In particular, damage of the endothelium, wall remodeling, the involvement of chronic inflammation, lipid accumulation and oxidation, and neovascularization occur in both diseases. The main difference between the two diseases is that atherosclerosis primarily affects the intima, whereas in IAs the entire vessel wall is subject to structural remodeling.<sup>44</sup>

In this entire process, the aneurysm experiences significant morphological changes. Because of the initially homogenous structure of the vessel wall, the aneurysm displays a mostly regular structure at early stages of the disease. The ongoing wall remodeling leads to histological fragmentation of the vessel wall and an increased heterogenization of the mechanical properties of the arterial wall, enabling the formation of new focal outpouchings or corrugation of the aneurysmal wall. In this stage, the aneurysm wall often displays considerable variation in thickness.<sup>46</sup> The surrounding anatomical structures may contribute to the absorption of mechanical stress as the size of the aneurysm increases. The biomechanical properties of the encasing tissue also determine the most likely directions of growth of the aneurysm, which contributes to the overall appearance of the aneurysm. Finally, the presence of intraluminal thrombosis also affects the morphological appearance of the aneurysm in angiographic imaging, which can be recognized as indentations or beveling of the original shape.

## 1.2 Management strategies for unruptured IAs

*Historical notes.* The treatment alternatives have developed greatly in recent times. It is believed that aneurysms have been known across different cultures since the ancient days.<sup>47,48</sup> The first scientific description of an IA is attributed to the Italian anatomist Giovanni Battista Morgagni in 1761.<sup>49</sup> The British surgeon Sir Victor Horsley performed the first surgical treatment of an IA (by arterial ligation) on a human patient in 1885.<sup>48</sup> Surgical and imaging methods were further pioneered until surgical clipping became established as treatment method for IAs in the 1940s.<sup>48,50</sup> Endovascular methods began to emerge in the 1960s, with Guido Guglielmi accomplishing the first saccular occlusion using detachable coils in 1988.<sup>51</sup> With the advent of modern medical imaging techniques (CT became clinically available in the 1970s, MRI in the 1980s) and the subsequent increased detection of unruptured IAs, preventive interventions became more common.

*Treatment today.* Ruptured and unruptured IAs are today treated by means of either surgical clipping or endovascular coiling.<sup>1,2</sup> Both management strategies aim at protecting the aneurysm sac from further exposure to hemodynamic stresses. *Surgical clipping* involves the placement of special clips at the neck of the aneurysm. To gain access to the affected vessel segment, a craniotomy and a subsequent retraction of the brain is required,<sup>6,52</sup> thus making the intervention an arduous experience for the patient even in the favorable case where no (pre- or postsurgical) subarachnoidal bleeding is involved. The aim of *endovascular coiling* is to occlude the aneurysm lumen by means of platinum coils. To this end, a guiding catheter is used that is normally inserted at a femoral artery and navigated through the arterial circulation to the aneurysm site.<sup>52</sup> For wide aneurysm necks, or to prevent thromboembolic complications, flow-diverters or stents are sometimes used.<sup>53</sup> Endovascular treatment is less invasive than surgical clipping, but the aneurysm recurrence rate is higher.<sup>6</sup>

*Costs of treatment.* The morbidity, mortality, and thus the overall costs of aneurysm treatment are difficult to determine as they depend on various factors such as rupture prior to treatment, patient age, type of treatment, aneurysm location, aneurysm size and several more.<sup>1,54,55</sup> A meta-study on the surgical repair of unruptured IAs based on data from totally 9'845 patients measured a

mortality as 1.7% and a morbidity (i.e. the rate of unfavorable outcome including death) as 6.9%.<sup>54</sup> However, these rates were subject to temporal change and study quality. Regardless of these difficulties, treatment comes with risks, which must be weighed against the risks of not treating the aneurysm.

*Costs of non-treatment.* The yearly risk of rupture of a saccular IA on average is about 1% per year,<sup>28</sup> even though this rate varies considerably with the aneurysm location.<sup>28,33</sup> Hence, IA rupture is relatively rare. However, the immediate consequence of an IA rupture, a subarachnoid hemorrhage, is known for a high mortality and morbidity among survivors.<sup>8</sup> 10-15% of the patients die before receiving any medical attention, and about 50% of the patients die within the first 6 months after rupture.<sup>9</sup> Survivors of SAH often suffer from long-term neurological disabilities and systemic complications, and are at risk of rebleeding.<sup>9</sup>

*Angiographic imaging.* The primary source of information for the diagnosis, assessment and treatment of ruptured or unruptured aneurysms is 3D angiographic imaging.<sup>1</sup> Three common modalities are available: magnetic resonance angiography (MRA), computed tomography angiography (CTA), and 3D rotational angiography (3DRA).<sup>56-58</sup> 3DRA achieves the best spatial resolution, at the cost of higher radiation doses. The actual spatial resolution ( $\neq$  voxel size) of an image depends on the available device, imaging protocol and the presence of optical distortions such as motion blur. The resolution falls roughly into the following ranges – 3DRA: 0.3-0.6 mm, CTA: 0.6-1.2 mm, MRA: 0.8-1.4 mm. It is common for all modalities to use a contrast agent to increase the contrast between the arterial lumen and the surrounding tissue; a catheter injection of the contrast agent can further improve the contrast. Unruptured IAs are typically diagnosed incidentally using CTAs and MRAs, while 3DRA imaging is employed in the context of aneurysm intervention.

*Clinical decision problem.* The combination of relatively low risk of rupture, high costs of SAH, nonnegligible treatment costs, a complex pathobiology and a lack of reliable markers for disease status makes the clinical decision making difficult. Clinicians are therefore in need for treatment guidelines or prognostic tools to support them in making their decision.

## UIATS: Delphi Consensus Score

		W	+	-	
Patient	<b>Age (single)</b>	< 40 years	4		
		40-60 years	3		
		61-70 years	2		
		71-80 years	1		
		> 80 years	0		
	<b>Risk factor incidence (multiple)</b>	Previous SAH from a different aneurysm	4		
		Familial intracranial aneurysms or SAH	3		
		Japanese, Finnish, Inuit ethnicity	2		
		Current cigarette smoking	3		
		Hypertension (systolic BP > 140 mmHg)	2		
		Autosomal-polycystic kidney disease	2		
		Current drug abuse (cocaine, amphetamine)	2		
	Current alcohol abuse	1			
<b>Clinical Symptoms related to uIA (multiple)</b>	Cranial nerve deficit	4			
	Clinical or radiological mass effect	4			
	Thromboembolic events from the aneurysm	3			
	Epilepsy	1			
<b>Other (multiple)</b>	Reduced quality of life due to fear of rupture	2			
	Aneurysm multiplicity	1			
<b>Life expectancy due to chronic/malignant disease (single)</b>	< 5 years	4			
	5-10 years	3			
	> 10 years	1			
<b>Comorbid disease (multiple)</b>	Neurocognitive disorder	3			
	Coagulopathies, thrombophilic diseases	2			
	Psychiatric disorder	2			
Aneurysm	<b>Maximum diameter (single)</b>	< 4 mm	0		
		4.0-6.9 mm	1		
		7.0-12.9 mm	2		
		13.0-24.9 mm	3		
		≥ 25mm	4		
	<b>Morphology (multiple)</b>	Irregularity or lobulation	3		
		Size ratio ≥ 3 or aspect ratio > 1.6	1		
	<b>Location (single)</b>	BA bifurcation	5		
		Vertebral/basilar artery	4		
		ACoM/A or PCoM/A	2		
	<b>Other (multiple)</b>	Aneurysm growth on serial imaging	4		
		Aneurysm de novo formation on serial imaging	3		
		Contralateral stenooclusive vessel disease	1		
Treatment	<b>Age-related risk (single)</b>	< 40 years	0		
		41-60 years	1		
		61-70 years	3		
		71-80 years	4		
		> 80 years	5		
	<b>Aneurysm size-related risk (single)</b>	< 6.0 mm	0		
		6.0-10.0 mm	1		
		10.1-20.0 mm	3		
		> 20 mm	5		
	<b>Aneurysm complexity-related risk (single)</b>	High	3		
	Low	0			
<b>Intervention-related risk</b>	Constant	5			
<b>Unruptured Intracranial Treatment Score: A – B A B</b>					

## PHASES: Regression

	W	+	
<b>Population</b>	EU/USA/CA	0	
	Japanese	3	
	Finnish	5	
<b>Hypertension</b>	No	0	
	Yes	1	
<b>Age</b>	<70 years	0	
	≥70 years	1	
<b>Size of aneurysm</b>	<7.0 mm	0	
	7.0-9.9 mm	3	
	10.0-19.9 mm	6	
	≥ 20 mm	10	
<b>Earlier SAH from another IA</b>	No	0	
	Yes	1	
<b>Site of aneurysm</b>	ICA	0	
	MCA	2	
	ACA/PCoM/A/post.	4	
<b>PHASES Score</b>			

## ELAPSS: Regression

	W	+	
<b>Earlier SAH</b>	Yes	0	
	No	1	
<b>Location of IA</b>	ICA/ACA/ACoM/A	0	
	MCA	3	
	PCoM/A/posterior	5	
<b>Age</b>	≤ 60 years	0	
	> 60 years (per 5 years)	1	
<b>Population</b>	USA/CND/EU/China	0	
	Japan	1	
	Finland	7	
<b>Size of aneurysm</b>	1.0-2.9 mm	0	
	3.0-4.9 mm	4	
	5.0-6.9 mm	10	
	7.0-9.9 mm	13	
	≥ 10 mm	22	
<b>Shape of aneurysm</b>	Regular	0	
	Irregular	4	
<b>ELAPSS Score</b>			

**Figure 1.5:** Recent scoring schemes for IA risk assessment. The unruptured intracranial aneurysm treatment score (UIATS) employed the Delphi consensus method to identify and prioritize the factors that a panel of clinical experts considered relevant for deciding in favor of (red column with sub-score *A*) or against (green column with sub-score *B*) treatment.<sup>59</sup> If  $A > B + 2$ , treatment is advised, if  $A < B - 2$ , conservative management is recommended. The PHASES score<sup>28</sup> and ELAPSS score<sup>20</sup> both are based on Cox regression analyses of data from multicentric patient cohorts. While the PHASES scoring scheme is a model for aneurysm *rupture*, ELAPSS aims at predicting aneurysm *growth*. Weights (columns *W*) indicate the contribution to the score if a criterion applies.

## 1.3 Tools for clinical decision making

### 1.3.1 Treatment guidelines and scoring models

Clinical treatment guidelines for unruptured intracranial aneurysms suggest as to when and how the disease should be treated. Such recommendations are based on the study of risk factors for aneurysm formation and rupture, but also take into account practical aspects such as the availability of treatment procedures or the associated costs of treatment. Treatment guidelines for uIAs have developed considerably in recent years, and increasingly reflect the multifactorial nature of the disease.

For instance, aneurysm (lumen) size was soon identified as an indicator of wall morbidity and risk of rupture (e.g., Wiebers et al.<sup>60</sup> 1987, and again Wiebers et al.<sup>61</sup> in 2003). Today, this parameter is widely used as an assessment criterion, not least because of its property of being relatively easy to determine via angiographic imaging. Yet the determination of a specific treatment threshold is the subject of an ongoing scientific debate.<sup>61-63</sup> If the threshold is set too high, rupture-prone aneurysms remain untreated, while too low a threshold results in overtreatment of uIAs (false positive vs. false negative trade-off). Furthermore, the use of a single criterion may not adequately reflect the different facets of the disease, such as the varying characteristics of IA at different anatomical locations, or the exposure of the patient to congenital or modifiable risk factors. However, it quickly becomes difficult for a clinician to weigh multiple criteria against each other, as it requires the knowledge of hazard ratios or associated costs for an objective assessment.

To better accommodate the multifactorial nature of the disease, scoring schemes, or more generally multivariate prediction models, have been developed to support clinical decision making. Scoring schemes assign scores  $\beta$  to specific disease and/or patient characteristics  $x_i$  (per case  $i$ ) and map the result to a fixed scale based on the chosen calculation method. This can be formally written as  $s_i = f(x_i|\beta)$ , where  $x_i$  represents a  $d$ -dimensional vector of (categorical) observations for case  $i$ , and  $\beta$  the collection of all scores  $\beta_{c_j}$  for any value  $c \in \mathcal{C}_j$  of all sets of categorical variables  $\mathcal{C}_j, j \in \{1, \dots, d\}$ :

$$\beta = (\beta_{c_j} | \forall c \in \mathcal{C}_j, \forall j \in \{1, \dots, d\})$$

The parameters  $\beta$  are derived using a training procedure with the objective that the resulting scores best reflect the risks associated with the disease status. The model ideally is validated on the basis of clinical data. An action, or more specifically, treatment, is recommended if the resulting score  $s_i$  exceeds a certain threshold  $\delta$ :

$$y_i = g(x_i|\beta, \delta) = f(x_i|\beta) < \delta$$

Figure 1.5 illustrates three recent scoring tools. In the Unruptured Intracranial Treatment Score (UIATS) by Etminan et al.<sup>59</sup> expert knowledge from 39 specialists was condensed into a scoring scheme and externally validated by 30 additional specialists. This method weighs up criteria for aneurysm repair against those for conservative treatment. The adopted Delphi consensus method, however, has been criticized for not being scientific.<sup>64</sup> Furthermore, UIATS was associated with overtreatment in a validation study.<sup>65</sup> The PHASES score by Greving et al.<sup>28</sup> was derived from a pooled analysis of data from 8'382 patients in six prospective cohort, followed up over a total of 29'166 patient-years. A Cox regression analysis was performed to model the outcome (rupture/no rupture) based on a selection of available characteristics about the aneurysm or patient, resulting in a scoring model for rupture. Backes et al.<sup>20</sup> applied a methodologically very similar approach for the development of the ELAPSS scoring model for aneurysm growth. The model is based on data from 1'909 uIAs in 1'507 patients monitored over 5'782 patient-years. The scoring scheme could help to adjust the follow-up examination periods. Both PHASES and ELAPSS have been externally validated, with mixed results.<sup>31,66–71</sup>

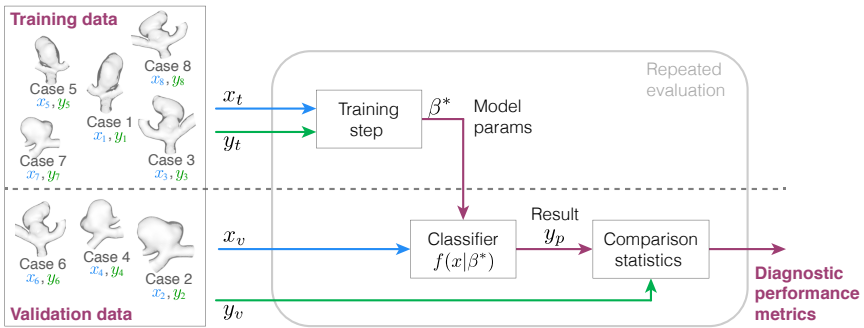
### 1.3.2 Prediction models for aneurysm disease status

The above scoring schemes can be regarded as special cases of the more general class of binary classification models (in the following also referred to as prediction models, predictive models or diagnostic models). Like scoring models, general prediction models also aim at describing a functional relationship  $g(\cdot|\beta, \delta)$  between observations  $x_i$  of the case  $i$  and the corresponding most probable (binary) outcome variable  $y \in \{0,1\}$ :  $y = g(x|\beta, \delta)$ . The variables  $x$  are also known as predictor variables, independent variables, risk factors or *features*; the results  $y$  are alternatively called response



variable, predicted variable, dependent variable, or *label*.<sup>72</sup> The model function  $g(\cdot | \beta, \delta)$  is often referred to as *classifier*. In the context of this thesis, label  $y$  usually represents the disease status: stable/unstable, unruptured/ruptured.

The model parameters  $\beta$  are identified by means of supervised learning (Figure 1.6).<sup>73,74</sup> A central assumption in supervised learning is that the target label is known for the available data (as opposed to unsupervised learning where the target label is not known). For each round of training and validation, the dataset is partitioned randomly into training and validation sets. The goal of the training step is to identify the classification parameters  $\beta^*$  given the observations  $x_t$  and the vector of binary target labels  $y_t$  of the training dataset. This requires the solution of an optimization problem. The predictive capability of a trained model (classifier, represented by a function  $f(\cdot | \beta)$  and its parameters  $\beta$ ) is estimated based on the validation dataset  $\{(x_v, y_v)\}$ . The training procedure is repeated for multiple re-instantiations of training and validation partitions, which permits a statistical evaluation of the prediction performance. This validation procedure is also referred to as *internal validation*. A final model is trained on the entire dataset, yielding  $f(\cdot | \beta^*)$  with parameters  $\beta^*$ . For the *external validation*, this final model  $f(\cdot | \beta^*)$  is validated against a dataset acquired independently of the dataset used for model development. The same diagnostic performance metrics are computed as for the internal validation. In this thesis, bootstrap sampling of the validation dataset is used to estimate error bounds for these metrics.



**Figure 1.6:** Supervised learning of a (binary) classification model  $y_i = f(x_i|\beta)$ .

Different types of linear and nonlinear models exist with specific cost functions  $f(\cdot | \cdot)$  and/or assignment function  $g(\cdot | \cdot)$ . Most of the work of this thesis was based on regularized logistic regression models (LASSO-LR), for which  $f(x|\beta, \beta_0) = 1/(1 + e^{-\beta \cdot x - \beta_0})$ . Many other approaches are available, such as support vector machines, gradient boosting machines, or neural nets, which come with their own (implicit or explicit) cost functions.<sup>73</sup>

Compared to scoring schemes, typical classification models offer greater flexibility and exploit the available training data more effectively because they are based on more generic mathematical functions. Furthermore, some methods permit a probabilistic interpretation of a scoring or *cost function*  $f(x|\beta) = p[y = 0] \in [0,1]$ , which can be binarized by means of a decision threshold  $\delta$ :  $y = g(x|\beta, \delta) = f(x|\beta) < \delta$ . Contrary to many clinical scoring schemes, quantitative observations such as the aneurysm size or patient age are fed directly into the classifier without prior conversion into categorical variables (compare with Figure 1.5), resulting in more efficient data utilization.

Various attempts have been made to predict the rupture status of an aneurysm based on probabilistic prediction models. For instance, Xiang et al.<sup>75,76</sup> developed and externally validated multivariate logistic regression models based on hemodynamic and morphological parameter to predict the rupture status of aneurysms. Bisbal et al.<sup>77</sup> explored a rupture prediction model consisting of up to 294 predictors, including patient characteristics, morphological parameters and hemodynamic features. Even though the authors adopted dimensionality reduction techniques, the validity of the approach is questionable since the model was developed using only 157 aneurysms. Detmer et al.<sup>78,79</sup> developed a multivariate rupture prediction model using 26 clinical, hemodynamic and morphometric parameters from 1631 aneurysms. It is the currently most comprehensive and best validated model for rupture status prediction, achieving a prediction accuracy of 0.82 and an area under the receiver operating characteristic (ROC-AUC) of 0.84. Chapter 2 discusses several more predictive models with a focus on aneurysm morphology.

## 1.4 Quantitative assessment of aneurysms

Radiologists rely on angiographic imaging when diagnosing and assessing unruptured IAs. Some characteristics such as aneurysm diameter or neck size are measured interactively using dedicated software tools. However, most characteristics such as the aneurysm location and the overall impression of shape are characterized qualitatively. Various methods have been proposed for the quantitative and systematic analysis of imaging-derived data, with the goal to identify reliable markers for disease progression.

Direct *in vivo* imaging of the vessel wall is currently not deemed feasible for the accurate assessment of aneurysms due to the variable, potentially low thickness (30-400 $\mu$ m) of the lesioned aneurysmal wall.<sup>46</sup> Blood-suppressing MR imaging (black-blood MRI<sup>80,81</sup>) in combination with contrast enhancement permit to visualize pathological vessel wall segments, exploiting the increased permeability of the lesioned vessel wall for the contrast substance.<sup>82,83</sup> This so-called vessel wall enhancement has been associated with wall inflammation and therefore could provide quantifiable information about the IA disease status.<sup>84</sup> However, the method currently is not specific and robust enough to reliably indicate degenerative wall conditions.<sup>85</sup> Yet other approaches have aimed at measuring the wall motion in aneurysms, which has been associated with IA rupture, but is difficult to detect and quantify.<sup>86</sup>

Motivated by the observation that biomechanical processes are responsible for initiation and progress of IAs, a relatively large body of literature describe physical modelling approaches. In these, geometric models of the aneurysm and the surrounding arteries are extracted from angiographic images, which are used to simulate *in-silico* the dynamic processes in the vessel lumen. Computational fluid dynamics (CFD) simulations describe the blood flow through the arterial and aneurysmal lumen and estimate the forces exerted on the vessel wall. The comparison of different flow patterns in cohort studies led to the discovery of different hemodynamic indices indicative of aneurysm rupture or growth.<sup>87-91</sup> Other numerical models further incorporate the dynamic interactions between the pulsating blood and the vessel wall tissue (fluid-structure-interaction, constitutive tissue models, chemo-mechano-biological models).<sup>43,92,93</sup> Numerical models and quantitative characteristics derived from these models have a great potential for use as personalized risk indication.

However, they are based on a variety of assumptions that may influence the outcome significantly.<sup>94</sup>

Finally, aneurysms can be quantified by their size and shape. Various different morphometrics have been proposed, some of which will be discussed in more detail in Chapter 2. Ujie et al.<sup>95</sup> and Weir et al.<sup>96</sup> examined the aspect ratio (dome-to-neck ratio). Raghavan et al.<sup>97,98</sup> proposed a set of intuitive geometric indices to quantify aneurysm shape (non-sphericity, ellipticity, undulation, curvature indices). Dhar et al.<sup>99</sup> suggested metrics that relate the aneurysm dome to the parent vessel geometry (aneurysm angle, aneurysm-to-vessel size ratio). Other metrics compared the extent of the aneurysm dome to the neck geometry (volume-to-ostium area, bottleneck factor).<sup>100</sup> Lauric et al. contributed two novel methods for shape description, the centroid-radii model and writhe-based indices.<sup>101,102</sup> Rohde et al.<sup>103</sup> investigated shape irregularities based on 2D Fourier analysis. Millán et al.<sup>104</sup> employed 3D Zernike theory, a mathematical framework to decompose shapes into different modes. Berkowitz<sup>105</sup> reviewed recent morphological methods and augmented the list of geometric indices.

Metrics for aneurysm size have been used longest for the quantitative characterization of IAs. Aneurysm size usually refers to the largest diameter that can be inscribed to the aneurysm dome, but there are different definitions.

Numerous studies have reported size distributions of ruptured and unruptured aneurysms.<sup>60–63,106–108</sup> The international Study of Unruptured Intracranial Aneurysms<sup>61</sup> (ISUIA) examined the natural history of uIAs and the risks of treatment in an international cohort of 4060 prospectively selected patients. Potential guidelines based on size have been discussed controversially.<sup>61–63</sup> Quantitative metrics other than size so far have received relatively little attention, even though alternative metrics have been shown to be substantially more informative with respect to disease status than size alone.

## 1.5 Purpose and goals

Irregularly shaped aneurysms are more likely to grow. In a natural history study, Lindgren et al.<sup>29</sup> showed that in a cohort of 4074 (predominantly Finnish) patients presenting 5814 aneurysms (2718 ruptured), 92% of the ruptured aneurysms were associated with irregular shape, while only 22% of the unruptured aneurysms were considered irregular. Related studies from Björkman et al.<sup>109</sup> or Räisänen et al.<sup>110</sup> on patients with multiple aneurysms and young patients confirm this observation. The study of pathobiological mechanisms by Frösen et al.<sup>41</sup> further linked irregular shape with increased histological heterogeneity and wall instability. In a survey among clinical experts (UIATS), irregular shape was considered as one of the key factors in decision making.<sup>59</sup>

Evidence suggests that the shape of the aneurysm, and in particular morphological irregularity, are reflective of the disease status. Because the aneurysmal disease presents often with an unclear indication for treatment, clinicians are in need of prognostic tools that facilitate the decision making.

The purpose of this dissertation was to quantify the morphology of IAs as seen in 3D medical imaging and to assess its informativeness with respect to the aneurysm disease status using data-driven methods. Specifically, the principal goals of this thesis were: A) establish and evaluate a multicohort database of aneurysm morphology; B) benchmark state-of-the-art morphometrics by their predictive capacity for the aneurysm disease status; C) explore and quantify the subjective assessment of aneurysm shape by human raters; D) relate perceived and quantitative morphology to other clinically relevant factors such as the anatomical localization of the aneurysm, patient age or smoking status.

## 2 Morphometric analysis of aneurysm shape

This chapter focuses on the quantitative morphology of aneurysms. The AneuX morphology database is introduced, a multicentric database built and curated in the context of this dissertation. After a review of existing methods to describe the shape of IAs, these methods are applied to the AneuX morphology database and evaluated by their ability to correctly predict the aneurysm rupture status.

*Contributions: The conceptualization and implementation of the study are my own for the most part. Sabine Schilling supported me with methodological advice. Philippe Bijlenga provided oversight of the AneuX and provided the clinical data. The extraction of 3D geometries has been conducted by Diana Sapina in Zurich, as well as Vitor Mendes Pereira and Rafik Ouared with their team in Geneva. Ueli Ebnöther has contributed to the computation and analysis of ZMIs<sup>111</sup>. Preliminary results of this work have been presented at the European Congress on Computational Methods in Applied Sciences, ECCOMAS 2016 in Crete, Greece. The results reported here have been re-evaluated on the updated dataset in 2020.*

## 2.1 Abstract

**Background.** To date, it remains difficult for clinicians to reliably assess the disease status of intracranial aneurysms. As an aneurysm’s 3D shape is strongly dependent on the underlying formation processes, it is believed that the presence of certain shape features mirror the disease status of the aneurysm wall. Currently, clinicians associate irregular shape with wall instability. However, no consensus exists about which shape features reliably predict instability. In this extensive study, we present a classification pipeline that seeks to identify those shape features that offer the highest predictive power of aneurysm rupture status.

**Methods.** 3D models of aneurysms were extracted from medical imaging data (3D rotational angiographies) using a standardized protocol. A variety of established representations of the 3D shape were calculated for the extracted aneurysm segment. These included geometry indices such as undulation, ellipticity and non-sphericity, writhe- and curvature-based metrics, and indices based on Zernike moments. Statistical learning methods were applied to find associations between shape features and aneurysm disease status. This processing pipeline was applied to a clinical dataset of 750 aneurysms registered in the AneuX morpho database.

**Results.** Non-sphericity index  $NSI$  ( $AUC = 0.80$ ), normalized Zernike energies  $Z_N^{\text{surf}}$  ( $AUC = 0.80$ ) and the modified writhe-index  $\bar{W}_{\text{mean}}^{L_1}$  ( $AUC = 0.78$ ) exhibited the strongest association with rupture status. The combination of predictors further improved the predictive performance (without location:  $AUC = 0.82$ , with location  $AUC = 0.87$ ). The anatomical location was a good predictor for rupture status on its own ( $AUC = 0.78$ ). Different protocols to isolate the aneurysm dome did not affect the prediction performance. We identified problems regarding generalizability if trained models are applied to datasets with different selection biases.

**Conclusions.** Morphology presented a clear indication for the aneurysm disease status. Because rupture rates vary with aneurysm location, predictive models should be compared to a baseline model using only location as predictor.

## 2.2 Introduction

Intracranial aneurysms (IAs) are a common disease of cerebral arteries with a complex pathobiology. Confronted with an increased rate of incidentally diagnosed IAs, clinicians are in need of a marker for disease instability to better balance the risks of rupture against the risks of treatment. This marker could ideally be acquired non-invasively in the context of routine examinations.

In this context, aneurysm shape has been proposed as a candidate for such a marker for several reasons. Firstly, pathophysiological evidence suggests that structural changes in the aneurysmal wall are linked to macroscopic deformations of the wall.<sup>41,112</sup> The presence of vasa vasorum or the formation of organized luminal thrombosis, which frequently accompany IAs, do also leave an imprint in the vascular lumen as seen in contrast enhanced imaging.<sup>41,42,113</sup>

Secondly, shape can be seen as an expression of hemodynamic flow patterns. The local geometry of aneurysms governs the blood flow and the fluidic forces exerted on the vessel wall. Variations of these forces have been associated with wall damage, aneurysm initiation and growth.<sup>42,114–117</sup> Shape and flow dynamics are interrelated: changes in morphology influence the flow patterns in the vicinity of the aneurysm, which in turn can stimulate wall remodeling that eventually can lead to new morphological variations.<sup>42</sup>

Thirdly, imaging is a non-invasive utility readily available in clinics. It is the primary source of information for the diagnosis and treatment of IAs. In addition to its location, the anatomical embedding and the size of an aneurysm, radiologists can also infer its shape from medical images.

This wealth of evidence is contrasted by the paucity of guidelines that address morphology quantitatively. To date, the assessment of aneurysm shape is based mainly on the subjective opinion by the clinicians.

The purpose of this study is to investigate different methods to quantify aneurysm morphology and to examine how shape relates to the disease status. We carefully discuss the limitations of the complete processing pipeline and give possible reasons why morphological metrics have not established themselves in clinical practice yet.



## 2.2.1 Related works

Several methods to analyze the 3D shape of aneurysms have been proposed and used to predict aneurysm disease status. Raghavan and Ma<sup>97,98</sup> as well as Berkowitz<sup>105</sup> reviewed and proposed intuitive and computationally convenient geometric indices. Indices such as the size ratio (SR), the aspect ratio (AR), the non-sphericity index (NSI) and ellipticity index (EI), which all capture particular characteristics of an aneurysm's shape, have been associated with aneurysm rupture<sup>75,95,97,99,118</sup>. Metrics such as the SR, the vessel angle and the inclination angle, aim at incorporating the parent vessel geometry and its relationship to the aneurysm dome.<sup>99</sup> Other metrics are based on local surface properties that are aggregated by means of surface norms or histogram-based statistics. For instance, curvature-based metrics as suggested by Raghavan and Ma<sup>97,98</sup> fall into this category. Lauric et al. proposed the centroid-radii model<sup>101</sup> as well as the writhe-number based characterization of the aneurysm surface<sup>102</sup> that were both applied to rupture status prediction. Other approaches choose a different mathematical representation of surfaces (or volumes), as done by Millán et al.<sup>104</sup>. Based on the groundwork done by Canterakis<sup>119</sup> and Novotni et al.<sup>120</sup>, they suggested to describe aneurysms by means of Zernike moment invariants (ZMIs) and use them to compare the 3D geometries of aneurysms and the surrounding vasculature. As of now, Detmer et al.<sup>78</sup> have developed the largest multifactorial model for rupture status prediction. Their model is based on data of 1631 aneurysms, and accounts for predictors describing morphometric, hemodynamic and patient parameters. They also successfully validated their data on a subset of the HUG database.<sup>79</sup>

Table 2.1 presents a selection of recent studies examining the predictive capability of aneurysm morphology. If multiple parameters were examined, we report the model configuration with the highest predictive capacity. If multiple scores were reported for the same model configuration, we opted for the least optimistic (e.g., scores achieved on model validation instead of scores achieved during model development).

However, the overview of Table 2.1 does not adequately reflect the heterogeneities in the data and methods used for these studies. For instance, not all studies worked with consecutively recruited patients, or one study (Liu et al.<sup>121</sup>) distinguished between stable and unstable aneurysms, as opposed to

ruptured and unruptured aneurysms in all other studies. To highlight these methodological differences, the author assessed the studies by three loose quality criteria: (1) a complete reporting as suggested by the initiatives on the transparent reporting of diagnostic accuracy studies (STARD)<sup>122</sup> or multivariate predictive models (TRIPOD)<sup>123</sup>; (2) the dataset size in relation to the total number features examined by the study; (3) the use of a validation scheme that examines the generalizability of diagnostic models.

In the following, we focus on aneurysm morphology, which is also reflected by the selection of works in Table 2.1. It is worthwhile to notice that some studies examined morphology in combination with hemodynamical metrics or patient characteristics (such as sex or age).

Study	Dataset			Evaluation			Results		Comments	
	Year	n	y=0 y=1	Predictor(s)	Method	Quali.	PA	AUC		
Weir et al. <sup>96</sup>	2003	767	361	406	Aspect ratio	uni	ROC	++	-	0.79 Different imaging modalities
Raghavan et al. <sup>97</sup>	2005	27	18	9	Non-sphericity	uni	ROC		-	0.83 CT-angios only, very small dataset
Dhar et al. <sup>99</sup>	2008	45	25	20	Non-sphericity	uni	ROC		-	0.76
Dhar et al. <sup>99</sup>	2008	45	25	20	Size ratio	uni	ROC		-	0.83
Xiang et al. <sup>75,76</sup>	2011	119	81	38	Size ratio	uni	ROC	++	-	0.83 Model validated on new data
Hoffman et al. <sup>124</sup>	2019	331	172	159	Size ratio	uni	ROC	+	-	0.70
Lauric et al. <sup>101</sup>	2011	58	35	23	Centroid radii	uni	LR-ROC	+	0.80	0.85 Sidewall aneurysms only
Lauric et al. <sup>102</sup>	2010	58	35	23	Writhe	multi	LR-ROC		0.87	- Sidewall aneurysms only
Millán et al. <sup>104</sup>	2007	55	24	31	ZMI	multi	LDA		0.80	-
Liu et al. <sup>121</sup>	2019	420	*296	*124	GIs	multi	LR-ROC	+	-	0.85 The only study predicting stability
Niemann et al. <sup>125</sup>	2018	100	57	43	GIs	multi	GPLS	+	0.69	0.70 Small dataset, large number predictors
Detmer et al. <sup>78,79</sup>	2018	1631	1139	492	GIs+other	multi	LR-ROC	+++	0.76	0.82 Group lasso LR, validated on external data

**Table 2.1:** Some results from recent morphological studies on IA rupture prediction. The datasets are summarized by the total number of aneurysms  $n$ , and the number of unruptured/ruptured (or stable/unstable \*) cases. The adopted methods are outlined by the predictor(s) used to train univariate or multivariate rupture prediction models, and the classification/discrimination method: ROC (receiver-operating characteristic analysis), LR-ROC (ROC analysis after training logistic regression (LR) models), LDA (linear discriminant analysis) or GPLS (generalized partial least squares). The author rated the study quality by casting “+” marks (1) for a complete reporting in the spirit of the STARD initiative<sup>122</sup>, (2) for a reasonable dataset size, (3) for validating the model externally. Further abbreviations: PA – prediction accuracy; AUC – area under the ROC curve; GIs – geometry indices; CFD – computation fluid dynamics.

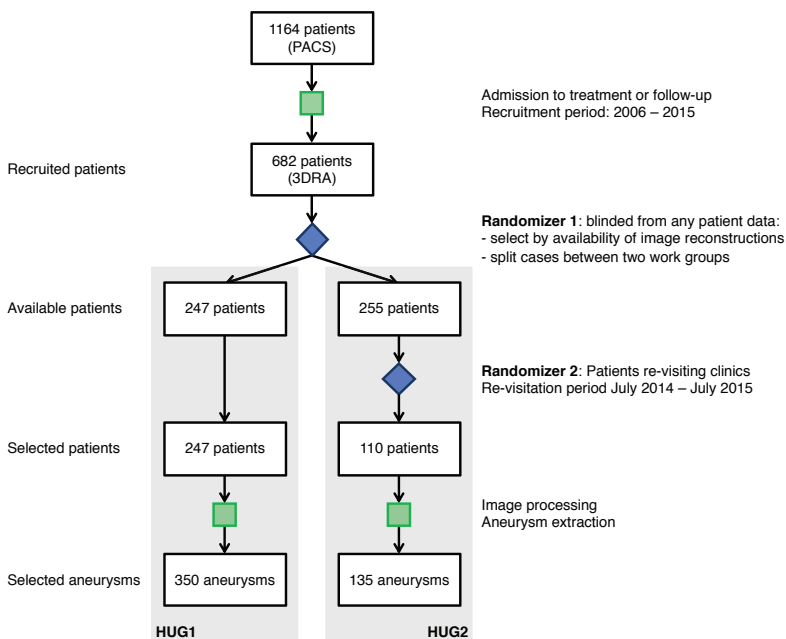
## 2.3 Material and methods

### 2.3.1 Imaging and patient data

Between September 2006 and July 2015 data from 1164 patients were collected prospectively and consecutively at the Geneva University Hospital (HUG), continuing the data collection scheme initiated and implemented during the @neurIST project<sup>33,126</sup>. A significant proportion of the cohort was only followed up using magnetic resonance imaging (MRI) or computed tomography (CT) imaging. 682 patients were identified as being at risk or suffered from a ruptured aneurysm and were therefore investigated by 3D rotational angiography (3DRA). The data were split randomly between two independent teams of data curators by ignoring any prior information about the cases (Figure 2.1). In 180 cases no reconstructions were available for the 3DRA and therefore were excluded. While team 1 processed the 3DRA from all assigned 247 patients, team 2 selected from the total of 255 the 110 patients that visited the HUG for aneurysm repair or post-treatment follow-up examinations (scheduled 6 weeks, 3 months, 1 year, 2 years or 5 years after treatment) during a fixed time frame of one year (Figure 2.1). Both teams processed only angiograms of aneurysms before treatment. If multiple examinations were available, the oldest image was included.

In addition to angiographic data, the datasets included sex, age, rupture status and anatomical location (per aneurysm) for all the cases.

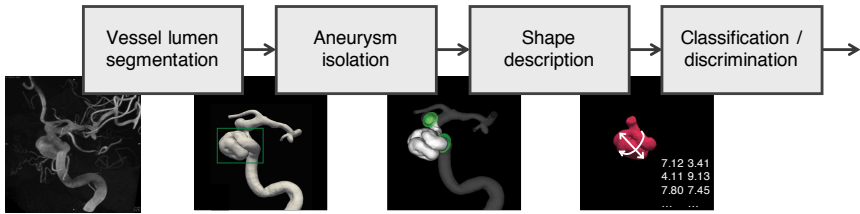
To test whether our findings generalize to other datasets, we expanded the database by two external datasets: From the @neurIST project<sup>126,127</sup>, we included 164 aneurysms (151 patients) acquired in Barcelona, Geneva and Sheffield. From the publicly available Aneurisk database<sup>128</sup> we used 101 aneurysms (97 patients) retrospectively collected at the Ca' Granda Hospital, Niguarda, Milano between 2002 and 2006.<sup>129</sup> The data processing was described by Piccinelli et al.<sup>130,131</sup>.



**Figure 2.1:** Acquisition process for the HUG dataset. Starting from the same set of recruited patients, two teams of data curators segmented the vascular structures in 3DRA images following similar protocols.

	HUG1	HUG2	@neurIST	Aneurisk	Overall
Number of patients	247	110	151	97	605
Sex	F: 197 (77%) M: 57 (33%)	F: 81 (74%) M: 29 (26%)	F: 109 (67%) M: 42 (33%)	F: 61 (63%) M: 36 (37%)	F: 445 (73%) M: 164 (27%)
Patient age in years (mean±SD)	F: 56.4±14.0 M: 54.3±13.8	F: 54.4±12.7 M: 50.6±12.2	F: 53.4±12.2 M: 49.8±10.6	F: 53.6±15.2 M: 55.5±10.8	F: 55.0±13.6 M: 52.8±12.2
Number of sIAs	350	135	164	101	750
Ruptured / unruptured	R: 87 (25%) U: 263 (75%)	R: 41 (30%) U: 79 (59%)	R: 89 (54%) U: 75 (46%)	R: 44 (44%) U: 57 (56%)	R: 261 (35%) U: 474 (65%)

**Table 2.1:** Summary of the cases included into the AneuX morphology database, stratified by data source. Note that for *HUG2*, the rupture status of 15 aneurysms was not available. Abbreviations: SD – standard deviation; sIAs – saccular intracranial aneurysms

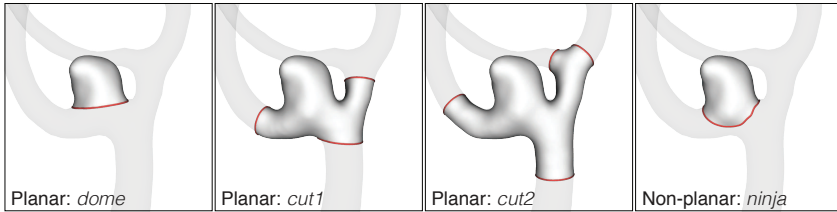


**Figure 2.2:** Overview of the data processing pipeline applied to all aneurysms in the AneuX morphology database.

### 2.3.2 Data processing

3D models of aneurysms were extracted from 3DRA images according to the same processing pipeline sketched in Figure 2.2. The exact processing varied slightly for the different data sources, but generally followed the protocol proposed by the @neurIST consortium<sup>127</sup>, which puts a special emphasis on the standardization of medical data collection.

In a first step, the data operators extracted 3D surface models of the aneurysm and surrounding vascular segments with the help of either @neuFuse<sup>126</sup> (HUG1, @neurIST) or GIMIAS<sup>132</sup> (HUG2), two highly related software tools developed for this type of problem. The primary segmentation method in use was non-parametric Geodesic Active Regions (GAR), a robust method optimized for the extraction of vascular structures in angiographies free of any tuning parameters<sup>133,134</sup>. The segmentation and surface extraction methods for the Aneurisk data are described in Antiga et al.<sup>135</sup> and Piccinelli et al.<sup>130,131</sup>. Similar to non-parametric GAR, it makes use of implicit deformable models, but relies on a semi-automatic initialization strategy with the goal to robustly segment the vessels of interest as indicated by the user. All methods make use of marching cubes<sup>136</sup> to construct a triangular surface mesh from the binary segmentation images. The data curators assessed the resulting meshes and, if required, manually fixed formations that they identified as segmentation or imaging artifacts using a mesh editor.



**Figure 2.3:** Cut configurations of the AneuX morphology database. Cut lines are shown in red.

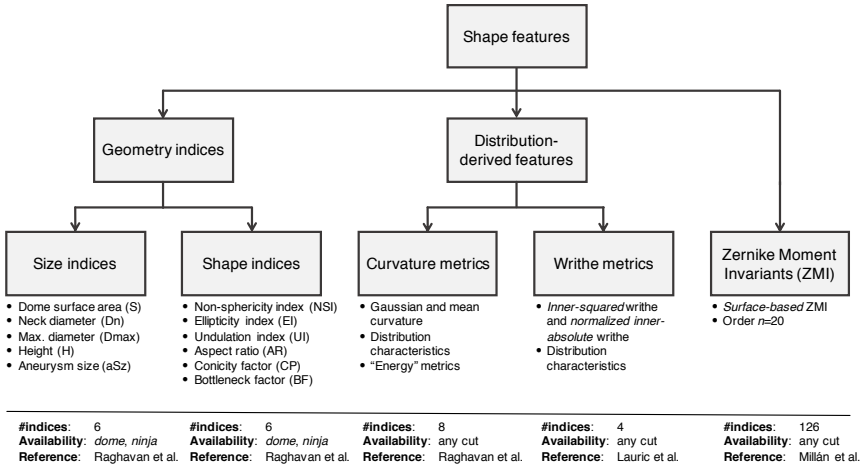
In a second step, an operator isolated the aneurysms using planar and non-planar cuts following a set of cut rules (Figure 2.3). The *dome* cut disjoins the aneurysm dome from the parent vasculature by one single planar cut. For *cut1* and *cut2*, cut planes are placed perpendicularly to the local centerline in one or two vessel diameters distance from the *dome*. If the rule could not be applied because of an adjacent bifurcation, the closest valid cut before or after the bifurcation was chosen. The non-planar *ninja* cut was placed along the boundary (the so-called neck) of the aneurysmal protrusion. Like the *dome* cut, a *ninja* cut captures the aneurysm dome, but permits a more natural isolation of the aneurysm as assessed by the operator. Related studies made use of similar isolation schemes (*dome*: Ma et al.<sup>98</sup>, *cut1*: Berti et al.<sup>127</sup>, *ninja*: Mut et al.<sup>137</sup>). For this processing step, an in-house cut tool was used.

To ensure similar mesh properties across different dataset sources and to sanitize the meshes from defects such as minuscule cracks, singular edges or orphaned cells<sup>138</sup> we re-meshed all geometries using VMTK<sup>135</sup>. The chosen target cell area of  $0.05\text{mm}^2$  corresponds roughly to the resolution of typical 3DRAs (with voxel sizes around  $300\mu\text{m}$ ). We applied Taubin smoothing<sup>139,140</sup> in preparation for the morphometric description of the aneurysm geometries. We employed the implementation provided by VTK<sup>141</sup> with passband 0.2 and 30 iterations. Note that Taubin’s method is a topological smoother that disregards spatial information. The smoothing parameters therefore require readjustment if the mesh resolution varies.

### 2.3.3 Morphometric description of the aneurysms

*Morphological features* or *morphometrics* quantitatively describe the shape of 3D objects, ideally meeting the following requirements: (1) invariance to translation and rotation of the reference coordinate frame, (2) efficiency in encoding relevant morphological information, and (3) robustness with respect to imaging or surface mesh quality. In this study we examined if the candidate features fulfill requirement (2) and (3). Requirement (1) was satisfied by all candidates.

The shape features considered for this study can be grouped into three different categories (Figure 2.4). *Geometry indices* (GIs) quantify specific geometrical properties of the aneurysm and are typically scalar-valued. *Distribution-derived features* include information on the variation of morphological properties across points (or mesh cells). *Zernike Moment Invariants* are based on a transformed representation of the 2D manifold allowing to derive a set of coefficients capturing the entirety of the geometry under observation.



**Figure 2.4:** Overview of the shape features considered for this study. Note that *dome* and *ninja* cuts as well as *cut1* and *cut2* are used interchangeably. If all cut configurations are taken into account (*dome/ninja*, *cut1/cut2*), a maximal number of 288 indices (12 GIs + 24 DDFs + 252 ZMIs) was computed. Most investigations were based on the *dome* cut alone, for which 150 (12GIs + 12 DDFs + 126 ZMIs) indices were available. Abbreviation: DDF – distribution-derived features



### 2.3.3.1 Geometry indices

Geometry indices (GIs) are designed to capture very specific properties of a 3D shape. Advantages of GIs are their geometric interpretability and their low computational complexity. For these reasons, some GIs such as the aneurysm size, neck diameter or aspect ratio are already used in the clinical context, where they are computed approximately from manual measurements. In this study, GIs previously reviewed by Ma, Raghavan and Berkowitz<sup>97,98,105</sup> are included: 6 indices for size (dome volume, dome surface area, neck diameter, maximum diameter, aneurysm height, aneurysm size) and 6 indices for shape (aspect ratio, ellipticity index, non-sphericity index, undulation index, conicity parameter, bottleneck factor). Definitions are given in Appendix 2.A.

Several metrics require a reference plane at the aneurysm neck. The intersection of this plane with the aneurysm is often referred to as *ostium*. For *dome* cuts, this reference plane coincides with the cut-plane. For the non-planar *ninja* cuts, we defined the neck plane as the best-fit plane through the cut line.

### 2.3.3.2 Distribution derived features

Distribution-derived features characterize the variation of local shape properties evaluated across points  $\mathbf{p}$  of a surface  $\mathcal{S}$ . For this study, we considered two such properties, *curvature* and *writhe*, both of which have been used already to characterize IAs<sup>97,98,102</sup>.

The *curvature* at a point  $\mathbf{p} \in \mathcal{S}$  can be expressed by means of Gaussian curvature  $K_G(\mathbf{p})$  and mean curvature  $K_M(\mathbf{p})$ . We used VTK<sup>141</sup> to compute the local curvature values for discrete surface meshes, which we subsequently aggregated as described in Appendix 2.B.1. A total of 8 different curvature-derived features are evaluated, which include the well-known metrics for total Gaussian and Mean curvature GLN and MLN<sup>98</sup>, and two novel metrics.

The *writhe number* measures surface asymmetries and “twisting forces” as seen from a surface point  $\mathbf{p}$ .<sup>102</sup> Originally introduced in knot-theory to characterize curves, the writhe number was generalized by Lauric et al.<sup>102</sup> for 3D surfaces. We distinguished between writhe  $W^{L_2}(\mathbf{p})$  and *normalized* writhe  $\bar{W}^{L_1}(\mathbf{p})$ , resulting in a total of 4 different writhe-based shape features (Appendix 2.B.2). Curvature and writhe features were evaluated for all cut types (*dome*, *ninja*, *cut1*, *cut2*).

### 2.3.3.3 Zernike Moment Invariants

3D Zernike moments (ZMs) and the Zernike moment invariants (ZMIs) were first described by Canterakis<sup>119</sup> and applied by Novotni et al.<sup>120</sup> in the context of 3D shape retrieval. Millán et al.<sup>104</sup> introduced the ZMI for the assessment of intracranial aneurysm morphology.

The goal of 3D Zernike transformation is to describe an input geometry in terms of the so-called Zernike basis: a set of (complex-valued) polynomials strongly related to spherical harmonics. The Zernike theory resembles Fourier theory in that a forward transformation yields a set of coefficients (the ZMs) that can be used as weights to reconstruct the original surface through a weighted summation of its basis functions (inverse transformation). The method permits to decompose a geometry into morphological “modes” of gradually increasing complexity. The maximum mode order  $n$  is chosen so as to capture enough morphological details by the ZMs. ZMs can be made invariant to translation and isotropic scaling,<sup>120</sup> but only an additional transformation yields the rotation invariant ZMIs, forming a viable shape descriptor (see Appendix 2.C).

For this study, we included ZMIs up to order  $n = 20$ , corresponding to a shape descriptor of 121 independent values. In addition, we computed ZMI-based “energies”  $Z_N^{\text{surf}}$  for five different maximal orders (Appendix 2.C). We limited the use to *surface-based* ZMIs (as opposed to *volume-based* ZMIs<sup>104,142</sup>) because they carried a slightly stronger signal in our experiments.

### 2.3.3.4 Shape descriptors and aneurysm location

We computed the above features for all 750 aneurysms and available cut types. We based our analysis primarily on features computed for *dome* cuts, unless otherwise noticed (Figure 2.4). Any collection of one or more morphological features is termed *shape descriptor*.

Motivated by the fact that morphology and the associated risks vary with the anatomical location of IAs<sup>28,33,63,78</sup>, we have added location as the single non-morphometric predictor to our feature pool. The locations are specified in Table 2.3. The categorical variable was represented in the numerical feature space using  $n_{\text{locs}}$  one-hot-encoded dummy variables, which are all zero, except for the one representing the sample’s location.

### 2.3.4 Analysis of the diagnostic capability of aneurysm shape

We examined the morphological features for a relation with the aneurysm rupture status. In a first step, we assessed the univariate properties of all morphometric features, and then trained and validated multivariate classification models for the prediction of the aneurysms' rupture status. Our setup was designed primarily as a benchmark to identify feature configurations that have the strongest association with rupture status.

We evaluated unpaired Student's  $t$ -tests between ruptured and unruptured aneurysms for each of the 150 features. The significance level was set to  $\alpha = 0.05/d$ , with Bonferroni corrector  $d = 150$  to correct for multiple testing.<sup>143</sup>

For better comparability of the results, we applied the same statistical learning scheme for both univariate and multivariate shape descriptors. All reported results are based on regularized (LASSO) logistic regression models, but were generally reproducible (with minor variations) also with other classification methods such as linear and non-linear support vector machines (SVM), decision tree-based methods (in particular LightGBM<sup>144</sup>) and neural nets.

We centered and scaled the morphometric features to 0-mean and a standard deviation of 1, which improved convergence rates during classifier training. The submatrix of dummy variables was not standardized. For multivariate models, we optionally reduced the feature space dimensionality by means of a principle component analysis (PCA), selecting the  $k$  first principle components retaining 90% of the total variance in the (training) data.

A feasible value for the regularization strength  $\lambda$ , the only tuning parameter of the LASSO cost function, was identified using a grid search. All logistic regression models were validated using a 5-fold cross-validation (CV) scheme with 20 repetitions, resulting in a total of 100 model evaluations. To avoid information leakage between training and test data, the parameters for feature space standardization and optional PCA were computed on training data only.

For all of the 100 models trained in this CV setup, we evaluated the ROC-AUC (the area under the receiver operating characteristic (ROC) curve) using the test data and report mean and standard deviation. We further calculated prediction accuracy, sensitivity, specificity and Cohen's  $\kappa$  at the optimal

classifier threshold, characterized by the point on the ROC curve closest to the point (0,1). Cohen’s  $\kappa$  baselines the model predictions with by-chance agreement, which is considered more robust with regard to imbalanced datasets than other metrics.

All training and (internal) validation of the classification models were performed using HUG data only. In a final step, we (externally) validated the multivariate prediction models using the @neurIST and Aneurisk datasets. We report the average AUC and Cohen’s  $\kappa$ .

This learning pipeline was applied to all univariate and multivariate models. Table 2.2 summarizes the multivariate models considered in this study. Besides the maximal model including all 150 morphometric features (with PCA), a multivariate model was assembled by selecting the best-performers in the univariate model with a  $AUC > 0.7$ .

Identifier	$d$	PCA	Details
MAX	10	yes	All morphometric features except for ZMIs of order $n \geq 10$
MAX+loc.	11	yes	Same as MAX, extended by anatomical location
BUP	12	no	Independent selection of the best univariate performers with an $AUC > 0.7$
BUP+loc.	13	no	Same as BUP, extended by anatomical location
NSI+loc.	2	no	NSI and location
$Z_6^{\text{surf}}$ +loc.	2	no	Normalized ZMI energy for maximum order 6
LOC	1	no	Location only

**Table 2.2:** Description of multivariate models considered in this study and their number  $d$  of predictors. BUP refers to the “best univariate performers” summarized in Table 2.4. Note that the categorical location predictor expands to 12 (hot-one-encoded) dummy variables.

### 2.3.5 Software tools

Most computations were performed in Python 3.6. Only the implementation of the ZMI is based on C++ code. For the mesh-based operations, we employed VTK (the Visualization Toolkit<sup>141</sup>) and VMTK (the Vascular Modelling Toolkit<sup>135</sup>). Several utilities to develop, compute and analyze morphometric descriptors have been assembled in our Geometric Modelling Toolkit (GMTK). For the statistical analysis and machine learning, we relied on the Python packages SciPy<sup>145</sup> (v1.3), scikit-learn<sup>146</sup> (v0.22) and statsmodels<sup>147</sup> (v0.11).

## 2.4 Results

We report here the results for our basic dataset configuration using the two *HUG* datasets consisting of 470 aneurysms (128 ruptured, 342 unruptured, Table 2.1), and the morphometric features evaluated for the *dome* cut. Statistical learning was performed using LASSO-LR and 5-fold cross-validation with 20 repetitions, resulting in 100 model evaluations. Deviations from this setup are marked explicitly. The dataset for external validation consisted of 265 cases from the @neurIST and Aneurisk projects. Table 2.3 summarizes all data used, stratified by aneurysm location and rupture status.

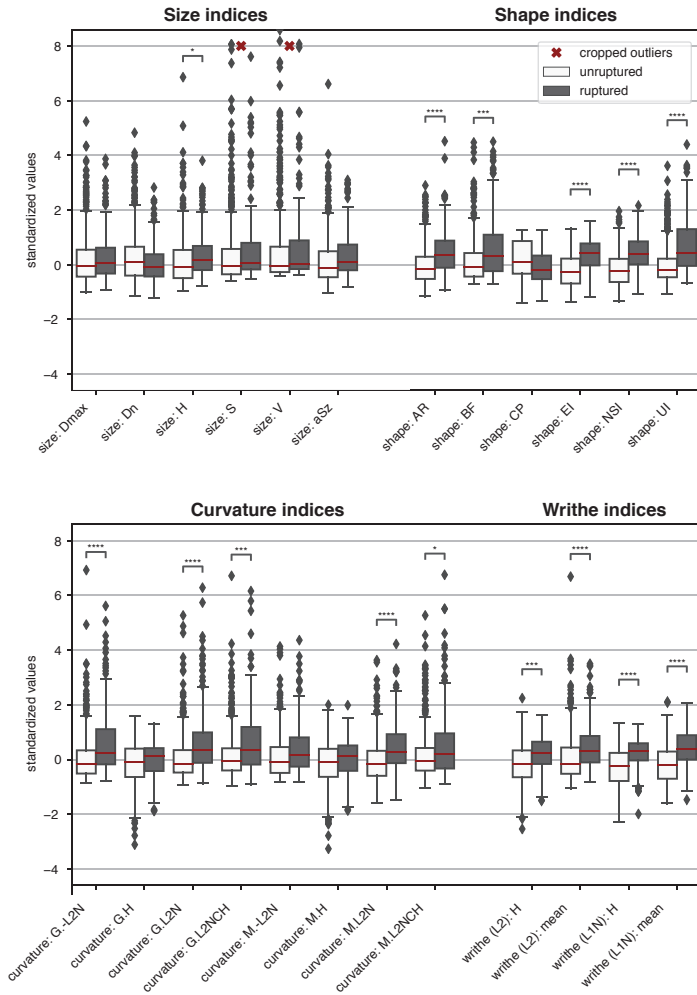
### 2.4.1 Univariate analysis

Figure 2.5 visualizes the morphometric data on the aneurysms stratified by rupture status (ZMI data was excluded for lack of space). Asterisks indicate if the class differences between the sample means were statistically significant.

Table 2.4 presents the predictive accuracy of the 12 best performing features plus aneurysm size ( $aSz$ ). For the sake of brevity, we refer to AUC as the principal comparison accuracy metric. Values for AUC ranged from  $0.80 \pm 0.06$  (for *NSI*) to  $0.40 \pm 0.08$  (for volume *V*).

Location	HUG1		HUG2		@neurIST		Aneurisk		Total		
	U	R	U	R	U	R	U	R	U	R	$\Sigma$
MCA bif	57	8	19	4	19	21	14	9	109	42	151
PCoMA	21	17	9	10	16	38	8	11	54	76	130
ACoMA	33	43	8	11	0	1	6	17	47	72	119
ICA oph	48	1	16	3	21	5	18	2	103	11	114
ICA bif	15	1	5	0	6	9	2	0	28	10	38
MCA	23	1	5	1	3	0	4	0	35	2	37
BA tip	11	4	4	3	2	7	3	3	20	17	37
ICA cav	28	0	3	0	2	0	1	0	34	0	34
ACA	9	5	5	3	1	3	0	1	15	12	27
VB other	10	2	3	4	1	2	0	1	14	9	23
ICA chor	7	4	2	1	3	2	1	0	13	7	20
PCA	1	1	0	1	1	1	0	0	2	3	5
<b>Total</b>	<b>263</b>	<b>87</b>	<b>79</b>	<b>41</b>	<b>75</b>	<b>89</b>	<b>57</b>	<b>44</b>	<b>474</b>	<b>261</b>	<b>735</b>

**Table 2.3:** Summary of all datasets stratified by aneurysm location and rupture status. See section Abbreviations and Figure 1.2 for an overview of location identifiers. U – unruptured; R – ruptured.



**Figure 2.5:** Boxplots summarizing the data of the 470 HUG samples stratified by rupture status. For easier comparison, each metric was centered and scaled such that the overall median and interquartile range mapped to 0 and 1, respectively. ZMI data was omitted. Single asterisks \*, double asterisks \*\*, triple asterisks \*\*\* and quadruple asterisks \*\*\*\* indicate significance for  $t$ -tests at the  $\alpha = 0.05, 0.01, 0.001$  and  $0.0001$  level, under consideration of the Bonferroni correction for multiple testing (correction factor 150)

Univariate models (internal validation, cut dome)							
Category	Predictor	AUC	Accuracy	Sensitivity	Specificity	Precision	Kappa
Shape	$NSI$ , non-sphericity	<b>0.80±0.05</b>	0.73±0.04	0.75±0.08	0.72±0.05	0.50±0.05	0.41±0.08
ZMI	norm. energy $Z_6^{\text{surf}}$	<b>0.80±0.05</b>	0.74±0.04	0.75±0.08	0.74±0.06	0.52±0.06	0.43±0.09
ZMI	norm. energy $Z_3^{\text{surf}}$	<b>0.78±0.04</b>	0.73±0.04	0.61±0.09	0.78±0.05	0.51±0.06	0.36±0.09
Writhe	$\overline{W}_{\text{mean}}^{L_1}$	<b>0.78±0.04</b>	0.72±0.04	0.71±0.09	0.72±0.05	0.49±0.05	0.37±0.07
Shape	$UI$ , undulation	<b>0.77±0.05</b>	0.74±0.04	0.61±0.10	0.79±0.05	0.52±0.06	0.38±0.09
Curvature	$GLN$	<b>0.75±0.05</b>	0.71±0.04	0.59±0.08	0.76±0.05	0.48±0.06	0.32±0.08
Curvature	$MLN$	<b>0.75±0.05</b>	0.69±0.04	0.63±0.08	0.71±0.05	0.45±0.05	0.31±0.08
Shape	$AR$ , aspect ratio	<b>0.75±0.05</b>	0.70±0.04	0.61±0.11	0.74±0.05	0.46±0.05	0.32±0.09
ZMI	$ZMI_{3,1}^{\text{surf}}$	<b>0.74±0.05</b>	0.66±0.04	0.71±0.09	0.64±0.06	0.42±0.04	0.29±0.07
ZMI	$ZMI_{5,1}^{\text{surf}}$	<b>0.72±0.05</b>	0.66±0.05	0.68±0.09	0.66±0.06	0.43±0.05	0.28±0.09
Writhe	$W_{\text{mean}}^{L_2}$	<b>0.72±0.05</b>	0.70±0.04	0.58±0.10	0.74±0.05	0.46±0.06	0.30±0.09
Size	$aSz$	<b>0.64±0.05</b>	0.65±0.04	0.46±0.10	0.72±0.06	0.38±0.06	0.16±0.09

**Table 2.4:** *Internal validation* results of the best univariate classification models, ordered by decreasing ROC-AUC. We only considered models with an AUC > 0.7 and removed highly correlated features (with a Pearson correlation  $\rho > 0.95$ ). The list is extended by the best performing size metric: aneurysm size. We report mean and standard deviation (mean ± std) for 100 model evaluations of our cross-validation scheme. The data compares to the blue lines in Figure 2.6.

Multivariate models (internal validation, cut dome)							
Category	#	AUC	Accuracy	Sensitivity	Specificity	Precision	Kappa
Max. model (+PCA)	10*	<b>0.82±0.05</b>	0.74±0.04	0.75±0.09	0.74±0.04	0.52±0.05	0.43±0.08
Best univariate	12	<b>0.82±0.05</b>	0.74±0.04	0.75±0.08	0.74±0.05	0.52±0.05	0.43±0.08
Location only	12	<b>0.78±0.04</b>	0.69±0.04	0.78±0.10	0.65±0.05	0.46±0.04	0.36±0.07
Max. model + loc. (+ PCA)	22*	<b>0.87±0.04</b>	0.79±0.04	0.78±0.08	0.80±0.04	0.60±0.06	0.53±0.08
Best univ. + location	24	<b>0.87±0.04</b>	0.80±0.04	0.77±0.09	0.80±0.05	0.60±0.06	0.53±0.09
NSI and location	13	<b>0.87±0.04</b>	0.79±0.04	0.79±0.08	0.79±0.04	0.59±0.06	0.52±0.08
$Z_6^{\text{surf}}$ and location	13	<b>0.87±0.04</b>	0.78±0.04	0.76±0.10	0.79±0.05	0.57±0.06	0.50±0.10

**Table 2.5:** *Internal validation* results of the multivariate classification models. Column # indicates the dimensionality of the models' feature space, or the (average) number of dimensions retained after PCA if marked with an asterisk. Note that location adds 12 dummy features (one for each location) but represents one single predictor. The data compares to the blue lines in Figure 2.6.

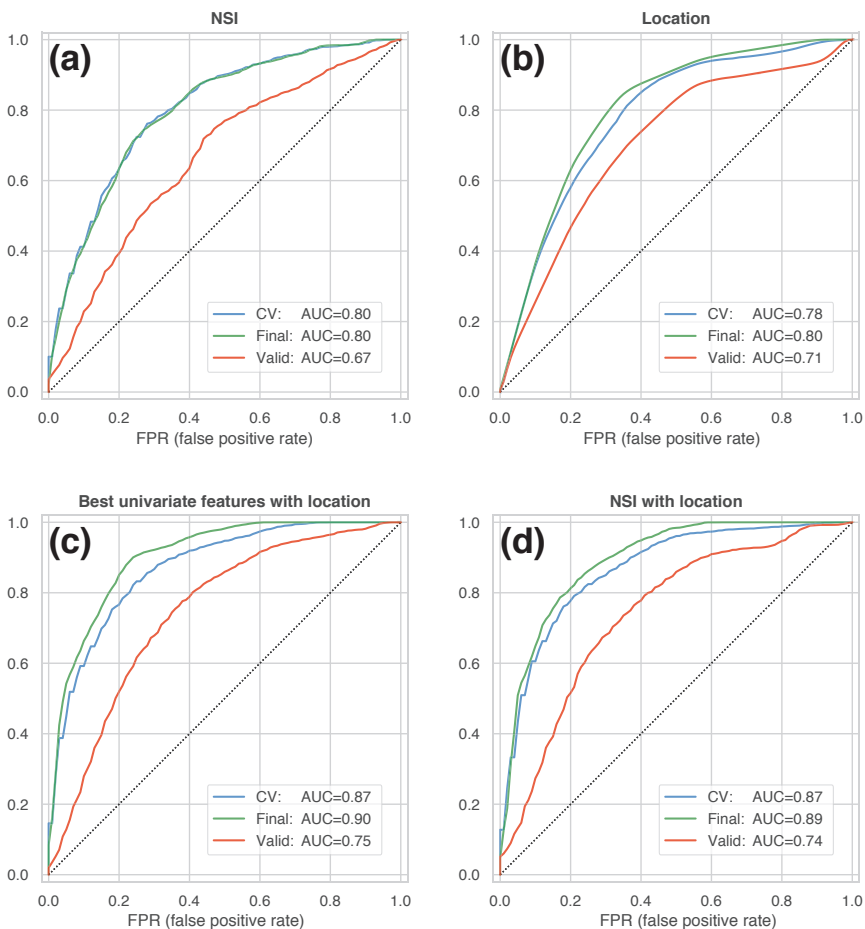
Univariate models (external validation, cut dome)							
Category	Predictor	AUC	AUC-diff	Accuracy	Sensitivity	Specificity	Kappa
Shape	$NSI$ , non-	<b>0.65±0.03</b>	-0.15	0.62±0.03	0.52±0.04	0.72±0.04	0.24±0.05
ZMI	norm. energy $Z_6^{\text{surf}}$	<b>0.67±0.03</b>	-0.14	0.61±0.03	0.50±0.04	0.73±0.04	0.23±0.05
ZMI	norm. energy $Z_3^{\text{surf}}$	<b>0.70±0.03</b>	-0.08	0.63±0.03	0.47±0.04	0.80±0.03	0.27±0.05
Writhe	$\overline{W}_{\text{mean}}^{L_1}$	<b>0.69±0.03</b>	-0.09	0.61±0.03	0.52±0.04	0.71±0.04	0.23±0.06
Shape	$UI$ , undulation	<b>0.66±0.03</b>	-0.11	0.60±0.03	0.44±0.05	0.76±0.03	0.21±0.05
Curvature	$GLN$	<b>0.59±0.04</b>	-0.16	0.56±0.03	0.39±0.04	0.73±0.04	0.13±0.06
Curvature	$MLN$	<b>0.57±0.04</b>	-0.17	0.54±0.03	0.39±0.04	0.69±0.04	0.08±0.06
Shape	$AR$ , aspect ratio	<b>0.61±0.04</b>	-0.14	0.57±0.03	0.46±0.05	0.69±0.04	0.15±0.06
ZMI	$ZMI_{3,1}^{\text{surf}}$	<b>0.71±0.03</b>	-0.03	0.64±0.03	0.64±0.05	0.65±0.04	0.28±0.06
ZMI	$ZMI_{5,1}^{\text{surf}}$	<b>0.61±0.03</b>	-0.11	0.58±0.03	0.51±0.04	0.65±0.04	0.16±0.05
Writhe	$W_{\text{mean}}^{L_2}$	<b>0.58±0.04</b>	-0.14	0.53±0.04	0.44±0.04	0.61±0.05	0.05±0.07
Size	$aSz$	<b>0.50±0.04</b>	-0.14	0.48±0.03	0.36±0.04	0.61±0.04	0.04±0.06

**Table 2.6:** External validation results of the same univariate predictors of Table 2.4. The univariate models trained on HUG data were here validated using the @neurIST and Aneurisk datasets. We report mean and standard deviation (mean ± std) for 100 bootstrap samples of the validation data. AUC-diff measures the differences between the AUC scores from the internal validation (Table 2.4) and the external validation. The data compares to the red lines in Figure 2.6.

Multivariate models (external validation, cut dome)							
Category	#	AUC	AUC-diff	Accuracy	Sensitivity	Specificity	Kappa
Max. model (+PCA)	10*	<b>0.67±0.03</b>	-0.15	0.63±0.03	0.66±0.04	0.61±0.04	0.27±0.06
Best univariate	12	<b>0.70±0.03</b>	-0.13	0.64±0.03	0.64±0.05	0.64±0.04	0.28±0.06
Location only	12	<b>0.71±0.03</b>	-0.07	0.67±0.03	0.66±0.04	0.69±0.04	0.34±0.05
Max. model + loc. (+ PCA)	22*	<b>0.73±0.03</b>	-0.14	0.67±0.03	0.62±0.04	0.73±0.04	0.34±0.06
Best univ. + location	24	<b>0.74±0.03</b>	-0.13	0.68±0.03	0.59±0.04	0.77±0.03	0.36±0.06
NSI and location	13	<b>0.74±0.03</b>	-0.13	0.70±0.03	0.62±0.04	0.78±0.04	0.39±0.06
$Z_6^{\text{surf}}$ and location	13	<b>0.73±0.03</b>	-0.13	0.68±0.03	0.61±0.05	0.76±0.04	0.36±0.06

**Table 2.7:** External validation results of the same multivariate models of Table 2.5. The models were trained on HUG data and validated with @neurIST and Aneurisk data. AUC-diff measures the difference between the AUC scores of the internal (Table 2.4) and external validation. The data compares to the red lines in Figure 2.6.





**Figure 2.6:** ROC curves summarizing the internal and external validation of four different model configurations: **(a)** non-sphericity NSI **(b)** anatomical location **(c)** best univariate features according to Table 2.4 with location **(d)** NSI with location. The blue lines represent the *internal* model validation and constitute the mean of 100 ROC curves (computed on test-data folds) during cross-validated training (blue line, CV). The green and red lines characterize the performance of the final model trained on the entire HUG dataset, which was validated on 100 bootstrap samples of the HUG dataset (the training dataset, green lines) and the *external validation* datasets from the @neurIST and Aneurisk projects (red lines).

## 2.4.2 Multivariate analysis

Table 2.5 summarizes the internal validation results for the multivariate models from Table 2.2. For better handling of the high dimensionality of the *MAX* models, PCA was applied, retaining 90% of the total variance present in the data. For the *BUP* models (best univariate performers), we included the features from Table 2.4. The *LOC* model used only location as predictor (Table 2.3) and served as reference.

## 2.4.3 Validation using external data

All univariate and multivariate models were trained and internally validated using HUG data only. After cross-validation based on subsets of training data, final models were computed including all data. These final models were then externally validated using the @neurIST and Aneurisk datasets. We report the resulting metrics of a bootstrapped ROC analysis (with 100 re-samplings of the validation dataset). Tables 2.6 and 2.7 summarize the external validation results for the univariate and multivariate models. Figure 2.6 compares the results with the internal validation using four exemplary models.

## 2.4.4 Dependency on the cut configuration

We computed the morphometrics for different “cut types” (Figure 2.3). Note that the geometry indices (GIs) were defined only for the two cut-configurations solely including the aneurysm dome (planar *dome*, non-planar *ninja*). Features based on curvature, writhe and ZMI were computed on all cuts (*dome*, *ninja*, *cut1*, *cut2*).

Because *dome* and *ninja* cuts both capture the aneurysm dome, the metrics computed for these cut configurations are directly comparable. Some metrics deviated considerably across different cuts. For instance, the aneurysm height  $H$  varied by up to +50% (for small aneurysms) and +10% in average when going from *dome* to *ninja* cuts. Other metrics also were susceptible to variations in the cut, most notably aspect ratio  $AR$  (measured as the height-to-neck ratio), the writhe metrics, and the ZMI (with larger differences for higher orders  $n$ ). Aneurysm size  $aSz$ ,  $NSI$ , and the important curvature metrics  $GLN$

and *MLN* were comparatively stable. The normalized ZMI energies  $Z_N^{\text{surf}}$  were considerably more stable with respect to alteration in the cutline than the single ZMIs. To summarize the differences  $d_{i,j} = x_{i,j}^{\text{dome}} - x_{i,j}^{\text{ninja}}$  of metric  $j$  evaluated for the two cut types per (HUG) dataset  $i$ , we computed the median-to-IQR ratio  $\delta_{j'}$  (IQR: interquartile range):

$$\delta_{j'} = \delta[d_{i,j=j'}] = \text{median}(d_{i,j=j'}) / \text{IQR}(d_{i,j=j'})$$

We report here the mean  $\bar{\delta}[\delta_{j'}]$  per feature category: GI size ( $\bar{\delta} = 0.10$ ), GI shape ( $\delta = 0.21$ ,  $\delta = 0.17$  without AR), curvature ( $\delta = 0.17$ ), writhe ( $\delta = 0.22$ ), ZMI ( $\delta = 0.43$ ), ZMI energies ( $\delta = 0.20$ ).

Despite the marked differences between the shape features evaluated for *dome* and *ninja* cuts, their univariate predictive capacity ( $\text{AUC}(x_{i,j}^{\text{dome}})$  vs.  $\text{AUC}(x_{i,j}^{\text{ninja}})$ ) was not significantly affected (unpaired  $t$ -tests, two-sided,  $\alpha = 0.05$ , adjusted for multiple testing). For the relevant predictors reaching an  $\text{AUC} > 0.7$  in the univariate models (cf. Table 2.4), differences in AUC amounted only to fractions of the AUC standard deviation.

Metrics based on cut configurations including segments of the parent vasculature (*cut1*, *cut2*) generally performed worse than metrics computed for *dome* and *ninja*. Curvature metrics, writhe indices and single ZMIs played no significant role in these experiments (*cut1*:  $\text{AUC} < 0.65$ , *cut2*:  $\text{AUC} < 0.60$ ). Only the normalized ZMI energies  $Z_N^{\text{surf}}$  (see Appendix 2.C) maintained their predictive ability, with  $Z_{10}$  reaching  $\text{AUC} = 0.77 \pm 0.06$  for *cut1* and  $\text{AUC} = 0.66 \pm 0.6$  for *cut2*.

## 2.5 Discussion

Here we examined different aspects of quantitative morphology with the goal to identify shape features that best reflect disease status. With a dataset comprising 470 ruptured and unruptured intracranial aneurysms, we were able to extend several findings from peer literature. We validated the generated univariate and multivariate models against external data provided by the @neurIST and Aneurisk dataset. These findings as well as the methodological setup *per se*, warrant careful discussion.

This section is structured as follows. In the first two subsections, we comment on the insights from the univariate analysis, mostly focusing on the quantitative description of the aneurysms. We then proceed to compare univariate and multivariate models. Finally, we address some concerns with respect to the methodology, and derive recommendations for future research.

Throughout this discussion, we use ROC-AUC, the area under the receiver operating characteristic curve, as the principal quality metric for diagnostic accuracy of the models. Other metrics such as prediction accuracy, sensitivity and specificity have also been provided in the result section to offer a more complete picture of the models' predictive power. In our subsequent reasoning we exploit the fact that the training/validation procedures were strictly the same in all experiments, thereby making the results comparable.

### 2.5.1 Which features encode disease status?

Univariate performance on rupture status predictions helped us identify those metrics that are most likely to indicate an aneurysm's rupture status.

*Geometry indices.* Of all 12 GIs, *NSI* most accurately predicted rupture (AUC =  $0.80 \pm 0.05$ ). Other shape metrics measuring elongation (*EI*, *AR*) and undulation (*UI*) were also potent univariate predictors for rupture, with AUC scores between 0.75 and 0.79 (Table 2.4). Metrics capturing the size of the aneurysm were associated with aneurysm rupture (*aSz*: AUC =  $0.64 \pm 0.05$ , *H*: AUC =  $0.64 \pm 0.05$ ), but to a significantly smaller degree than most of the shape metrics (with *BF* and *CP* being the exceptions). The neck diameter  $D_n$  (AUC =  $0.54 \pm 0.06$ ) was not linked to rupture status. These findings

underscore the insight that taking into account aspects of morphology other than size can substantially improve the assessment of aneurysms. This lies in contrast with the argumentative line of a previous debate on treatment guidelines, in which size was given more attention than morphology.<sup>29,61</sup>

*Curvature-based metrics.* Curvature metrics capture surface undulation and bending. Of all curvature metrics, the well-established *GLN* (AUC =  $0.75 \pm 0.05$ ) and *MLN* (AUC =  $0.75 \pm 0.05$ ) performed the best.

*Writhe-based metrics.* Writhe-based metrics can be related to surface asymmetry and twisting.<sup>102</sup> Our modified definition of surface writhe  $\bar{W}^{L_1}(\mathbf{p})$  is normalized by the surface area, produced better results than the non-normalized  $W^{L_2}(\mathbf{p})$  (cf. Appendix 2.B.2). This is due to  $\bar{W}^{L_1}$  characterizing only shape, whereas the signal contained in non-normalized  $W^{L_2}$  depends on both shape and size. The Pearson correlation coefficient  $\rho_P$  between aneurysm size  $aSz$  and  $W_{\text{mean}}^{L_2}$  was 0.93; and only 0.20 between  $aSz$  and  $\bar{W}_{\text{mean}}^{L_1}$ . Note that the unmodified definition for surface writhe by Lauric et al.<sup>102</sup> did not prove useful in our experiments (AUC =  $0.57 \pm 0.06$ ).

*Zernike moment invariants.* From 121 considered indices, only the indices  $Z_{2,0}^{\text{surf}}$ ,  $Z_{3,1}^{\text{surf}}$  and  $Z_{5,1}^{\text{surf}}$  exhibited consistent as well as significant inter-class differences. Higher-order moments yielded either less or no useful information with respect to rupture status. Low order ZMI can be computed with less effort and are more robust with respect to mesh variations than high order ZMI.

*Normalized Zernike energies.* The  $Z_N^{\text{surf}}$  were good predictors for rupture status. All five ( $Z_1^{\text{surf}}$ ,  $Z_2^{\text{surf}}$ ,  $Z_3^{\text{surf}}$ ,  $Z_6^{\text{surf}}$ ,  $Z_{10}^{\text{surf}}$ ) achieved (univariately) an AUC larger than 0.7. We found that some  $Z_N^{\text{surf}}$  were strongly correlated (Spearman) with undulation/elongation (*NSI*, *EI*, *UI*), *AR* and surface writhe ( $\bar{W}^{L_1}$ ), with correlation coefficients  $\rho_{Sp}$  between 0.85 and 0.90. They were also associated with perceived irregularity ( $\rho_{Sp} = 0.76$ , see Chapter 3, Table 3.6). However, mathematical analysis of these metrics was beyond the scope of this study.

*Summary:* All categories except size metrics were well represented among the best performing candidates (Table 2.4). We recommend using the modified definition of surface writhe  $\bar{W}^{L_1}(\mathbf{p})$ , for which the index  $\bar{W}_{\text{mean}}^{L_1}$  provided the best results. Our suggested Zernike energies  $Z_N^{\text{surf}}$  proved to be indicative of the rupture status, which was, along with *NSI* (and its sibling *EI*) among the best predictors.

## 2.5.2 How relevant is the cut configuration?

The cut line separates the aneurysm from the surrounding vasculature and has a bearing on most of the morphometric parameters. Since cutting was performed manually by operators, we investigated to what extent it affected the results of this study. To this end, we considered two different cut configurations involving the aneurysm dome only: a planar one (identifier *dome*) and non-planar one (*ninja*). The two sets of rules for separating the aneurysm from the nearby vasculature were applied independently by two operators (one rule for each operator). Naturally, this led to substantial differences in the neck region of the aneurysm geometries (illustrated exemplary in Figure 2.3).

Albeit these differences, our analysis revealed that the particular choice of the neckline on average had little impact on the metrics' capacity to predict rupture status, indicating that the selected metrics are fairly robust with respect to the cut type. Even though the *ninja* cut has a better physiological justification than the *dome* cut, it did not substantially improve the prediction outcome.

Metrics involving segments of the parent vasculature (*cut1*, *cut2*) consistently produced worse results compared to *dome* and *ninja* cuts. The more of the parent vasculature was included in the cut, the less accurately the diagnostic models performed (*dome* > *cut1* > *cut2*). For cut types including vascular segments, metrics based on curvature, writhe and ZMI failed to discern ruptured from unruptured aneurysms, contradicting results reported by Lauric et al.<sup>102</sup> and Millán et al.<sup>104</sup>. The normalized ZMI energies  $Z_N^{\text{surf}}$  maintained the diagnostic capacity also for cut types *cut1* and *cut2*:  $Z_{10,\text{cut1}}^{\text{surf}}$ : AUC =  $0.77 \pm 0.05$ ,  $Z_{10,\text{cut2}}^{\text{surf}}$ : AUC =  $0.66 \pm 0.06$ ), while  $Z_{10,\text{dome}}^{\text{surf}}$ : AUC =  $0.77 \pm 0.05$ ).

The lack of predictiveness in some of the metrics for *cut1*- and *cut2*-geometries does not, however, imply that the parent vessel geometry is irrelevant for disease status prediction. Our pool of features lacks metrics that explicitly describe the parent vessel geometry relative to the aneurysm dome (for instance size ratio (SR), vessel angle or inclination angle<sup>99</sup>). Notably, SR has previously been associated with aneurysm rupture.<sup>148</sup>

*Summary:* Both *dome* and *ninja* cuts enabled equal predictive performance of morphometrics. However, metrics available for other *cut1* and *cut2* showed a weaker association with rupture status than their dome-only counterparts.

### 2.5.3 Do multiple shape predictors lead to a better model?

The combination of multiple predictors moderately improved the prediction accuracy (Table 2.4 and Table 2.5), with no signs of excessive model overfitting (green vs. blue lines in Figure 2.6). However, the net improvement of the multivariate models over the univariate models was relatively small: The best univariate predictor (*NSI*) achieved an AUC of  $0.80 \pm 0.05$ .

The maximal model MAX (AUC =  $0.82 \pm 0.05$ ) and the BUP model using a selection of best univariate performers (AUC =  $0.82 \pm 0.05$ ) achieved the same performance. This indicates that the combination of many weak univariate predictors (MAX) does not provide more information about the disease status than a selection of best performers (BUP). This also held true for nonlinear models. We explored non-linear classification models such as support vector machines with a Gaussian kernel, gradient boosted decision trees and basic neural nets (multilayer perceptrons).<sup>73</sup>

That the MAX and BUP models performed equally well is indicative of redundancy in the descriptors. To assess the level of redundancy, we applied a PCA of the (standardized) feature matrix for the dome cut (470 samples vs. 150 features). A PCA retaining 50%, 75%, 90%, 95% and 99% of the total variance required 5, 19, 44, 62 and 98 of maximally 150 principal components.

This observed redundancy is in part a consequence of how the metrics are mathematically defined and the nature of the geometries under inspection, which leads to particular aneurysm features being captured in multiple ways by different metrics. For instance, we measured a Spearman correlation coefficient  $\rho_{Sp}(GLN, MLN) = 0.99$  between *GLN* and *MLN*, or  $\rho_{Sp}(V, S) = 1.00$  between volume *V* and surface area *S*, which is not unexpected for saclike geometries. Other correlations lack such a simple geometric interpretation, such as the high correlation index between *NSI* and  $Z_{10}^{\text{surf}}$   $\rho_{Sp}(NSI, Z_{10}^{\text{surf}}) = 0.95$  or  $\rho_{Sp}(GLN, W_{\text{mean}}^{L2}) = 0.91$ . Some data redundancy could be attributed to the physiological processes that underlie aneurysm formation. For instance, larger aneurysms were more likely to show irregular structures (blebs, lobules), which was also reflected in our data:  $\rho_{Sp}(aS_z, GLN) = 0.82$ .

Due to these high correlations, we were able to further reduce the number of predictors to four: *NSI*,  $\bar{W}_{\text{mean}}^{L1}$ , *GLN* and *aS<sub>z</sub>*. This model performed about the same as the BUP model: AUC =  $0.82 \pm 0.04$ .

*Summary:* Multivariate models (Table 2.5) performed only slightly better on the HUG dataset than the univariate models, even though the entirety of shape features captured a relatively wide range of morphological characteristics, despite any data redundancy. This corroborates the value of  $NSI$  and  $Z_N^{\text{surf}}$ , but also of  $\bar{W}_{\text{mean}}^{L_1}$  and  $GLN$ , as efficient indicators of those IA shape characteristics that are relevant for distinguishing the rupture status.

#### 2.5.4 What is the effect of location as predictor?

Because aneurysm morphology and associated risks vary significantly with the anatomical location<sup>33</sup>, we included location as the only non-morphometric predictor to our models. This resulted in a substantial increase of diagnostic accuracy (AUC =  $0.869 \pm 0.038$  for the MAX model and AUC =  $0.875 \pm 0.037$  for the BUP model, Table 2.5). Two minimal models ( $NSI$  + location,  $Z_6^{\text{surf}}$  + location) performed both with essentially the same diagnostic accuracy: AUC =  $0.867 \pm 0.038$  ( $NSI$ ) and AUC =  $0.873 \pm 0.038$  ( $Z_6^{\text{surf}}$ ).

We trained also a location-only model, which performed with an AUC =  $0.780 \pm 0.043$  (Figure 2.6b). Aneurysm location alone is therefore about as informative about an aneurysm's rupture status as its morphology.

The probability of rupture varies considerably with location. Adding location as a predictor incorporates therefore prior information about the probability of rupture into the classifier (Table 2.3), since probability of rupture varies considerably with location. For instance, it enables the classifier to assign a low score to aneurysms at the cavernous or ophthalmic segments of the internal carotid (locations at which aneurysms rupture very rarely), regardless of their morphology. If only cases were considered at locations with a roughly balanced amount of ruptured and unruptured aneurysms (ACoMA, PComA, ACA, ICA chor, compare with Table 2.3), the prediction scores returned to their original values of the models without location as predictor

*Summary:* The addition of aneurysm location improved the predictive accuracy substantially because this enables the classifier to account for varying rupture probabilities. A model relying on  $NSI$  and location as predictors excelled other models in terms of AUC, prediction accuracy and parsimony.



### 2.5.5 External validation results

On the HUG dataset, models based on morphometry and anatomical location lead to results comparable to other recent studies on IA morphology (Table 2.1). However, all investigated models performed markedly worse on the external datasets @neurIST and Aneurisk (Table 2.6 and Table 2.7). This indicates that the models do not generalize well to these datasets.

A closer inspection of the two external datasets revealed several differences that may explain this loss of predictive accuracy. Both @neurIST and Aneurisk datasets exhibited a relatively balanced ratio of ruptured and unruptured cases (Table 2.3). In total, the validation dataset consisted of 132 unruptured and 133 ruptured aneurysms (1:1), as opposed to 342 unruptured and 128 (3:1) in the HUG training dataset. Furthermore, the distribution of the different locations differed between training and validation datasets. Most notably, comprised only one AComA case, and an equally disproportionate number of PComA cases. Aneurisk matched the HUG datasets in terms of location distribution more closely. However, its unruptured cases were about 50% larger than the average of all unruptured HUG cases (Table 2.8). Aneurisk's unruptured aneurysms were even larger than the ruptured ones, which was not the case for the HUG datasets.

All this indicates that the validation dataset (@neurIST + Aneurisk) differed significantly in its structural composition and characteristics from the training dataset (HUG1 + HUG2), with strong repercussions for diagnostic accuracy. To further substantiate this finding, we repeated the entire analysis using HUG1 as the training and HUG2 as the external validation dataset. Even though HUG1 and HUG2 were processed by different persons, the medical data were collected by the same medical staff in the same period of time, which is likely to have led to a very comparable case selection. This structural data homogeneity translated into substantially improved predictive accuracy, with  $AUC = 0.84 \pm 0.04$  for the bivariate model  $NSI+location$ ,  $AUC = 0.88 \pm 0.03$  for  $Z_6^{surf}+location$ , and  $AUC = 0.72 \pm 0.05$  for the location-only model.

*Summary:* To ensure predictive accuracy, models require that the data they process for prediction possess the same characteristics as the data they have been trained with. However, the HUG datasets and the validation datasets differed in various key characteristics.

Dataset	#	<i>aSz</i>		<i>AR</i>		<i>NSI</i>	
		U	R	U	R	U	R
HUG1	350	5.58±3.98	6.82±3.86	1.01±0.56	1.43±0.77	0.12±0.09	0.20±0.08
HUG2	120	5.82±3.07	7.41±4.28	1.03±0.38	1.37±0.56	0.11±0.07	0.21±0.09
@neurIST	164	5.93±3.44	6.83±4.08	1.07±0.64	1.33±0.86	0.14±0.11	0.19±0.09
Aneurisk	101	8.78±5.47	6.92±4.90	1.28±0.68	1.39±0.57	0.15±0.09	0.19±0.07
Overall	735	5.91±4.22	6.93±4.17	1.04±0.56	1.38±0.68	0.13±0.09	0.20±0.09

**Table 2.8:** Summary statistics for the entire AneuX morpho database, stratified by dataset and rupture status. We used here median±IQR because the metrics were not normally distributed. Abbreviations: IQR – interquartile range; U/R – unruptured/ruptured; *aSz* – aneurysm size; *AR* – aspect ratio; *NSI* – non-sphericity index

## 2.5.6 Comparison with the Detmer model

The work by Detmer et al.<sup>61</sup> is currently the most comprehensive rupture status prediction model published. Based on data of 1631 aneurysms, they developed a multifactorial model consisting of 26 morphometrical, hemodynamical and clinical parameters. In a subsequent study<sup>79</sup>, they successfully validated their model for the Aneurisk and HUG2 datasets. In the following, we compare our results with those of Detmer et al.

As in our present study, Detmer et al. also observed an impaired prediction performance when validating only with Aneurisk data. This loss was less pronounced ( $AUC = 0.82$ , compared to  $AUC = 0.87$  for the internal validation) as in our case. As the HUG2 dataset more closely resembled the training dataset, used in their study, predictive performance attained roughly the same scores as in their internal validation.<sup>78,79</sup>

The most striking difference between the study of Detmer et al. and our approach lies in the model complexity. Whereas Detmer used 26 predictors, our minimal model required two: one for shape, and one for location. Yet, it is worth noting that the Detmer model was trained on a multicentric dataset four times larger than the one used in this study, and was possibly better able to average out the kind of dataset heterogeneity as described above. The multifactorial basis of their model may also have provided additional

robustness against dataset heterogeneity by distributing the predictive signal across several predictors.

A predictive model with fewer predictors, however, reduces the requirements on the data acquisition process. In this regard, the suggested parsimonious model configuration consisting of one morphometric variable and location has advantages over the Detmer model. The anatomical location of an aneurysm is routinely determined by clinicians and would thus not require an additional data collection step. The morphometric description of a previously extracted aneurysm geometry is typically not very demanding, either. Furthermore, radiomic methods exist to compute morphometric parameters on imaging data directly<sup>149</sup> and therefore would not require a mesh extraction. This last aspect in our view is pivotal for the successful deployment of predictive models to a clinical environment.

Differences also exist in the principal motivation for developing predictive models. Detmer et al. envisioned a tool for risk assessment that assigns to each tested aneurysm a score (probability) indicating how much it resembles the ruptured aneurysms of the training dataset. The working assumption for their approach is that the resemblance of a given unruptured aneurysm to ruptured cases in the training dataset captures that aneurysm's risk for future rupture. The primary purpose of our study was to examine the informativeness of various morphometric parameters and combinations thereof for the discrimination of ruptured and unruptured aneurysms. Because our benchmark was based on a very similar statistical learning procedure, our results can be compared with those of Detmer et al.

*Summary:* Our minimal models (morphometry + location) compare well with the Detmer model and outperform it in terms of model parsimony, but are likely less robust in presence of different dataset biases due to reliance on fewer predictor variables.

## 2.5.7 Limitations

This study adopts an approach that has recently experienced broader use: Statistical learning schemes are deployed to identify a functional relationship between the quantitative descriptions of aneurysms and a probabilistic assignment of their disease status.<sup>75,77,97,99,102,104,121,125</sup> If such predictive models generalize well to data other than the one used for the model development, they offer potential use as assistive tools for clinical decision making.

The predictive power of such models depends crucially on the data available for model development. A problem common to many recent studies on aneurysm morphology is related to the use of data from 3DRAs, an angiographic method usually employed only in the context of treatment. Unruptured IAs in such databases likely have been assessed previously by a clinician as at risk to rupture and treatment was recommended. Therefore, such datasets likely do not adequately reflect the natural distribution of IA characteristics in the general population. In particular, those cases that have been classified by clinicians as less dangerous are underrepresented in 3DRA datasets. This effect is mitigated to some degree by the fact that some patients develop multiple aneurysms, which were not the primary target of treatment. About 30% of the patients in the HUG datasets developed such multiple, co-occurring aneurysms.

The characteristics of aneurysm datasets vary over time. For instance, the increased availability of imaging facilities has increased the number of incidentally diagnosed unruptured IAs. As a consequence, the ratio of ruptured to unruptured aneurysms in clinical databases has decreased over time. Likewise, the treatment guidelines have evolved considerably in the last 30 years, which also affected the selection of cases available for such studies. These trends contribute to the above data disparities observed in this study between the HUG, @neurIST and Aneurisk datasets.

Due to these reasons, this approach is often criticized, not for the method *per se*, but for the data that are used to train the models.<sup>150</sup> In particular, it is doubted whether the insights gained from analyzing the differences between unruptured and ruptured aneurysms can serve as the basis for reliable proxies of “risk” or “instability”. We therefore refrained from using such terms in our

study and focused on assessing the sensitivity of morphological features with respect to the aneurysm’s rupture status.

A future study could investigate whether the insights of this study remain valid for distinguishing stable and unstable aneurysms (which is clinically more relevant than the ruptured/unruptured dichotomy) and how large the differences must be to detect instability. A dataset based on follow-up data would be very advantageous for a study like the one carried out here. However, as discussed by Ramachandran et al.<sup>151</sup>, such datasets can also suffer from selection biases.

Finally, it was conjectured that the morphology of aneurysms might change as a result of rupture.<sup>60,152,153</sup> While this cannot be excluded in general, several studies have suggested that for the majority of ruptured cases this does not apply.<sup>29,62,63,97,154</sup>

## 2.6 Conclusions

We have conducted a comprehensive study to examine the potency of morphology to encode the disease status of IAs. Based on the AneuX morphology database consisting of 470 aneurysms acquired at the HUG and 265 additional cases from external databases, we investigated how various aspects of the morphometric description of aneurysms relate to rupture status.

Morphology is a good predictor for the aneurysm disease status. Metrics such as  $NSI$ ,  $Z_N^{\text{surf}}$  and  $\overline{W}^{L_1}$  are able to capture relevant shape characteristics to distinguish between ruptured and unruptured cases. Location is an important cofounding factor for the aneurysm shape. It is therefore beneficial to assess the morphology per location. Predictive models depend strongly on the datasets used for training. We found that clinical datasets are prone to different selection bias.

## 2.A Appendix: Geometry Indices

The following metrics are defined for *dome* and *ninja* cuts. Some metrics use the neck plane as reference, which is naturally given for planar dome cuts, and is defined as the best-fit plane through the cut-line for non-planar *ninja* cuts.

### 2.A.1 Indices measuring the aneurysm size

Index	Symbol	Unit	Details
Volume	$V$	mm <sup>3</sup>	Volume of aneurysm dome.
Surface area	$S$	mm <sup>2</sup>	Surface area of the aneurysm dome (without neck area).
Neck diameter	$D_n$	mm	Characteristic diameter of the contour in the neck plane: $D_n = 4 \cdot S_n / P_n$ , where $P_n$ is the perimeter of the neck contour.
Max. diameter	$D_{max}$	mm	Diameter of the largest cross section parallel to the neck plane.
Aneurysm height	$H$	mm	Maximal extent perpendicular to the cut plane.
Aneurysm size	$aSz$	mm	Diameter of the minimum bounding sphere containing the dome

### 2.A.2 Indices measuring the aneurysm shape

Index	Symbol	Unit	Details
Aspect ratio	$AR$	–	Ratio between height and neck diameter: $AR = H / D_n$
Bottleneck factor	$BF$	–	Ratio between max. diameter and neck diameter: $BF = D_{max} / D_n$
Non-sphericity index	$NSI$	–	Measure for the deviation from a semi-spherical shape: $NSI = 1 - (18\pi)^{1/3} \cdot V^{2/3} / S, \quad NSI \in [0,1]$ Captures both elongation and undulation. $NSI = 0$ holds for a perfect hemi-sphere.
Ellipticity index	$EI$	–	Measure for the elongation of the aneurysm dome. Given the volume $V_{CH}$ and surface area $S_{CH}$ of the aneurysm's convex hull, $EI = 1 - (18\pi)^{1/3} \cdot V_{CH}^{2/3} / S_{CH}, \quad EI \in [0,1].$ $EI = 0$ holds for a shape with hemi-spherical convex hull.
Undulation index	$UI$	–	Measure for the amount of surface undulation $UI = 1 - (V / V_{CH}), \quad UI \in [0,1],$ where $V_{CH}$ is the volume of the convex hull. $UI = 1$ indicates that the aneurysm is perfectly convex.
Conicity parameter	$CP$	–	Location of the max. diameter relative to the aneurysm height: $CP = 0.5 - H_{max} / H, \quad CP \in [-0.5, +0.5].$ $H_{max}$ is the height at which $D_{max}$ occurs. $CP = 0.5$ if $D_{max}$ is measured in the neck plane. Sensitive to the choice of the neck plane.

## 2.B Appendix: Distribution-derived metrics

The metrics here presented describe the distribution of surface properties  $f(\mathbf{p}): \mathcal{S} \rightarrow \mathbb{R}$  that map points  $\mathbf{p} \in \mathcal{S}$  of the input surface  $\mathcal{S}$  to scalar values. We distinguish between curvature and writhe metrics, which can be computed on four cut configurations (*dome*, *ninja*, *cut1*, *cut2*).

### 2.B.1 Curvature-based features

We consider two different definitions of curvature at a point  $\mathbf{p} \in \mathcal{S}$  for a sufficiently regularly shaped surface  $\mathcal{S}$ . *Gaussian* and *mean curvature* are defined as

$$K_G(\mathbf{p}) = \kappa_1(\mathbf{p}) \cdot \kappa_2(\mathbf{p})$$

$$K_M(\mathbf{p}) = \frac{1}{2}(\kappa_1(\mathbf{p}) + \kappa_2(\mathbf{p})).$$

$\kappa_1(\mathbf{p})$  and  $\kappa_2(\mathbf{p})$  are the two principal curvatures at point  $\mathbf{p}$ . For triangular meshes, the curvature can be approximated by *discrete* Gaussian and Mean curvature, implementations of which are available in VTK<sup>141</sup>.

Ma and Raghavan et al.<sup>98</sup> compute the total curvature using surface norms

$$\mathcal{E}_{q,\star} = \left( \int_{\mathbf{p} \in \mathcal{S}} |K_\star(\mathbf{p})|^q d\mathcal{S} \right)^{1/q},$$

where  $q = 2$  and  $K_\star(\mathbf{p})$  is either the Gaussian (G) or mean curvature (M). The authors normalize  $\mathcal{E}_{2,G}$  and  $\mathcal{E}_{2,M}$  with the total curvature of a sphere with equal surface area ( $\mathcal{E}_{2,\star}^\circ$ ), yielding their metrics GLN and MLN. For a discretized surface with curvature  $K_\star(\mathbf{p}_i) = K_{\star,i}$  and local surface patch area  $\Delta S_i$  at mesh point  $\mathbf{p}_i$ , the GLN and MLN can be written as

$$\text{GLN} = \frac{\mathcal{E}_{2,G}}{\mathcal{E}_{2,G}^\circ} = \frac{\sqrt{\sum_{\Delta S_i} K_{G,i} \cdot \Delta S_i}}{\sqrt{4\pi \cdot 1/r_o^2}} = \frac{1}{4\pi} \sqrt{\frac{\sum_{\Delta S_i} \Delta S_i \cdot \sum_{\Delta S_i} K_{M,i} \cdot \Delta S_i}{\sum_{\Delta S_i} \Delta S_i}}$$

$$\text{MLN} = \frac{\mathcal{E}_{2,M}}{\mathcal{E}_{2,M}^\circ} = \frac{\sqrt{\sum_{\Delta S_i} K_{M,i} \cdot \Delta S_i}}{\sqrt{4\pi}} = \sqrt{\frac{1}{4\pi} \sum_{\Delta S_i} K_{M,i} \cdot \Delta S_i}$$

Index	Symbol	Unit	Details
Total curvature, normalized	$G. L_2N,$ $M. L_2N$	–	Corresponds to GLN and MLN, as defined above.
Total negative curvature, normalized	$G. -L_2N,$ $M. -L_2N$	–	Like $G. L_2N$ and $M. L_2N$ , but only considering negative curvature: $\frac{\mathcal{E}_{2,*}^-}{\mathcal{E}_{2,*}^0} = \left( \int_{\mathbf{p} \in \mathcal{S}, K_*(\mathbf{p}) < 0}  K_*(\mathbf{p}) ^2 d\mathcal{S} \right)^{1/2} / \mathcal{E}_{2,*}^0$ <p>Motivated by the observation that “complex” dome geometries exhibit larger regions with negative curvature (saddles for G, dominantly concave for M) exist. Metric for shape irregularity.</p>
Total L2 curvature, normalized by convex hull	$G. L_2NCH,$ $M. L_2NCH$	–	Total curvature normalized by total curvature of the convex hull: $\mathcal{E}_{2,*} / \mathcal{E}_{2,*}^{CH}$ <p>Measures the undulation or blebbyness of an aneurysm. The convex hulls should be re-meshed prior to measuring curvature.</p>
Entropy of curvature	$G. H, M. H$	–	Given a kernel-based approximation $q_*(x)$ for the value distribution of $K_*(\mathbf{p})$ , the (differential) entropy $h_* = - \int_{x \in X} q_*(x) \log q_*(x) dx$ <p>can be seen as measure for the spread of <math>q_*(x)</math>.<sup>102</sup></p>

The computation of curvature metrics is very sensitive to mesh irregularities. We therefore re-meshed and moderately smoothed all surfaces as described in Section 2.3.2. Furthermore, some caution should be taken when evaluating the curvature at points near the edge  $\partial\mathcal{S}$ .

## 2.B.2 Writhe-based features

The *writhe number* measures surface asymmetries and “twisting forces” as seen from a surface point  $\mathbf{p}$ . Originally invented to characterize curves in knot-theory, the writhe number  $W(\mathbf{p})$  was generalized by Lauric et al.<sup>102</sup> to surfaces:

$$w(\mathbf{p}, \mathbf{q}) = \frac{[\mathbf{n}_p, \mathbf{q} - \mathbf{p}, \mathbf{n}_q]}{\|\mathbf{n}_p\| \cdot \|\mathbf{q} - \mathbf{p}\| \cdot \|\mathbf{n}_q\|}$$

$$W(\mathbf{p}) = \int_{\mathbf{q} \in \mathcal{S}} w(\mathbf{p}, \cdot) \cdot d\mathcal{S}$$



where  $\mathcal{S}$  is the surface,  $\mathbf{n}_x$  is the normal vector at a point  $\mathbf{x}$ , and  $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$  represents the triple product of vectors  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ . Discretization is straightforward:

$$W_i = \sum_{0 < j < N} w_{ij} \cdot \Delta S_j \approx W(\mathbf{p}_i)$$

where  $w_{ij} = w(\mathbf{p}_i, \mathbf{p}_j)$  for two mesh vertices  $\mathbf{p}_i, \mathbf{p}_j$ , and  $\Delta S_j$  the area of a differential surface patch around vertex  $\mathbf{p}_j$ , which we computed as 1/3 of the total area covered by all triangles containing  $\mathbf{p}_j$ . It holds that  $S = \sum_j \Delta S_j$ .

We found that modifying the summation of  $W$  significantly improved the signal in terms of rupture prediction and perceived irregularity (see Chapter 3).

$$W^{L_1}(\mathbf{p}) = \int_{q \in \mathcal{S}} |w(\mathbf{p}, \mathbf{q})| \cdot d\mathcal{S}$$

$$W^{L_2}(\mathbf{p}) = \left( \int_{q \in \mathcal{S}} w(\mathbf{p}, \mathbf{q})^2 \cdot d\mathcal{S} \right)^{1/2}$$

In addition, we normalized these metrics by the surface area, which makes the writhe number invariant to isotropic scaling of the input surface:

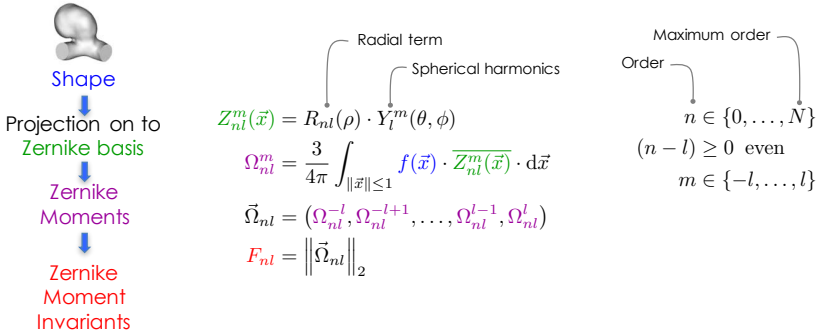
$$\bar{W}^{L_1}(\mathbf{p}) = W^{L_1} / \|\mathcal{S}\|, \quad \bar{W}^{L_2}(\mathbf{p}) = W^{L_2} / \|\mathcal{S}\|$$

Lauric et al.<sup>102</sup> suggested to describe the writhe number distribution of a mesh using histogram statistics, such as central moments, cumulants or entropy. We considered only mean and empirical (differential) entropy, which we found to work best in our context. (Higher order moments and cumulants, as used in Lauric et al.<sup>102</sup>, tend to be numerically unstable.)

Index	Symbol	Unit	Details
Mean writhe (L2)	$W_{\text{mean}}^{L_2}$	mm <sup>2</sup>	Empirical mean of writhe numbers $W_i^{L_2}$
Writhe entropy (L2)	$W_H^{L_2}$	-	Empirical entropy of writhe numbers $W_i^{L_2}$
Mean writhe (L1, normalized)	$\bar{W}_{\text{mean}}^{L_1}$	-	Empirical mean of writhe numbers $\bar{W}_i^{L_1}$
Writhe entropy (L1, norm.)	$\bar{W}_H^{L_1}$	-	Empirical entropy of writhe numbers $\bar{W}_i^{L_1}$

## 2.C Appendix: Zernike Moment Invariants

*Zernike moments* (ZMs) and the *Zernike moment invariants* (ZMIs), originally described in 2D and generalized to 3D by Canterakis<sup>119</sup>, have been adopted by Millán et al.<sup>155</sup> to assess IA morphology. Figure 2.B.1 illustrates the principal steps to compute the ZMIs for an input geometry. *Zernike basis* functions  $\{Z_{nl}^m\}$  can be understood as radially modulated spherical harmonics  $\{Y_l^m\}$ . The *Zernike moments*  $\Omega_{nl}^m$  are obtained by projecting the input surface onto these basis functions  $\{Z_{nl}^m\}$ . Finally, the ZMIs  $F_{nl}$  are computed as the  $L_2$ -norm of the  $\Omega_{nl}^m$  for all  $m \in [-l, l]$ . Since the Zernike basis is defined only within the unit sphere, the geometry needs to be transformed to fit within the unit sphere. This naturally ensures invariance of ZMIs (and ZMs) to (isotropic) scaling and translation of the input surface. It can be shown that ZMIs are also invariant to object rotation, which is a necessary requirement for a viable shape descriptor.



**Figure 2.B.1:** Overview on the computation of Zernike Moment Invariants (ZMI) for an input shape (a 2D manifold in 3D space). The surface is projected onto the Zernike basis  $\{Z_{nl}^m\}$ , which yields the Zernike Moments. From these, the Zernike Moment Invariants  $F_{nl}$  are computed, which are invariant to translation, rotation and (isotropic) scaling of the input surface.

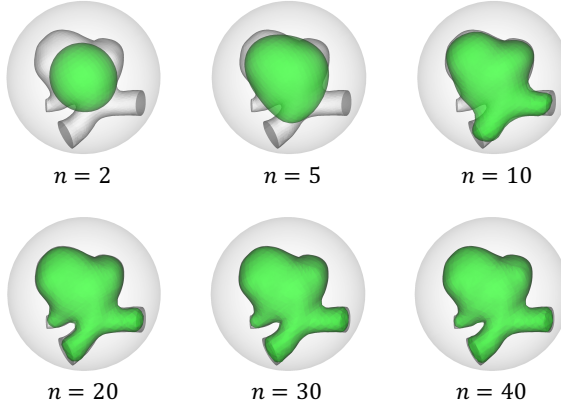
The ZMs can be computed as a function of *geometry moments*. There are two variants: *volume-based* moments assume the mass to be evenly distributed throughout the object volume, while *surface-based* moments consider the object mass to be condensed on the surface.<sup>142</sup> Accordingly, we distinguish

between  $ZMI_{\text{surf}}$  and  $ZMI_{\text{vol}}$ . Our implementation is based on a revised version of the algorithm provided by Novotni et al.<sup>120</sup>, adopting ideas by Pozo et al.<sup>142</sup> to optimize the computation. See details, refer to Ebnöther.<sup>111</sup>

The resulting vector of ZMIs represents a multi-dimensional shape descriptor that can be used to compare different shapes with one another. Similar shapes are supposed to have a relatively small distance in the ZMI feature space, although the relationship between particular macroscopic morphological changes of the input object and its effect on the ZMI values are very unintuitive. The parameter  $n$  is called the *order* of the Zernike decomposition. The index  $l$  satisfies  $0 \leq l \leq n$ , and  $n - l$  even. Low order ZMI  $F_{nl}$  contain information about coarse morphological properties of the input shape whereas the  $F_{nl}$  for higher values of  $n$  represent finer details about the shape (Figure 2.B.2). A ZMI descriptor for  $n = 20$  is composed of 121 different coefficients, whereas for  $n = 40$  a total of 441 indices exist. The computational costs grow with  $\mathcal{O}(n^6)$ .

A numerical analysis revealed that the ZMI can be computed robustly with double floating-point precision up to order  $n = 20$ , whereas quadruple floating-point precision (long double) or higher should be used for higher orders  $n > 20$ , which increases the computational costs. By experimentation, we found that ZMs and ZMIs are able to capture all the major morphological characteristics for  $n = 20$  with sufficient accuracy, as illustrated in Figure 2.B.2. Furthermore, we limited this study to  $ZMI_{\text{surf}}$ , as they had a slight (but insignificant) advantage over  $ZMI_{\text{vol}}$  in our experiments.

ZMs and ZMIs are defined for all of our cut configurations. The ability to encode the geometry of the surrounding vasculature and to capture the aneurysm geometry as a whole make this method attractive for the morphological description of aneurysms. On the downside, ZMIs are computationally expensive to acquire and very difficult to interpret geometrically. Furthermore, because ZMs have a so-called global support, locally confined changes in the input surface may alter the all ZMIs in a non-intuitive way. This makes them vulnerable to variations in the cut-plane selection. Also, it is unlikely that ZMIs can be used to detect the presence of certain morphological structures of interest, such as blebs or necks. We believe that ZMIs are most useful for the identification of morphologically *similar* objects.



**Figure 2.B.2:** The original shape can be reconstructed from the Zernike Moments  $\Omega_{nl}^m$ . The more ZMs that are taken into account, the more accurate the reconstruction. Low-order ZMIs characterize the basic shape, whereas high-order ZMI contribute ever finer details of the shape.

Finally, to condense the ZMI of a geometry  $\mathcal{S}$  into a single index, we aggregated all ZMI up to order  $n = N$  according to the following formula

$$Z_N = \frac{\sum_{n \leq N, l} F_{nl}^2}{f},$$

where  $f = V_{\mathcal{S}} / (\frac{4}{3}\pi r_s^3)$  is the volumetric fill ratio between the geometry  $\mathcal{S}$  and its bounding sphere with radius  $r_s$ . In our experiments with *volume-based* ZMIs, we found that the  $Z_N$  reach a plateau for growing  $N$ . The author conjectures that  $Z_N$  converges this plateau faster for “regular” objects  $\mathcal{S}$  than irregular ones. Sampling  $Z_N$  for some fixed value  $N$  (e.g.,  $N = 6$ ) therefore yields a candidate metric for irregularity. Providing a mathematical analysis of the  $Z_N$  was out of scope of this study.

Index	Symbol	Details
ZMIs	$ZM_{n,l}^{\text{surf}}$	Surface-based ZMI, $n \leq 20$ and $l$ such that $n - l > 0$ and even
Normalized ZMI energy	$Z_N^{\text{surf}}$	Squared sum of ZMIs, normalized by fill ratio, with $N = 1, 2, 3, 6, 10$

## 2.D Appendix: Contributions

### 2.D.1 AneuX morphology database

All studies described in this thesis made use of the AneuX morphology database. It currently consists of 750 aneurysm geometries from two internal and two external sources. The internal sources HUG1 and HUG2 have been acquired at University Hospitals in Geneva (HUG) and consist of totally 485 aneurysms from 365 patients. This internal dataset was complemented by data from two external projects: @neurIST<sup>126,127</sup> and Aneurisk<sup>128</sup>, adding another 265 aneurysms from 248 patients. Details about the different sources are provided in Section 2.3.1. All geometric models were processed according to the protocol described in Section 2.3.2. The dataset consists of a geometric representation of the aneurysm in four different cut configurations (*dome*, *ninja*, *cut1*, *cut2*, cf. Figure 2.3) and different mesh resolutions (target cell area of  $0.01\text{mm}^2$ ,  $0.05\text{mm}^2$ ,  $0.10\text{mm}^2$  and  $0.25\text{mm}^2$ ). For all but few exceptions, patient sex and age, and the aneurysm's rupture status and anatomical location was known. The dataset is summarized in Table 2.1.

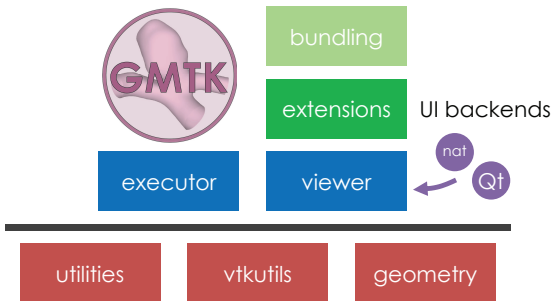
The author has laid out the AneuX morphology database and has implemented and performed all mesh-based processing. The surface geometries have been provided by Vitor Mendes Pereira, Rafik Ouared and team (HUG1); Philippe Bijlenga and Diana Sapina (HUG2); and Christoph M. Friedrich on behalf of the @neurIST consortium. The Aneurisk dataset has been downloaded from the public website: AneuriskWeb.

## 2.D.2 The Geometric Modelling Toolkit

An extensive number of tools were required for the processing of the aneurysm geometries and their subsequent analysis. These tools have been collected in the Geometric Modelling Toolkit (GMTK), an open source project written Python (version 3.6+) that extends the functionality of VTK<sup>141</sup> and VMTK<sup>135</sup>. Its architecture is illustrated in Figure 2.D.1.

Central design goals for the toolkit were (A) the efficient automated or interactive processing of hundreds of cases; (B) possibility to debug and visualize surface meshes; and (C) auto-documentation functionality to simplify reproducibility of the results.

Various commercial and open-source tools were available for operating with surface meshes. However, these tools did not directly support customized workflows optimized for the processing of hundreds of geometries.



**Figure 2.D.1:** Software architecture of the Geometric Modelling Toolkit (GMTK). The functionality is implemented in submodules `utilities`, `vtkutils` and `geometry`. The executor interface ensures a standardized interface. The viewer application (with native or Qt-based backend) is built on top of GMTK.

### **2.D.3 GMTK File Viewer**

The File Viewer is a minimalistic, cross-platform viewer for 3D surface geometries based on GMTK and VTK. It is extensible via an extension API and can be bundled with data, permitting an easy distribution of the tool.

The rating data for thy psychometric studies (Chapters 3 and 4) was acquired by means of a File Viewer extension (Figure 3.1). The rating task required the frequent comparison of different aneurysms, for which the workflow was optimized. The tool was key for the successful collection of rating data from 39 raters (comprising over 30'000 single morphological assessments).

The cut-tool, another File Viewer extension, was of great use to place the planar and non-planar cuts (Figure 2.3). More than 6000 different cuts were specified with this tool within a single working week.

Finally, the GMTK File Viewer was used extensively for studying, visualizing and debugging aneurysm morphology. All screenshots showing aneurysms have been created with this tool.

### 3 Quantification of perceived irregularity

To encounter the methodological difficulties discovered in the previous chapter, we followed up on a novel rater-based approach to describe the shape of IAs. The goal was to better understand what constitutes an irregular aneurysm as seen by a human rater. While this chapter focuses on the quantification of *perceived* morphology, the next chapter 4 examines its clinical relevance.

**Contributions:** *The conceptualization and implementation of this study are my own. The acquisition of rating data involved the development of a standalone rating tool consisting of a viewer for 3D geometries and a rating form. Sabine Schilling gave essential inputs on how to process the rating data statistically. Stefan Glüge assisted me in the application of machine learning methods. Eliisa Netti, Daniel Rüfenacht and Philippe Bijlenga were of great help for the recruitment of study participants. The aneurysm geometries used in this study are part of the AneuX morpho database. Preliminary results<sup>156,157</sup> have been presented at the 5<sup>th</sup> and 6<sup>th</sup> International Conferences on Computational and Mathematical Biomedical Engineering, CMBE 2017 (in Pittsburgh, PA, USA), and CMBE 2019 (in Sendai, Japan). The content of this chapter is completely taken from Juchler et al.<sup>158</sup>:*

*Juchler N, Schilling S, Glüge S, Bijlenga P, Rüfenacht D, Kurtcuoglu V, Hirsch S: Radiomics approach to quantify shape irregularity from crowd-based qualitative assessment of intracranial aneurysms. Computational Methods in Biomechanics and Biomedical Engineering. 2020*



### 3.1 Abstract

The morphological assessment of anatomical structures is clinically relevant, but often falls short of quantitative or standardized criteria. Whilst human observers are able to assess morphological characteristics qualitatively, the development of robust shape features remains challenging. In this study, we employ psychometric and radiomic methods to develop quantitative models of perceived irregularity of intracranial aneurysms (IAs). First, we collect morphological characteristics (e.g. irregularity, asymmetry) in imaging-derived data and aggregated the data using rank-based analysis. Second, we compute regression models relating quantitative shape features to the aggregated qualitative ratings (ordinal or binary). We apply our method for quantifying perceived shape irregularity to a dataset of 134 IAs using a pool of 179 different shape indices. Ratings given by 39 participants show good agreement with the aggregated ratings (Spearman rank correlation  $\rho_{sp} = 0.84$ ). The best-performing regression model based on quantitative shape features predicts the perceived irregularity with  $R^2 = 0.84 \pm 0.05$ .

### 3.2 Introduction

Linking disease phenotype to image-derived features for computer-aided diagnosis is a central aim in radiomics. While morphological assessment of anatomical structures plays an important role in clinical practice, it is typically based on qualitative, subjective descriptions of phenotypic characteristics. For the clinical use-case of intracranial aneurysm (IA) assessment, we present an approach to translate a qualitative diagnostic judgment of a morphological characteristic into a quantitative metric.

IAs are focal malformations of cerebral arteries, prevalent in 2-5% of the population.<sup>11</sup> On average, IAs rupture with an incidence rate of about 1% per year.<sup>27</sup> Ruptures lead to hemorrhagic stroke, associated with high mortality and morbidity.<sup>8,159</sup> Today, disease status is assessed subjectively, as is the need to treat an aneurysm. An increasing body of literature links irregular aneurysm shape with pathologic wall biology<sup>41,112</sup> and increased rate of rupture.<sup>29,160</sup> Some clinicians have hypothesized this association all along, integrating it into

their subjective mental model for making treatment decisions.<sup>32,160,161</sup>

“Irregularity” is a vague concept: the medical community has neither developed a common vocabulary to describe irregularity, nor established a standard irregularity rating scheme. As a result, assessments differ between clinicians.<sup>162,163</sup>

To address these issues, we have developed a method for morphological assessment of IAs that can be generalized to other psychometric quantification problems. Based on data collected with our interactive rating tool for 3D geometries, we show how to aggregate perceived irregularity and judge the degree of consent (Spearman rank correlation). We compare sub-cohorts of raters (e.g. laypersons vs. clinicians) to assess the test-setup or the inclusion criteria of the raters. Using a pool of geometric shape features, we derive and validate regression models to reproduce the aggregated irregularity ratings. Finally, we break down perceived irregularity into particular morphological attributes (presence of blebs, lobules, rough surface, asymmetry, complex parent vasculature) and again model these quantitatively.

### **3.3 Materials and methods**

Given 3D models of the structure under observation, we relate qualitative ratings of morphology to quantitative descriptions of shape through correlation analysis and multivariate regression. In the following, we elucidate our method for the assessment of IA irregularity.

#### **3.3.1 Imaging and patient data**

Our dataset comprised 134 saccular IAs (41 ruptured, 78 unruptured, 15 with unknown rupture status) of 110 patients from the University Hospital Geneva (HUG). We extracted geometric models of the aneurysms and the surrounding vasculature from 3D rotational angiographies (3DRAs, voxel sizes in the range of 200-350 $\mu$ m) by applying vessel lumen segmentation (geodesic active regions,<sup>134</sup> implemented in the software package GIMIAS<sup>132</sup>). Standard marching cubes<sup>164</sup> was used to convert binary segmentation images into surface meshes. We re-meshed all surfaces using VMTK<sup>135</sup> for a target mesh-cell area

of  $0.01\text{mm}^2$ . This step led to more regular meshes and improved overall quality. We assessed the stability of our pipeline regarding different mesh resolutions by comparing the quality of the regression models, exemplarily shown for target mesh area of  $0.01\text{mm}^2$  and  $0.05\text{mm}^2$ .

### 3.3.2 Quantitative shape description

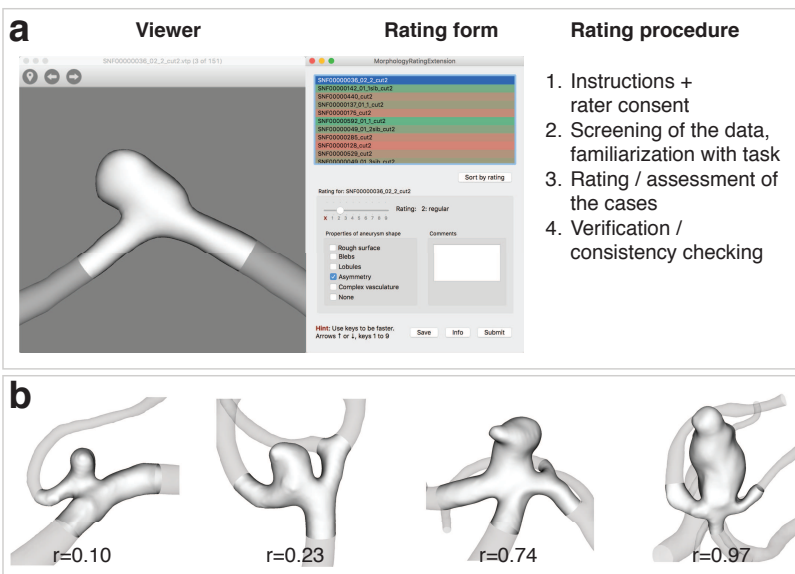
We isolated the IA dome with a single planar cut (e.g. Raghavan et al.<sup>97</sup>) and computed a set of morphological indices falling into three different types (cf. Table 3.1). *Geometry indices* (GIs) capture specific geometric characteristics of the aneurysm dome in a single number. We considered 12 different GIs that primarily measure size or shape.<sup>97,99,105</sup> Metrics computed from local surface properties are termed *distribution-derived indices*. We included features based on *curvature* (both Gaussian and mean curvature<sup>97</sup>) and *surface writhe*.<sup>102</sup> Curvature features measure the “bending” and “tortuosity” of the surface, while writhe-based features can be interpreted as a measure of surface asymmetry. Finally, *moment-based descriptors* decompose the surface into different modes. We included Zernike Moment Invariants (ZMI),<sup>120</sup> which are related to spherical harmonics and compactly represent a 3D surface geometry. Being invariant to scale, translation and rotation of surfaces, ZMIs are well-suited as a basis for comparison of 3D objects. For this study, we used *surface-based* ZMIs<sup>104</sup> up to order  $n = 20$ , resulting in 121 different indices. In total, the pool comprised  $d = 179$  different shape indices.

Type	Sub-type	Details	#indices
<b>Geometry Indices</b>	Size indices	<ul style="list-style-type: none"> <li>- Dome volume</li> <li>- Dome surface area</li> <li>- Neck diameter</li> <li>- Max. diameter</li> <li>- Height</li> <li>- Aneurysm size</li> </ul>	6
	Shape indices	<ul style="list-style-type: none"> <li>- Non-sphericity index</li> <li>- Ellipticity index</li> <li>- Undulation index</li> <li>- Aspect ratio</li> <li>- Conicity factor</li> <li>- Bottleneck factor</li> </ul>	6
<b>Distribution-derived features</b>	Curvature metrics	<ul style="list-style-type: none"> <li>- Gaussian and mean curvature</li> <li>- Distribution characteristics</li> <li>- Total curvature, normalized by surface area</li> </ul>	22
	Writhe metrics	<ul style="list-style-type: none"> <li>- Free writhe and normalized inner-squared writhe</li> <li>- Distribution characteristics</li> </ul>	24
<b>Moment-based descriptors</b>	Zernike Moment	- Surface-based	121
	Invariants (ZMI)	- Order $n = 20$	
<b>Total</b>			<b>179</b>

**Table 3.1:** Composition of the feature pool for the morphological assessment of IAs. #indices indicates the number of indices that a particular type contributes to the pool.

### 3.3.3 Qualitative shape assessment

The rating tool consisted of two elements: a 3D viewer to examine the object interactively using computer mouse and keyboard, and a rating form to collect the ratings (cf. Figure 3.1a). The written task description emphasized the *qualitative* assessment of shape without providing further clinical information. The raters confirmed having carefully read and understood the instructions before starting the inquiry.



**Figure 3.1: (a)** Screenshot of the rating tool for interactive display of 3D geometries and rating acquisition, here for IA morphology: perceived irregularity (ordinal assessment) and a list of morphological attributes (binary assessments). The tool facilitates the efficient comparison of cases and rating verification. **(b)** Exemplary IA geometries ordered by increasing perceived irregularity from very regular ( $r'_i = 0$ ) to very irregular ( $r'_i = 1$ ).

The raters assessed each aneurysm in terms of its shape irregularity on a 9-point rating scale, from “1 – very regular” to “9 – very irregular” (cf. Figure 3.1b). We intentionally refrained from specifying the properties of a perfectly regular/irregular aneurysm. Instead, we relied on the common-sense understanding of geometry and the intuitive nature of the inquiry. To familiarize themselves with the dataset, the participants had to skim through all cases first. After case-by-case assessment with randomized order, the participants could sort the geometries by increasing irregularity rating and adjust their initial assessment.

We chose a 9-point rating scale to strike a balance between task complexity, rater consistency and informational value: Additional irregularity

levels permit a more fine-grained ordering of the cases, but also impair the rater’s ability to consistently sort the cases by increasing irregularity.

As a secondary task, we asked the raters to decide whether the aneurysm under examination exhibited one of the following five morphological attributes: a rough (non-smooth) surface, blebs or lobules, an asymmetric appearance, a complex configuration of the parent vasculature/bifurcation, or none of those (cf. Table 3.2). We refer to this part of the inquiry as the (binary) *assessment of morphological attributes*.

A cohort of 39 participants was recruited for the inquiry, which all passed an outlier test (see next section). For each participant, the inquiry resulted in a rough ordering of the cases by perceived irregularity, measured in 9 levels. A subset of 26 raters additionally provided assessments for morphological attributes.

<b>Attribute</b>	<b>Descriptions</b>
<i>Rough surface</i>	Does the surface show an overall rough, non-smooth surface? Does it show structures that do not qualify as blebs or lobules?
<i>Blebs</i>	Are any blebs visible? A bleb is any localizable elevation of the dome surface whose volume is <i>smaller than 25%</i> of the primary dome compartment.
<i>Lobules</i>	Are any lobules visible? A lobule is any localizable elevation of the dome surface whose volume is <i>larger than 25%</i> of the primary dome compartment.
<i>Asymmetry</i>	Does the aneurysm appear asymmetric? Geometric asymmetry applies if the aneurysm dome lacks axes of symmetry.
<i>Complex vasculature</i>	Does the surrounding vasculature look complex such that it affects the overall perceived complexity of the aneurysm?
<i>Nothing applies</i>	None of the options above apply.

**Table 3.2:** Descriptions of the morphological attributes used in this study.

### 3.3.4 Processing of the rating data

#### 3.3.4.1 Ordinal rating of irregularity

The varying shape of the rating distributions for each rater (cf. Figure 3.2a) reflects rater subjectivity. To correct for this effect, we ranked the ordinal ratings per rater, where the average ordinal rank for ratings of equal value (tied rank) was computed.

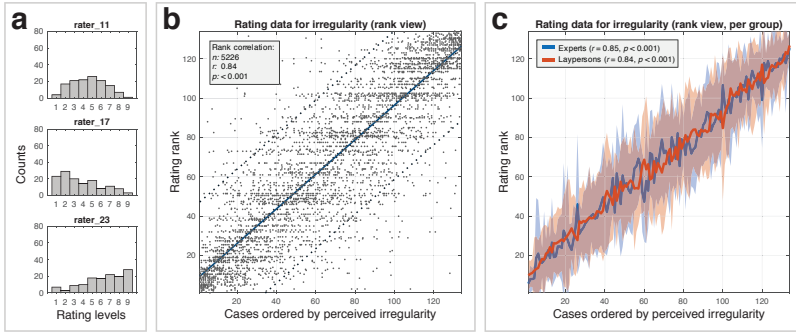
Next, we aggregated the ranked irregularity ratings by computing their means per case. The rating aggregates  $r_i$  for case  $i$  take values in the range  $[1, n]$  where  $n = 134$  is the sample size. To normalize this range, we mapped the rating aggregates  $r_i$  linearly onto  $r'_i \in [0, 1]$ , with 0 and 1 standing for “very regular” and “very irregular”, respectively. Hereinafter, we will refer to these normalized, rater-bias adjusted aggregates  $r'_i$  as *perceived irregularities*. As a measure of collective agreement, we computed the Spearman rank correlation  $\rho_{Sp}$  between perceived irregularities  $r'_i$  and the original rating ranks of every rater. To characterize the rater cohort and to test for potential problems with the rating acquisition, we analysed the contribution  $\epsilon_j = \sigma_j^2 / \sigma_{tot}^2$  of each rater  $j$  to the overall variance

$$\sigma_{tot}^2 = \sum_{i=1}^n \frac{1}{m-1} \sum_{j=1}^m \left( r'_{ij} - \mu(r'_{ij}) \right)^2 \quad (1)$$

in the data ( $m$ : number of raters,  $n$ : number of cases,  $r'_{ij}$ : normalized rank for rating  $i$  of rater  $j$ ). We applied a robust  $z$ -score analysis on the  $\epsilon_j$  following.<sup>165</sup> A rater  $j$  was defined to be an outlier if the modified  $z$ -score

$$z_{mod}(\epsilon_j) = (\epsilon_j - \tilde{\epsilon}) / \tilde{\sigma}_\epsilon = 0.6745 \cdot (\epsilon_j - \tilde{\epsilon}) / \text{MAD}_\epsilon \quad (2)$$

was larger than 4.0, where  $\tilde{\sigma}_\epsilon$  represents a robust estimator for the standard deviation of the  $\epsilon_j$ ,  $\tilde{\epsilon}$  and  $\text{MAD}_\epsilon$  denote the median and the median absolute deviation of  $\epsilon_j$ , respectively, and 0.6745 is the 75<sup>th</sup> percentile of the standard normal distribution.



**Figure 3.2:** (a) Exemplary histograms summarizing the ordinal irregularity ratings of three different raters, demonstrating different rating biases. (b) Scatter plot showing the ratings by the 39 included raters for the 134 aneurysms ( $n = 5226$  data points, ranked per rater, Spearman rank correlation  $\rho_{sp} = 0.84$  ( $p < 0.001$ ) between the individual rating ranks and the aggregated). The plot also shows the regression line and its 95% tolerance- and confidence intervals (dotted lines). (c) Data stratified by rater sub-cohort (clinical experts vs. instructed laypersons). Solid lines: mean rating ranks per aneurysm. Shaded areas:  $\pm$  standard deviation of rating ranks.

### 3.3.4.2 Binary ratings of morphological attributes

For each case  $i$  and morphological attribute  $k$ , we computed the relative counts  $q'_{ik}$  of votes in favour of that attribute, normalized by the number of raters. Similar to perceived irregularity, this metric captures how strongly the rater cohort agrees in recognizing a particular morphological attribute. Note that the aggregates  $q'_{ik}$  have similar properties to the perceived irregularities  $r'_i$  and therefore can be used interchangeably in the subsequent analysis.

Like in the case of perceived irregularity, we also assessed the collective agreement for the ratings of morphological attributes. We considered two methods to assess the average rater agreement for *binary* ratings of morphological attributes. Fleiss' kappa  $\kappa_F$  measures the agreement within the entire rater cohort, which we evaluated for each morphological attribute separately. Because this first approach ignores any rater-dependent subjectivity, we used a second approach in which we compare the binary ratings  $q_{ijk}$  of rater  $j$ ,



attribute  $k$ , and cases  $i$  with the binarized aggregates  $q_{ik}^{\text{bin},j} = \text{sign}(q'_{ik} - \tau_{jk}^*)$ . The binarization threshold  $\tau_{jk}^*$  is computed for each rater and attribute such that Cohen’s kappa  $\kappa_C$  (a measure for inter-rater agreement) between rater  $j$  and “binarized average rater” is maximal. In this context,  $\tau_{jk}^*$  can be interpreted as a perceptual threshold for a rater  $j$  to accept the presence of a particular attribute  $k$ . Table 3.3 summarizes the average  $\kappa_C$  and  $\tau_{jk}^*$  for all 26 raters. Both  $\kappa_C$  and  $\tau_{jk}^*$  can be used to identify outlier raters using a similar procedure as described in the main article. No such outliers were found in our data.

Morphological attribute	Hard comparison	Soft comparison	Threshold $\tau_{jk}^*$ (mean $\pm$ std)
	Fleiss' kappa $\kappa_F$	Cohen's kappa $\kappa_C$ (mean $\pm$ std)	
Asymmetry	0.173 (slight-fair)	0.535 $\pm$ 0.148 (moderate-substantial)	0.464 $\pm$ 0.170
Rough surface	0.316 (fair-moderate)	0.659 $\pm$ 0.094 (substantial)	0.397 $\pm$ 0.187
Blebs	0.274 (fair)	0.625 $\pm$ 0.075 (substantial)	0.453 $\pm$ 0.191
Lobules	0.282 (fair-moderate)	0.647 $\pm$ 0.117 (substantial)	0.438 $\pm$ 0.223
Complex vasculature	0.175 (slight-fair)	0.523 $\pm$ 0.143 (moderate)	0.322 $\pm$ 0.171

**Table 3.3:** Average agreement for the binary ratings on the morphological attributes, evaluated using hard and soft comparisons of raters (see text). The data comprises ratings for 134 cases from 26 different raters (16 instructed laypersons, 10 clinical experts). Our results suggest that the raters substantially agree if the rater subjectivity is taken into account, and that agreement varies across different attributes.

### 3.3.4.3 Association of qualitative ratings and quantitative features

We performed a multivariate analysis to identify “crowd-sourced” shape models that capture perceived morphological characteristics. The size of the feature pool was first reduced by several means: Either we applied principal component analysis (PCA) to identify directions in the feature space with maximal information content, or we ranked and selected relevant features based on univariate linear metrics (correlation coefficients between features and perceived characteristics) or *feature importance*. Feature importance is a statistical measure of how relevant a predictor was in training a potentially nonlinear relationship between the predictor variables (shape features) and

response (ratings) with decision trees. To estimate feature importance and to compute non-linear regression, we made use of gradient boosting machines (GBM) provided through the LightGBM framework.<sup>144</sup>

Next, we computed multivariate regression models for four different configurations (cf. Table 3.4).  $\mathcal{F}_{univ}$  represents the set of best performing features from the univariate analysis,  $\mathcal{F}_{imp}$  signifies the set of most important features (“importance” as defined above), accounting for 80% of the total importance. For the PCA, the  $d^*$  principal components in the (ranked) data space are used, where  $d^* < d = 179$  is the number of features that preserve 90% of the overall variance in the data. Instead of ordinary least squares (OLS) regression, we relied on support vector regression (SVR), which is more robust and performed better on our data for higher dimensional feature spaces.

We trained and validated the multivariate models with 5-fold cross-validation and  $q = 50$  repetitions. The average root-mean-square error (RMSE) and the coefficient of determination ( $R^2$ ), computed over the  $q$  repetitions, were used as performance metrics to compare the different regression models.

Model	Regressor	Feature space configuration			Motivation
		Selection	$d$	Repr.	
Ref.	SVRlin	Best univariate feature	1	ranked/metric	Reference model using the best performing univariate feature of the pool.
A1	SVRlin	$\mathcal{F}_{univ}$	19	ranked	Combine statistically independent predictors with good univariate prediction in a multivariate model.
A2	SVRlin	$\mathcal{F}_{univ} \cup \mathcal{F}_{imp}$	31	ranked	
B	SVRlin	PCA, 90% of total variance	6	ranked	Reduce problem complexity by reducing redundancy in the data space. This assumes an (approximately) linear relationship.
C	GBM	$\mathcal{F}_{imp}$	31	metric	A nonlinear regression model may capture complex relationships between explanatory and predicted variables more accurately

**Table 3.4:** Overview of the model configurations used in this study.  $d$  represents the number of dimensions of the reduced feature space;  $\mathcal{F}_{univ}$  and  $\mathcal{F}_{imp}$  are the set of features with the best univariate and most important candidates, respectively. SVRlin: support vector regression with linear kernel. GBM: gradient boosting machine. PCA: principal component analysis. Ref.: reference model. Repr.: data representation.

## 3.4 Results

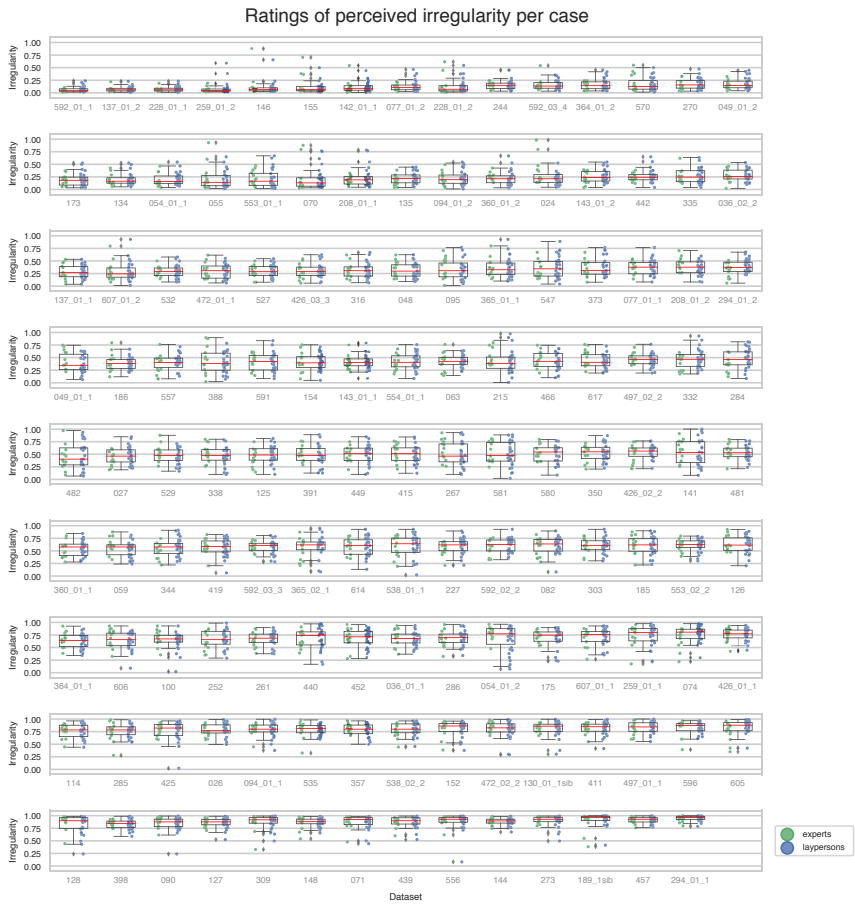
### 3.4.1 Rating data

We acquired rating data for perceived irregularity of 39 raters from Japan, USA and Europe, all of which passed the outlier test base on the robust z-score. This resulted in a pairwise Spearman rank correlation  $\rho_{Sp} = 0.84$  ( $p < 0.001$ ), where  $\rho_{Sp}$  was computed between perceived irregularities  $r'_i$  and the original ratings, ranked per rater,  $r'_{ij}$ .

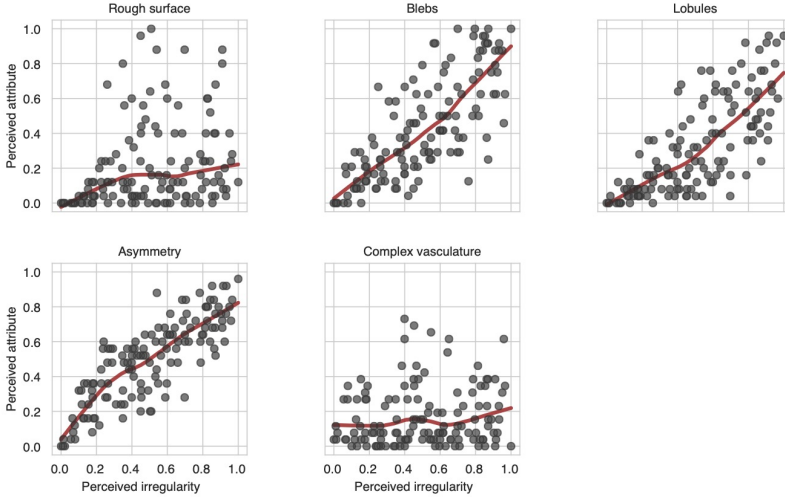
We also compared the ratings of rater sub-cohorts stratified by professional background. While *clinical experts* rated morphological irregularity on average by 0.467 rating points higher than the *instructed laypersons* (the difference is significant, paired-sample *t*-test,  $p < 0.001$ ), the resulting *rank*-based aggregate for perceived irregularity cannot be discriminated statistically (paired-sample *t*-test,  $p = 0.967$ ). As a consequence, the perceived irregularity  $r'_i$  is very similar for experts and laypersons, as seen in Figure 3.2c.

The level of agreement *per case*  $i$ , measured here as the standard deviation  $\sigma'_i$  of (per-rater) ranked irregularity ratings  $r'_{ij}$ , varied across cases. A low standard deviation implies a good interrater agreement.  $\sigma'_i$  ranged between 0.050 and 0.261, with a mean of 0.152 (measured in the scale of perceived irregularity  $r'_i \in [0,1]$ ). The agreement was higher between experts than between laypersons ( $\bar{\sigma}'_{i,exp} = 0.146$ , vs.  $\bar{\sigma}'_{i,lay} = 0.151$ ), but the difference did not reach statistical significance (paired *t*-test,  $p = 0.16$ ). The best agreement among the raters was observed for extreme cases; very regularly or very irregularly shaped aneurysms were rated the most consistently (cf. Figure 3.3).

Figure 3.4 shows the aggregates  $q'_{ik}$  for the morphological attributes in relation to the perceived irregularity  $r'_i$ . Interpolation curves (locally weighted scatterplot smoothing, LOWESS) reveal that perceived irregularity is associated with perceived presence of asymmetry, blebs and lobulations. This trend, however, was not distinguishable for rough surface and complex vasculature.



**Figure 3.3:** The irregularity rating ranks by all 39 raters for all 134 cases. The cases are sorted by increasing mean. By comparing the data spread, one can observe that the inter-rater agreement varies considerably between different cases. As a trend, the agreement is high for the extreme cases (very regular, very irregular).



**Figure 3.4:** Aggregated ratings for the morphological attributes. The plots show the data (dots) for the 134 cases, comparing the perceived irregularity (abscissa) with the aggregated ratings (relative counts) of the following six attributes (multiple choices allowed): asymmetry, rough surface, blebs, lobules, complex parent-vasculature and nothing (if none of the characteristics applied). We also show LOWESS regression curves (with smoothing factor 0.2) to identify possible trends in the ratings.

### 3.4.2 Multivariate quantitative model for perceived irregularity

Given the rating aggregates (explained variable) and the pool of shape descriptors (predictor variables), we trained statistical models that map feature vectors to ratings. We devised four model configurations (A1, A2, B, C, Table 3.4), for which we report the RMSE as performance metric, which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\tilde{r}'_i - r'_i)^2}$$

as performance metric (cf. Table 3.5). RMSE measures the average difference between predicted  $\tilde{r}'_i$  and measured perceived irregularity  $r'_i$ . We also report the coefficient of determination  $R^2$ , which measures the proportion of the total variance (in the predicted variable) explained by the model. For reference, we

give the best-performing *univariate* model, based on the curvature-metric area-normalized  $L_2$ -norm of Gaussian curvature, also known as GLN<sup>97</sup>. This model was trained with the same cross-validation setup for both ranked and metric, non-ranked data. Generally, the inclusion of additional predictors reduced the RMSE score. On ranked data, the prediction error was diminished by about 11% on average, and by about 28% on the metric data. The models predicting the aggregated assessments of morphological attributes generally resulted in a lower prediction performance.

Predicted variable	Data	Model	RMSE			R <sup>2</sup>		
			mean	std	p-val	mean	std	p-val
Perceived irregularity	Ranked	<i>Reference</i>	0.129	0.016	Ref.	0.788	0.067	Ref.
		<i>Model A1</i>	0.122	0.016	< 0.001	0.809	0.067	< 0.001
		<i>Model A2</i>	0.113	0.015	< 0.001	0.836	0.051	< 0.001
		<i>Model B</i>	0.129	0.016	< 0.001	0.786	0.068	< 0.001
	Metric	<i>Reference</i>	0.150	0.018	Ref.	0.677	0.085	Ref.
		<i>Model C</i>	0.109	0.012	< 0.001	0.829	0.055	< 0.001
Rough surface	Ranked	<i>Reference</i>	0.228	0.026	Ref.	0.464	0.138	Ref.
		<i>Model A2</i>	0.216	0.027	< 0.001	0.513	0.144	< 0.001
Blebs	Ranked	<i>Reference</i>	0.203	0.023	Ref.	0.511	0.133	Ref.
		<i>Model A2</i>	0.189	0.024	< 0.001	0.577	0.125	< 0.001
Lobules	Ranked	<i>Reference</i>	0.203	0.037	Ref.	0.510	0.170	Ref.
		<i>Model A2</i>	0.174	0.024	< 0.001	0.638	0.109	< 0.001
Asymmetry	Ranked	<i>Reference</i>	0.202	0.022	Ref.	0.492	0.141	Ref.
		<i>Model A2</i>	0.172	0.020	< 0.001	0.627	0.114	< 0.001
Complex vasc.	Ranked	<i>Reference</i>	0.300	0.027	Ref.	0.032	0.137	Ref.
			0.293	0.027	> 0.05	0.070	0.179	> 0.05

**Table 3.5:** Summary of the prediction performances for the different multivariate model configurations used to predict the perceived irregularity (upper half) and the morphological attributes (lower half). The models were trained and validated in a nested cross-validation scheme with 50 repetitions. For perceived irregularity, the best performing *univariate* model (based on the curvature metric GLN) is given as reference. We evaluated the models for ranked and non-ranked data representation, where both explanatory and predicted variables were ranked prior to training. Root mean squared error (RMSE) and the coefficient of determination ( $R^2$ ) are provided. For the morphological attributes, we report the results of the best-performing univariate and multivariate models.

## 3.5 Discussion

In this study, we have collected and aggregated qualitative, ordinal and binary ratings for aneurysm shape. For instance, the perceived irregularity  $r'_i$  reflects the collective opinion on the morphological irregularity. The single irregularity ratings per case can vary strongly between participants (cf. Figure 3.3), but *rank*-based analysis (Spearman correlation  $\rho_{Sp} = 0.84$ ) suggests that raters agree, on average, with the *ordering* of the cases.

This result is robust to local permutations in the ordering of the cases or the exclusion of some raters. Using the aggregated metric allows correction for the inherent subjectivity that comes with irregularity ratings. The results from the subsequent analysis are thus equally robust by design.

The pronounced spread of the ratings around the average is a consequence of the open task formulation, the inconsistency typical of subjective assessment (*intra*-rater disagreement), and the heterogeneous composition of the rater cohort (*inter*-rater disagreement). However, our rank-based method deals robustly with the amount of rater variability.

The level of agreement varies considerably between different cases: extreme cases (very regular, very irregular) are rated more consistently than cases in-between. This variability would ideally be addressed with quantitative criteria to evaluate morphological irregularity. To determine how professional qualification affects ratings, we also compared sub-cohorts of participants. Our results suggest that clinical experience did not affect the judgment of perceived irregularity.

Finally, we developed statistical models to predict perceived irregularity. Such models map quantitative morphological metrics to the subjective assessment of shape, a task that can be considered cognitively complex, involving intuition, experience and conscious thinking. So far, no quantitative metric exists that specifically measures irregularity of aneurysm shape. A tool to quantify irregularity will help clinicians to assess aneurysms while removing rater subjectivity.

A combination of multiple shape features performed better than univariate models to predict perceived irregularity (cf. Table 3.5). A larger model uncertainty (standard deviation of RMSE, Table 3.5), as a result of an increased number of model predictors, is overcompensated by increased prediction

accuracy. In the case of ranked and metric data, the RMSE improved by 11% and 28%, respectively.

We repeated the analysis for other morphological characteristics for which it is equally difficult to specify robust, quantitative rules. The prediction performances of these models for the aggregates  $q'_{ik}$ , however, are poorer. This might be partially explained by the binary assessments carrying less information than ordinal ratings. Binary rating data leads to graded aggregates  $q'_{ik}$ , with repercussions on the prediction metrics. Furthermore, the shape features included only insufficiently describe the IA attributes. The development of specific features for these attributes was outside the scope of this study. Regardless of the lower prediction power, we demonstrated that the method can also be applied to binary rating data.

The morphological assessment of anatomical structures is not only relevant for IAs. More generally, the morphology of tissue, bones, organs or vessels, plays an important role in the management of various diseases. We argue that the proposed methodology to capture, normalize and inspect the collective opinion of a rater cohort is equally applicable in other clinical contexts as well. There are two principal requirements for our methodology: 1) The morphology must be assessable by visual inspection, either from 3D surface geometries as in our case, or from 2D or 3D intensity images. 2) A set of quantitative metrics must be computable from the input data (feature pool) that are thought to capture the qualitative metric (e.g. asymmetry, irregularity, tortuosity).

When working with morphological metrics derived from imaging data, we recommend examining their mesh and resolution dependency. In our use-case, the reduction of mesh resolution (we assessed two surface meshes with average cell areas of  $0.01\text{mm}^2$  and  $0.05\text{mm}^2$ ) did have a small but noticeable effect on single features (Table 3.6 of the supplemental material section). While most metrics are unaffected, curvature metrics are sensitive to mesh resolution. The lower mesh resolution of  $0.05\text{mm}^2$  yields slightly better correlation coefficients. Fine tuning the mesh size in respect to the imaging resolution holds potential to incrementally improve the model performance.

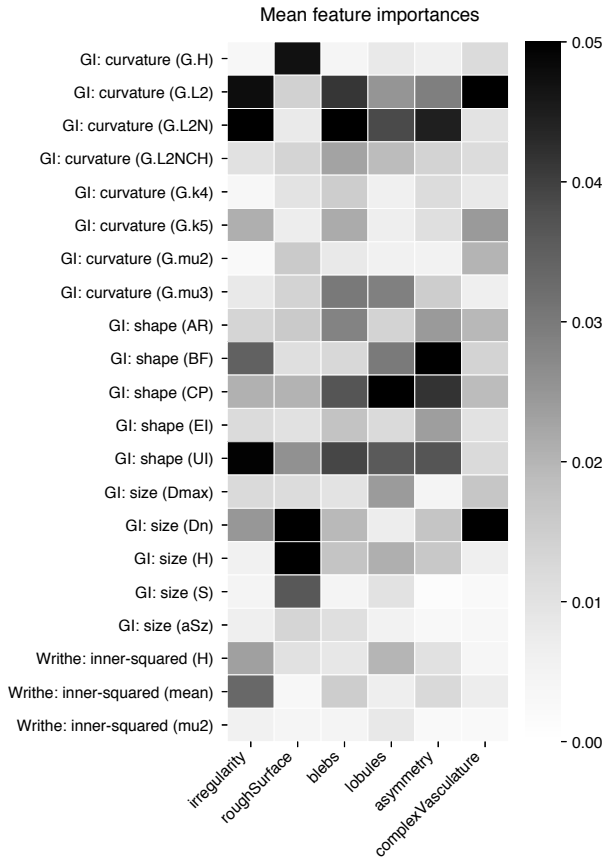
Putatively, the model accuracies will further increase with a higher number of raters and cases. Although we consider our dataset well-balanced in terms of morphological attributes, it is possible that some characteristics are



over- or underrepresented. The features available in our pool might therefore not encompass all morphological attributes that raters take into account, and it is conceivable that metrics exist that encode perceived irregularity more efficiently than the ones we used. We disregard other factors that may have an influence on the morphology of the structure under assessment. In the use-case of IAs presented, for example, a stratification of the aneurysms by location would be an interesting aspect for a follow-up study.

Predictor	Correlation $\rho_{Sp}$	
	$A = 0.01 \text{ mm}^2$	$A = 0.05 \text{ mm}^2$
GI: curvature ( $GLN$ )	0.89	0.92
GI: curvature ( $MLN$ )	0.88	0.91
writhe: inner squared ( $W_H^{L_2}$ )	0.84	0.83
writhe: inner squared ( $W_{\text{mean}}^{L_2}$ )	0.84	0.83
GI: curvature ( $G. -L_2N$ )	0.84	0.84
GI: shape ( $NSI$ )	0.80	0.80
GI: shape ( $EI$ )	0.76	0.76
ZMI: normalized energy ( $Z_{30}^{\text{surf}}$ )	0.74	0.73
ZMI: normalized energy ( $Z_{10}^{\text{surf}}$ )	0.74	0.74
ZMI: normalized energy ( $Z_{20}^{\text{surf}}$ )	0.74	0.73
GI: curvature ( $G. L_2NCH$ )	0.73	0.77
ZMI: normalized energy ( $Z_5^{\text{surf}}$ )	0.73	0.73
GI: shape ( $BF$ )	0.72	0.72
GI: size ( $aSz$ )	0.70	0.70
GI: shape ( $UI$ )	0.67	0.71
GI: curvature ( $M. L_2NCH$ )	0.66	0.76

**Table 3.6:** Best performing univariate predictors for perceived irregularity, evaluated for two different average mesh cell areas  $A = 0.01\text{mm}^2$  and  $A = 0.05\text{mm}^2$ . We included only metrics with Spearman correlation  $\rho_{Sp} > 0.7$ . The overall ordering of the features appears relatively stable for the two different mesh sizes examined. Only curvature metrics yielded systematically larger coefficients. *Abbreviations:* Curvature  $L_2N$  – total curvature ( $L_2$ -norm), normalized by the surface area; curvature  $L_2NCH$  – same as curvature  $L_2N$  but further normalized by the total curvature ( $L_2N$ ) of the convex hull; writhe mean,  $H$ : mean and entropy or second statistical moment of the writhe values for a surface; GI – geometry indices; NSI – non-sphericity index; EI – ellipticity index; UI – undulation index; BF – bottleneck factor; aSz – aneurysm size; ZMI: Zernike moment invariants.



**Figure 3.5.** Comparison of the mean feature importance (FI) for the prediction of perceived characteristics, averaged over the 1000 gradient boosting machines (GBMs) trained in the feature selection step. FI measures how valuable a feature was when training GBMs. Black and white colour indicate high and low FI, respectively. The listing is freed from highly redundant features and features that show low importance in all morphological characteristics. Abbreviations: see caption of Table 3.6.

## 3.6 Conclusions

We successfully applied our method to the assessment of IA morphology, for which we trained novel quantitative models for irregularity using qualitative assessments of shape. The inspection of qualitative morphological assessment across multiple raters offers possibilities i) to develop new consensus-based rating-schemes, and ii) to design quantitative tools for the judgement of morphological characteristics. Since the elements of our method do not depend on the particular use-case, our methodology can be useful for the assessment of anatomical structures other than aneurysms.

## 4 Diagnostic value of irregularity

When assessing the shape of unruptured IAs as seen in medical imaging, clinicians rely mostly on their subjective impression. In chapter 3, we introduced the methodology to acquire ratings on perceived morphology from human participants with the goal to find a quantitative surrogate for the human perception of irregularity. In this second rater-based study, we use the same rating data to examine how perceived irregularity and other morphological attributes relate to clinical factors.

**Contributions:** *The conceptualization and implementation of this study are my own. Philippe Bijlenga helped to coin this study for a clinical audience. Sabine Schilling provided guidance in statistical matters. The content of this chapter is completely taken from Juchler et al.<sup>166</sup>:*

*Juchler N, Schilling S, Bijlenga P, Morel S, Rüfenacht D, Kurtcuoglu V, Hirsch S: Shape irregularity of the intracranial aneurysm lumen exhibits diagnostic value. Acta Neurochirurgica. 2020*

## 4.1 Abstract

**Background:** Morphological irregularity is linked to intracranial aneurysm wall instability and manifests in the lumen shape. Yet there is currently no consent on how to assess shape irregularity. The aims of this work are to quantify irregularity as perceived by clinicians, to break down irregularity into morphological attributes, and to relate these to clinically relevant factors such as rupture status, aneurysm location, and patient age or sex.

**Methods:** 13 clinicians and 26 laypersons assessed 134 aneurysm lumen segmentations in terms of overall perceived irregularity and five different morphological attributes (presence/absence of a rough surface, blebs, lobules, asymmetry, complex geometry of the parent vasculature). We examined rater agreement and compared the ratings with clinical factors by means of regression analysis or binary classification.

**Results:** Using rank-based aggregation, the irregularity ratings of clinicians and laypersons did not differ statistically. Perceived irregularity showed good agreement with curvature (coefficient of determination  $R^2 = 0.68 \pm 0.08$ ) and was modeled very accurately using the five morphological rating attributes plus shape elongation ( $R^2 = 0.95 \pm 0.02$ ). In agreement with previous studies, irregularity was associated with aneurysm rupture status ( $AUC = 0.81 \pm 0.08$ ); adding aneurysm location as an explanatory variable increased the AUC to  $0.87 \pm 0.09$ . Besides irregularity, perceived asymmetry, presence of blebs or lobules, aneurysm size, non-sphericity, and curvature were linked to rupture. No association was found between morphology and any of patient sex, age, and history of smoking or hypertension. Aneurysm size was linked to morphology.

**Conclusions:** Irregular lumen shape carries significant information on the aneurysm's disease status. Irregularity constitutes a continuous parameter that shows a strong association with the rupture status. To improve the objectivity of morphological assessment, we suggest examining shape through six different morphological attributes, which can characterize irregularity accurately.

## 4.2 Introduction

Intracranial aneurysms (IAs) are focal deformations of cerebral arteries, prevalent in 2-5% of the population.<sup>11</sup> IAs normally remain stable, yet they rupture with a lesion incidence rate of 1.2% per patient year.<sup>27</sup> The resulting hemorrhagic stroke is catastrophic with high mortality and morbidity.<sup>8,159</sup> IAs are increasingly detected due to improving imaging technology and its frequent regular use. As most diagnosed IAs are deemed to be stable, clinicians have to take complex disease management decisions. Meanwhile, biomarkers expressing the instability of the detected IA are still lacking today.

Recent studies on the pathogenic processes of wall remodeling suggest that the biological status of the IA manifests in shape changes. Irregularly shaped aneurysms have been associated with instable wall conditions<sup>41,112</sup> and higher risk of rupture.<sup>29,32,160</sup> Morphological wall characteristics such as irregular protrusions, flattened and slightly curly surfaces, or indentations are assumed to indicate destructive remodeling processes within the vessel wall, thrombus formations or vessel wall hyperplasia.<sup>6,41,112</sup> IA wall remodeling is thought to be a progressive process, where with further deterioration of the wall more irregularities appear. Hence, radiologists often appraise the shape of the aneurysm lumen as a proxy for wall remodeling.

Aneurysm shape irregularity has recently been added to the risk assessment of IAs.<sup>20,59,167</sup> Although used as a descriptive category, shape irregularity is not formally defined, leading to inconsistent shape assessments between raters.<sup>162,163</sup> A considerable number of metrics exists to quantify IA morphology,<sup>97,99,105,168</sup> but no robust criteria to distinguish between regular and irregular shapes have been established so far.

We have previously developed a quantitative model for lumen irregularity that matches the human perception of shape.<sup>158</sup> We employed a psychometric method to measure the perceived irregularity of IA domes from human raters assessing highly resolved representations of IA lumens. We reproduced aggregated shape assessments accurately by using a multivariate model of quantitative shape features that can be computed automatically from image data.

In this study, we examine the diagnostic value of lumen irregularity in IA domes in three steps: (a) evaluate the consistency of clinical experts and instructed laypeople in their assessment of irregularity; (b) identify the selection of morphological attributes that reflects perceived irregularity best; and (c) determine whether irregularity is associated with the known clinical risk factors aneurysm location, patient's sex, age, smoking status, and history of arterial hypertension, as well as the aneurysm's rupture status.

## **4.3 Materials and methods**

### **4.3.1 Imaging and patient data**

Between September 2006 and July 2015, information on 1164 patients was collected prospectively and consecutively in the @neurIST study<sup>33</sup> at the Geneva University Hospital. A significant proportion of the cohort was only followed up using magnetic resonance imaging (MRI) or computed tomography (CT) imaging. A total of 593 patients were identified as being at risk or suffered from a ruptured aneurysm and were therefore investigated by 3D rotational angiography (3DRA). From these cases, we selected 110 patients through a two-stage randomized process (first step: subset of 255 patients for which 3D reconstructions were accessible; second step: subset of 110 patients that visited the clinic between July 2014 and July 2015 for treatment or aftercare), harboring a total of 134 saccular IAs (41 ruptured, 78 unruptured, 15 with uncertain rupture status). In addition to angiographic data, the dataset included sex, age, rupture status (per aneurysm), history of smoking, and history of hypertension for a subset of the patients (Table 4.1).

We included both ruptured and unruptured aneurysms to compare the morphology between these two subcohorts and to benchmark our findings with existent literature. While the rupture status reflects aneurysm wall instability approximatively, the comparison shall not be overinterpreted as a prediction of the rupture risk.

Overview: patients			Unknown
Sex	81 females (age: 54.4y)	29 males (age: 50.6y)	0
Aneurysm rupture status	41 ruptured	56 unruptured	13
Smoking status	57 with smoking history	35 non-smokers	18
Hypertension status	40 with hypertension history	56 without hypertension	14

Overview: aneurysms		ruptured	unruptured	unknown	total
<b>Total</b>		41	78	15	134
<b>Patient sex</b>	female	26	62	10	98
	male	15	16	5	36
<b>Locations</b>	MCA M1	5	25	2	32
	ICA oph	3	16	4	23
	PCoMA	10	9	4	23
	ACoMA	11	8	1	20
	ACA A2-Per	2	5	1	8
	BA tip	3	4	0	7
	ICA bif	0	5	1	6
	others	7	6	2	15
<b>Size/Shape</b> (mean±std)	aSz [mm]	8.6±3.9	6.3±3.1	6.2±2.2	7.0±3.4
	AR [-]	1.5±0.5	1.1±0.4	1.2±0.6	1.2±0.5
	NSI [-]	0.21±0.06	0.13±0.06	0.15±0.10	0.16±0.07
	GLN [-]	6.9±3.2	4.2±1.9	5.1±3.1	5.1±2.8

**Table 4.1:** 110 patients harboring 134 aneurysms were included in this study. Smoker – former or current smoker after the estimated consumption of 300 or more cigarettes. Arterial hypertension – blood pressure greater than 140/90 mmHg, independent of any treatment for hypertension. MCA M1 – M1 segment of the middle cerebral artery; ICA oph – ophthalmic segment of the internal carotid artery; PComA – posterior communicating artery; AComA – anterior communicating artery; ACA A2-Per – pericallosal segment of the anterior cerebral artery; BA tip – tip of the basilar artery; ICA bif – ICA bifurcation. aSz – aneurysm size, AR – aspect ratio, NSI – non-sphericity index, GLN – total Gaussian curvature (normalized)

### 4.3.2 Morphometric quantification of the IA lumen

For the assessment of the morphology, we extracted geometric 3D models of the aneurysms and the surrounding vasculature from the 3D angiographies by applying vessel lumen segmentation (geodesic active regions<sup>134</sup>, implemented in the software package GIMIAS<sup>132</sup> by CISTIB, University of Sheffield).



For the automated radiomic description, we derived quantitative morphometric data for each aneurysm dome.<sup>97,98</sup> Most notably, we computed aneurysm size (aSz), the non-sphericity index (NSI), and the normalized total Gaussian curvature (GLN). aSz is the maximum diameter of the aneurysm dome. NSI captures elongation and surface undulation of the dome. It assumes values between 0 and 1, where NSI = 0 holds for a perfect half-sphere. GLN is a measure for the total Gaussian curvature of the IA dome, normalized by the total curvature of a sphere with equal volume. We employed inhouse software written in Python for these morphological computations, as well as for all subsequent statistical analyses.

### 4.3.3 Morphological assessment of IA lumen by human raters

A total of 39 raters were included in this study: 13 *clinical experts* with an average experience of 12.0 years in researching or treating IAs, 26 *instructed laypersons* with a biological or technical background and at least a general understanding of the disease.

The raters assessed each IA in terms of shape irregularity on a 9-point rating scale, from “1 – very regular” to “9 – very irregular”. The task description emphasized the subjective assessment of shape. No clinical information on the cases was provided. A subset of 26 participants (10 clinical experts, 16 instructed laypersons) assessed the presence/absence of five *morphological attributes* for each IA: rough (non-smooth) surface, blebs, lobules, asymmetric appearance, complex configuration of the parent vasculature/bifurcation, or none of these.

To aggregate the ordinal irregularity ratings per aneurysm, we ranked the rating data per rater (to adjust for rater bias) and computed the means per case, normalized to the range [0, 1]. For the binary ratings of the morphological attributes, we computed the number of votes in favor of that attribute, normalized by the total number of assessments for that case. All rating aggregates  $p'_{ik}$ , for aneurysm  $i$  and attribute  $k$ , take on values between 0 and 1, the limits standing for “very weakly perceived” and “very strongly perceived”,

respectively. We refer to these metrics  $p'_{ik}$  as *perceived* (e.g., perceived irregularity, perceived asymmetry).

We examined the consistency of the ratings as follows. For the *ordinal* irregularity ratings, we calculated the mean Spearman rank correlation between perceived irregularity of the entire cohort and the corresponding ratings computed per rater. For the *binary* attribute ratings, we computed Cohen's kappa between each rater and the individually binarized average rater.

Missing *rating* data was handled by exclusion. Our method is generally robust with respect to missing or outlying rating data.<sup>158</sup> Missing *clinical* data was also handled by exclusion, under the assumption that the misses occurred at random and independent of the property under examination. The numbers of valid cases per property are reported in Table 4.2.

#### **4.3.4 Relationship between irregularity and morphological attributes**

We assessed the relationship of the morphological attributes with perceived irregularity both univariately and multivariately. We applied 5-fold cross-validated, ordinary least squares regression, for which we report the coefficient of determination ( $R^2$ ) and the root mean square error (RMSE) as performance metrics, averaged over 100 randomly shuffled re-instantiations of our dataset of ratings (mean and standard deviation of totally 500 samples). Finally, we expanded the multivariate regression model by quantitative metrics such as aSz, NSI, and GLN to examine if the proposed set of morphological attributes possibly requires extension.

Data		Rupture		Sex		Smoking		Hypertension		IA size (aSz)		Patient age				
		78 unruptured 41 ruptured 119 total	36 male 98 female 134 total	40 non-smokers 74 smokers 114 total	70 non-hypertens 49 hypertensive 119 total	134 aneurysms Mean: 6.9mm Std.: 3.4mm	104 aneurysms Mean: 53.0y Std.: 12.1y									
Characteristic	Sample	RF	Agreem.	AUC	p-val	AUC	p-val	AUC	p-val	AUC	p-val	$\rho_{Sp}$	p-val			
Rough surface	26 raters	0.21	$\kappa$	0.66	0.59	0.12	0.55	0.41	0.51	0.90	0.50	0.97	0.68	***	0.03	0.76
Blebs	26 raters	0.43	$\kappa$	0.63	0.71	**	0.54	0.53	0.56	0.32	0.52	0.70	0.50	***	0.10	0.33
Lobules	26 raters	0.31	$\kappa$	0.65	0.79	***	0.64	0.01	0.52	0.68	0.52	0.69	0.40	**	0.02	0.83
Asymmetry	26 raters	0.47	$\kappa$	0.54	0.81	***	0.59	0.10	0.51	0.85	0.53	0.62	0.49	***	0.09	0.39
Complex vasc.	26 raters	0.17	$\kappa$	0.52	0.51	0.85	0.56	0.28	0.52	0.76	0.57	0.17	0.13	0.28	0.03	0.76
Irregularity	39 raters	-	$\rho_{Sp}$	0.84	0.81	***	0.58	0.14	0.50	0.90	0.51	0.79	0.71	***	0.08	0.42
Aneurysm size	134 IAs	-	-	-	0.71	**	0.62	0.03	0.57	0.18	0.52	0.24	-	-	0.14	0.16
Non-sphericity	134 IAs	-	-	-	0.83	***	0.60	0.07	0.55	0.38	0.50	0.97	0.61	***	0.01	0.93
Curvature	134 IAs	-	-	-	0.78	***	0.58	0.15	0.55	0.34	0.52	0.78	0.81	***	0.09	0.38
Patient age	110 pats	-	-	-	0.58	0.21	0.59	0.18	0.55	0.48	0.74	*	0.14	0.16	-	-

**Table 4.2:** Morphological attributes, aneurysm size (aSz), non-sphericity (NSI), total curvature (GLN), and patient age (green columns) and their relationship to binary or binarized factors (blue columns) and scalar factors (red columns). Relative frequency (RF); number of times a (binary) attribute was identified divided by the number of cases and raters. Rater agreement for binary assessments is given as the mean Cohen's  $\kappa$  between the rater and the individually binarized average rater, and as the mean Spearman correlation coefficient  $\rho_{Sp}$  for ordinal rating data. The association strength of the characteristics with respect to binary clinical factors (blue columns) was measured as the AUC and the  $p$ -values of Mann-Whitney  $U$ -tests for the per-class differences. Spearman coefficient  $\rho_{Sp}$  correlates morphological attribute and the clinical factor (red columns), along with its  $p$ -value. Single asterisks \*, double asterisks \*\* and triple asterisks \*\*\* indicate significance at the  $\alpha = 0.05, 0.01, \text{ and } 0.001$  level, under consideration of the Bonferroni correction. All significant associations are shaded in yellow

### 4.3.5 Relationship between morphology and clinical factors

We examined the aggregated ratings of morphological characteristics for a relationship with the following clinically relevant factors: (i) for the aneurysm: rupture status, size, and location; (ii) for the patient: sex, age, history of smoking (former or current smoker), history of hypertension (patient with treated or untreated hypertension).

For *continuous* variables (aneurysm size and patient age), we report Spearman rank correlation coefficients between these variables and the aggregated ratings. For *binary* parameters (rupture status, sex, smoking status, and history of hypertension), we computed univariate classification models (logistic regression, 5-fold cross-validated, 100 data shuffles). The area under the receiver operating characteristic curve (AUC) served as a primary performance measure. We report the mean and standard deviation over all 500 re-instantiations of the test dataset:  $AUC = \text{mean} \pm \text{std}$ . Additionally, we tested the per-class differences for statistical significance using two-sided Mann-Whitney  $U$ -tests (significance level  $\alpha = 0.05$ ). We applied conservative Bonferroni correction for multiple pairwise testing, for which we set the correction factor to 60 (four binary and two continuous parameters times 10 characteristics to be examined).

To assess *location dependency*, we restricted the analysis to locations with at least 20 samples: the M1 segment of the middle cerebral artery (MCA M1), ophthalmic segment of the internal carotid artery (ICA oph), and aneurysms at the posterior and anterior communicating artery (PComA and AComA). These locations cover 73% of the cases (Table 4.1).

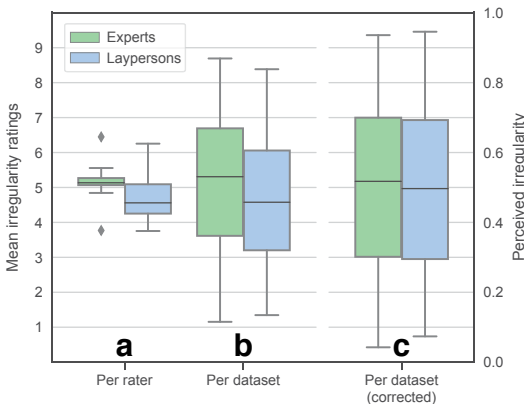
We computed *multivariate* classification models (logistic regression) to examine the relationship between morphology and rupture status, combining perceived irregularity with morphometrics (aSz, NSI, GLN) and location. One-hot encoding was used to represent categorical variables (location) in a metric feature space. We applied the same model validation scheme as for the univariate case.

## 4.4 Results

### 4.4.1 Rating data

We investigated the consistency among the raters and between experts and laypersons for both irregularity and morphological attributes. This ensures that the aggregated metrics represent the collective opinion sufficiently well. The rating data comprised of 5219 ordinal irregularity ratings (39 raters, 134 cases, 7 misses) and 17'420 binary ratings (26 raters, 134 cases, 5 attributes, 0 misses).

*Perceived irregularity*: The individual orderings of the cases by increasing irregularity rating agree well with the collective opinion on irregularity, indicated by a mean Spearman correlation coefficient of  $\rho_{Sp} = 0.84$  ( $p < 0.001$ ) between the absolute ratings and perceived irregularity. On an absolute scale, clinical experts rated the cases significantly higher than instructed laypersons by 0.47 (paired-sample  $t$ -test,  $p < 0.001$ ,  $n = 134$ ). The distributions of ratings were more consistent within clinical raters (indicated by a narrow interquartile range for the mean ratings per raters in Figure 4.1a). However, the rater agreement per case (measured as the standard deviation of ratings) was similar for both clinicians and laypersons (paired-sample  $t$ -test,  $p = 0.89$ ,  $n = 134$ ). Also, the rater subcohorts cannot be discriminated statistically if the bias-adjusted (ranked) metric *perceived irregularity* is used (paired-sample  $t$ -test,  $p = 0.97$ , Figure 4.1c).



**Figure 4.1:** Rating characteristics stratified by cohort. **(a)** and **(b)** visualize the mean absolute ratings per rater and per dataset, respectively. **(c)** shows the same data as **(b)** after correcting for rater bias: mean of ranked ratings, normalized to the range [0,1]

	aSz	NSI	GLN
irregularity	0.70	0.80	0.89
asymmetry	0.47	0.73	0.71
rough surface	0.65	0.22	0.49
blebs	0.51	0.56	0.72
lobules	0.41	0.70	0.65
vasculature	0.13	-0.13	-0.01

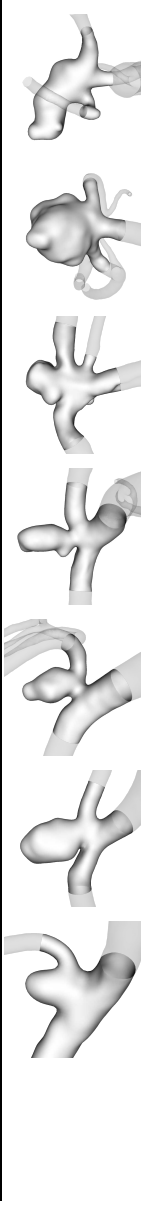
**Table 4.4:** Spearman correlation coefficients  $\rho_{sp}$  between the six morphological characteristics examined in this study and the metrics for aneurysm size (aSz), non-sphericity/elongation (NSI), and total Gaussian curvature (GLN). The color maps values between 0 (red) and 1 (green)

*Perceived morphological attributes:* Based on the mean Cohen’s  $\kappa$  between raters and the binarized average rater, we observe substantial rater agreement for the assessment of rough surface, blebs, and lobules, and moderate-to-substantial agreement for asymmetry and complex vasculature (Table 4.2). Again, no statistically significant difference in the bias-adjusted agreement metric was identified between clinical experts and instructed laypersons. The rating aggregates are illustrated for a selection of aneurysms in Table 4.3.

*Relationship with quantitative metrics:* We measured Spearman correlation coefficients  $\rho_{sp}$  of 0.70, 0.80, and 0.89 between perceived irregularity and the quantitative metrics aSz, NSI, and GLN, respectively (Table 4.4).

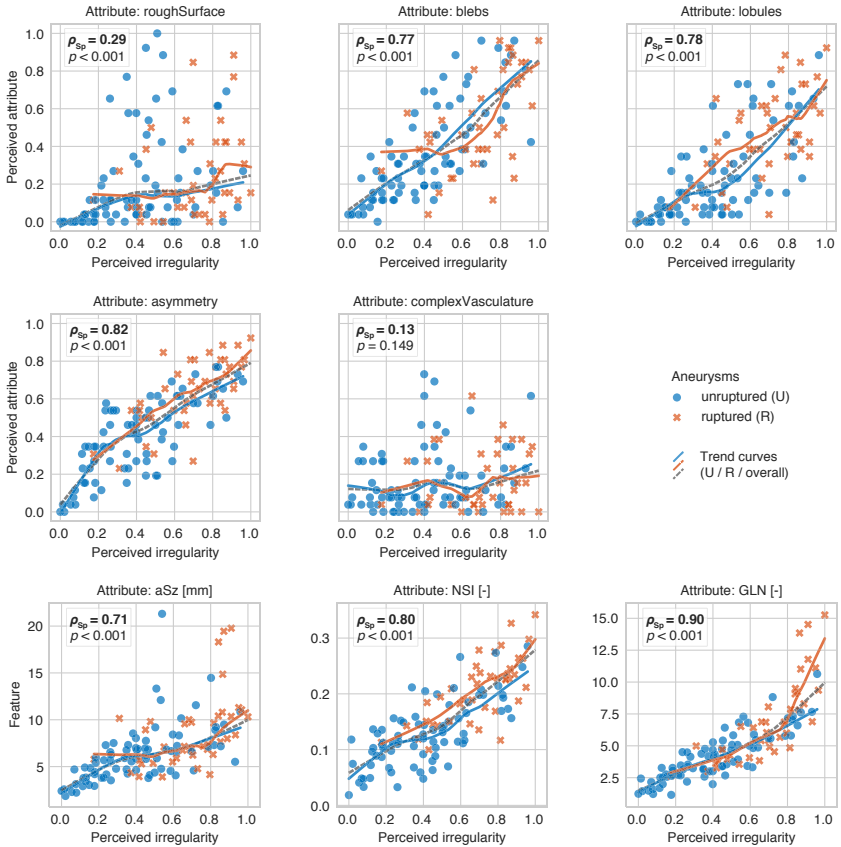
#### 4.4.2 Relationship between morphological attributes and irregularity

The univariate relationships between perceived irregularity and the five morphological attributes are illustrated in Figure 4.2. The multifactorial linear model combining all morphological attributes revealed  $R^2 = 0.92 \pm 0.03$  and  $RMSE = 0.075 \pm 0.011$  for a total of 100 re-evaluations (5-fold cross validation, 20 repetitions). Adding NSI as a factor improved the model accuracy significantly:  $R^2 = 0.95 \pm 0.02$ ,  $RMSE = 0.061 \pm 0.008$  (see Figure 4.3a). Expanding the quantitative metrics for aneurysm size (aSz) and curvature (GLN) did not improve the model further. Examining different classes of morphometrics (geometry features<sup>97</sup>, writhe-based features<sup>102</sup>, Zernike moment invariants<sup>169</sup>), we observed that metrics measuring dome *elongation* (NSI, ellipticity index, aspect ratio<sup>97</sup>) improved the model accuracy the best.



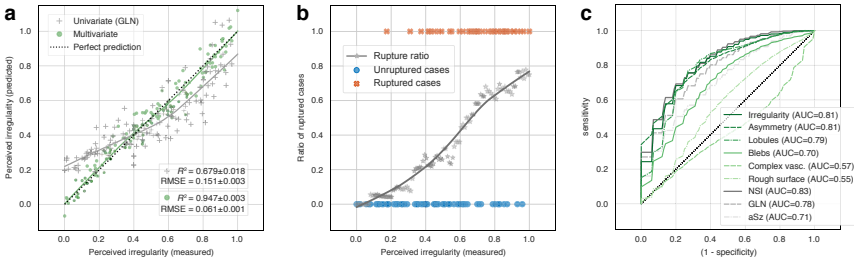
<b>Aneurysm</b>	<b>MCA M1</b> unruptured female, 54y	<b>ICA bif</b> unruptured male, 40y	<b>PICA</b> ruptured female, 42y	<b>ICA bif</b> unruptured female, 24y	<b>BA tip</b> ruptured female, 66y	<b>AComA</b> unruptured female, 62y	<b>AComA</b> ruptured male, 78y
<b>Irregularity</b>	0.08	0.41	0.60	0.79	0.81	0.87	0.93
<b>Rough surf.</b>	0.00	0.32	0.16	0.16	0.24	0.68	0.08
<b>Blebs</b>	0.04	0.33	0.50	0.96	0.83	1.00	0.92
<b>Lobules</b>	0.00	0.08	0.40	0.16	0.44	0.24	0.64
<b>Asymmetry</b>	0.12	0.40	0.56	0.48	0.68	0.52	0.92
<b>Vasculature</b>	0.27	0.04	0.00	0.04	0.46	0.15	0.15
<b>NSI</b>	0.05	0.21	0.18	0.27	0.12	0.16	0.26
<b>aSz [mm]</b>	3.14	6.88	5.50	7.11	6.29	9.49	8.39

**Table 4.3:** Selected IAs and their qualitative and quantitative characteristics. The cases are sorted from left to right by increasing perceived irregularity. AComA – anterior communicating artery; aSz – aneurysm size; BA tip – tip of the basilar artery; ICA bif – ICA bifurcation; MCA M1 – M1 segment of the middle cerebral artery; NSI – non-sphericity index; PICA – posterior inferior cerebellar artery



**Figure 4.2:** Perceived irregularity in relation with morphological characteristics (rough surface, presence of blebs or lobules, asymmetry, or complex vasculature) and selected quantitative features (aSz – aneurysm size, NSI – non-sphericity index, GLN – total Gaussian curvature). The colors/markers encode the rupture status, interpolation curves indicate trends. The Spearman correlation coefficients  $\rho_{sp}$  and the corresponding  $p$ -values are also provided





**Figure 4.3: (a)** Perceived irregularity measured for the 134 aneurysms in relation with the irregularity reproduced by a univariate linear regression model based on curvature index GLN and a multivariate linear regression based on six predictors: asymmetry, rough surface, blebs, lobules, complex vasculature, and non-sphericity (NSI). Interpolation curves indicate trends. **(b)** Relationship between perceived irregularity and rupture. The ratio of ruptured aneurysms is computed in a sliding window of width 0.2 for perceived irregularity (grey markers and trendline). The rupture status of each case is marked (blue and red markers). **(c)** Mean receiver operating characteristic (ROC) curves illustrate the ability of the morphological characteristics considered to discriminate the aneurysm rupture status. A larger area under the curve (AUC) signifies higher discriminative power.  $R^2$  – coefficient of determination; RMSE – root mean square error; aSz – aneurysm size

#### 4.4.3 Relationship between morphology and clinical factors

*Rupture status:* Of the 119 aneurysms with known rupture status, 41 were ruptured and 78 unruptured. The average irregularity rating ranks were significantly higher for ruptured aneurysms (two-sided Mann-Whitney  $U$ -test,  $p < 0.001$ ). This observation held true also if the data was stratified by the two rater groups. To assess the discriminative capability of perceived irregularity, we visualized the ratio of ruptured versus unruptured cases as a function of perceived irregularity for a sliding window (width 0.2, Figure 4.3b). We also computed the mean AUC to be  $0.81 \pm 0.04$ . For the optimal point of the mean receiver operating characteristic (ROC) curve, this translates to a model accuracy of 0.74 (sensitivity: 0.73, specificity: 0.74). Besides overall perceived irregularity, we were able to find strong associations between perceived asymmetry, as well as perceived presence of blebs and lobules, aneurysm size (aSz), and non-sphericity (NSI) (see Figure 4.3b, c and Table 4.2). A

multivariate logistic regression model combining perceived irregularity, aSz and NSI outperformed the univariate model in terms of mean AUC ( $0.82 \pm 0.08$ ), but the difference did not reach statistical significance (unpaired Student's *t*-test,  $p = 0.13$ ). Combining irregularity with (one-hot encoded) location predictors, on the other hand, significantly increased the AUC to  $0.87 \pm 0.09$  (unpaired Student's *t*-test,  $p < 0.001$ ). In this model, we considered only the four locations with at least 20 samples (MCA M1, ICA oph, PComA, and AComA).

*Patient sex, smoking history, history of hypertension, and age:* Neither perceived irregularity nor any of the five morphological attributes considered in this study carried significant information on patient sex, smoking status, or hypertension (Table 4.2). Likewise, the analysis of the (Spearman) correlation between age and morphological characteristics did not reveal any statistically significant association. Male patients develop aneurysms with larger aSz and NSI, and with lobules being perceived more likely, but the observation did not reach statistical significance (Table 4.2).

*Aneurysm location:* Unruptured MCA M1 aneurysms tended to be more asymmetric than unruptured IAs of the other three locations. Likewise, unruptured ICA oph IAs were perceived as considerably more regular, exhibiting fewer blebs/lobules and, possibly related, a lower NSI than the rest of the unruptured cases. Note that in our dataset, the ICA oph and MCA M1 aneurysms exhibited a relatively strong imbalance between ruptured and unruptured IAs (rupture ratio 1:5), while for PComA and AComA IAs, the rupture ratio was nearly balanced.

## 4.5 Discussion

Shape assessments of clinical experts (neurosurgeons, interventional neuroradiologists) and instructed laypersons are statistically indistinguishable in terms of relative ordering of the cases. This suggests that the assessment of IA morphology with our experimental setup is guided by the intuitive, visual perception of geometry rather than prior knowledge about the disease. With regard to the observed difference of average irregularity ratings, we conjecture that clinicians rely on an individual mental model covering a wider spectrum of cases than the dataset included in this study.

Overall, we found good agreement of the raters with the collective aggregates for perceived irregularity and morphological attributes. The observed variability in the rating data is a consequence of the open task formulation, the inconsistency typical for human subjective assessment (perceptual and attentional differences), the heterogeneous composition of the rater cohort, and the number of rating levels. We measured a moderate-to-substantial interrater agreement for morphological attributes, comparable with Suh et al.<sup>162</sup> for the human discrimination of daughter sacs and lobulations.

We modeled perceived irregularity by means of all morphological attributes using multivariate linear regression. By considering the attributes asymmetry, rough surface, blebs, lobules, and complex parent vasculature, the perceived irregularity was explained already very accurately. This suggests that our rater cohort was inherently consistent with its ratings for irregularity and morphological attributes, and that our particular choice of morphological attributes reflects the various manifestations of perceived irregularity reasonably. Extending the regression model by quantitative metrics for size and morphology allowed us to identify elongation/non-sphericity as a sixth characteristic, which we had not considered *a priori* as a rating attribute.

The model predicting perceived irregularity by means of the measurements of the morphological attributes ( $R^2 = 0.95 \pm 0.02$ ) outperforms our model based on quantitative features considerably ( $R^2 = 0.84 \pm 0.05$ ; see Juchler et al.<sup>158</sup>). This suggests that our set of quantitative features was not comprehensive enough to capture the human perception with sufficient accuracy and requires further extension.

Attribute	Coefficient
Intercept	$c_0: -0.09 \pm 0.01$
Asymmetry	$c_1: 0.12 \pm 0.02$
Rough surface	$c_2: 0.28 \pm 0.01$
Blebs	$c_3: 0.34 \pm 0.01$
Lobules	$c_4: 0.38 \pm 0.02$
Complex vasculature	$c_5: 0.28 \pm 0.02$
Non-sphericity/elongation	$c_6: 0.30 \pm 0.02$

**Table 4.5:** Coefficients of the multivariate linear regression model for perceived irregularity including the five morphological attributes plus dome elongation, averaged over 100 repeated model evaluations (mean  $\pm$  std)

The regression model links morphological attributes to perceived irregularity. Its coefficients (Table 4.5) can be interpreted by how much the presence of a morphological trait contributed to the average rater’s perception of irregularity. Note that the different morphological attributes were not equally prevalent in our dataset. Complex parent vasculature or rough surface was identified only half as frequently as the presence of blebs or asymmetry (Table 4.2).

From the clinical risk factors considered, only aneurysm size was linked to morphology, which reflects the well-established fact that pathogenic wall processes stimulate both global growth of the aneurysm and the formation of morphological structure.<sup>41,42,62</sup> We found no association of morphology with patient sex and age, although these factors play a role in the prevalence of aneurysms.<sup>11</sup> The same holds for smoking, which is known to have an adverse effect on the vessel wall integrity and aneurysm formation and growth.<sup>170</sup>

Perceived irregularity spanned a continuum that is strongly linked to rupture status. With higher irregularity, the rupture ratio increased (Figure 4.3b). Irregularity alone discriminated ruptured from unruptured aneurysms relatively accurately (AUC: 0.81, prediction accuracy: 0.73). Combining perceived irregularity with location increased the association significantly (AUC: 0.87, prediction accuracy: 0.78), which confirms the widely accepted view that aneurysm morphology varies with location. The performance of our two-factor model (perceived irregularity, aneurysm location) was comparable with recent multifactorial models for aneurysm rupture status.<sup>75–78,121</sup>

We compared our results with the findings reported by Lindgren et al.<sup>29</sup>, where the shape irregularity of 5814 aneurysms was rated on a binary scale (single assessments only). For this, we binarized our irregularity ratings using the optimal ROC point (optimality based on Youden’s J statistic), resulting in

73% of the ruptured aneurysms and 26% of the unruptured aneurysms being classified as irregular, which is in good agreement with the corresponding rates of 92% and 22% reported by Lindgren et al.<sup>29</sup>. The difference can possibly be explained by the fact that stable aneurysms with a regular shape are underrepresented in our dataset, and that irregular structures are more easily perceived in 3DRA images than in MR or CT angiograms.

Aneurysm shape may change under rupture in some cases.<sup>29,171</sup> To examine whether the rater assessment varies with rupture status, we re-evaluated our study on the subset of unruptured aneurysms alone and were able to reproduce our key observations that raters assess irregularity consistently ( $\rho_{SP} = 0.81$  vs.  $\rho_{SP} = 0.84$  for the complete dataset) and that irregularity is decoded by morphological attributes (AUC for multivariate linear regression model:  $0.94 \pm 0.03$ , compared with  $0.95 \pm 0.02$ ). From this, we conclude that the assessment of morphology is independent of the rupture status. Unruptured IAs still exhibited variations in morphological characteristics, albeit to a lesser extent than ruptured aneurysms. This confirms the prevailing view that morphological structures develop gradually as a result of complex processes within the wall and therefore primarily reflect the stability of the lesion rather than the rupture event itself.

#### **4.5.1 Limitations**

3DRA data tend to be biased towards unstable cases. The inclusion of patients with multiple aneurysms (21 out of 110 patients) mitigates this limitation slightly because secondary aneurysms tend to have simpler geometries. Using MR or CT angiographic data would solve this problem partially but demands higher efforts in image processing and standardization.

To simplify the qualitative and quantitative assessment of the aneurysm geometry, we operated with 3D surface models derived from 3DRA. We do not see any obstacles to applying the suggested rating scheme on 3D medical imaging data directly, without the detour of extracting 3D surfaces.

The collocation of morphological attributes and pathological intramural processes or thrombus formation has not been addressed in this study, but concepts exist how to relate morphological with wall histological data.<sup>112,172</sup>

## 4.6 Conclusions

We showed that perceived irregularity of the lumen can be decomposed consistently into six different morphological attributes that reflect remodeling processes of the aneurysmal wall: presence of a rough surface, blebs, lobules, asymmetry, an elongated dome, and a complex geometry of the parent vasculature.

Morphology carries important information on the disease status of IAs. Perceived irregularity, asymmetry, the presence of lobules, and, to a lower extent, blebs, are more frequent in ruptured aneurysm. Our association model for rupture status based solely on aneurysm morphology and location achieves similar discrimination performance compared with recent studies but requires considerably fewer factors.

With the curvature metric GLN, and, to a slightly lower extent, non-sphericity NSI, being well correlated with perceived irregularity, these indices can be considered first-order quantitative proxies for morphological irregularity.

Irregularity is independent from the clinical factors patient sex, age, history of smoking, and hypertension. In contrast, aneurysm location and size have a significant impact on morphology. In particular the location dependency deserves further attention.

IAs constitute a vessel wall pathology that shows a great variability, which is also reflected in the shape. Structural wall heterogeneity is associated with morphological irregularity. The assessment of morphology from vessel lumen holds the potential for an automated shape analysis that establishes irregularity as a biomarker for vessel wall instability.

## 4.7 Acknowledgments

We would like to thank John Bennett for proofreading the manuscript as well as all study participants (in alphabetical order): Jonas Abeken, Sepideh Amin-Hanjani, Hitomi Anzai, Andrea Boraschi, Stefano Buoso, Marco Corniola, Claudia Danzer, Diane de Zélicourt, Felicitas J. Detmer, Michael J. Durka, Guilherme Fideles, Christian F. Freyschlag, Victor Garcia, Manuel Gehlen, Stefan Glüge, David M. Hasan, Nora Huuska, Kartik Jain, Keisuke Kadooka, Eva L. Leemans, Max Lehtinen, Filippo Molica, Manuel Nüesch, Eliisa Netti, Rahul Raj, Sandro Roth, Isabelle Rudolf, Karl Schaller, Andreas Spiegelberg, Vincent Tutino, Isabel Wanke, Kazuhiro Watanabe, Paul Watton, Stephan Wetzel, Karsten H. Wrede and Erich Zbinden.

## 5 Discussion

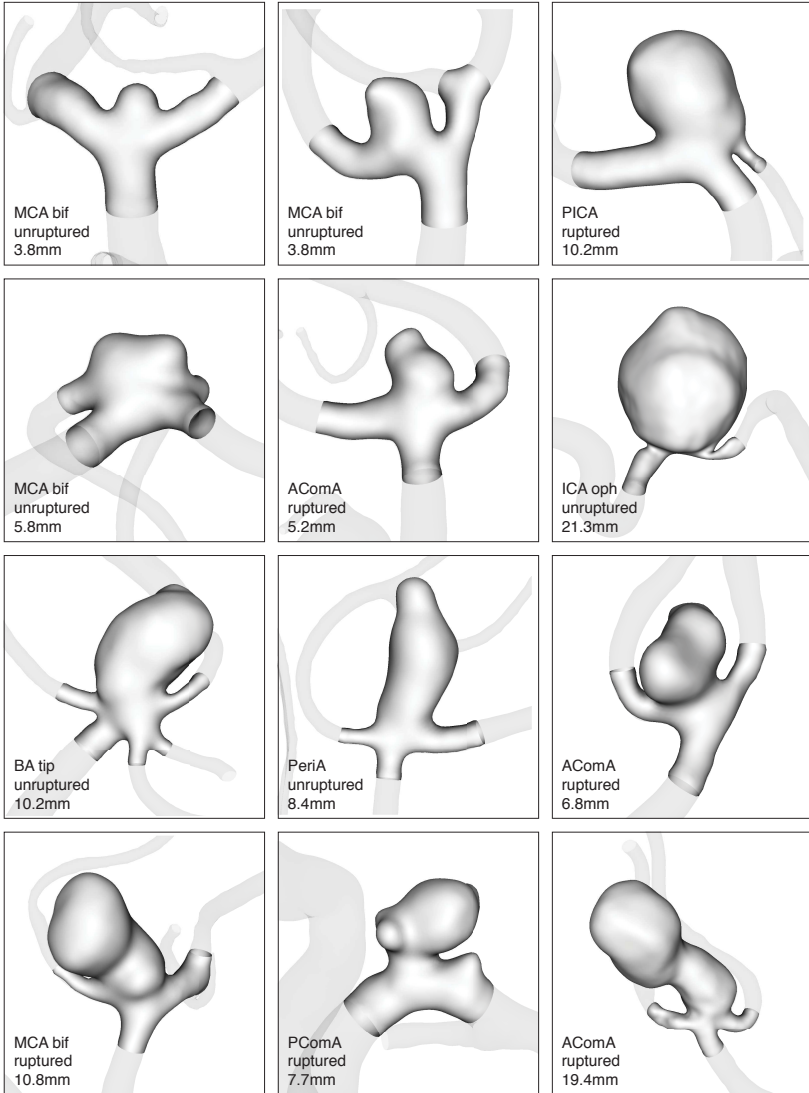
The formation of intracranial aneurysms progresses slowly over months and years without immediate consequences for the patient. Often, the impaired vessel region wall can stabilize, and growth comes to a rest. Temporarily at least, until the equilibrium of regenerative and degenerative processes is disturbed, initiating another episode of wall remodeling. Pulsatile hemodynamic forces acting upon the arterial wall and complex intramural processes degrading the laminar structure of the vessel wall over time affect the course of the disease. Episodes of acute and chronic inflammation, an increasing heterogenization of the vessel wall tissue and the formation of intraluminal thrombosis may further aggravate the structural condition of the aneurysmal wall, until, eventually, the vessel wall fails to bear the hemodynamic load. (Figure 1.4)

The complex interactions leading to the formation of the aneurysm are ultimately reflected by the property that is the focus of this dissertation: the shape of the aneurysm. The possibility to relate the morphological appearance of an IA in medical imaging to its pathobiology would facilitate medical decisions on the diagnosis of an unruptured IA. Shape is of great clinical importance because the pathological status of the aneurysmal wall cannot be examined non-invasively.

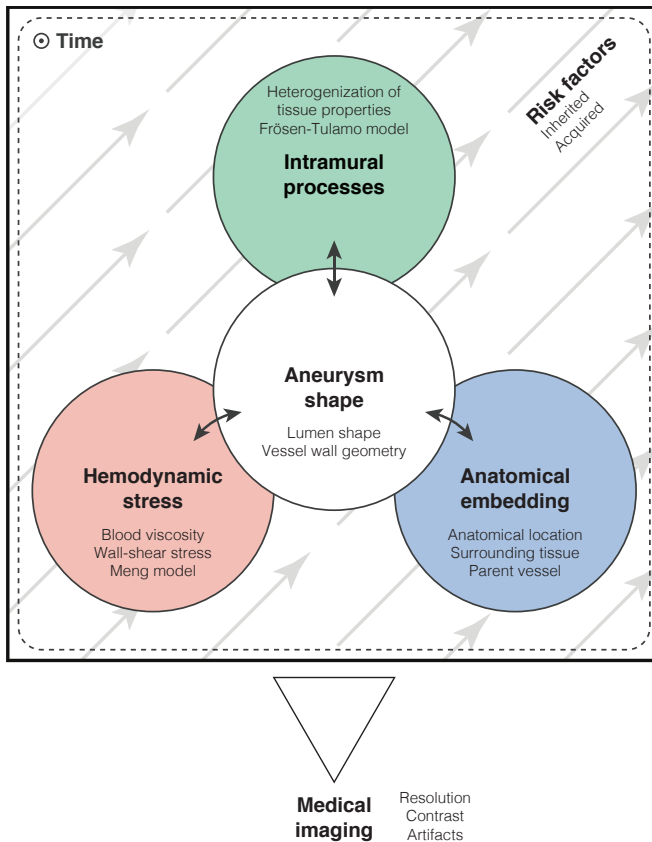
This thesis aimed at answering questions related to shape and its application as risk predictor: To which extent does aneurysm morphology relate to the disease status? How to quantify the morphology of an aneurysm? How specific and sensitive is shape as a risk predictor? How does shape relate to other risk factors?

The work presented followed a data-driven approach. For this purpose, the AneuX morphology database consisting of hundreds of aneurysm geometries was assembled and tools were developed to process and assess the data. A comprehensive study on the quantitative description of shape was performed in Chapter 2. Because shape to date is judged mainly subjectively by clinicians, the author has also explored the relationship between human assessments and quantitative metrics using psychometric methods (Chapter 3, Chapter 4).





**Figure 5.1:** A selection of cases from the HUG2 dataset illustrating the morphological variability of IAs. The cases are sorted by increasing (perceived) irregularity. The cases are labeled with location, rupture status and aneurysm size.



**Figure 5.2:** Factors affecting the aneurysm shape. Hemodynamic stress (Meng model<sup>42)</sup>, the interplay of regenerative and degenerative processes within the vessel wall (Frösen-Tulamo model<sup>41)</sup>) and the anatomical embedding of the aneurysm contribute to the shape of the aneurysm. All processes are continuously affected by internal and external risk factors. The processes, risk factors and thus the aneurysm shape are subject to temporal changes. Medical imaging allows to visualize the shape, but the observation is limited by the resolution, contrast and artifacts.

## 5.1 Morphological variability of aneurysm shape

As illustrated in Figure 5.1, the shapes of IAs vary greatly. But how exactly does the shape of an aneurysm form? Figure 5.2 summarizes the main factors. Organic remodeling of the vessel wall is mainly driven by complex intramural processes.<sup>41</sup> Hemodynamic flow patterns provide important stimuli and codetermine how uniform the aneurysm grows.<sup>42</sup> Morphological irregularity results from a gradually increasing heterogenization of tissue properties.<sup>41</sup> Macroscopic changes in shape, on the other hand, influence the hemodynamic flow patterns and may provide further stimuli for tissue remodeling. The surrounding anatomical structures can contribute to the aneurysm's shape by imposing resistance to growth. The caliber of the parent vessel, the position of the aneurysm relative to the parent vessel and nearby vascular structures (bifurcations, bends, perforators) that impact the flow patterns in the vicinity of the aneurysm affect the observable shape of the aneurysm, too. Furthermore, all processes are dependent on numerous external *risk factors*, which themselves can vary over time. Finally, the cerebral vascular system is subject to pronounced morphological and topological variability,<sup>35,36</sup> contributing to the observed variation of shape.

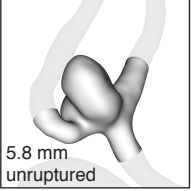
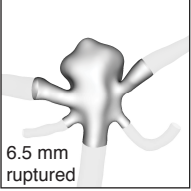
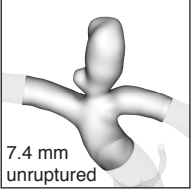
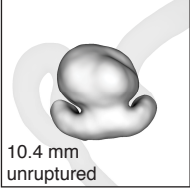
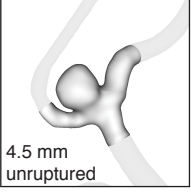
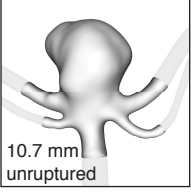
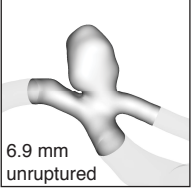
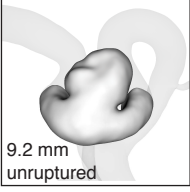
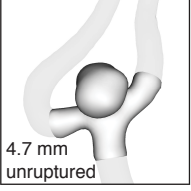
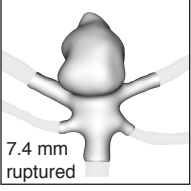
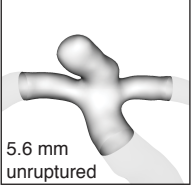
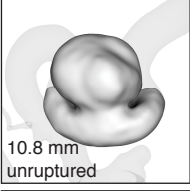
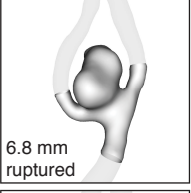
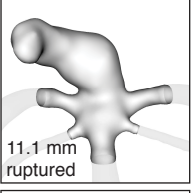
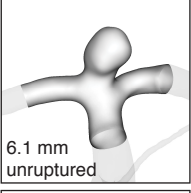
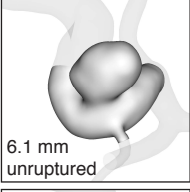
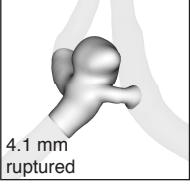
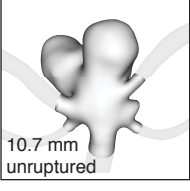
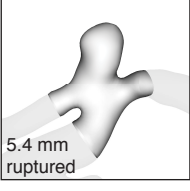
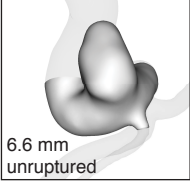
Despite this variability, some morphological patterns occurred repeatedly in our data. This work has focused on *saccular* aneurysms (as opposed to *non-saccular* aneurysms).<sup>6</sup> Accordingly, all aneurysms had a saclike primary compartment in common. In our data, smaller aneurysms were predominantly of spherical shape, though protruding from the parent vessel to varying degrees. An elongated shape or the presence of secondary outpouchings (blebs and lobules) were very common. The aneurysms often grew at an angle (non-perpendicularly) to the parent artery or developed asymmetric characteristics. A beveling of the surface or dents were observed occasionally and may be associated with intraluminal thrombus. Some aneurysm geometries displayed an undulated, "rough" surface. This type of characteristic could be related to an atherosclerotic contamination of the arterial wall (leading to calcifications and irregular wall thickening). Furthermore, aneurysms observed at the same anatomical location often exhibited resembling characteristics (pronounced for BA tip, ICA cav or AComA, Figure 5.3).

These morphological patterns measurably reflect the disease status, as demonstrated in both the quantitative (Chapter 2) and psychometric studies (Chapters 4) presented above.

In Chapter 4, we used these characteristics, assessed qualitatively by human raters, to refine the notion of *morphological irregularity*, a term used in clinical practice as a diffuse characterization of IA morphology.<sup>29,167</sup> In agreement with other sources (e.g., Lindgren et al.<sup>29</sup>), *perceived irregularity* was strongly linked to aneurysm rupture status (Section 4.4.2). Moreover, we have demonstrated that the relationship between irregularity and rupture status was continuous: increasing irregularity is associated with an increased rate of ruptured aneurysms (Figure 4.3b). This observation fits well the model of gradual wall degradation described above. And it underscores the value of morphology as a risk indicator.

In the univariate analysis of our quantitative study (Chapter 2), in which a wide range of different morphometric parameters was examined for an association with the aneurysm rupture status, candidates from all morphometric categories (Figure 2.4) were linked to the aneurysm rupture status (Table 2.4), with the non-sphericity index (*NSI*) and normalized Zernike energies ( $Z_N^{\text{surf}}$ ) performing best ( $\text{AUC} = 0.80 \pm 0.05$ ). The combination of multiple metrics revealed an even stronger association with rupture status (Table 2.5) Since the morphometrics can be attributed to different qualitative characteristics such as asymmetry, undulation or the presence of blebs/lobules (Appendix 2 and Chapter 3), the observations of the morphometry-only study of Chapter 2 were consistent with the findings of the psychometric study (Chapter 4).

The aneurysm size is associated with morphological irregularity but is a considerably weaker predictor than morphometric parameters (Chapter 2: aSz:  $\text{AUC} = 0.64$  vs. NSI:  $\text{AUC} = 0.80$ , Chapter 4: aSz = 0.71 vs irregularity:  $\text{AUC} = 0.81$ ). This is of relevance as the aneurysm size is used frequently in clinical guidelines for treatment (e.g., PHASES<sup>28</sup> or UIATS<sup>59</sup>, Chapter 1), while there are no equivalent objective criteria for assessing the shape of the aneurysm.

AComA	BA tip	ICA bif	ICA cav
 5.8 mm unruptured	 6.5 mm ruptured	 7.4 mm unruptured	 10.4 mm unruptured
 4.5 mm unruptured	 10.7 mm unruptured	 6.9 mm unruptured	 9.2 mm unruptured
 4.7 mm unruptured	 7.4 mm ruptured	 5.6 mm unruptured	 10.8 mm unruptured
 6.8 mm ruptured	 11.1 mm ruptured	 6.1 mm unruptured	 6.1 mm unruptured
 4.1 mm ruptured	 10.7 mm unruptured	 5.4 mm ruptured	 6.6 mm unruptured

**Figure 5.3:** Selected aneurysms with similar morphological characteristics at four different anatomical locations. The rupture status and the aneurysm size (largest diameter) are indicated. The aneurysms tend to resemble each other, though the morphological variability increases with the size of the aneurysm. Abbreviations: AComA – anterior communicating artery; BA tip – tip of basilar artery; ICA – internal carotid artery; ICA bif – terminal bifurcation of ICA; ICA cav – cavernous segment of ICA.

## 5.2 The value and practicality of morphometric indices

The study presented in Chapter 2, among other objectives, aimed to answer the question of which morphometric features encode the disease status most efficiently. For this purpose, a relatively broad range of state-of-the-art methods was considered (geometry indices, writhe- and curvature-based metrics, Zernike moment invariants) and extended (Zernike energies, modified writhe-based metrics).

### 5.2.1 Non-sphericity

Throughout all experiments (Chapters 2-4), the non-sphericity index *NSI* was among the best performing morphometric candidates. It was one of the best univariate predictors for aneurysm rupture (Table 2.4) and related well to perceived irregularity (Table 3.6). The index has several advantageous properties that are of practical importance: It is easy to compute; as a measure for elongation and undulation, it has a very intuitive geometric interpretation; *NSI* naturally normalizes to values between 0 and 1; it is already well established as a morphometric; it is not very sensitive to the cut used to separate the aneurysm dome from the parent vessel; and its definition not only applies to surface meshes but also holds for aneurysm domes represented as binarized voxel image.<sup>149</sup>

Non-sphericity as defined by Raghavan et al.<sup>97</sup> normalizes the volume-to-area ratio  $r = V^{2/3}/S$  of a surface by  $r_{\text{ref}}$  of a half-sphere (see also Appendix 2.A.2). This choice was motivated by the observation that a perfectly regular aneurysm resembles that of a half-sphere. In a future study, one could elaborate with alternative definitions, where the volume-to-area ratio  $r$  of an object is normalized by  $r_{\text{ref}}$  of its convex hull or another reference geometry (e.g. the average of a size- and location-matched subset of cases in a database of aneurysms).

### 5.2.2 Normalized Zernike energies $Z_N^{\text{surf}}$

The  $Z_N^{\text{surf}}$  are computed as the squared sum of ZMI up to maximum order  $N$ , normalized by the object’s fill-ratio after transformation into the unit sphere (Appendix 2.C). Even though the mathematical properties of  $Z_N^{\text{surf}}$  have not yet been fully described, these novel morphological indices have performed well in our experiments.  $Z_N^{\text{surf}}$  with  $N = 6$  was among the best two univariate predictors for rupture status and was correlated with perceived irregularity (Table 3.6), though to a lower degree than  $NSI$ .

The main advantage of the ZMI *energies* over the ZMI (as proposed by Millán et al.<sup>104</sup> for use as aneurysm shape descriptors) is that they represent a *single* parameter rather than a high-dimensional shape descriptor. (Recall: for  $N = 10$  or  $20$  the number of ZMIs is 36 and 121, respectively, see Appendix 2.C). Furthermore, in our experiments, it was sufficient to compute  $Z_N^{\text{surf}}$  only up to  $N = 10$ , which is beneficial, since the computational costs for computing the ZMI increase very quickly with  $\mathcal{O}(N^6)$ . The fact that only low-order ZMI (representing the low-frequency modes of the aneurysm shape) were required to compute the  $Z_N^{\text{surf}}$  makes them applicable also for binary volumetric images and modalities with lower resolution and contrast than 3DRA.

ZMI-based metrics are mathematically involved, therefore non-intuitive and less established than other metrics. Normalized Zernike energies simplify the handling of ZMI metrics substantially and generally performed well for predicting the disease status. However, further studies are needed to better understand the mathematical and practical properties of this novel metric.

### 5.2.3 Other metrics

Our quantitative study identified more metrics that exhibited a similarly strong relationship with the aneurysm’s disease status: The index  $\overline{W}_{mean}^{L_1}$  based on a modified definition of the writhe-number, the ellipticity index  $EI$  (which was strongly correlated to  $NSI$  in our data) and the undulation index  $UI$ .

The curvature metrics  $GLN$  and  $MLN$  were also good indicators for disease/rupture status. Furthermore, they were the best indicator for perceived irregularity (Table 3.6). However, curvature metrics have less favorable properties compared to geometry indices or ZMI-based predictors since they depend strongly on the preprocessing and the quality of the surface meshes.

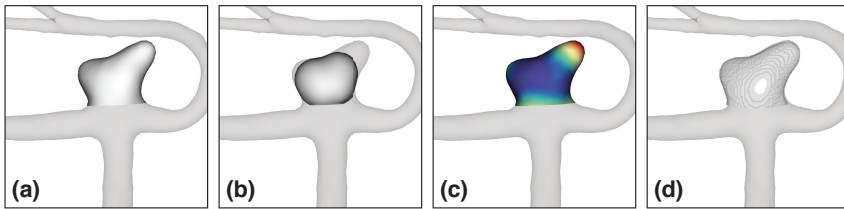
Metrics describing only the dome geometry performed the best in our experiments based on the AneuX morphology database. Other metrics that included parts of the parent vasculature (*cut1*, *cut2* configuration, available for writhe-and curvature-based metrics and ZMI) were linked less strongly with IA rupture status. The only exception were the normalized Zernike energies ( $Z_N^{\text{surf}}$ ), which maintained their predictive accuracy to a good extent for cut types other than *dome* and *ninja* (Figure 2.3).

Note that metrics describing the aneurysm in relation to the parent vessel geometry were underrepresented in this work. Most notably the size ratio (aneurysm size / parent vessel radius) but also the inclination angle have been associated with rupture in earlier studies,<sup>75,99,124</sup> but were lacking in our analyses.

#### 5.2.4 Morphometrics and qualitative characteristics

As mentioned already earlier (Sections 2.5.2 and 3.5), our list of features might not have encompassed all morphological characteristics to completely capture the variability present in the data. Also, the metrics were possibly not perfectly sensitive to the characteristics of interest. For instance, *NSI* is able to measure both dome elongation and generic undulation. But it is neither very sensitive nor specific to the presence of blebs/lobules (illustrated in Table 4.3). Likewise, the indices based on the writhe number did not relate well to the intuitively perceived notion of asymmetry (Chapter 3), even if the writhe number mathematically can be regarded as a measure for asymmetry.<sup>102</sup> The author believes that the morphological assessment of aneurysms would benefit from metrics that mimic perceived characteristics of the shape, such as blebs or lobules (see Table 3.2 for a distinction). These characteristics were very indicative for perceived irregularity (Table 4.5, Figure 4.2). Lobules are sometimes referred to as daughter sacs or secondary aneurysms, while blebs can be seen as “small” lobules. Blebs and lobules represent two partitions of the entire “spectrum” of focal surface protrusions. Figure 5.4 outlines the nature of a possible *blebbiness* metric. For this metric, a reference surface is computed (for instance the Voronoi diagram core<sup>118</sup>, a low-order ZM reconstruction<sup>120</sup> or the convex hull of the aneurysm dome). A distance transform from the input surface to the reference surface reveals the pointwise distances (Figure 5.4c), which can be used to isolate blebs or lobules.





**Figure 5.4:** Illustration of a possible blebbiness metric. Based on the input geometry **(a)**, a reference geometry is computed representing the core of the aneurysm **(b)**. A distance transform could reveal the “blebbiness structure” of an aneurysm **(c)**. Various methods are available to extract a representative aneurysm core and to identify/quantify a bleb. Depending on the choice, such a blebbiness metric could in principle also be applied to volumetric images directly **(d)**.

## 5.2.5 Hemodynamic indices

Because hemodynamic stress plays an important role in the formation and growth of intracranial aneurysms, indices describing blood flow patterns and the total wall shear stress (or wall shear stress variations) in and around the IA could serve as risk indicators, much like morphological indices. For instance, mean and maximum wall shear stress as well as oscillatory shear index have been associated with aneurysm rupture and growth.

CFD-based methods and morphological approaches have in common that a geometric 3D representation of the aneurysm must first be extracted from angiographic imaging data. Therefore, a similar protocol described in Section 2.3.2 applies for both approaches. While the region of interest for the morphometric description is typically limited to the aneurysm dome or its immediate vicinity, hemodynamic models require considerably larger vascular segments, which results in additional manual extraction work or increases the technical requirements for an automated preprocessing scheme (Section 2.3.2).

Furthermore, hemodynamic models are based on modeling assumptions and parameters (most notably the boundary conditions and blood viscosity), which cannot be validated easily on a per-patient basis. This constitutes an additional source of uncertainty. Studies comparing different hemodynamic models have identified a relatively large disparity within simulated results.<sup>173,174</sup> However, it was pointed out that absolute differences in the flow field predictions, caused by different model assumptions, not necessarily

compromise the ability to discriminate pathological from physiological flow-conditions by means of characteristic indices.<sup>173</sup>

Hemodynamics and the aneurysm shape stand in a reciprocal relationship to each other. The vessel geometry confines the blood flow and is therefore a determinant of flow patterns, while conversely, flow-induced stresses through biomechanical processes eventually results in morphological alterations. Hemodynamic and morphological parameters are thus believed to provide redundant information regarding the aneurysm disease status, a view supported by a quantitative rank-based analysis of Retarekar et al.<sup>175</sup> A comparative study by Xiang et al.<sup>75</sup> reported 4 morphological and 5 hemodynamic parameters associated with aneurysm rupture, with the size ratio (AUC = 0.83) and oscillatory shear index (AUC = 0.81) being the best univariate morphological and hemodynamic predictors for rupture, respectively. Another recent study by Ashkezari et al.<sup>176</sup> observed that the presence of blebs was an indicator for unstable flow patterns and concentrated inflow jets. The logistic regression model for aneurysm rupture by Detmer et al.<sup>78</sup> made use of hemodynamic, morphological and clinical predictors. They employed a training procedure with feature selection in which morphological and hemodynamic parameters were selected with comparable frequency. In summary, recent studies suggest that morphological predictors can be seen as proxies for pathological flow-conditions.

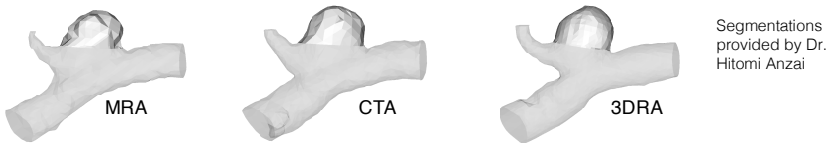
Regardless of these observations, hemodynamic metrics still introduce a physical dimension that is complementary to morphology. The study of blood-flow patterns, alongside with biomechanical models that also involve the dynamical properties of the vessel wall (for example Teixeira et al.<sup>92</sup> or Aparício et al.<sup>93</sup> for two recent examples), deserve their merits by pushing forward the mechanistic understanding about the disease, which a purely descriptive approach (using morphology) would not have been capable of.

### 5.2.6 Robustness and cross-modal applicability

As previously mentioned, not all morphometrics are equally robust with regard to imaging insufficiencies such as low resolution, low contrast, low signal-to-noise ratio, or the presence of imaging artifacts. The aspect of robustness is particularly relevant with regard to the applicability of morphometric methods to MR- and CT-angiography. In a recent rater-based study by Kwak et al.<sup>177</sup> based on 652 small unruptured IAs (< 7mm) examining the shape assessment of radiologists based on MRA and 2D/3D digital subtraction angiography (DSA), it was found that the sensitivity of the raters to identify irregular IAs was low for MRAs (especially for very small IAs smaller than 3mm).

To gauge this effect for the quantitative analysis of shape, we investigated for a subset of our metrics how strongly these metrics differ for the same aneurysms seen in different imaging modalities. Figure 5.5 illustrates the results for one example. The IA geometry can vary considerably, notably in the neck region, but also smaller characteristics such as blebs were less discernible for CTA and MRA, mostly due to their lower resolution. Curvature-based metrics were particularly sensitive.

The focus of this dissertation was the assessment of the information content of morphology in general. To this end, we relied entirely on 3DRA data because resolution and contrast are highest for this imaging modality.



**Figures:** Reconstructions of the same aneurysm from three different imaging modalities.  
**Table:** A selection of Geometry Indices (GI) and their relative change to the 3DRA model.

Model	V [mm <sup>3</sup> ]		Dn [mm]		aSz [mm]		NSI [-]		UI [-]		AR [-]		Color code
3DRA	18.95	0.0%	3.42	0.0%	4.11	0.0%	0.06	0.0%	0.02	0.0%	0.80	0.0%	error  < 10%
CTA	18.80	0.8%	3.56	4.2%	4.70	14.5%	0.07	9.4%	0.04	48.4%	0.71	11.6%	error  < 20%
MRA	16.11	15.0%	3.28	4.1%	4.50	9.7%	0.09	46.8%	0.07	181.9%	0.75	5.5%	error  > 20%

**Figure 5.5:** Morphometric comparison of an aneurysm seen in different imaging modalities. The geometric models derived from 3DRA images served as baseline. Note that previous smoothing of the surface extracted from the MRA would reduce the corresponding error for NSI and UI significantly.

### 5.3 Human perception of aneurysm shape

Today, clinicians usually assess the shape of IAs on the basis of qualitative criteria and subjective judgement. To bridge the gap between the current clinical practice and the purely quantitative assessment of aneurysm shape (Chapter 2), the author conceived a psychometric approach to measure within a group of raters how the aneurysm shape was perceived. To this end, the raters had to visually inspect and morphologically assess 134 aneurysm geometries shown in a 3D viewer in randomized order. Aggregated assessments of the global appearance of the aneurysm (perceived irregularity) and five particular morphological attributes (surface roughness, blebs, lobules, asymmetry, complex parent vasculature) were measured and compared to the quantitative, morphometric description of the aneurysms. While in Chapter 3, the focus was on identifying quantitative metrics that best relate to perceived irregularity, Chapter 4 aimed to explore how perceived irregularity (a continuous, cohort-normalized metric) is linked to the aneurysm disease status and clinical risk factors. Furthermore, it was investigated, which morphological attributes constitute “shape irregularity”.

Previous studies on the qualitative assessment of aneurysm morphology use definitions of particular morphological structures that remain often vague and address the intuitive perception of shape by a human. For instance, in the work by Lindgren et al.<sup>29</sup> that was based on morphological assessments of radiologists, an aneurysm was considered *irregular* if it presented blebs or multiple lobes, without further characterizing those secondary structures. So far, there is no standardized taxonomy for morphological structures. For example, blebs or lobules are also referred to as irregular protrusion, lobulation, lobes, daughter sacs, bubbles, loculi or secondary outpouchings.<sup>1,12,29,59,160</sup> The lack of unequivocal definitions or dedicated metrics for morphological structures complicates the comparative analysis in multi-cohort studies. With the presented studies, the author provided new means to quantify these characteristics of an aneurysms, which are normally assessed qualitatively.

The visual perception of shape by the brain and the underlying neurological processes are not well established. Several studies have identified the brain regions in the visual and inferotemporal cortices that are involved in the perception of shape (along the so-called *ventral stream*).<sup>178–180</sup> Several

perceptual models have been proposed on how the human brain processes and structures visual stimuli to obtain information about the spatial shape of an object. These concepts involve the edges or occlusion contours of the object; curvature extrema; illumination effects such as shading and specular reflections; 3D shape inference from motion, binocular disparity or optical texture; perceptual constancy principles such as symmetry, planarity or occlusion characteristics; and the veridical perception of shape (that is the direct perception of the shape as an object's intrinsic physical attribute, independent of the view point of the observer or illumination characteristics).<sup>179,181-184</sup> The reconstruction of 3D information from 2D retinal projections is inherently ambiguous. However, by combining multiple cues, the human brain is able to constrain these ambiguities such that the relevant aspects of 3D shape still can be recovered uniquely.<sup>179</sup> Moreover, according to Gestalt psychology and more recent psychophysical studies, the human brain is efficient at detecting deviations from regular patterns or differences between similar objects.<sup>183,184</sup> Those characteristics of human shape perception were highly relevant when the raters had to assess morphological attributes such as the irregularity or asymmetry of an aneurysm relative to other samples in the dataset.

In our setup, the primary cues for the perception of the spatial object originated from motion (interactive rotation) of the geometry in the 3D viewer and the shades drawn onto the surface by ambient light. Optically deceptive effects were avoided in this setup. The performance of the human brain in anomaly/similarity detection is affected by factors such as the orientation of the principal stimulus or illumination properties.<sup>183</sup> The smooth aneurysm geometries thus were all shown in an "upright" orientation with a fixed ambient illumination (compare with Figure 5.1). The aneurysms were scaled to fit best in the field of view. Accordingly, the size of the aneurysms could not be inferred immediately by looking at the geometries. The cases were presented in randomized order to compensate for recency bias (raters are likely to base their evaluation of new cases on their evaluation of previous cases). All this ensured reasonable conditions to examine the morphological characteristics, well aligned with current practices in cognitive sciences for psychophysical measurements of perceived shape.<sup>179,181</sup>

In chapter 4, we have shown that the ratings of perceived irregularity from laypersons and clinicians cannot be distinguished statistically if the ratings were

ranked per participant. This suggests that the (relative) assessment of irregularity is led by the naturally acquired, intuitive perception of shape, and not by the specific training of the participants.

The absolute irregularity ratings, however, did differ significantly (experts:  $\bar{r}_{exp} = 5.16$ , laypersons  $\bar{r}_{lay} = 4.69$ , on a scale ranging from 1 to 9). This difference between laypersons and clinicians probably reflects the bias towards the more severe cases in our dataset and the experience in working with aneurysms. While laypersons based their assessment on just the 134 cases included in this study, some clinicians might have baselined their assessment with the cohort they experience in everyday practice. While being speculative on the cause for this absolute difference, we showed with our rank-based analysis that the study outcome was not affected by it, since the *relative ordering* of the cases was very similar across both cohorts.

Even though the agreement on perceived irregularity was good among raters on average, the agreement varied considerably for different cases (Figure 3.3). The standard deviation of (normalized) irregularity rating ranks  $\sigma'_i$  for case  $i$  (see Section 3.4.1), a measure for rater agreement, was ranging from 0.050 (good agreement) to 0.261 (low agreement), with a mean of 0.152. Figure 3.3 further reveals that the agreement varies systematically. For very regular and very irregular shapes, the spread of rating ranks was small, whereas aneurysms of medium irregularity exhibited the largest spread. Hence, raters were able to “anchor” the extreme cases on the scale while they struggled to rank the intermediate cases in between. This demonstrates a limitation of human assessment (lack of coherence) and encourages the use of quantitative metrics for irregularity.

In summary, the rater-based approach enables a refined qualitative assessment of morphology by clinicians, for instance by means of a scoring scheme for morphology. Or it may help to identify new metrics able to quantify more specifically perceived irregularity or morphological attributes such as blebs, lobules or asymmetry.

## 5.4 IA morphology and anatomical location

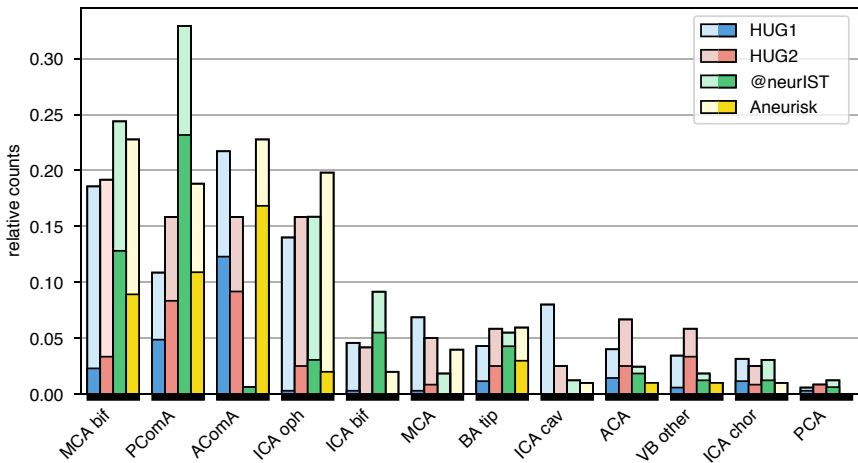
In Chapters 2 and 4, it was demonstrated that the combination of morphological data with information about the anatomical location of the aneurysm substantially increases the predictive accuracy of these parameters for the disease status (rupture status). On the one hand, the typical morphological characteristics of an aneurysm vary with the location (Figure 5.3). The use of a categorical variable for aneurysm location permits the classifier to compensate for location-specific differences. On the other hand, the risk of rupture deviates strongly for different locations.<sup>1,28,33</sup>

Figure 5.6 illustrates for each of the cohorts in the AneuX morphology database (HUG, @neurIST, Aneurisk) the proportion of ruptured and unruptured aneurysms at different anatomical locations. Three sources of variation can be observed. First, the relative frequencies of aneurysms at different locations varies considerably, with AComA, PComA, MCA bif and ICA oph being the most frequent locations. Second, the rate of rupture aneurysms depends strongly on the anatomical site at which the aneurysm was observed. Aneurysms at ICA oph, ICA cav are relatively unlikely to rupture, whereas aneurysms at the AComA, PComA or the posterior circulation are much more likely to rupture. Third, the relative number of ruptured and unruptured cases may vary between different datasets, which can be caused by small numbers of cases observed at certain locations, or because of different selection criteria/biases in data collection (see next section).

In Chapter 2, it was shown that already the aneurysm location carries information about the probable rupture status of the observed aneurysm. Multivariate predictive models that use the aneurysm location as predictor variable should therefore be baselined on a model that considers only location as predictor.

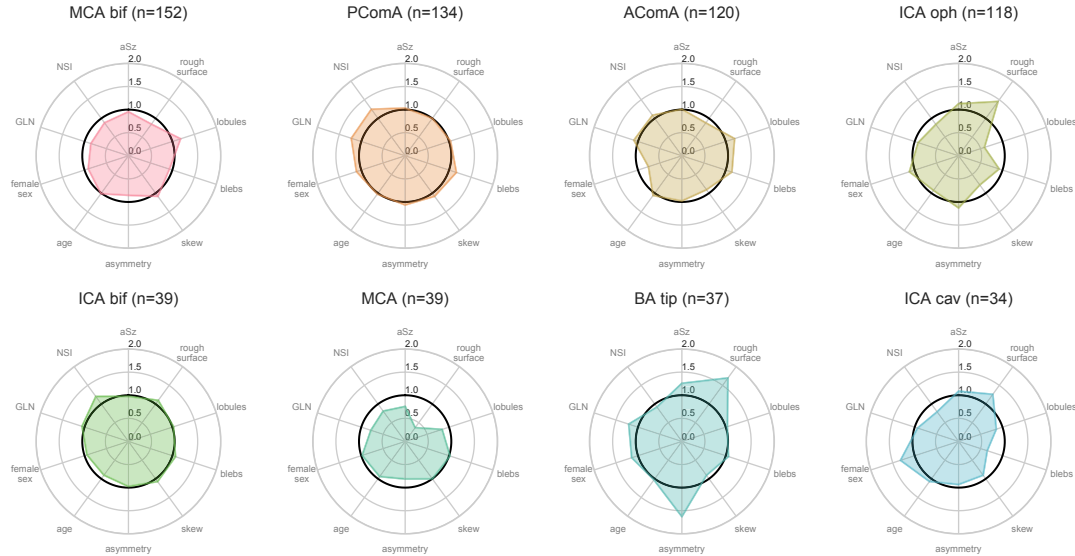
Figure 5.8 depicts the average values of A) morphological metrics (size  $aSz$ , non-sphericity  $NSI$ , curvature  $GLN$ , as well as the rate of morphological attributes (blebs, lobules, asymmetry, skew) as assessed by an operator) and B) clinical characteristics (patient sex, age). All these characteristics are evaluated and visualized for each of the different location using polar graphs. For better readability, the values are normalized by the mean values of the complete dataset, including all anatomical locations. If a characteristic is

underrepresented or smaller than in the complete collection of aneurysms, the measurement falls inside the unit circle, otherwise is larger. By means of those “location-profiles” it can be seen that aneurysms at the PComA tend to be more elongated than aneurysms at other locations, that aneurysms at the BA tip tended to be more asymmetric, and AComA aneurysms were slightly less prevalent in females than in males, and so forth. Figure 5.9 repeats the same graphical method, but further stratifies the data by the rupture status.

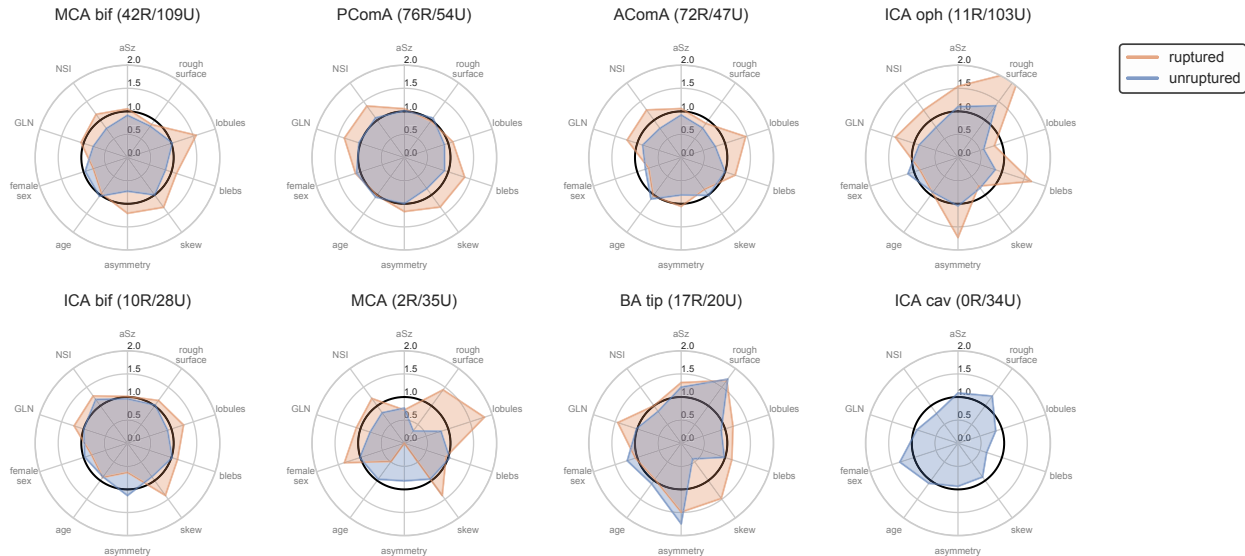


**Figure 5.6:** Relative frequencies of aneurysms per anatomical location and dataset. The bars for each dataset add up to 1. The ratio of ruptured (bright patches) and unruptured (light patches) cases is indicated by different shadings.





**Figure 5.7:** Location profiles for the 8 most prevalent aneurysm locations (compare with Figure 5.6). The axes of the spider charts indicate characteristics of the aneurysms at a certain location normalized by the overall mean of the complete AneuX morphology cohort. A value larger/smaller than one indicates that the particular characteristic is over-/under-expressed at this location. The data axes are: aneurysm size ( $aSz$ ), non-sphericity ( $NSI$ ), area-normalized total Gaussian curvature ( $GLN$ ), the ratio of female patients as well as patient age, and the prevalence of five morphological attributes that have been assessed (binary yes/no) by a single rater for all 750 cases of AneuX morphology database: surface roughness, presence of blebs/lobules, a skewed angle with respect to the parent vessel and asymmetry. See Chapter Abbreviations and Figure 1.2 for the meaning of the location acronyms.



**Figure 5.8:** Same data as in Figure 5.7 stratified further by aneurysm rupture status. Not only the average morphology varies per aneurysm location, but also the differences between ruptured and unruptured aneurysms. The number of ruptured (R) and unruptured (U) cases is indicated in the subtitles of the plots. For some locations, only very few ruptured cases were available.

## 5.5 Methodological considerations

The amount of health data being collected is increasing at a breathtaking pace.<sup>185</sup> The rapid digitization in healthcare bears the potential of improving clinical outcomes while at the same time reducing costs. The emerging wealth of clinical data also offers new means to expand our understanding about pathologies by screening the data for risk factors and pathogenic markers.

In this context, it is not surprising that *machine learning* (ML) has become very popular in medical sciences. ML offers tools capable of robustly and efficiently learning underlying patterns in “dirty” real-world data for the prediction of new events. The field of ML has evolved rapidly in recent years, and various tools and software platforms have emerged, which certainly have contributed to the popularity of ML also in the medical community. Compared to other methodological frameworks, ML methods make little assumptions about the data-generating processes, which is particularly helpful in medical data.<sup>186</sup> The central objective in ML is to create accurate predictions based on previous observations.

Traditionally, medical data science has been relying on *statistics* as the primary methodological framework to study clinical data. The focus in statistics is primarily on inference and testing: Inference refers to the probabilistic modelling of the data-generating processes, and testing to the assessment of how new observations relate to these processes.<sup>186,187</sup>

There is no need to emphasize the differences between these two fields. ML has substantial overlap with statistics. Both share a common theoretical framework (statistical learning theory<sup>74</sup>), and various tools such as logistic regression, bootstrapping or cross-validation are used in both “worlds”. Some proponents even argue that the two fields cannot be disentangled.<sup>186,187</sup> While study design patterns (see below) are usually attributed to the field of statistics, the proper planning and execution of a data collection process is equally relevant for both statistical and ML analyses. After all, the same noble dictum holds for both disciplines: “*garbage in – garbage out*”. If the available data is flawed, findings based on these data will likely be inaccurate or wrong.

Accordingly, data-driven models heavily depend on the data used for their development. In this respect, study design plays a central role and can affect the interpretation of a study fundamentally.

In a *prospective* cohort study, patients are selected based on previously specified criteria before the outcome (e.g., rupture/no rupture, or growth/no growth) is known and then are monitored over time. In a *retrospective* study, the outcome is known at the time patients are selected. A common goal for both study types is to identify risk factors and measure their effects on the outcome. Studies based on historical cohorts are sometimes also referred to as prospective, provided that any follow-up moves forward in time.<sup>188</sup> As pointed out by Euser et al.<sup>189</sup>, both study designs can have advantages and disadvantages:

*The major strength of a prospective cohort study is the accuracy of data collection with regard to exposures, confounders, and endpoints, but this is realized at the cost of an inevitable loss of efficiency, for this design is both expensive and time-consuming because of a usually long follow-up period. Vice versa, the retrospective design is a very time-efficient and elegant way of answering new questions with existing data, but one has no choice other than to work with what has been measured in the past, often for another purpose (e.g. patient care) than the one under investigation. (Euser et al.<sup>189</sup>, 2009)*

Datasets consisting of 3DRA images of IAs are often *cross-sectional*, where the angiography was acquired at a specific point in time: admission to hospital for treatment of a ruptured or unruptured aneurysm. In contrast, *longitudinal* (or follow-up) studies monitor patients over time. A *natural history* study is a special instance of longitudinal studies where patients at risk for developing a disease are followed up without clinical intervention. A patient cohort is *consecutive* if all eligible patients are admitted to the cohort who visit a clinical facility within the study period. Non-consecutive patient cohorts may suffer from additional selection bias. *Randomized* trials would offer the best grounds for scientific analyses, but are deemed infeasible in the context of clinical assessment of aneurysms for practical or ethical reasons.<sup>190</sup>

The implementation of a study design and the avoidance of selection biased is proving difficult because it often competes with the primary goal of healthcare providers, which is to ensure the best possible treatment for the patients. Furthermore, it involves the education of the clinical staff, adjustment of documentation procedures and the construction of a data infrastructure. Recommendations for a feasible study design in the context of research on IAs have been compiled by Korja and Kaprio.<sup>150</sup>

The majority of the AneuX morphology datasets (HUG + @neurIST) were collected prospectively, though with different selection criteria. For example, aneurysms at the AComA were excluded in the @neurIST dataset for reasons related to the CFD study for which the dataset was created. The Aneurisk data was collected retrospectively,<sup>129</sup> which may partly explain the lower prediction performance of the logistic regression models in the external validation experiment of Chapter 2 (Table 2.7).

The morphological studies listed in Table 2.1 pursued a retrospective, cross-sectional design (with the exception of Liu et al.<sup>121</sup>). Retrospective studies typically suffer from stronger selection bias than prospective studies, for which they are often criticized. In contrast, the models trained in this thesis were based on prospectively selected, consecutive data (HUG datasets, Figure 2.1), thus avoiding potential selection bias. The work of Rahman et al.<sup>148</sup> on the size ratio provides another example of a prospective study design. Because of the increasing digitalization in hospitals, the costs for prospective studies have dropped. This favors a shift from retrospective study designs towards prospectively collected patient data in future.

The cross-sectional nature of many morphological studies involving ruptured and unruptured aneurysms has been criticized.<sup>152</sup> Since the cases were observed only at single points in time, no information is available on the natural course of the aneurysms. Such studies examine a potential relationship between predictor variables (shape parameters) and the outcome variable (rupture, no rupture) by applying statistical tests (or other methods) that compare the predictors of one class with those of the other class. Problematic is the interpretation of the results in terms of the predictive value of a morphological parameter: If an unruptured aneurysm has morphological features similar to those of ruptured aneurysms, it is likely to rupture. In this interpretation, observations from the comparison of ruptured and unruptured aneurysms

(rupture status dichotomy) are extrapolated to the problem of distinguishing stable from unstable aneurysms (stability status dichotomy). This is frequently contested, since it remains unclear whether the ruptured aneurysms underwent morphological changes when they rupture, thereby increasing the difference between the two subcohorts in the predictor variables.<sup>60,152,153</sup> However, several studies refuted the criticism by providing evidence that in most cases the shape remains stable after rupture.<sup>29,62,63,97,154</sup>

The studies in this dissertation are grounded on the observation that significant differences can be measured between ruptured and unruptured aneurysms that cannot be fully explained because of rupture. Even if the above criticism is true, it is still possible to extract information from the comparison of ruptured and unruptured aneurysms. But one should keep in mind that the discrimination of the rupture status (“which of the cases in the datasets were ruptured”) represents a surrogate problem that only approximates the clinically relevant stability prediction problem (“which aneurysms are unstable and need treatment?”). The author therefore is of the opinion that the term “rupture risk” should be used with caution in connection with probabilistic models. In this thesis, such models were used primarily to assess the discriminative power of predictor variables. Metrics assessing the predictive performance of probabilistic models were not interpreted as ability to predict future rupture, but as ability to distinguish between ruptured and unruptured aneurysms.

Study designs other than prospective/cross-sectional also suffer from limitations. A prospective, longitudinal study design is often considered most appropriate to assess the predictive capability of a morphometric parameter.<sup>150,152</sup> However, such studies are costly<sup>189</sup> and, as discussed by Ramachandran et al.<sup>151</sup> can suffer from significant selection bias. Referring to the validity of the PHASES score<sup>28</sup>, but applicable also to other studies, Darsaut et al.<sup>191</sup> criticized natural history studies in which aneurysm patients are monitored over time. Usually, such studies only include patients who either refuse treatment or for whom clinicians considered the risk of rupture to be low. Since the patient selection is based on existing but possibly false beliefs about the disease, it is likely that the outcome of a study based on this patient selection will be biased towards these beliefs, rendering the method unscientific. Randomized trials offer a much better perspective for scientific knowledge gain. However, in the context of IA research, such trials often are

deemed impractical because of ethical concerns, because such trials will suffer from significantly lower case numbers and larger costs, and because patient randomization rarely is perfect.<sup>190</sup> To the author’s knowledge, there is currently only one randomized study related to IAs (on the assessment of endovascular treatment options).<sup>192</sup>

When working with interventional 3DRA datasets, unruptured aneurysms tend to be “shifted” towards the unstable (that is growing and rupture-prone) cases, because clinicians already recommended treatment for these cases. It is therefore likely that the unruptured cases have more characteristics in common with the class of ruptured cases than would be the case with a more general collection of aneurysms. Although the problem is mitigated by including untreated secondary aneurysms in patients with multiple aneurysms, it will be beneficial to extend CTA and MRA data to future morphology databases.

Data ages. The increased availability of imaging data as well as the improved understanding of IAs and the associated risk factors have shifted dataset characteristics notably. Today, unruptured IAs are diagnosed more frequently than it was the case 30 years ago. Accordingly, recent prospectively collected datasets are expected to exhibit a larger proportion of unruptured IAs. Finally, treatment guidelines and interventional methods also develop over time (Chapter 1) and therefore may affect selection biases.

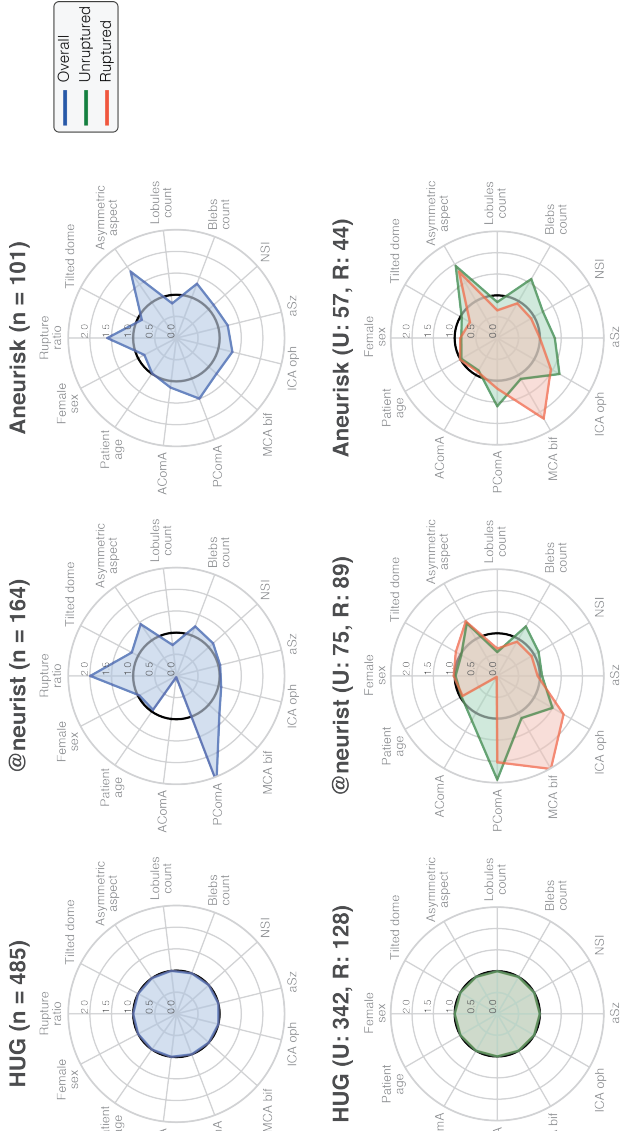
In Chapter 2 it was reported that the datasets used for validation (Aneurisk, @neurIST) differed from the datasets used for training (HUG). For instance, it was pointed out in Chapter 2 that the Aneurisk dataset differed in several characteristics from the HUG datasets used for developing a classification model. In particular, the proportion of ruptured aneurysms was higher (Aneurisk: 44%, HUG: 27%, Table 2.1) in the Aneurisk dataset, and the unruptured aneurysms were significantly larger (median aneurysm size  $aSz$ , Aneurisk: 8.78mm, HUG: 5.6mm, Table 2.8). Since larger aneurysms tend to be more unstable (the risk of rupture increases with the aneurysm size),<sup>1,31</sup> the unruptured aneurysms in the Aneurisk dataset are more likely to resemble the ruptured aneurysms from the training dataset. Furthermore, larger aneurysms were shown to be more irregular and elongated (Table 4.2, or Ashkezari et al.<sup>193</sup>), properties associated with aneurysm rupture. As a consequence, the distinction of the ruptured and unruptured aneurysms in the Aneurisk dataset based on evidence from the reference/training dataset (HUG) failed more often

(Tables 2.6 and 2.7). But what caused the aforementioned differences in the datasets? Differences in the patient population is unlikely to be the reason, as both datasets originate from European hospitals (Milano and Geneva). A more plausible explanation is that the datasets were created using different case selection procedures. While for the HUG dataset, patients were selected prospectively, it is possible that different selection criteria were in place for the retrospective Aneurisk dataset (no details were provided on AneuriskWeb<sup>128</sup>). A similar reasoning applies for the observed discrepancy in @neurIST dataset.

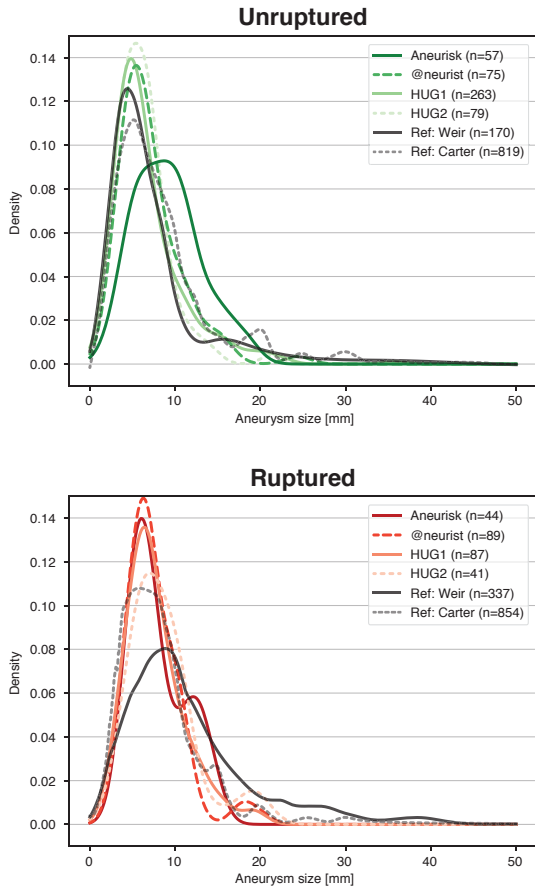
Note that when Aneurisk and @neurIST data was used for the model development according to the same training procedure as described in Chapter 2, the average prediction performance did increase for these datasets. However, this came at the cost of a reduced accuracy on the HUG datasets and an overall drop in prediction performance, further indicating a dataset incompatibility.

In conclusion, the data-driven analysis of diseases as complex as intracranial aneurysms as well as the assessment of possible treatment options is chronically short of reliable data. The relatively large size of potential risk- and confounding factors in combination with the large variability of the aneurysmal disease make the search for “data patterns” difficult. Large, multicentric databases will help to compensate for dataset-specific biases. However, a clear specification the data collection is of utmost importance. Because clinical selection criteria may change over time, and the implementation of reliable data collection procedures is difficult, it is vital to examine the differences between datasets. Besides summary data and statistical testing, the graphical comparison of dataset properties offers an effective tool to identify potential flaws in the data (Figure 5.6, Figure 5.9, Figure 5.10).





**Figure 5.9:** Dataset profiles for the visual inspection of dataset characteristics. For better interpretability, the characteristics are normalized by the corresponding values of the reference dataset (here HUG). Values close to 1 indicate good agreement with the reference dataset. Using this graphical tool, it can be verified quickly that the ratio of ruptured cases was larger for the Aneurisk and @neurIST datasets; and that the @neurIST is lacking cases at the AComA location, whereas PComA aneurysms were overrepresented. These dataset profiles can also be used to compare the subsets of ruptured (R) and unruptured (U) cases. Details: The values labeled with AComA, PComA, MCA bif and ICA oph measure the proportion of aneurysms at this location (weighed by the corresponding ratio of the reference dataset). Dome tilt, asymmetric aspect and the blebs and lobules count were evaluated based on assessments of a human rater (details are omitted).



**Figure 5.10:** Distribution of the aneurysm size in different datasets. The curves represent kernel-density estimates (for Gaussian kernels) which have been computed for size-histograms. For comparison, the distributions reported by Weir et al.<sup>106</sup> and Carter et al.<sup>107</sup> were also added to this plot. It is important to relate new data to existing data to identify potential data incompatibilities. For instance, the Aneurisk dataset consisted of unusually large unruptured aneurysms, whereas the dataset from Weir et al. is skewed towards larger ruptured aneurysms.

In a controversial editorial comment, Kallmes<sup>194</sup> expressed some concerns regarding the usefulness of computational methods from a clinical viewpoint. Observing a growing number of computational risk indices, Kallmes pointed out a certain level of incoherence within the scientific literature and a possible overinterpretation of by-chance associations. Even though addressing primarily the community of biomechanical modelers, the points of concern certainly apply to the morphological analysis of aneurysms as well.

As already mentioned in Section 2.2., the completeness of the reports on aneurysm morphology varies considerably. For instance, multiple definitions exist for such elemental metrics as aneurysm size, neck diameter or aspect ratio, and it is not always clear which ones apply. Also, the statistical validation of findings is sometimes rather weak, as indicated in Table 2.1, which may lead to associations that do not generalize.

In this light, an important contribution of this thesis was to reevaluate existing methods and to benchmark them against each other, with the goal to assess their robustness based on a considerably larger dataset than peers have used previously. Furthermore, we have demonstrated in Chapter 2 the problem of generalizability and varying selection biases, aspects that future data-driven studies must take care off better. This also signifies the primary motivation to publish the AneuX morphology database for scientific reuse.

In regard to the methodological critique by Kallmes, this author would also like to refer to the constructive responses by Cebra and Meng<sup>195</sup> as well as Robertson and Watton<sup>196</sup>, who suggested to better streamline dataset acquisition and processing, who foresaw the increase of dataset sizes and motivated interdisciplinary studies. Indeed, the community since has evolved in these directions. The quantitative analysis of aneurysms must be seen as work in progress, or as Daniel Rufenacht once said: “Science [on intracranial aneurysms] is like shooting on a moving target”. The current knowledge is rapidly evolving, affecting the evidence/data, based on which new knowledge is created.

## 6 Conclusion and outlook

Both quantitative and rater-based analysis of IA morphology confirmed that the shape is predictive of the disease status. Shape irregularity measured either by quantitative metrics associated with irregularity (such as NSI, Zernike energies or new metrics for blebiness or asymmetry) or by a scoring scheme for morphology can help to refine the clinical assessment of aneurysms.

The author has demonstrated that combining morphometric parameters with the aneurysm location in multivariate prediction models significantly increases the informativeness of morphometry with respect to the aneurysm disease status parameters. Future research could further integrate morphology into larger multifactorial models, provided that sufficient data is collected.

The conclusions drawn from data-driven analyses depend heavily on the data used. This thesis has shown that observations valid for one dataset may not generalize to other cohorts due to different selection biases or data processing methods. To facilitate the interpretation of new results by future research, the author suggests establishing reference datasets to which new datasets can be compared. The AneuX morphology database could be part of such a reference.

The data used for training the models in this thesis were based on prospectively selected, consecutive cases. The inclusion of only interventional 3DRA displaying ruptured and unruptured IAs shifted the selection towards unstable cases. Future work on quantitative morphology should incorporate CTA and MRA sources from general examinations that will more closely reflect the general population. To complement current treatment guidelines (which often use IA size as decision criterion), future studies of aneurysm morphology could focus specifically on small and medium-sized aneurysms. Finally, the prospectively collected time-series data, though costly and prone to selection bias, may further increase the clinical value of shape.

In this thesis, the morphological properties of aneurysms were evaluated based on mesh-based geometric models. Direct assessment of morphology using volumetric imaging data could simplify the processing and further exploit the additional information contained in the intensity gradients of the images. Furthermore, by applying deep-learning or radiomics, the relatively laborious development of shape features could be avoided.



## 7 References

1. Etminan, N. & Rinkel, G. J. Unruptured intracranial aneurysms: development, rupture and preventive management. *Nat. Rev. Neurol.* (2016). doi:10.1038/nrneurol.2016.150
2. Brown, R. D. & Broderick, J. P. Unruptured intracranial aneurysms: Epidemiology, natural history, management options, and familial screening. *Lancet Neurol.* **13**, 393–404 (2014).
3. Sakalihan, N., Limet, R. & Defawe, O. D. Abdominal aortic aneurysm. *Lancet* **365**, 1577–89 (2005).
4. Kuzmik, G. A., Sang, A. X. & Elefteriades, J. A. Natural history of thoracic aortic aneurysms. *J. Vasc. Surg.* **56**, 565–571 (2012).
5. Humphrey, J. D. & Taylor, C. A. Intracranial and Abdominal Aortic Aneurysms : Similarities , Differences , and Need for a New Class of Computational Models. *Annu. Rev. Biomed. Eng.* 221–248 (2008). doi:10.1146/annurev.bioeng.10.061807.160439
6. Krings, T. *et al.* Intracranial aneurysms: From vessel wall pathology to therapeutic approach. *Nat. Rev. Neurol.* **7**, 547–559 (2011).
7. Yong-Zhong, G. & Van Alphen, H. A. M. Pathogenesis and histopathology of saccular aneurysms: Review of the literature. *Neurol. Res.* **12**, 249–255 (1990).
8. Nieuwkamp, D. J. *et al.* Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: a meta-analysis. *Lancet Neurol.* **8**, 635–642 (2009).
9. van Gijn, J., Kerr, R. S. & Rinkel, G. J. Subarachnoid haemorrhage. *Lancet* **369**, 306–318 (2007).
10. Suarez, J. I., Tarr, R. W. & Selman, W. R. Aneurysmal Subarachnoid Hemorrhage. *N. Engl. J. Med.* **354**, 387–396 (2006).
11. Vlak, M. H. M., Algra, A., Brandenburg, R. & Rinkel, G. J. E. Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: A systematic review and meta-analysis. *Lancet Neurol.* **10**, 626–636 (2011).
12. Crompton, M. R. Mechanism of Growth and Rupture in Cerebral Berry Aneurysms. *Br. Med. J.* **1**, 1138–1142 (1966).
13. McDowell, M. M. *et al.* Demographic and clinical predictors of multiple intracranial aneurysms in patients with subarachnoid hemorrhage. *J. Neurosurg.* **128**, 961–968 (2017).
14. Juvela, S., Poussa, K. & Porras, M. Factors affecting formation and growth of intracranial aneurysms: A long-term follow-up study. *Stroke* **32**, 485–491 (2001).
15. Boulouis, G. *et al.* Unruptured intracranial aneurysms: An updated review of current concepts for risk factors, detection and management. *Rev. Neurol. (Paris)*. **173**, 542–551 (2017).

16. Backes, D., Rinkel, G. J. E., Laban, K. G., Algra, A. & Vergouwen, M. D. I. Patient-and aneurysm-specific risk factors for intracranial aneurysm growth: A systematic review and meta-analysis. *Stroke* **47**, 951–957 (2016).
17. Vlak, M. H. M., Rinkel, G. J. E., Greebe, P. & Algra, A. Independent risk factors for intracranial aneurysms and their joint effect: A case-control study. *Stroke* **44**, 984–987 (2013).
18. Brown, R. D. *et al.* Screening for brain aneurysm in the Familial Intracranial Aneurysm study: Frequency and predictors of lesion detection. *J. Neurosurg.* **108**, 1132–1138 (2008).
19. Alg, V. S., Sofat, R., Houlden, H. & Werring, D. J. Genetic risk factors for intracranial aneurysms: A meta-analysis in more than 116,000 individuals. *Neurology* **80**, 2154–2165 (2013).
20. Backes, D. *et al.* ELAPSS score for prediction of risk of growth of unruptured intracranial aneurysms. *Neurology* **88**, 1600–1606 (2017).
21. Chien, A. *et al.* Unruptured intracranial aneurysm growth trajectory: Occurrence and rate of enlargement in 520 longitudinally followed cases. *J. Neurosurg.* **132**, 1077–1087 (2020).
22. Koffijberg, H., Buskens, E., Algra, A., Wermer, M. J. H. & Rinkel, G. J. E. Growth rates of intracranial aneurysms: Exploring constancy. *J. Neurosurg.* **109**, 176–185 (2008).
23. Rinkel, G. J. E., Djibuti, M., Algra, A. & van Gijn, J. Prevalence and Risk of Rupture of Intracranial Aneurysms : A Systematic Review. *Stroke* **29**, 251–256 (1998).
24. Murayama, Y. *et al.* Risk Analysis of Unruptured Intracranial Aneurysms: Prospective 10-Year Cohort Study. *Stroke* **47**, 365–371 (2016).
25. Juvela, S., Poussa, K., Lehto, H. & Porras, M. Natural history of unruptured intracranial aneurysms: A long-term follow-up study. *Stroke* **44**, 2414–2421 (2013).
26. Korja, M., Lehto, H. & Juvela, S. Lifelong rupture risk of intracranial aneurysms depends on risk factors: A prospective Finnish cohort study. *Stroke* **45**, 1958–1963 (2014).
27. Wermer, M. J. H., Van Der Schaaf, I. C., Algra, A. & Rinkel, G. J. E. Risk of rupture of unruptured intracranial aneurysms in relation to patient and aneurysm characteristics: An updated meta-analysis. *Stroke* **38**, 1404–1410 (2007).
28. Greving, J. P. *et al.* Development of the PHASES score for prediction of risk of rupture of intracranial aneurysms: A pooled analysis of six prospective cohort studies. *Lancet Neurol.* **13**, 59–66 (2014).
29. Lindgren, A. E. *et al.* Irregular Shape of Intracranial Aneurysm Indicates Rupture Risk Irrespective of Size in a Population-Based Cohort. *Stroke* **47**, 1219–1226 (2016).
30. Broderick, J. P. *et al.* Greater rupture risk for familial as compared to sporadic unruptured intracranial aneurysms. *Stroke* **40**, 1952–1957 (2009).
31. Backes, D. *et al.* PHASES Score for Prediction of Intracranial Aneurysm Growth. *Stroke* **46**, 1221–1226 (2015).
32. Kleinloog, R., Mul, N. De, Post, J. A. & Rinkel, G. J. E. Risk Factors for

- Intracranial Aneurysm Rupture: A Systematic Review. *Neurosurgery* **82**, 431–440 (2017).
33. Bijlenga, P. *et al.* Risk of Rupture of Small Anterior Communicating Artery Aneurysms Is Similar to Posterior Circulation Aneurysms. *Stroke* **44**, 3018–3026 (2013).
  34. Alpers, B. J., Berry, R. G. & Paddison, R. M. Anatomical Studies of the Circle of Willis in Normal Brain. *Arch. Neurol. Psychiatry* **81**, 409–418 (1959).
  35. Dimmick, S. J. & Faulder, K. C. Normal variants of the cerebral circulation at multidetector CT angiography. *Radiographics* **29**, 1027–1043 (2009).
  36. Iqbal, S. A Comprehensive Study of the Anatomical Variations of the Circle of Willis in Adult Human Brains. *J Clin Diagn Res* **7**, 2423–2427 (2013).
  37. Bogunovic, H., Pozo, J. M., Cardenes, R., Roman, L. S. & Frangi, A. F. Anatomical labeling of the circle of willis using maximum a posteriori probability estimation. *IEEE Trans. Med. Imaging* **32**, 1587–1599 (2013).
  38. Tanaka, H. *et al.* Relationship between variations in the circle of Willis and flow rates in internal carotid and basilar arteries determined by means of magnetic resonance imaging with. *AJNR Am J Neuroradiol* (2006).
  39. Kayembe, K. N. T., Sasahara, M. & Hazama, F. Cerebral aneurysms and variations in the circle of Willis. *Stroke* **15**, 846–850 (1984).
  40. Harteveld, A. A. *et al.* Data on vessel wall thickness measurements of intracranial arteries derived from human circle of Willis specimens. *Data Br.* **19**, 6–12 (2018).
  41. Frösen, J. *et al.* Saccular intracranial aneurysm: Pathology and mechanisms. *Acta Neuropathol.* **123**, 773–786 (2012).
  42. Meng, H., Tutino, V. M., Xiang, J. & Siddiqui, A. High WSS or Low WSS? Complex interactions of hemodynamics with intracranial aneurysm initiation, growth, and rupture: Toward a unifying hypothesis. *Am. J. Neuroradiol.* **35**, 1254–1262 (2014).
  43. Robertson, A. M. & Watton, P. N. Chapter 8 - Mechanobiology of the Arterial Wall. *Transp. Biol. Media* 275–347 (2013). doi:http://dx.doi.org/10.1016/B978-0-12-415824-5.00008-4
  44. Ollikainen, E. *Atherosclerotic and inflammatory changes in saccular intracranial aneurysms.* (2018).
  45. Starke, R. *et al.* The Role of Oxidative Stress in Cerebral Aneurysm Formation and Rupture. *Curr. Neurovasc. Res.* **10**, 247–255 (2013).
  46. Kadasi, L. M., Dent, W. C. & Malek, A. M. Cerebral aneurysm wall thickness analysis using intraoperative microscopy: Effect of size and gender on thin translucent regions. *J. Neurointerv. Surg.* **5**, 201–206 (2013).
  47. Stehbens, W. E. History of aneurysms. *Med. Hist.* **2**, 274–280 (1958).
  48. Fox, J. L. Historical Aspects of Intracranial Aneurysms. in *Intracranial Aneurysms* 1–14 (Springer, 1983). doi:10.1007/978-1-4612-5437-9\_1
  49. McDonald, C. & Korb, M. Intracranial Aneurysms. *Arch. Neurol. Psychiatry* **42**, 298–328 (1939).
  50. Cohen-Gadol, A. A. & Spencer, D. D. Harvey W. Cushing and cerebrovascular surgery: Part II, vascular malformations. *J. Neurosurg.* **101**, 553–559 (2004).



51. Guglielmi, G. History of endovascular endosaccular occlusion of brain aneurysms: 1965-1990. *Interv. Neuroradiol.* **13**, 217–224 (2007).
52. Ahmed, A., Kalani, M. Y. S., Nakaji, P. & Spetzler, R. F. Aneurysms, Intracranial. *Encycl. Neurol. Sci.* **1**, 177–182 (2014).
53. D’Urso, P. I., Lanzino, G., Cloft, H. J. & Kallmes, D. F. Flow diversion for intracranial aneurysms: A review. *Stroke* **42**, 2363–2368 (2011).
54. Kotowski, M. *et al.* Safety and occlusion rates of surgical treatment of unruptured intracranial aneurysms: A systematic review and meta-analysis of the literature from 1990 to 2011. *J. Neurol. Neurosurg. Psychiatry* **84**, 42–48 (2013).
55. Naggara, O. N. *et al.* Endovascular treatment of intracranial unruptured aneurysms: systematic review and meta-analysis of the literature on safety and efficacy. *Radiology* **256**, 887–97 (2010).
56. Sailer, A. M. H., Wagemans, B. A. J. M., Nelemans, P. J., De Graaf, R. & Van Zwam, W. H. Diagnosing intracranial aneurysms with mr angiography : Systematic review and meta-analysis. *Stroke* **45**, 119–126 (2014).
57. Westerlaan H.E. *et al.* Intracranial Aneurysms in Patients with Subarachnoid Hemorrhage : CT Angiography as a Primary Examination Tool for Diagnosis — Systematic Review and Purpose : Methods : Results : *Radiology* **258**, 134–45 (2011).
58. Van Rooij, W. J., Sprengers, M. E., De Gast, A. N., Peluso, J. P. P. & Sluzewski, M. 3D rotational angiography: The new gold standard in the detection of additional intracranial aneurysms. *Am. J. Neuroradiol.* **29**, 976–979 (2008).
59. Etminan, N. *et al.* The unruptured intracranial aneurysm treatment score. *Neurology* **85**, 881–889 (2015).
60. Wiebers, D. O., Whisnant, J. P., Sundt, T. M. & O’Fallon, W. M. The significance of unruptured intracranial saccular aneurysms. *J. Neurosurg.* **66**, 23–29 (1987).
61. Wiebers, D. O. *et al.* Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment. *Lancet* **362**, 103–110 (2003).
62. Beck, J. *et al.* Difference in configuration of ruptured and unruptured intracranial aneurysms determined by biplanar digital subtraction angiography. *Acta Neurochir. (Wien)*. **145**, 861–865 (2003).
63. Forget, T. R. *et al.* A Review of Size and Location of Ruptured Intracranial Aneurysms. *Neurosurgery* **49**, 1322–1326 (2001).
64. Fahed, R. & Darsaut, T. E. The Delphi Oracle and the management of aneurysms. *J. Neurointerv. Surg.* **9**, e1–e2 (2017).
65. Ravindra, V. M. *et al.* Validation of the unruptured intracranial aneurysm treatment score: Comparison with real-world cerebrovascular practice. *J. Neurosurg.* **129**, 100–106 (2018).
66. Bijlenga, P. *et al.* PHASES Score for the Management of Intracranial Aneurysm: A Cross-Sectional Population-Based Retrospective Study. *Stroke* **48**, 2105–2112 (2017).
67. Neyazi, B., Sandalcioğlu, I. E. & Maslehaty, H. Evaluation of the risk of rupture

- of intracranial aneurysms in patients with aneurysmal subarachnoid hemorrhage according to the PHASES score. *Neurosurg. Rev.* **42**, 489–492 (2019).
68. Foreman, P. M. *et al.* PHASES score applied to a prospective cohort of aneurysmal subarachnoid hemorrhage patients. *J. Clin. Neurosci.* **53**, 69–73 (2018).
  69. van Kammen, M. S. *et al.* External validation of the ELAPSS score for prediction of unruptured intracranial aneurysm growth risk. *J. Stroke* **21**, 340–346 (2019).
  70. Brinjikji, W. *et al.* PHASES and ELAPSS Scores Are Associated with Aneurysm Growth: A Study of 431 Unruptured Intracranial Aneurysms. *World Neurosurg.* **114**, e425–e432 (2018).
  71. Hilditch, C. A. *et al.* Application of PHASES and ELAPSS scores to ruptured cerebral aneurysms: how many would have been conservatively managed? *J. Neurosurg. Sci.* (2018). doi:10.23736/S0390-5616.18.04498-3
  72. Dodge, Y. & Commenges, D. *The Oxford dictionary of statistical terms.* (2006).
  73. Bishop, C. M. *Pattern Recognition and Machine Learning Springer Mathematical notation Ni.* (2006).
  74. Hastie, T., Tibshirani, R. & Friedman, J. *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction.* (2009).
  75. Xiang, J. *et al.* Hemodynamic-morphologic discriminants for intracranial aneurysm rupture. *Stroke* **42**, 144–152 (2011).
  76. Xiang, J. *et al.* Hemodynamic-morphological discriminant models for intracranial aneurysm rupture remain stable with increasing sample size. *J. Neurointerv. Surg.* **8**, 104–110 (2016).
  77. Bisbal, J., Engelbrecht, G., Villa-Uriol, M. C. & Frangi, A. F. Prediction of cerebral aneurysm rupture using hemodynamic, morphologic and clinical features: A data mining approach. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **6861 LNCS**, 59–73 (2011).
  78. Detmer, F. J. *et al.* Development and internal validation of an aneurysm rupture probability model based on patient characteristics and aneurysm location, morphology, and hemodynamics. *Int. J. Comput. Assist. Radiol. Surg.* (2018). doi:10.1007/s11548-018-1837-0
  79. Detmer, F. J. *et al.* External validation of cerebral aneurysm rupture probability model with data from two patient cohorts. *Acta Neurochir. (Wien).* 2425–2434 (2018). doi:10.1007/s00701-018-3712-8
  80. Edelman, R. R., Chien, D. & Kim, D. Fast selective black blood MR imaging. *Radiology* **181**, 655–660 (1991).
  81. Fayad, Z. A. *et al.* Noninvasive in vivo human coronary artery lumen and wall imaging using black-blood magnetic resonance imaging. *Circulation* **102**, 506–510 (2000).
  82. Dieleman, N. *et al.* Imaging intracranial vessel wall pathology with magnetic resonance imaging current prospects and future directions. *Circulation* **130**, 192–201 (2014).
  83. Mandell, D. M. *et al.* Intracranial Vessel Wall MRI: Principles and Expert Consensus Recommendations of the American Society of Neuroradiology.

- AJNR. Am. J. Neuroradiol.* [Epub ahead of print] (2016).  
doi:10.3174/ajnr.A4893
84. Matouk, C. C. *et al.* Vessel Wall Magnetic Resonance Imaging Identifies the Site of Rupture in Patients With Multiple Intracranial Aneurysms: Proof of Principle. *Neurosurgery* **72**, 492–496 (2013).
  85. Cornelissen, B. M. W. *et al.* Vessel wall enhancement of intracranial aneurysms: Fact or artifact? *Neurosurg. Focus* **47**, 1–9 (2019).
  86. Vanrossomme, X. A. E., Eker, O. F., Thiran, J. P., Courbebaisse, G. P. & Zouaoui Boudjeltia, K. Intracranial aneurysms: Wall motion analysis for prediction of rupture. *American Journal of Neuroradiology* **36**, 1796–1802 (2015).
  87. Cebal, J. R., Castro, M. A. & Soto, O. Blood-flow models of the circle of Willis from magnetic resonance data. *J. Eng. Math.* 369–386 (2003).  
doi:10.1023/B:ENGI.0000007977.02652.02
  88. Cebal, J. R. *et al.* Characterization of cerebral aneurysms for assessing risk of rupture by using patient-specific computational hemodynamics models. *Am. J. Neuroradiol.* **26**, 2550–2559 (2005).
  89. Cebal, J. R. *et al.* Efficient pipeline for image-based patient-specific analysis of cerebral aneurysm hemodynamics: Technique and sensitivity. *IEEE Trans. Med. Imaging* **24**, 457–467 (2005).
  90. Sforza, D. M., Putman, C. M. & Cebal, J. R. Hemodynamics of Cerebral Aneurysms. *Annu. Rev. Fluid Mech.* **41**, 91–107 (2009).
  91. Sforza, D. M. *et al.* Hemodynamics in growing and stable cerebral aneurysms. *J. Neurointerv. Surg.* 1–6 (2015). doi:10.1136/neurintsurg-2014-011339
  92. Teixeira, F. S., Neufeld, E., Kuster, N. & Watton, P. N. Modeling intracranial aneurysm stability and growth: an integrative mechanobiological framework for clinical cases. *Biomech. Model. Mechanobiol.* (2020). doi:10.1007/s10237-020-01351-2
  93. Aparicio, P., Thompson, M. S. & Watton, P. N. A novel chemo-mechanobiological model of arterial tissue growth and remodelling. *J. Biomech.* **49**, 2321–2330 (2016).
  94. Valen-Sendstad, K. *et al.* Real-World Variability in the Prediction of Intracranial Aneurysm Wall Shear Stress: The 2015 International Aneurysm CFD Challenge. *Cardiovasc. Eng. Technol.* **9**, 544–564 (2018).
  95. Ujiie, H., Tamano, Y., Sasaki, K. & Hori, T. Is the aspect ratio a reliable index for predicting the rupture of a saccular aneurysm? *Neurosurgery* **48**, 495–503 (2001).
  96. Weir, B. *et al.* The aspect ratio (dome/neck) of ruptured and unruptured aneurysms. *J. Neurosurg.* **99**, 447–451 (2003).
  97. Raghavan, M. L., Ma, B. & Harbaugh, R. E. Quantified aneurysm shape and rupture risk. *J. Neurosurg.* **102**, 355–62 (2005).
  98. Ma, B., Harbaugh, R. E. & Raghavan, M. L. Three-dimensional geometrical characterization of cerebral aneurysms. *Ann. Biomed. Eng.* **32**, 264–273 (2004).
  99. Dhar, S. *et al.* Morphology parameters for intracranial aneurysm rupture risk assessment. *Neurosurgery* **63**, 185–196 (2008).

100. Yasuda, R. *et al.* Aneurysm Volume-to-Ostium Area Ratio. *Neurosurgery* **68**, 310–318 (2011).
101. Lauric, A., Miller, E. L., Baharoglu, M. I. & Malek, A. M. Rupture status discrimination in intracranial aneurysms using the centroid-radii model. *IEEE Trans. Biomed. Eng.* **58**, 2895–2903 (2011).
102. Lauric, A., Miller, E. L., Baharoglu, M. I. & Malek, A. M. 3D shape analysis of intracranial aneurysms using the writhe number as a discriminant for rupture. *Ann. Biomed. Eng.* **39**, 1457–1469 (2011).
103. Rohde, S. *et al.* Fourier analysis of intracranial aneurysms: Towards an objective and quantitative evaluation of the shape of aneurysms. *Neuroradiology* **47**, 121–126 (2005).
104. Millán, R. D., Dempere-Marco, L., Pozo, J. M., Cebal, J. R. & Frangi, A. F. Morphological characterization of intracranial aneurysms using 3-D moment invariants. *Med. Imaging, IEEE Trans.* **26**, 1270–1282 (2007).
105. Berkowitz, B. M. Development of metrics to describe cerebral aneurysm morphology. (University of Iowa, 2016).
106. Weir, B., Disney, L. & Karrison, T. Sizes of ruptured and unruptured aneurysms in relation to their sites and the ages of patients. *J. Neurosurg.* **96**, 64–70 (2002).
107. Carter, B. S., Sheth, S., Chang, E., Sethl, M. & Ogilvy, C. S. Epidemiology of the Size Distribution of Intracranial Bifurcation Aneurysms: Smaller Size of Distal Aneurysms and Increasing Size of Unruptured Aneurysms with Age. *Neurosurgery* **58**, 217–223 (2006).
108. Zheng, Y. *et al.* Size, Aspect Ratio and Anatomic Location of Ruptured Intracranial Aneurysms: Consecutive Series of 415 Patients from a Prospective, Multicenter, Observational Study. *Cell Transplant.* **28**, 739–746 (2019).
109. Björkman, J. *et al.* Irregular Shape Identifies Ruptured Intracranial Aneurysm in Subarachnoid Hemorrhage Patients with Multiple Aneurysms. *Stroke* **48**, 1986–1989 (2017).
110. Räisänen, S. *et al.* Impact of Young Age on the Presentation of Saccular Intracranial Aneurysms: Population-Based Analysis of 4082 Patients. *Neurosurgery* **0**, 1–9 (2017).
111. Ebnöther, U. Shape Understanding of Intracranial Aneurysms. (ETH Zurich, Zurich University of Applied Sciences, 2015).
112. Morel, S. *et al.* Correlating Clinical Risk Factors and Histological Features in Ruptured and Unruptured Human Intracranial Aneurysms: The Swiss AneuX Study. *J. Neuropathol. Exp. Neurol.* **77**, 555–566 (2018).
113. Cebal, J. *et al.* Flow conditions in intracranial aneurysm lumen associate with inflammation and degenerative changes of the aneurysm wall. *Am. J. Neuroradiol.* **in press**, 1–8 (2016).
114. Meng, H. *et al.* Complex hemodynamics at the apex of an arterial bifurcation induces vascular remodeling resembling cerebral aneurysm initiation. *Stroke* **38**, 1924–1931 (2007).
115. Kulcsár, Z. *et al.* Hemodynamics of cerebral aneurysm initiation: The role of wall shear stress and spatial wall shear stress gradient. *Am. J. Neuroradiol.* **32**, 587–594 (2011).

116. Frösen, J., Cebral, J., Robertson, A. M. & Aoki, T. Flow-induced, inflammation-mediated arterial wall remodeling in the formation and progression of intracranial aneurysms. *Neurosurg. Focus* **47**, 1–13 (2019).
117. Ferguson, G. G. Physical factors in the initiation, growth, and rupture of human intracranial saccular aneurysms. *J. Neurosurg.* **37**, 666–677 (1972).
118. Piccinelli, M. *et al.* Automatic neck plane detection and 3d geometric characterization of aneurysmal sacs. *Ann. Biomed. Eng.* **40**, 2188–2211 (2012).
119. Canterakis, N. 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition. *11th Scand. Conf. Image Anal.* **In 11th Sc**, 85–93 (1999).
120. Novotni, M. & Klein, R. 3D Zernike Descriptors for Content Based Shape Retrieval. *Proc. eighth ACM Symp. Solid Model. Appl.* 216–225 (2003). doi:10.1145/781636.781639
121. Liu, Q. *et al.* Prediction of Aneurysm Stability Using a Machine Learning Model Based on PyRadiomics-Derived Morphological Features. *Stroke* **50**, 2314–2321 (2019).
122. Cohen, J. F. *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* **6**, e012799 (2016).
123. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Br. J. Surg.* **102**, 148–158 (2015).
124. Hoffman, H. *et al.* A retrospective comparison of sac and lobe morphology between ruptured and unruptured intracranial aneurysms. *J. Clin. Neurosci.* **60**, 88–92 (2019).
125. Niemann, U. *et al.* Rupture Status Classification of Intracranial Aneurysms Using Morphological Parameters. *Proc. - IEEE Symp. Comput. Med. Syst.* **2018-June**, 48–53 (2018).
126. Villa-Uriol, M. C. *et al.* @neurIST complex information processing toolchain for the integrated management of cerebral aneurysms. *Interface Focus* **1**, 308–319 (2011).
127. Berti, G. *et al.* *Integrated biomedical informatics for the management of cerebral aneurysms - D23v2 - Analysis Protocols Version 2. Aneurist* (2010).
128. Aneurisk-Team. AneuriskWeb project website. *Emory University, Department of Math & CS* (2012). Available at: <http://ecm2.mathcs.emory.edu/aneuriskweb>. (Accessed: 1st July 2017)
129. Ford, M. D., Hoi, Y., Piccinelli, M., Antiga, L. & Steinman, D. A. An objective approach to digital removal of saccular aneurysms : technique and applications. *Br. J. Radiol.* **82**, 55–61 (2009).
130. Piccinelli, M., Veneziani, A., Steinman, D. A., Remuzzi, A. & Antiga, L. A framework for geometric analysis of vascular structures: Application to cerebral aneurysms. *IEEE Trans. Med. Imaging* **28**, 1141–1155 (2009).
131. Piccinelli, M. *et al.* Geometry of the internal carotid artery and recurrent patterns in location, orientation, and rupture status of lateral aneurysms: An image-based computational study. *Neurosurgery* **68**, 1270–1285 (2011).

132. Larrabide, I. *et al.* GIMIAS: An open source framework for efficient development of research tools and clinical prototypes. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **5528**, 417–426 (2009).
133. Hernandez, M. & Frangi, A. F. Non-parametric geodesic active regions: Method and evaluation for cerebral aneurysms segmentation in 3DRA and CTA. *Med. Image Anal.* **11**, 224–241 (2007).
134. Bogunović, H. *et al.* Automated segmentation of cerebral vasculature with aneurysms in 3DRA and TOF-MRA using geodesic active regions: An evaluation study. *Med. Phys.* **38**, 210–222 (2010).
135. Antiga, L. *et al.* An image-based modeling framework for patient-specific computational hemodynamics. *Med. Biol. Eng. Comput.* **46**, 1097–1112 (2008).
136. Lorensen, W. E. & Cline, H. E. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. **21**, 163–169 (1987).
137. Mut, F. *et al.* Computational hemodynamics framework for the analysis of cerebral aneurysms. *Int. j. numer. method. biomed. eng.* **27**, 822–839 (2011).
138. Campen, M. & Kobbelt, L. Practical Guide to Polygon Mesh Repairing. *Eurographics* 1–55 (2012). doi:10.2312/conf/EG2012/tutorials/t4
139. Taubin, G. Curve and surface smoothing without shrinkage. *IEEE International Conference on Computer Vision* 852–857 (1995). doi:10.1109/icc.1995.466848
140. Taubin, G., Zhang, T. & Golub, G. *Optimal Surface Smoothing as Filter Design. Teaching Statistics* **16**, (1994).
141. Schroeder, W., Martin, K. & Lorensen, B. *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*. (Kitware, 2006).
142. Pozo, J. M., Villa-Urriol, M.-C. & Frangi, A. F. Efficient 3D Geometric and Zernike Moments Computation from Unstructured Surface Meshes. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 471–484 (2011).
143. Bonferroni, C. E. Teoria statistica delle classi e calcolo delle probabilità. *Pubbl. del R Ist. Super. di Sci. Econ. e Commer. di Firenze* 3–62 (1936).
144. Ke, G. *et al.* LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* 3149–3157 (2017).
145. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
146. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
147. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proc. 9th Python Sci. Conf.* 57 (2010).
148. Rahman, M. *et al.* Size Ratio Correlates With Intracranial Aneurysm Rupture Status: A Prospective Study. *Stroke* **41**, 916–920 (2010).
149. Van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
150. Korja, M. & Kaprio, J. Controversies in epidemiology of intracranial aneurysms and SAH. *Nat. Rev. Neurol.* **12**, 50–55 (2016).
151. Ramachandran, M. *et al.* Assessment of image-derived risk factors for natural course of unruptured cerebral aneurysms. *J. Neurosurg.* **124**, 288–95 (2016).

152. Juvela, S. & Korja, M. Intracranial Aneurysm Parameters for Predicting a Future Subarachnoid Hemorrhage: A Long-Term Follow-up Study. *Neurosurgery* **81**, 432–440 (2017).
153. Skodvin, T. O., Johnsen, L. H., Gjertsen, Ø., Isaksen, J. G. & Sorteberg, A. Cerebral Aneurysm Morphology before and after Rupture: Nationwide Case Series of 29 Aneurysms. *Stroke* **48**, 880–886 (2017).
154. Yasui, N., Suzuki, A., Nishimura, H., Suzuki, K. & Abe, T. Long-term Follow-up Study of Unruptured Intracranial Aneurysms. *Neurosurgery* 1155–1160 (1997). doi:10.1097/0006123-199706000-00007
155. Millán, R. D., Dempere-Marco, L., Pozo, J. M., Cebral, J. R. & Frangi, a F. Morphological characterization of intracranial aneurysms using 3-D moment invariants. *IEEE Trans. Med. Imaging* **26**, 1270–82 (2007).
156. Juchler, N. *et al.* Using shape descriptors to categorize intracranial aneurysms. in *CMBE2017: 5th International Conference on Computational and Mathematical Biomedical Engineering* (eds. Nithiarasu, P. & Robertson, A. M.) **2**, 693–695 (2017).
157. Juchler, N. *et al.* Understanding morphological irregularity in cerebral aneurysms – a rater-based study. in *CMBE2019: 6th International Conference on Computational and Mathematical Biomedical Engineering* (eds. Nithiarasu, P., Ohta, M. & Oshima, M.) 583–586 (2019).
158. Juchler, N. *et al.* Radiomics approach to quantify shape irregularity from crowd-based qualitative assessment of intracranial aneurysms. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* 1–9 (2020). doi:10.1080/21681163.2020.1728579
159. Karamanakos, P. N. *et al.* Risk factors for three phases of 12-month mortality in 1657 patients from a defined population after acute aneurysmal subarachnoid hemorrhage. *World Neurosurg.* **78**, 631–639 (2012).
160. UCAS Japan Investigators. The natural course of unruptured cerebral aneurysms in a Japanese cohort. *N. Engl. J. Med.* **366**, 2474–82 (2012).
161. Ujiie, H. *et al.* Effects of Size and Shape (Aspect Ratio) on the Hemodynamics of Saccular Aneurysms: A Possible Index for Surgical Treatment of Intracranial Aneurysms. *Neurosurgery* **45**, 119–130 (1999).
162. Suh, S. H., Cloft, H. J., Huston, J., Han, K. H. & Kallmes, D. F. Interobserver variability of aneurysm morphology: Discrimination of the daughter sac. *J. Neurointerv. Surg.* **8**, 38–41 (2014).
163. Forbes, G., Fox, A. J., Huston III, J., Wiebers, D. O. & Torner, J. *Interobserver Variability in Angiographic Measurement and Morphologic Characterization of Intracranial Aneurysms: A Report from the International Study of Unruptured Intracranial Aneurysms.* *AJNR Am J Neuroradiol* **17**, (1996).
164. Lorensen, W. E. & Cline, H. E. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Comput. Graph.* **21**, 163–169 (1987).
165. Iglewicz, B. & Hoaglin, D. Volume 16: How to Detect and Handle Outliers. in *The ASQC Basic References in Quality Control: Statistical Techniques* **16**, (1993).
166. Juchler, N. *et al.* Shape irregularity of the intracranial aneurysm lumen exhibits

- diagnostic value. *Acta Neurochir. (Wien)*. (2020). doi:10.1007/s00701-020-04428-0
167. Wanke, I. & Rüfenacht, D. A. Zufallsbefund intrakranielles Aneurysma. *Swiss Med. Forum* **17**, 80–82 (2017).
  168. Piccinelli, M. *Characterization of cerebral aneurysm morphology*. (Technische Universiteit Eindhoven, 2012). doi:<https://doi.org/10.6100/IR741500>
  169. Millan, R. D., Dempere-Marco, L., Pozo, J. M., Cebal, J. R. & Frangi, A. F. Morphological Characterization of Intracranial Aneurysms Using 3-D Moment Invariants. *IEEE Trans. Med. Imaging* **26**, 1270–1282 (2007).
  170. Juvela, S., Porras, M. & Poussa, K. Natural history of unruptured intracranial aneurysms: probability of and risk factors for aneurysm rupture. *J Neurosurg* **93**, 379–387 (2000).
  171. Schneiders, J. J. *et al.* Rupture-associated changes of cerebral aneurysm geometry: High-resolution 3D imaging before and after rupture. *Am. J. Neuroradiol.* **35**, 1358–1362 (2014).
  172. Cebal, J. *et al.* Flow Conditions in the Intracranial Aneurysm Lumen Are Associated with Inflammation and Degenerative Changes of the Aneurysm Wall. *Am. J. Neuroradiol.* **38**, 119–26 (2017).
  173. Valen-Sendstad, K. & Steinman, D. A. Mind the gap: Impact of computational fluid dynamics solution strategy on prediction of intracranial aneurysm hemodynamics and rupture status indicators. *Am. J. Neuroradiol.* **35**, 536–543 (2014).
  174. Valen-Sendstad, K. *et al.* Real-World Variability in the Prediction of Intracranial Aneurysm Wall Shear Stress : The 2015 International Aneurysm CFD Challenge. *Cardiovasc. Eng. Technol.* (2018). doi:10.1007/s13239-018-00374-2
  175. Retarekar, R. *et al.* Stratification of a population of intracranial aneurysms using blood flow metrics. *Comput. Methods Biomech. Biomed. Engin.* **18**, 1072–1082 (2015).
  176. Salimi Ashkezari, S. F., Mut, F., Chung, B. J., Robertson, A. M. & Cebal, J. R. Hemodynamic conditions that favor bleb formation in cerebral aneurysms. *J. Neurointerv. Surg.* neurintsurg-2020-016369 (2020). doi:10.1136/neurintsurg-2020-016369
  177. Kwak, Y., Son, W., Kim, Y.-S., Park, J. & Kang, D.-H. Discrepancy between MRA and DSA in identifying the shape of small intracranial aneurysms. *J. Neurosurg.* 1–7 (2020). doi:10.3171/2020.4.JNS20128
  178. Pasupathy, A. Neural basis of shape representation in the primate brain. in *Progress in Brain Research* **154**, 293–313 (2006).
  179. Todd, J. T. The visual perception of 3D shape. *Trends Cogn. Sci.* **8**, 115–121 (2004).
  180. Welchman, A., Deubelius, A., ... V. C.-N. & 2005, undefined. 3D shape perception from combined depth cues in human visual cortex. *nature.com*
  181. Fleming, R. W., Torralba, A. & Adelson, E. H. Specular reflections and the perception of shape. *J. Vis.* **4**, 798–820 (2004).
  182. Ramachandran, V. S. Perception of shape from shading. *Am. J. Ophthalmol.* **105**, 446 (1988).



183. Giannouli, V. Symmetry perception in humans. *ENCEPHALOS* **50**, 31–42 (2013).
184. Pizlo, Z. *3D Shape: Its Unique Place in Visual Perception*. (The MIT Press, 2008). doi:10.7551/mitpress/7705.001.0001
185. Raghupathi, W. & Raghupathi, V. Big data analytics in healthcare: promise and potential. *Heal. Inf. Sci. Syst.* **2**, 1–10 (2014).
186. Bzdok, D., Altman, N. & Krzywinski, M. Points of Significance: Statistics versus machine learning. *Nat. Methods* **15**, 233–234 (2018).
187. StackExchangeCommunity. The Two Cultures: statistics vs. machine learning? - CrossValidated. *stats.stackexchange.com* Available at: <http://stats.stackexchange.com/questions/6/the-two-cultures-statistics-vs-machine-learning?newreg=7fc901aec1144035bd9e0ac16d297e70>. (Accessed: 20th July 2020)
188. Vandenbroucke, J. P. Prospective or retrospective: What’s in a name? *Br. Med. J.* **302**, 249–250 (1991).
189. Euser, A. M., Zoccali, C., Jager, K. J. & Dekker, F. W. Cohort studies: Prospective versus retrospective. *Nephron - Clin. Pract.* **113**, (2009).
190. Rinkel, G. J. E., Algra, A., Greving, J. P., Vergouwen, M. D. I. & Etminan, N. PHASES and the natural history of unruptured aneurysms: Science or pseudoscience? *J. Neurointerv. Surg.* **9**, 618 (2017).
191. Darsaut, T., Fahed, R. & Raymond, J. PHASES and the natural history of unruptured aneurysms: Science or pseudoscience? *J. Neurointerv. Surg.* **9**, 527–528 (2017).
192. Raymond, J., Darsaut, T. E. & Molyneux, A. J. A trial on unruptured intracranial aneurysms (the TEAM trial): Results, lessons from a failure and the necessity for clinical care trials. *Trials* **12**, 1–14 (2011).
193. Salimi Ashkezari, S. F. *et al.* Blebs in intracranial aneurysms: prevalence and general characteristics. *J. Neurointerv. Surg.* neurintsurg-2020-016274 (2020). doi:10.1136/neurintsurg-2020-016274
194. Kallmes, D. F. Point: CFD - Computational fluid dynamics or confounding factor dissemination. *Am. J. Neuroradiol.* **33**, 395–396 (2012).
195. Cebal, J. R. & Meng, H. Counterpoint: Realizing the clinical utility of computational fluid dynamics - Closing the gap. *Am. J. Neuroradiol.* **33**, 396–398 (2012).
196. Robertson, A. M. & Watt. Computational Fluid Dynamics in Aneurysm Research: Critical Reflections, Future Directions. *Am. J. Neuroradiol.* **33**, 992–995 (2012).