Journal Pre-proof

Data Driven Models Merging Geometric, Biomechanical, and Clinical Data to Assess the Rupture of Abdominal Aortic Aneurysms

Marta Alloisio, Antti Siika, Joy Roy, Sebastian Zerwes, Alexander Hyhlik-Duerr, T. Christian Gasser

Please cite this article as: Alloisio M, Siika A, Roy J, Zerwes S, Hyhlik-Duerr A, Gasser TC, Data Driven Models Merging Geometric, Biomechanical, and Clinical Data to Assess the Rupture of Abdominal Aortic Aneurysms, *European Journal of Vascular & Endovascular Surgery* (2025), doi: https://doi.org/10.1016/j.ejvs.2025.06.002.

Data Driven Models Merging Geometric, Biomechanical, and Clinical Data to Assess the Rupture of Abdominal Aortic Aneurysms

Marta Alloisio [a], Antti Siika [b], Joy Roy [b], Sebastian Zerwes [c], Alexander Hyhlik-Duerr [c], T. Christian Gasser [a,*]

[a] Department of Engineering Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden

[b] Department of Molecular Medicine and Surgery, KI Karolinska Institute, Stockholm, Sweden

[c] Vascular Surgery Medical Faculty, University of Augsburg, Augsburg, Germany

[*] Corresponding author. KTH Solid Mechanics, Osquars Backe, Stockholm SE10044, Sweden.

E-mail address: gasser@kth.se (T. Christian Gasser).

Running title odd: Can Machine Learning Better Assess the Rupture Risk of AAA?

Running title even: Marta Alloisio et al.

WHAT THIS PAPER ADDS

The clinical diameter based decision on whether to operate on a patient with abdominal aortic aneurysm (AAA) leads to many surgeries that could have been postponed. This work aims to identify ruptured and intact AAAs through the integration of geometric, clinical, and biomechanical variables by machine learning (ML) models. The developed approach greatly improves the specificity of AAA classification without compromising its sensitivity. Ranking of risk factors in the ML models allowed the discernment of the most critical variables in AAA rupture.

Objective: Despite elective repair of a large portion of stable abdominal aortic aneurysms (AAAs), the diameter criterion cannot prevent all small AAA ruptures. Since rupture depends on many factors, this study explored whether machine learning (ML) models (logistic regression [LogR], linear and non-linear support vector machine [SVM-Lin and SVM-Nlin], and Gaussian Naïve Bayes [GNB]) might improve the diameter based risk assessment by comparing already ruptured (diameter 52.8 – 174.5 mm) with asymptomatic (diameter 40.4 – 95.5 mm) aortas.

Methods: A retrospective case–control observational study included ruptured AAAs from two centres (2010 – 2012) with computed tomography angiography images for finite element analysis. Clinical patient data and geometric and biomechanical AAA properties were fed into ML models, whose output was compared with the results from intact cases. Classifications were explored for all cases and those having diameters below 70 mm. All data trained and validated the ML models, with a five fold cross-validation. SHapley Additive exPlanations (SHAP) analysis ranked the factors for rupture identification.

Results: One hundred and seven ruptured (20% female, mean age 77 years, mean diameter 86.3 mm) and 200 non-ruptured aneurysmal infrarenal aortas (22% female, mean age 74 years, mean diameter 57 mm) were investigated through cross-validation methods. Given the entire dataset, the diameter threshold of 55 mm in men and 50 mm in women provided a 58% accurate rupture classification. It was 99% sensitive (AAA rupture identified correctly) and 36% specific (intact AAAs identified correctly). ML models improved accuracy (LogR 90.2%, SVM-Lin 89.48%, SVM-Nlin 88.7%, and GNB 86.4%); accuracy decreased when trained on the ≤ 70 mm group (55/50 mm diameter threshold 44.2%, LogR 82.5%, SVM-Lin 83.6%,

SVM-Nlin 65.9%, and GNB: 84.7%). SHAP ranked biomechanical parameters other than the diameter as the most relevant.

Conclusion: A multiparameter estimate enhanced the purely diameter based approach. The proposed predictability method should be further tested in longitudinal studies.

Keywords: Abdominal aortic aneurysm, Aortic rupture, Artificial intelligence, Machine learning, Prognosis, Surgery

4

INTRODUCTION

Abdominal aortic aneurysm (AAA) is a localised dilatation of the infrarenal aorta, resulting in a diameter greater than 30 mm.[1] On the basis of the diameter of the aorta, AAA can be classified as small (not considered for repair, < 55/50 mm for men/women) or large (≥ 55/50 mm) when surgical repair can be considered.[2] The formation and progression of an AAA are tightly linked to proteolytic degradation of elastin and collagen in the aortic wall.[3] At progressive expansion, the risk of aortic rupture increases, an event that is fatal in the majority of cases.[4] It remains difficult to predict when a patient with an AAA will require surgery[5] to prevent rupture. Even in surveillance programmes, a small proportion of patients suffer from rupture prior to reaching the indication for repair, while on the other hand a significant proportion can reach large aneurysm diameters without rupturing during their lifetime.[6]

Large AAA diameter is not the only risk factor, and rupture has also been associated with biomechanical factors,[7] AAA shape[8] and volume, ethnicity, female sex, family susceptibility, mean arterial pressure,[9] expansion rate,[10] fludeoxyglucose uptake,[11] intraluminal thrombus (ILT),[12] smoking,[13] and many others.[3] Regardless of the diameter being a strong surrogate measure of AAA rupture risk, it has clear limitations in the prediction of the risk for aortic rupture in the individual patient.

Aortic rupture is the result of the interaction of multiple risk factors. The wall ruptures at the site where wall stress overcomes wall strength,[14] a spatial location that is often not at the level of the largest diameter.[16] Besides challenging classical statistical methods, the strong dependence among the individual risk factors for AAA rupture also hinders the study of their isolated contribution to the cumulative rupture risk.

Machine learning (ML) techniques (a subclass of artificial intelligence based models) have recently garnered attention in biomedical and mechanical fields.[16] They can process large numbers of data in a reasonable time and potentially establish complex relationships between input and output. Therefore, ML based models can integrate risk factors proposed by different research disciplines into a multidimensional AAA growth[17] and rupture risk assessment and allow individual risk factors to be ranked according to their importance.

The current observational study combines general patient characteristics with AAA geometric and biomechanical parameters into a single holistic factor determining the relationship to AAA rupture. This new method of identifying the potential for AAA rupture may potentially enhance AAA rupture risk assessment, leading to higher accuracy compared with diameter based clinical decision making.

MATERIALS AND METHODS

Patient and image data acquisition

Patients from Karolinska University Hospital (Stockholm, Sweden) and University Hospital Augsburg (Augsburg, Germany) with ruptured or intact AAA were considered for this retrospective observational case–control study. Ruptured AAAs were from clinical records from 2010 to 2021. Inclusion required contrast enhanced computed tomography angiography (CTA) scan (1 – 5 mm slice thickness), excluding approximately 50% owing to biomechanical analysis limitations (e.g., large rupture site, highly swirled contrast medium, poor image quality, tortuous anatomy). Patients

with non-ruptured AAA within the same CTA scan slice thickness (from AAA surveillance at Karolinska University Hospital in 2012 – 2013, or treated at Augsburg University Hospital in 2010 – 2021) were considered at a 2:1 intact/rupture ratio, age and sex matched, with < 10% excluded for similar reasons. CTA in plane resolution ranged from 0.39 mm to 0.8 mm. Blood pressure was recorded before CTA in intact cases or at the last admission before rupture. If unavailable (51%), the blood pressure was set to 140/80 mmHg.

The use of anonymised data was approved by the local ethics committees of both involved clinics.

As the patient group included cases of very large diameters, cases with diameters ≤ 70 mm were also uniquely analysed, thus resulting in a group where AAA rupture risk estimation would be clinically more relevant. Diameter matched groups were intentionally not used in order to not *a priori* exclude or compromise the diameter as a risk factor.

The Shapiro Wilk test was used to assess normality. Student's *t* test (normal data) or Mann–Whitney *U* test (non-normal data) were used to evaluate statistically significant differences of each feature in the rupture and non-ruptured cases ($p < .050$).

Image data post-processing

Commercial software (A4clinics Research Edition; VASCOPS GmbH, Graz, Austria) was used to extract geometric and biomechanical features from CTA images. A

semiautomatic segmentation[18] identified the lumen, the ILT, and the outer contour of the vessel wall. Segmentation and analyses were performed between the lowest main renal artery and aortic bifurcation, ignoring accessory renal arteries. Given the 3D model, geometric features (AAA diameter, maximum luminal diameter, maximum ILT thickness, vessel volume, lumen volume, ILT volume) were automatically extracted. Diameters were computed perpendicularly to the anatomical centreline and thus the centreline of the AAA outer contour.

The finite element method was used to calculate wall stress using the patient specific reconstructions, assuming hyperelastic, isotropic, and incompressible[19,20] tissue. Stress was derived from wall strength estimates based on sex, family history, ILT thickness, and the ratio between local aneurysm diameter and the expected normal value.[18,21] The finite element method model, loaded with the patient specific mean arterial pressure and fixed at the lowest renal artery and the aortic bifurcation levels, was used to compute biomechanical properties (peak wall stress, mean wall stress, peak wall rupture risk index [PWRI], rupture risk equivalent diameter [RRED], mean wall rupture risk index [MWRI], peak ILT stress, mean ILT stress, peak ILT rupture risk, and mean ILT rupture risk) (Supplementary Table S1). In this study, all stress measures refer to the von Mises stress,[22] and risk indices are formed by stress/strength ratio.[18,23]

PWRI is the highest stress to strength ratio where wall stress is based on the individual finite element computation,[14] while wall strength considers age, sex, ILT layer thickness, and relative diameter expansion.[21] RRED specifies the hypothetical

8

AAA diameter at which PWRI matches the non-ruptured AAA population.[23] Supplementary Table S1 lists a detailed definition of all parameters.

The analysis took approximately 15 minutes for intact and 30 minutes for rupture cases, respectively. Figure 1 illustrates a typical AAA after image processing.

Machine learning analyses and prediction models

The outcome (AAA rupture) was classified by use of several ML models, including logistic regression (LogR),[25] linear and non-linear support vector machine[26] (SVM-Lin and SVM-Nlin), and Gaussian Naïve Bayes (GNB).[27] The models were realised and trained in Scikit-learn v.1.4.0.[28]

Data statistics and cleansing. As the accuracy of any ML algorithm strongly depends on the quality and homogeneity of the input, all variables were normalised and scaled (mean = 0, variance = 1), a measure known to particularly avoid the major artificial influence of variables with extensive variations. A principal component analysis, a high dimensional visualisation technique, revealed that ruptured and intact cases could not easily be separated (Supplementary Figure S1), and one case was regarded as an outlier and therefore removed.

Given that highly correlated variables promote ML overfitting (i.e., an excessively complex model accurately predicting the training data but poorly the validation data[29]), the pairwise correlation amongst the 18 features listed in Table 1 was explored. Given that the correlation (Spearman's rank order correlation coefficient ρ) among two features exceeded 0.80, the one with the lower area under the receiver

9

operating characteristic (ROC) curve (AUROC) was disregarded (Supplementary Table S2). The specific threshold was set to keep an adequate number of features in relation to the number of cases in the group.

As there was a significant disparity in the occurrence of intact compared with ruptured AAA cases, misclassification made by the ruptured class was given a greater penalty, thus preventing the models from favouring the intact cases.

Group partitioning, model training, and validation. The ML models utilised the aforementioned features to classify AAA rupture and were developed either on the entire group or the ≤ 70 mm group. A five fold cross-validation was applied to train and validate the ML models, and within each fold data for training and validation represented approximately 80% and 20%, respectively. The analysis was repeated three times to eliminate potential biases caused by random dataset shuffling, thus producing multiple models. Consequently, the overall ML classification performance was based on the average metrics received from all 30 models.

Classification metrics. Given the number of correct (TP) and incorrect (FP) classification of AAA rupture, as well as the number of correct (TN) and incorrect (FN) classification of intact cases, the following metrics were used to quantify ML model performance: accuracy = (TP + TN)/(TP + TN + FP + FN); sensitivity = TP/(TP + FN); specificity = TN/(TN + FP); and AUROC = area under the ROC curve (relation between true positive rate [i.e., sensitivity] and false positive rate [i.e., 1 – specificity]) computed using the trapezoidal rule.

Regardless of the group used to develop the models (i.e., either the entire group or the $\leq 70\,mm$ group), the aforementioned metrics were only tested in the $\leq 70\,mm$ group as it holds clinical relevance. Additionally, SVM results were calibrated (via another LogR model) to retain probability outputs.[30]

As the primary AAA rupture assessment goal was to increase specificity while maintaining high sensitivity, the thresholds of the ML based rupture probability (initially 50%) were varied and the corresponding classifications were explored.

Interpretability. To overcome the drawback of data driven algorithms of hiding which and how input variables influence the prediction, the SHapley Additive exPlanations (SHAP) method[31] was employed to rank the AAA rupture factors. SHAP uses the classic Shapley values from game theory to assign a contribution to each feature in a prediction, scoring its impact on the model output. In this study, a Python based software package (https://shap.readthedocs.io/en/latest/) was used. Given a ML model, the average from all realisations (five folds times three samples) determined the final SHAP outcome and thus the relative influence of the individual features on the rupture classification. The SHAP value was also evaluated using the best trained models.

11

RESULTS

Patient groups

Data from 107 ruptured and 200 non-ruptured aneurysmatic infrarenal aortas in 242 men and 65 women were retrospectively considered (Table 2). Diameters ranged from 52.8 – 174.5 mm for ruptured cases and from 40.4 – 95.5 mm for intact cases.

The Shapiro–Wilk test showed no normally distributed feature ($p < .050$) on the entire group. The ≤ 70 mm group instead was more homogeneous, and the vessel volume, maximum luminal diameter, and mean wall stress were normally distributed.

In the entire group and the ≤ 70 mm group, many geometric and mechanical properties differed between ruptured and intact cases (Tables 2 and 3).

Age differed statistically significantly between ruptured and non-ruptured groups ($p = .012$) in the entire group but not in the ≤ 70 mm group. The female:male ratio was similar (21% female in ruptured cases *vs.* 22% in intact cases) in the entire group, while it was higher in the ruptured cases ≤ 70 mm (32% *vs.* 22%). No difference was seen in systolic or diastolic pressure, and the smoking status was similar in both groups, albeit unknown in 76%.

The diameter threshold

Given the entire group, the diameter threshold of 55 mm in men and 50 mm in women provided a 58% accurate rupture classification. It was 99% sensitive (AAA rupture identified correctly) and 36% specific (intact AAAs identified correctly). Its

12

classification accuracy decreased in the ≤ 70 mm group to 44.2% (sensitivity 96.7%, specificity 37.7%).

The classification performance of single features

Figure 2 illustrates ROC curves for all investigated features in the entire patient group as well as the ≤ 70 mm group. For all classifiers, the accuracy decreased from the entire group to the ≤ 70 mm group.

Although the AUROC values decreased significantly from the entire group to the ≤ 70 mm group, in both the external diameter, MWRI, RRED, and PWRI result in the most robust classifiers. As ILT volume and ILT thickness do not provide any information (AUROC < 50%) in the ≤ 70 mm group, they were excluded from the corresponding ML analysis.

Machine earning model classification performance

Compared with single feature classification, ML models improved the classification accuracy remarkably (LogR 90.2%, SVM-Lin 89.48%, SVM-Nlin 88.7%, and GNB 86.4%). As with single features, ML based classification accuracy also decreased in the ≤ 70 mm group (LogR 82.5%, SVM-Lin 83.6%, SVM-Nlin 65.9%, and GNB 84.7%). Table 4 reports ML based validation metrics in the ≤ 70 mm group in comparison with the diameter threshold, i.e., 55 mm in men and 50 mm in women. Data represent the mean from the 15 validation folds, i.e., the cases not used to train the models. Although ML models trained on the entire group showed higher

13

accuracy and specificity, model training in the ≤ 70 mm group resulted in a major improvement in sensitivity.

The LogR and SVM-Lin models emerged as the two best performing ML methods. Figure 3 illustrates the corresponding rupture probabilities in the ≤ 70 mm group.

Lowering the probability threshold for rupture, the best trained LogR model (trained on the ≤ 70 mm group) increased its sensitivity at the expense of a moderate decrease in specificity (Fig. 4B). At a rupture probability of 35%, the model displayed 96% sensitivity and 75% specificity in the ≤ 70 mm group. In comparison, Figure 4A provides a similar analysis for the diameter threshold, where the threshold in females was constantly 5 mm lower than in males. The performance of the other ML models is illustrated in Supplementary Figure S2. Models trained on the entire group displayed similar properties, with the sensitivity, however, not reaching these levels (Supplementary Figure S3).

The importance of individual risk factors

Figure 5 summarises the SHAP analysis and indicates the relative importance of individual features (rupture factors). No factor dominated, indicating that the ML models used all input information for prediction. The MWRI, the external diameter, and the RRED were key in predicting AAA rupture in the best LogR and SVM-Lin models. Supplementary Figure S4 shows the SHAP variable importance averaged over all models.

14

DISCUSSION

The risk of AAA rupture depends on multiple factors, making it difficult to predict surgery timing. Various ML models, integrating geometric, biomechanical, and basic patient characteristics into a single holistic factor, were tested. The approach analysed data from patients having either intact or already ruptured AAA. It clearly outperformed the diameter threshold in classifying intact and ruptured cases and delivered promising results for the assessment of rupture risk.

As the patient group in this study covered an extensive diameter range and ruptured AAAs are commonly larger, the diameter is *a priori* a sensitive classifier. Focusing on the clinically more interesting group with diameter ≤ 70 mm, the sensitivity of the diameter threshold and the ML models diminished considerably (Table 4).

Compared with the 55/50 mm maximum diameter, the LogR model correctly identified twice as many intact AAAs while at the same time maintaining the ability of the diameter thresholds to identify ruptures. As such, it resulted in 96% sensitivity but in a specificity of 75% against 38% of the diameter based method (Fig. 4).

The SHAP algorithm demonstrated that no single feature holds all the information, but several parameters are needed for accurate classification (Fig. 5; Supplementary Figure S4). Furthermore, ML models other than the GNB consistently ranked the most relevant features.

These models could not provide a 100% correct rupture classification, likely because of incomplete input information. Albeit probably important, patient data, such as

height, body surface area, body mass index, and comorbidities, were not available in this study. Unfortunately, the biochemical and histological status of the AAA wall were also unavailable, but both might have influenced wall strength and contributed to AAA rupture. Future studies should incorporate the aforementioned information, although larger patient groups would be needed for training, which is challenging given the scarcity of ruptures. Diameter matched subgroups have also not been considered, as including them would reduce the sample size too much, disregarding 63% of intact cases in the ≤ 70 mm group.

The current study design focused on larger cases to recruit a sufficiently large number of ruptured cases. Even the ≤ 70 mm group included large cases, rendering it less applicable to the clinically relevant population with diameters up to 55/50 mm.

Small datasets can lead to inaccurate predictions in unseen data. Regardless of this risk, all the data were enrolled for training and validation, and further studies are necessary to test the models. External validation from different hospitals and countries should also be evaluated. Furthermore, supervised learning requires manually labelled datasets, reflecting human errors and biases.

This study compared intact with already ruptured cases. As most non-ruptured cases underwent AAA repair after CTA examination, information on whether they would have ruptured if left untreated is lacking, limiting the inference of the results. Additionally, recruiting patients over the same time period led to more non-ruptured than ruptured cases, forcing many intact cases to be disregarded. Also, not all intact AAAs from 2010 – 2021 were screened, where ruptured cases were recruited, which

16

would have been necessary for a cross-sectional study. Regardless, the patient groups had many similarities, but hidden factors may have impacted the analysis. A recently reported study[32] overcame the aforementioned drawbacks, but screening 1 219 patients resulted in only 15 pairs of ruptured and non-ruptured cases, data that are clearly insufficient to train ML models.

The CTA scan analysis required the reconstruction and estimation of the pre-rupture AAA geometry. Excluding cases with large rupture sites, the segmentation algorithm reasonably "closed" the rupture, but it might have been influenced by the user's indications. As the pre-rupture anatomy is missing, the associated error remains unknown. In addition, large slice thickness (> 2 mm) complicated the segmentation and may require more manual interaction, but the results (geometric and biomechanical data) remain relatively unaffected by this input inconsistency. Finally, the study design neglects time to rupture, a possibly important parameter in the risk assessment. A cross-sectional evaluation and an alternative retrospective comparison, where all cases are intact at baseline and then followed up,[33] could have addressed these drawbacks.

Other ML classifiers, e.g., decision trees,[34] K-nearest neighbours (KNN),[35] random forests, and XGBoost,[36] demonstrated strong capabilities in disorder identification. ML algorithms, including LogR, SVM, KNN, decision trees, and linear discriminant, predicted the aortic aneurysm growth rate based on geometric features. Additional classifiers could give further insights into AAA rupture.[17] In recent years, deep learning and neural network (NN) models have become increasingly popular in data driven modelling.[37] Among these, convolutional NN and recurrent NN approaches

have attracted interest in medical applications.[38,39] As recurrent NNs aim to predict future events based on the patient's history, they require a plethora of longitudinal data, limiting their clinical use. Yet a convolutional NN approach integrating multiphysical features enhanced the AAA growth prediction by analysing 54 patients.[40] Since the current study involved already ruptured cases with a wide range of biomechanical and geometric characteristics, large datasets[41] would be necessary for the NN development. The data groups in this study were too small for accurate predictions, and NN models were not further considered.

Limitations

Despite the encouraging results, this study has several limitations. First, the retrospective design included already ruptured cases, representing a severe drawback. A cross-sectional study evaluating rupture risk in follow ups would have provided a more robust basis for assessing predictive performance. Second, the relatively small cohort limits the findings, underscoring the need for validation in larger datasets. More data would also allow considering additional potential risk factors, with genetic and proteomics being interesting inputs. Finally, the developed models require further testing on previously unseen data to evaluate their robustness.

Conclusion

This study found that a multiparameter estimate significantly enhanced purely diameter based decision making for AAA rupture identification. It therefore suggests that risk factors other than the diameter may be equally important. The predictability

of the proposed ML approach should be further tested in longitudinal studies to
scrutinise the presented approach.

CONFLICTS OF INTEREST

T.C.G. is a scientific advisor and shareholder of VASCOPS GmbH (Graz, Austria)
and Artec Diagnosis AB (Täby, Sweden). The authors declare no other conflict of
interest.

FUNDING

REFERENCES

1   McGregor JC, Pollock JG, Anton HC. The value of ultrasonography in the diagnosis of abdominal aortic aneurysm. *Scott Med J* 1975;**20**:133–7.

2   Wanhainen A, Van Herzeele I, Bastos Goncalves F, Bellmunt Montoya S, Berard X, Boyle JR, et al. Editor's Choice – European Society for Vascular Surgery (ESVS) 2024 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. *Eur J Vasc Endovasc Surg* 2024;**67**:192–331.

3   Sakalihasan N, Michel JB, Katsargyris A, Kuivaniemi H, Defraigne JO, Nchimi A, et al. Abdominal aortic aneurysms. *Nat Rev Dis Primers* 2018;**4**:34.

4   Sweeting MJ, Thompson SG, Brown LC, Powell JT, RESCAN Collaborators. Meta-analysis of individual patient data to examine factors affecting growth and rupture of small abdominal aortic aneurysms. *Br J Surg* 2012;**99**:655–65.

5   Chaikof EL, Dalman RL, Eskandari MK, Jackson BM, Lee WA, Mansour MA, et al. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J Vasc Surg* 2018;**67**:2–77.

6   Siika A, Lindquist Liljeqvist M, Zommorodi S, Nilsson O, Andersson P, Gasser TC, et al. A large proportion of patients with small ruptured abdominal aortic aneurysms are women and have chronic obstructive pulmonary disease. *PLoS One* 2012;**14**:e0216558.

7   Chung TK, da Silva ES, Raghavan SML. Does elevated wall tension cause aortic aneurysm rupture? Investigation using a subject-specific heterogeneous model. *J Biomech* 2017;**64**:164–71.

8 Doyle BJ, Callanan A, Burke PE, Grace PA, Walsh MT, Vorp DA, et al. Vessel asymmetry as an additional diagnostic tool in the assessment of abdominal aortic aneurysms. *J Vasc Surg* 2009;**49**:443–54.

9 Erhart P, Hyhlik-Dürr A, Geisbüsch P, Kotelis D, Müller-Eschner M, Gasser TC, et al. Finite element analysis in asymptomatic, symptomatic, and ruptured abdominal aortic aneurysms: in search of new rupture risk predictors. *Eur J Vasc Endovasc Surg* 2015;**49**:239–45.

10 Martufi G, Lindquist Liljeqvist M, Sakalihasan N, Panuccio G, Hultgren R, Roy J, et a. Local diameter, wall stress, and thrombus thickness influence the local growth of abdominal aortic aneurysms. *J Endovasc Ther* 2016;**23**:957–66.

11 Siika A, Talvitie M, Gasser TC, Hultgren R, Roy J. Peak wall stress and peak wall rupture index are associated with time to rupture in abdominal aortic aneurysms. *Aorta* 2022;**10**:A1–56.

12 Kazi M, Thyberg J, Religa P, Roy J, Eriksson P, Hedin U, et al. Influence of intraluminal thrombus on structural and cellular composition of abdominal aortic aneurysm wall. *J Vasc Surg* 2003;**38**:1283–92.

13 Svensjö S, Björck M, Wanhainen A. Update on screening for abdominal aortic aneurysm: a topical review. *Eur J Vasc Endovasc Surg* 2014;**48**:659–67.

14 Gasser TC. Biomechanical rupture risk assessment. *Aorta* 2016;**4**:42–60.

15 Roy J, Labruto F, Beckman MO, Danielson J, Johansson G, Swedenborg J. Bleeding into the intraluminal thrombus in abdominal aortic aneurysms is associated with rupture. *J Vasc Surg* 2008;**48**:1108–13.

16 Abdolmanafi A, Forneris A, Moore RD, Di Martino ES. Deep-learning method for fully automatic segmentation of the abdominal aortic aneurysm from computed tomography imaging. *Front Cardiovasc Med* 2023;**9**:1040053.

17 Geronzi L, Haigron P, Martinez A, Yan K, Rochette M, Bel-Brunon A, et al. Assessment of shape-based features ability to predict the ascending aortic aneurysm growth. *Front Physiol* 2023;**14**:1125931.

18 Gasser TC, Auer M, Labruto F, Swedenborg J, Roy J. Biomechanical rupture risk assessment of abdominal aortic aneurysms: model complexity versus predictability of finite element simulations. *Eur J Vasc Endovasc Surg* 2010;**40**:176–85.

19 Raghavan ML, Vorp DA. Toward a biomechanical tool to evaluate rupture potential of abdominal aortic aneurysm: identification of a finite strain constitutive model and evaluation of its applicability. *J Biomech* 2000;**33**:475–82.

20 Gasser TC, Görgülü G, Folkesson M, Swedenborg J. Failure properties of intraluminal thrombus in abdominal aortic aneurysm under static and pulsating mechanical loads. *J Vasc Surg* 2008;**48**:179–88.

21 Vande Geest JP, Wang DH, Wisniewski SR, Makaroun MS, Vorp DA. Towards a noninvasive method for determination of patient-specific wall strength distribution in abdominal aortic aneurysms. *Ann Biomed Eng* 2006;**34**:1098–106.

22 Gasser TC. *Vascular Biomechanics. Concepts, Models, and Applications*. Cham, Switzerland: Springer, 2021.

23 Gasser TC, Nchimi A, Swedenborg J, Roy J, Sakalihasan N, Böckler D, et al. A novel strategy to translate the biomechanical rupture risk of abdominal aortic aneurysms to their equivalent diameter risk: method and retrospective validation. *Eur J Vasc Endovasc Surg* 2014;**47**:288–95.

24 Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;**16**:321–57.

25 Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform* 2002;**35**:352–9.

26 Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;**20**:273–97.

27 Domingos P, Pazzani M. On the optimality of the simple Bayesian classifier under zero-one loss. *Mach Learn* 1997;**29**:103–30.

28 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thiron B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;**12**:2825–30.

29 James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: With Applications in R*. New York, NY: Springer, 2014.

30 Platt JC. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: *Advances in Large Margin Classifiers* (Smola AJ, Bartlett, PL, Schölkopf B, Schuurmans D, eds). Cambridge, MA: MIT Press, 1999:61–74.

31 Shapley L. Value for n-person games. In: *Contributions to the Theory of Games*, Volume II (Kuhn H, Tucker A, eds). Princeton, NJ: Princeton University Press, 1953:307–17.

32 Grassl K, Gasser TC, Enzmann FK, Gratl A, Klocker J, Wippel D, et al. Early prediction of abdominal aortic aneurysm rupture risk using numerical biomechanical analysis. *Diagnostics (Basel)* 2024;**15**:25.

33 Siika A, Talvitie M, Lindquist Liljeqvist M, Bogdanovic M, Gasser TC, Hultgren R, et al. Peak wall rupture index is associated with risk of rupture of abdominal aortic aneurysms, independent of size and sex. *Br J Surg* 2024;**111**:znae125.

34 Palacios-Ortega M, Guerra-Galán T, Jiménez-Huete A, García-Aznar JM, Pérez-Guzmán M, Mansilla-Ruiz MD, et al. Dissecting secondary immunodeficiency: identification of primary immunodeficiency within B-cell lymphoproliferative disorders. *J Clin Immunol* 2025;**45**:32.

35 Madura Meenakshi R, Padmapriya N, Venkateswaran N, Shperling S, Leshno A. Bispectrum analysis of thermal images for the classification of retinal vascular diseases. *Biomed Signal Process Control* 2025;**99**:106878.

36 Li X, Tian Y, Li S, Wu H, Wang T. Interpretable prediction of 30-day mortality in patients with acute pancreatitis based on machine learning and SHAP. *BMC Med Inform Decis Mak* 2024;**24**:328.

37 Stonko DP, Hicks CW. Mature artificial intelligence- and machine learning-enabled medical tools impacting vascular surgical care: a scoping review of late-stage, US Food and Drug Administration-approved or cleared technologies relevant to vascular surgeons. *Semin Vasc Surg* 2023;**36**:460–70.

38 Tomihama RT, Dass S, Chen S, Kiang SC. Machine learning and image analysis in vascular surgery. *Semin Vasc Surg* 2023;**36**:413–18.

39 Su P, Ding XR, Zhang YT, Liu J, Miao F, Zhao N. Long-term blood pressure prediction with deep recurrent neural networks. *IEEE EMBS Int Conf Biomed Health Inform* 2018; doi: 10.48550/arXiv.1705.04524.

40 Kim S, Jiang Z, Zambrano BA, Jang Y, Baek S, Yoo S, et al. Deep learning on multiphysical features and hemodynamic modeling for abdominal aortic aneurysm growth prediction. *IEEE Trans Med Imaging* 2023;**42**:196–208.

41 Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol* 2017;**69**:2657–64.

Figure 1. Reconstructed and biomechanically analysed abdominal aortic aneurysm cases. The image shows the colour coded von Mises stress of an intact female case of 55.6 mm in diameter (left) and a ruptured male case of 70.5 mm in diameter (right).

Figure 2. Performance of geometric and biomechanical features to assess the potential of abdominal aortic aneurysm (AAA) rupture. The area under the receiver operator characteristics (ROC) curve (AUROC; given in %) quantifies the classification accuracy in (A) the entire group and (B) the group of AAAs below 70 mm in diameter. True and false positive rates express the probability of correct and wrong classification of ruptured AAAs, respectively. DE = maximum external abdominal aortic aneurysm diameter; DL = maximum luminal diameter; TT = intraluminal thrombus thickness; VV = vessel volume; VL = lumen volume; VT = intraluminal thrombus volume; RRED = rupture risk equivalent diameter; PWS = peak wall stress; MWS = mean wall stress; PWRI = peak wall rupture risk index; MWRI = mean wall rupture risk index; PTS = peak intraluminal thrombus stress; MTS = mean intraluminal thrombus stress; PTRI = peak intraluminal thrombus rupture risk; MTRI = mean intraluminal thrombus rupture risk.

Figure 3. Rupture probability *vs.* maximum external aortic abdominal aneurysm (AAA) diameter in the $\leq$ 70 mm group according to (A) logistic regression (LogR) and (B) calibrated linear support vector machine (SVM-Lin) models.

25

Figure 4. Performance of the classification at varying threshold levels in the ≤ 70 mm diameter abdominal aortic aneurysm (AAA) group: (A) diameter threshold in females/males; and (B) logistic regression (LogR) model.

Figure 5. SHapley Additive exPlanations (SHAP) ranking of feature importance in the subgroup of the best trained models: (A) logistic regression (LogR) model; and (B) support vector machine with linear function (SVM-Lin) model. MWRI = mean wall rupture risk index; DE = maximum external abdominal aortic aneurysm diameter; RRED = rupture risk equivalent diameter; VV = vessel volume; DL = maximum luminal diameter; MTS = mean intraluminal thrombus stress; PWS = peak wall stress; MTRI = mean intraluminal thrombus rupture risk; PTRI = peak intraluminal thrombus rupture risk.

26

Table 1. Variables considered ("yes") and neglected ("no") in the classification of the entire group and the ≤ 70 mm maximum external abdominal aortic aneurysm diameter group.

Variable	Entire group	≤ 70 mm group
Age	Yes	Yes
Sex (female/male)	Yes	Yes
Abdominal aortic aneurysm diameter	Yes	Yes
Smoking status	No	No
Maximum luminal diameter	No	Yes
Lumen volume	No	No
Vessel volume	No	Yes
Intraluminal thrombus volume	No	No
Intraluminal thrombus thickness	Yes	No
Rupture risk equivalent diameter	Yes	Yes
Peak wall stress	No	Yes
Mean wall stress	Yes	No
Peak wall rupture risk index	No	No
Mean wall rupture risk index	No	Yes
Peak intraluminal thrombus stress	No	No
Mean intraluminal thrombus stress	Yes	Yes
Peak intraluminal thrombus rupture risk	Yes	Yes
Mean intraluminal thrombus rupture risk	Yes	Yes

Table 2. Features representing the analysed ruptured and intact aortic abdominal aneurysms (AAAs) in the entire group and the corresponding statistical differences.

Feature	Ruptured AAA (*n* = 108)	Intact AAA (*n* = 200)	*p* value [*]
Sex, female	20.56	21.50	.90
Age – y	76.70 ± 9.6	73.99 ± 8.3	.008
Smoking status			.25
Smoker	32	1	
Non-smoker	42	95	
Maximum AAA external diameter – mm	86.37 ± 20.8	56.58 ±7.5	<.001
Maximum luminal diameter – mm	62.44 ± 19.2	40.19 ± 8.7	<.001
Maximum ILT thickness – mm	25.54 ± 15.6	17.82 ± 9.1	<.001
Vessel volume – mL	405.57 ± 232	158.50 ± 57.4	<.001
Lumen volume –mL	201.91 ± 158.5	75.23 ± 33.5	<.001
ILT volume – mL	167.45 ± 135.7	60.92 ± 41.1	<.001
Peak wall stress – kPa	312.02 ± 100.1	202.55 ± 46.6	<.001
Mean wall stress – kPa	147.16 ± 45.4	105.77 ± 20.1	<.001
Peak wall rupture risk index	1.04 ± 0.6	0.42 ± 0.1	<.001
Rupture risk equivalent diameter – mm	99.79 ± 45.4	49.43 ± 12.3	<.001
Mean wall rupture risk index	0.50 ± 0.3	0.22 ± 0.1	<.001
Peak ILT stress – kPa	41.01 ± 28.6	26.67 ± 15.3	<.001
Mean ILT stress – kPa	8.75 ± 2.3	7.05 ± 1.2	<.001
Peak ILT rupture risk	0.70 ± 0.5	0.44 ± 0.2	<.001
Mean ILT rupture risk	0.17 ± 0.1	0.11 ± 0.0	<.001

Data are presented as % or mean ± standard deviation. AAA = abdominal aortic aneurysm, ILT = intraluminal thrombus.

[*] Non-parametric Mann–Whitney *U* test.

Table 3. Features representing the analysed ruptured and intact aortic abdominal aneurysms (AAAs) in the group of cases below 70 mm in maximum diameter and corresponding statistical analysis.

Feature	Ruptured AAA ($n =$ 25)	Intact AAA ($n =$ 192)	p value
Sex, female	32	22.40	.44[*]
Age – y	74.92 ± 8.4	73.77 ± 8.2	.50[*]
Smoking status			.26[*]
Smoker	32	1	
Non-smoker	40	95	
Maximum AAA external diameter – mm	63.51 ± 4.8	55.75 ± 6.2	<.001[*]
Maximum luminal diameter – mm	48.38 ± 10.8	39.67 ± 8.2	<.001[†]
Maximum ILT thickness – mm	14.84 ± 9.5	17.23 ± 8.7	.23[*]
Vessel volume – mL	178.61 ± 46.7	153.08 ± 47	.015[*]
Lumen volume – mL	98.14 ± 46.7	73.65 ± 31.9	.014[*]
ILT volume – mL	57.22 ± 42.1	57.58 ± 36.5	.81[*]
Peak wall stress –kPa	246.10 ± 63.5	200.32 ± 45.7	<.001[*]
Mean wall stress – kPa	120.95 ± 28.1	105.17 ± 20.1	.011[†]
Peak wall rupture risk index	0.65 ± 0.2	0.41 ± 0.1	<.001[*]
Rupture risk equivalent diameter – mm	69.17 ± 18.5	48.76 ± 11.8	<.001[*]
Mean wall risk rupture index	0.33 ± 0.1	0.22 ± 0.1	<.001[*]
Peak ILT stress – kPa	37.10 ± 40.9	26.30 ± 15.2	.066[*]
Mean ILT stress – kPa	8.80 ± 4.0	7.02 ± 1.2	<.001[*]
Peak ILT rupture risk	0.61 ± 0.7	0.43 ± 0.2	.043[*]
Mean ILT rupture risk	0.14 ± 0.1	0.11 ± 0.0	.002[*]

Data are presented as % or mean ± standard deviation. AAA = abdominal aortic aneurysm; ILT = intraluminal thrombus.

[*] Non-parametric Mann–Whitney *U* test (non-normal data).

[†] Student's *t* test (normal data).

Table 4. Performance of machine learning (ML) models and diameter thresholds to classify ruptured and intact abdominal aortic aneurysms (AAAs). ML models were trained either using the entire group or the group of cases with external diameter ≤ 70 mm, while performance data were acquired from the ≤ 70 mm AAA diameter group that had not been used to train the respective model.

Performance	Results from validation on the ≤ 70 mm group				
	LogR	SVM-Lin	SVM-Nlin	GNB	55/50 mm DE
Accuracy					
Entire group	90.2	89.8	87.7	86.4	
≤ 70 mm group	82.5	83.6	65.9	84.7	44.2
Specificity					
Entire group	96.1	95.8	94.9	93.4	
≤ 70 mm group	83.9	85.2	68.7	90	37.7
Sensitivity					
Entire group	38.9	40.3	26.2	28.5	
≤ 70 mm group	70.9	70.3	64.3	41.4	96.7
AUROC					
Entire group	83.9	84.3	83.1	74.1	
≤ 70 mm group	83.6	84.2	87.3	81.3	67.2

Data are presented as %. LogR = logistic regression model; SVM-Lin = support vector machine model with linear kernel; SVM-Nlin = support vector machine model with non-linear kernel; GBN = Gaussian Naïve Bayes model; 55/50 mm DE = maximum external abdominal aortic aneurysm diameter threshold criterion of 55 mm in men and 50 mm in women; AUROC = area under the receiver operating characteristic curve.