Individual risk assessment for rupture of abdominal aortic aneurysm using artificial intelligence

Skovbo, Joachim Sejr; Andersen, Nicklas Sindlev; Obel, Lasse Møllegaard; Laursen, Malene Skaarup; Riis, Andreas Stoklund; Houlind, Kim Christian; Pyndt Diederichsen, Axel Cosmus; Lindholt, Jes Sanddal

Published in:
Journal of Vascular Surgery

DOI:
10.1016/j.jvs.2024.11.017

Publication date:
2025

Document version:
Final published version

Individual risk assessment for rupture of abdominal aortic aneurysm using artificial intelligence

Joachim Sejr Skovbo, MD,[a,b] Nicklas Sindlev Andersen, PhD,[c] Lasse Møllegaard Obel, MD,[a,b] Malene Skaarup Laursen, MD,[a] Andreas Stoklund Riis, MD,[a] Kim Christian Houlind, PhD, MD,[d,e] Axel Cosmus Pyndt Diederichsen, PhD, MD,[f] *and* Jes Sanddal Lindholt, PhD, MD,[a,b]
Odense and Kolding, Denmark

ABSTRACT

Objective: This study aimed to develop a prediction tool to identify abdominal aortic aneurysms (AAAs) at increased risk of rupture incorporating demographic, clinical, imaging, and medication data using artificial intelligence (AI).

Design: A development and validation study for individual prognosis using AI in a case-control design.

Methods: From two Danish hospitals, all available ruptured AAA cases between January 2009 and December 2016 were included in a ratio of 1:2 with elective surgery controls. Cases with previous AAA surgery or missing preoperative scans were excluded. Features from computed tomography angiography scans and hospital records were manually retrieved. The sample was divided randomly and evenly into developmental and internal validation groups. A SHapley Additive exPlanations Feature Importance Rank Ensembling (SHAPFire) AI tool was developed using a gradient boosting decision tree framework. The final SHAPFire AI model was compared with models using (1) solely infrarenal anterior-posterior diameter and (2) all available features.

Results: The study included 637 individuals (84.8% men, mean age 73 ± 7 years, 213 ruptured AAAs). The SHAPFire AI incorporated 20 of 68 available features, and aneurysm size, blood pressure, and relationships between height and weight were given highest rankings. The receiver operating characteristic curve for the SHAPFire AI model displayed a significant increase in accuracy identifying ruptured AAA cases compared with the conventional model based solely on diameter with areas under the curves of 0.86 ± 0.04 and 0.74 ± 0.03 ($P = .008$), respectively. SHAPFire AI was comparable in performance with the model using all features.

Conclusions: This study successfully developed a SHAPFire AI tool to identify AAAs at increased risk of rupture with significantly higher accuracy than diameter alone. External validation of the model is warranted before clinical implementation. (J Vasc Surg 2025;81:613-22.)

Keywords: Abdominal aortic aneurysm; Artificial Intelligence; Risk assessment; Ruptured abdominal aortic aneurysm; Case-control study

The optimal timing for elective surgery to prevent future rupture in abdominal aortic aneurysms (AAAs) remains a subject of ongoing debate, necessitating consideration of the balance between the risks of rupture (rAAA) and the potential complications associated with surgery. The current optimal medical practice is to recommend elective repair when the AAA diameter is equal to or exceeds 5.5 cm for men and 5.0 for women.[1] Although these guidelines are straightforward, exploring additional factors in decision-making could potentially improve the precision of identifying suitable candidates for surgery as the number needed to treat to prevent a single death caused by rupture is 2.[2] Such approach may lead to a more cost-effective determination of both whom to operate on and the optimal timing for intervention, especially as the current cutoff diameters still leave around a tenth of ruptures occurring in smaller aneurysms.[3] In a Finnish study of rAAA, it was found that

From the Department of Cardiac, Thoracic and Vascular Surgery, Odense University Hospital,[a] the Department of Clinical Research,[b] and Department of Mathematics and Computer Science,[c] University of Southern Denmark, Odense; the Department of Vascular Surgery, Lillebaelt Hospital, Kolding[d]; the Department of Regional Health Research, University of Southern Denmark,[e] and the Department of Cardiology, Odense University Hospital,[f] Odense.

Additional material for this article may be found online at www.jvascsurg.org.

Correspondence: Joachim Sejr Skovbo, MD, Department of Cardiology, Odense University Hospital, J. B. Winsløws Vej 4, Odense 5000, Denmark (e-mail: Joachim.sejr.s.kristensen@rsyd.dk).

The editors and reviewers of this article have no relevant financial relationships to disclose per the JVS policy that requires reviewers to decline review of any manuscript for which they may have a conflict of interest.

0741-5214

https://doi.org/10.1016/j.jvs.2024.11.017

5.6% of men had a rupture at <55 mm and 11.5% of women at <52 mm.[4]

In all, the use of elective surgery to prevent rAAA could be improved with better tools for predicting rupture. Applying machine learning (ML) to identify risk factors and develop prediction tools could aid in individualizing the treatment with expected significant personal and socioeconomic benefits. Although studies on ML in the vascular field are still in their infancy, ML has already been applied in the field of AAA in various ways, such as predicting outcomes after AAA repair[5] and diagnosing AAA on a computed tomography angiography (CTA) scan.[6]

One limitation of ML is the difficulty in understanding which elements it uses to arrive at its final result. To address this issue, tools like SHapley Additive exPlanations Feature Importance Rank Ensembling (SHAPFire) have been developed. SHAPFire helps identify correlated features and rank them according to their predictive value, providing more transparency into the model's decision-making process.[7] Therefore, we hypothesized that using a SHAPFire ML approach incorporating various features would be superior in identifying cases of rAAA, compared with the conventional diameter-based approach.

The aim of this study was to develop a precision-decision tool for rAAA using artificial intelligence (AI) that analyzes variables from demographic, CTA imaging, medical, and clinical data in a subset of our study population. We then internally validated the SHAPFire AI model and compared its performance with using only maximum abdominal aortic anterior-posterior diameter and all available features.

ARTICLE HIGHLIGHTS

- **Type of Research:** Multicenter, retrospective analysis of prospectively collected registry data, case-control study
- **Key Findings:** A total of 637 patients who underwent either elective or acute surgery were included in the study. Using a model based on machine learning and 20 selected features derived from anamnestic information and computer tomography measurements was more accurate in identifying cases of ruptures than using solely the anterior-posterior diameter and equally effective as using all available features.
- **Take Home Message:** We developed and internally validated a machine learning algorithm, which successfully identified cases with good accuracy using just 20 features.

METHODS

This study is reported after the AI-expanded transparent reporting of a multivariable prediction model for individual prognosis or diagnosis.[8,9]

Study design. The study followed a case-control design.

Study population. The study population, including both cases and controls, was sourced from the Danish Vascular Registry.[10,11] Cases included all patients registered in the Danish Vascular Registry who underwent surgery for rAAA at the two vascular sites Odense University Hospital and Kolding Hospital in Denmark from January 2009 to December 2016. All registered cases were manually reviewed, and those with incorrect classification of rAAA, previous AAA surgery, residency outside the Southern Denmark region, or missing preoperative scans were excluded. For each rAAA case, two control patients were matched on year of surgery from the Danish Vascular Registry. These controls underwent elective surgery for AAA and met the same exclusion criteria. In addition, they were asymptomatic AAAs; that is, no lower back pain, pulsating abdominal mass, or abdominal tenderness was present. No prior sample size calculation was performed. The choice to match two controls per case was driven by the goal of maximizing statistical power. However, the case-to-control ratio was constrained by the finite pool of eligible patients, preventing us from expanding the control group further. No other matching criteria were applied aside from the time period.

Sources of data. The study population was identified using the Danish Vascular Registry.[10] The Danish Vascular Registry is a prospective nationwide registry monitoring all vascular procedures in Denmark with the main aim of improving the quality of treatment and enabling vascular research. The Danish Vascular Registry was internationally validated in 2019 by the Vascunet group and found to have high-quality data, especially regarding endpoints, being superior to local administrative data.[11]

Cases and controls were identified using data from the Danish Vascular Registry including type of surgery, date of surgery, sex, weight, height, and smoking status. Our main outcome was whether the patient was in the ruptured AAA or elective AAA group. This was measured by clinicians recording the data in the central database of the Danish Vascular Registry and, as a further validation, through the review of medical records and CTA scans. From medical records, we retrospectively isolated information regarding comorbidities, blood pressure before surgery and before rupture, medical use at the date of surgery, and blood tests (if available preoperatively, otherwise, 3-24 months after surgery). The preoperative CTA scan was taken within 24 hours for the acute cases and within 1 month for the elective controls as per conventional clinical practice in the region of Southern Denmark.

The preoperative CTA scans were retrospectively evaluated by two trained pregraduate medical students, and measurements regarding the abdominal aneurysm were recorded. Supervised measurements of the aortic diameter were performed by a board-certified cardiologist in CT cardiac imaging in the first 50 cases for each reviewer. If all paired differences were below 2 mm, unsupervised examination was allowed. If discrepancies between students (above 2 mm) after training and the difference could not be reconciled by the students, the diameter was decided by a senior examiner. All personnel received the same training in measurements to limit interobserver variability. All measurements were performed inner-to-inner edge as recommended in guidelines.[1]

The extensive list of all variables, definitions, and origin can be seen in Supplementary Table I (online only). All data were collected and combined in a database using the REDCap system.[12]

Artificial intelligence. In the following, we adopt ML terminology and will refer to variables as features. In this context, SHAPFire is an automated, wrapper-based feature importance ranking and selection tool that leverages the principles of SHAPFire.[7] In simple terms, SHAPFire allows us to correlate features with each other and rank them according to the predictive value. We only used the training dataset to identify the features of importance. This separation ensured that the feature selection process did not influence the training data, reducing the risk of overfitting and yielding more reliable performance estimates. After employing SHAPFire, we used a gradient boosting decision tree (GBDT) as our ensemble model.[13] This method incorporates boosting, making it more complex than simpler models such as logistic regression or a single decision tree. GBDTs are recognized for their state-of-the-art performance on tabular datasets, including ours, and they excel even with smaller datasets. An added advantage is that GBDTs inherently handle missing values. Although GBDTs are less complex than deep learning models, using deep learning for a small dataset like ours was deemed excessive and likely would have resulted in overfitting. We trained multiple GBDT models using different feature subsets to extract feature importance values. A twofold cross-validation was particularly suitable for our smaller dataset, as more folds could result in too few data points for stable performance estimates. See Supplementary Methods (online only) and Supplementary Fig 1 (online only) for a more detailed description of the ML process applied in the present study.

Statistical methods. The study population was equally and randomly divided into two subgroups using stratified random sampling ensuring an equal proportion of cases to controls at a ratio of 1:2. This method ensured proportional representation of the outcome variable's classes (nonruptured and ruptured AAAs) in both groups. Therefore, the two subgroups were comparable in terms of setting, eligibility criteria, outcome, and predictors. SHAPFire used the first group to rank and select key features in the dataset. The second group was then employed to train and test a GBDT model using SHAPFire-selected features. This model's training and testing involved a repeated stratified twofold cross-validation loop to obtain an unbiased estimate of the receiver operating characteristic (ROC) curves with area under the curve (AUC) scores, that is, the dataset is repeatedly split into two stratified folds, ensuring that each class's proportion is consistent across both folds, and this process is repeated twice and each time the ROC AUC score is computed using the hold-out fold.

To evaluate our AI tool's predictive accuracy, we compared three different ROC curves derived from training and testing: (1) a GBDT model using SHAPFire-selected features, (2) an AI model that incorporates all available features without any exclusions, and (3) a model based solely on the maximal anterior-posterior abdominal aortic diameter. Across cross-validation folds, we assessed various metrics, including probability, mean accuracy, specificity, sensitivity, and Youden's index.

Baseline characteristics for the entire subpopulation stratified by subgroups are presented. Categorical variables are reported with frequency and percentages of the population. Continuous variables are presented as means with standard deviation if normally distributed, otherwise, median and interquartile ranges. The Wilcoxon rank-sum test and the Pearson χ^2 test were used to test for the difference between the groups for numerical and categorical variables, respectively. For comparing ROC AUC, we used a parametric paired *t*-test. The significance level was set to 5%. No imputation was used for missing data points, as advanced tree-based ML methods, including the gradient boosting machines used in SHAPFire, handle missing values independently. Missing data are reported for each variable.

Ethics. This study was conducted with approval from the National Board of Patient Safety (Styrelsen for Patientsikkerhed, No. 3-3013-1944-1) and the Data Protection Authorities (No. 16/13774).

Data and code sharing. Data cannot be shared in accordance with Danish law. However, the code is available on request.

RESULTS

Description of the population. In this study, we included 637 individuals (84.8% men, mean age 73 ± 7 years) who were randomly subdivided into a training population and a validation population of 318 and 319, respectively. For an overview of the inclusion of patients, see Fig 1. The 213 (33.4%) rAAA cases were evenly distributed among the subgroups. As seen in Table I, the training and validation populations were homogeneous

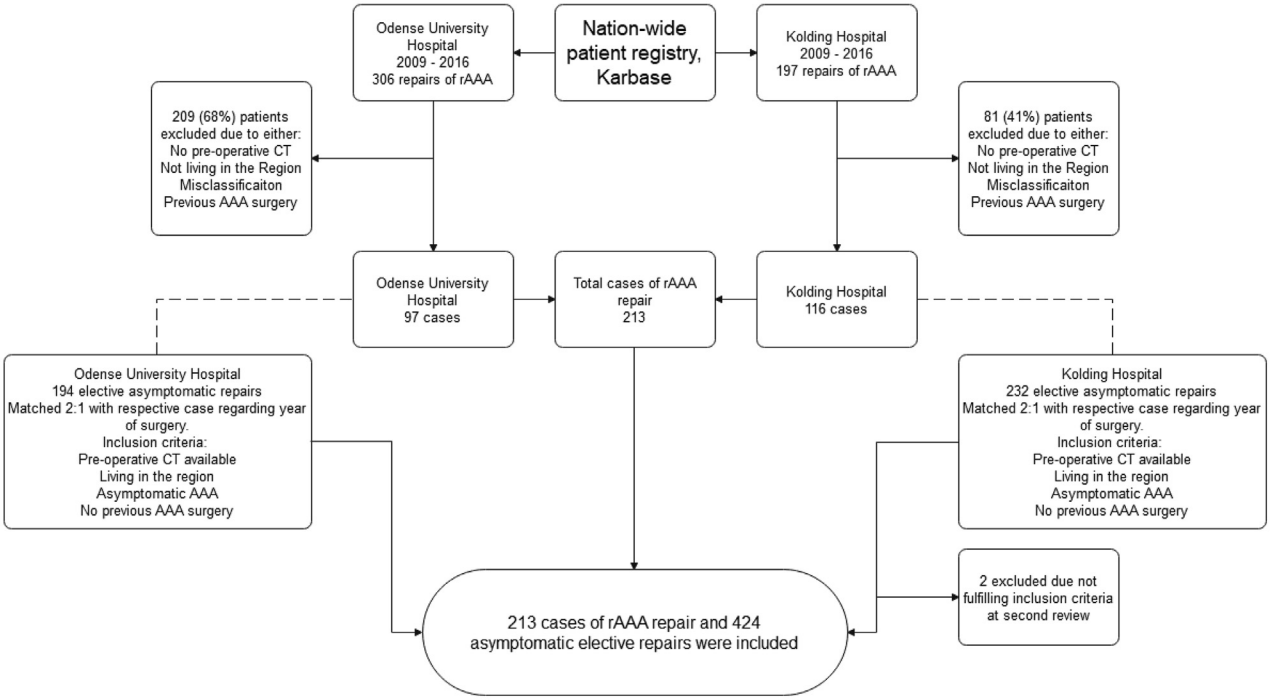


Fig 1. Flowchart of the inclusion of participants from two Danish sites to develop a machine learning (ML) tool for identifying ruptured abdominal aortic aneurysms (*rAAA*) in a case-control study. *CT*, Computed tomography.

regarding variables usually associated with AAA. A total overview of all variables available for AI can be seen in Supplementary Table I (online only).

Features of interest. The result of the SHAPFire's hierarchical ranking can be seen in Fig 2. From each cluster, the selected features used by the SHAPFire AI can be seen in teal in Table II. Twenty (29%) features were selected out of the 68 available features. The three clusters with the greatest affinity for the prediction of rupture were in descending order size of the aneurysm, blood pressure, and the relationship between weight and height. In addition, SHAPFire AI also included various measurements expressing aortic torsion and elongation, iliac size, measurements of calcification and intraluminal thrombus, statins, and antiplatelets. Hypertension as diagnosis, age, and smoking were selected, however, of lower rankings by SHAPFire. Sex was not selected as an important feature. Features associated with increased risk of rupture were maximal transverse diameter, pulse pressure, body surface area, maximal luminal diameter excluding the mural thrombus, anterior wall thickness, iliac artery diameters, distance from the renal artery to the aortic bifurcation (expressing elongation), distances from the iliac bifurcations to the aortic bifurcation, distance from the aortic bifurcation to os sacrum (expressing aortic torsion), age, and smoking, whereas AAA wall and thrombus calcification, use of statins and antithrombotic, and present AAA-neck thrombus among others decreased the risk of rupture (Table II)

Precision. The three ROC curves can be seen in Fig 3. The AUC was 0.86 (±0.04) for the SHAPFire AI model, whereas the one using all available features presented an AUC of 0.85 (±0.05) with no statistical difference ($P = .305$). The ROC curve using only the maximal anterior-posterior abdominal aortic diameter had an AUC of 0.74 (±0.03) and was significantly inferior to the SHAPFire AI ($P = .008$).

The different accuracies, sensitivities, and specificities for the three models can be seen in Supplementary Fig 2 (online only). The optimal sensitivities and specificities defined by maximal Youden's index can be seen in Supplementary Table II (online only).

DISCUSSION

In this case-control study, we developed and internally validated an ML algorithm for the identification of rAAA among controls of elective AAA surgery based on anamnestic and radiologic findings. The SHAPFire AI tool identified 20 features of importance. There was a significant gap in AUC and Youden's index in favor of the SHAPFire model compared with solely using the maximal abdominal aortic diameter, while being comparable to the model using all 68 features.

Interpretation. As deciphering AI-generated results poses challenges, we must exercise caution against excessive overanalyses, even though some of the findings proved intriguing. First, sex was not identified as an important feature, while it in general is recognized as

Table I. Baseline characteristics of the study population stratified by subgroup in an internal validation study using artificial intelligence (AI) to predict ruptured abdominal aortic aneurysm (rAAA)

Variables	Total (N = 637)	Training (n = 318)	Validation (n = 319)	*P* value
Age, years	73 (68-77) [0]	73 (68-77) [0]	73 (68-78) [0]	.4835
Male sex	540 (84.8) [0]	266 (83.6) [0]	274 (85.9) [0]	.4303
Smoking	[22]	[11]	[11]	.5413
Never	75 (11.8)	36 (11.3)	39 (12.2)	
Former	301 (47.3)	145 (45.6)	156 (48.9)	
Active	239 (37.5)	126 (39.6)	113 (35.4)	
Familiar disposition to abdominal aortic aneurysm	24 (3.8) [3]	15 (4.7) [0]	9 (2.8) [3]	.1743
BMI	27.1 (4.45) [76]	27.1 (4.48) [36]	27 (4.44) [40]	.3621
BSA (DuBois)	1.97 (0.212) [76]	1.96 (0.212) [36]	1.97 (0.212) [40]	.8844
Systolic blood pressure	144.5 (20.92) [81]	143.1 (20.7) [41]	146 (21.07) [40]	.1218
Diastolic blood pressure	82.7 (13.63) [81]	82.4 (13.51) [41]	83 (13.78) [40]	.9604
Medication				
Platelet inhibitors	373 (58.6) [13]	169 (53.1) [6]	191 (59.9) [7]	.0746
Anticoagulatory	64 (10) [12]	29 (9.1) [6]	35 (11) [6]	.4365
Statin	375 (58.9) [14]	184 (57.9) [7]	191 (59.9) [7]	.6005
NSAID	30 (4.7) [14]	16 (5) [6]	14 (4.4) [8]	.7149
β-Blockers	199 (31.2) [14]	98 (30.8) [7]	101 (31.7) [7]	.8178
Thiazide	150 (23.5) [14]	66 (20.8) [7]	70 (21.9) [7]	.7138
ACE/AT2	283 (44.4) [14]	150 (47.2) [7]	133 (41.7) [7]	.1602
Calcium antagonists	186 (29.2) [14]	91 (28.6) [7]	95 (29.8) [7]	.7459
Bronchodilator	88 (13.8) [13]	51 (16) [6]	37 (11.6) [7]	.1073
Comorbidities				
Hypertension	414 (65) [3]	212 (66.7) [1]	202 (63.3) [2]	.70
Diabetes	84 (13.2) [3]	48 (15.1) [1]	36 (11.3) [2]	.3716
Ischemic heart disease	186 (29.2) [1]	88 (27.7) [0]	98 (30.7) [1]	.3321
Former stroke	14 (2.2) [9]	9 (2.8) [4]	5 (1.6) [5]	.498
Chronic obstructive lung disease	111 (17.4) [2]	61 (19.2) [1]	50 (15.7) [1]	.2821
Former aortic dissection	3 (0.5) [4]	2 (0.60) [1]	1 (0.3) [3]	.6051
Autoimmune disease	29 (4.6) [9]	13 (4.1) [6]	16 (5) [3]	.8075
Maximal anterior-posterior diameter, cm	6.14 (5.48-7.39) [30]	6.16 (5.48-7.43) [14]	6.1 (5.48-7.34) [16]	.6491
Transverse diameter, cm	6.28 (5.51-7.65) [31]	6.32 (5.53-7.63) [15]	6.23 (5.46-7.75) [16]	.6823
Circumference, cm	19.66 (17.48-23.93) [31]	19.69 (17.48-23.93) [15]	19.61 (17.45-24.12) [16]	.7722
Ruptures	213 (33.4) [0]	106 (33.3) [0]	107 (33.5) [0]	.9554

ACE, Angiotensin-converting enzyme inhibitor; *AT2*, angiotensin II receptor antagonists; *BMI*, body mass index; *BSA*, body surface area; *NSAID*, nonsteroidal anti-inflammatory drug.
All dichotomous variables are reported with numbers and percentage of group population in parentheses. All continuous variables are reported with mean and standard deviation or median and interquartile range, dependent on the normality of data. For each variable, the number of missing data is reported in square brackets. To compare between groups, nonparametric testing was used; χ^2 for dichotomous and Wilcoxon rank-sum for continuous. All measurements of the abdominal aorta were inner-to-inner.

one of the most influential risk factors for AAAs and ruptures. One interpretation could be that sex is unimportant and strengthens the argument that the real correlation is between the ratio of the body size and the size of the aneurysm.[14-16] This argument is further supported by the fact that the most important feature was the transverse diameter of the AAA and the third most important feature was body surface area. An alternative perspective suggests that sex is inherently implied within the chosen features. Second, very few medications and comorbidities were selected. As recently published in a systematic review and meta-analysis by our research group, there is no compelling evidence for any known medication to halt the growth rate, risk of rupture, or risk of repair.[17] However, statins were one of two drugs to be a candidate for slowing growth, which the SHAPFire

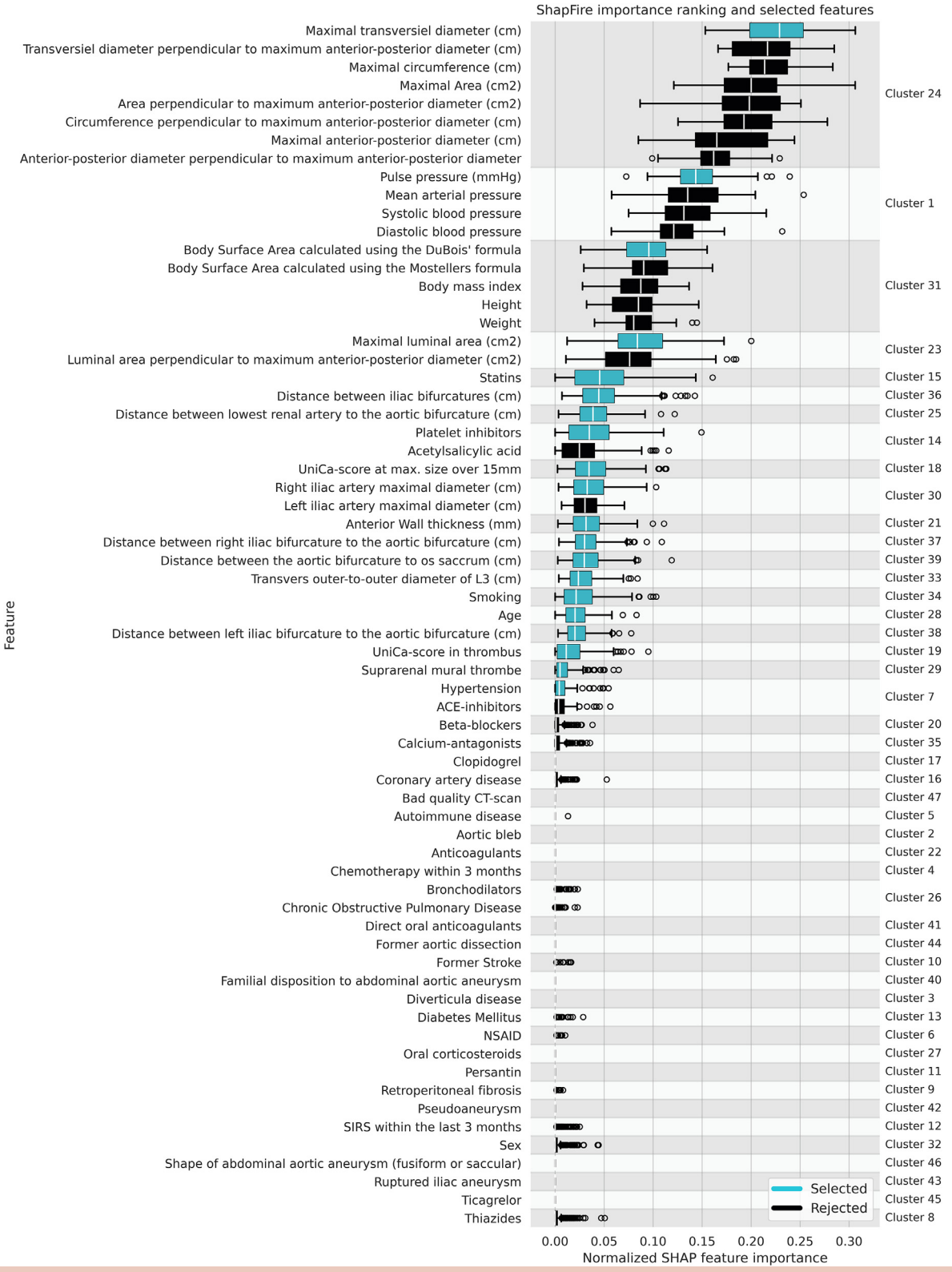


Fig 2. Ranking of each feature's boxplot stratified by clusters in an internal validation study of using artificial intelligence (AI) to predict ruptured abdominal aortic aneurysm (*rAAA*). Each cluster is identified to the right and by the shifting background color of gray and white. Each feature description is located to the left with a corresponding boxplot of ranked importance. For a comprehensive explanation of features, see Supplementary Table I (online only). The teal boxplots identify the SHapley Additive exPlanations Feature Importance Rank Ensembling (*SHAPFire*)-selected features (variables of most importance), whereas the black ones are rejected. *ACE,* Angiotensin-converting enzyme inhibitor; *CT,* computed tomography; *NSAID,* nonsteroidal anti-inflammatory drug.

Table II. Description of selected features for the SHapley Additive exPlanations Feature Importance Rank Ensembling (SHAPFire) artificial intelligence (AI) model stratified by elective or rupture status in an internal validation study on AI prediction of ruptured abdominal aortic aneurysm (rAAA)

Variables	Total (N = 637)	Elective (n = 424)	Ruptures (n = 213)	*P* value
Age, years	73 (68-77) [0]	73 (68-77) [0]	73 (68-78) [0]	.3537
Smoking	[22]	[4]	[18]	.0082
Never	75 (11.8)	43 (10.1)	32 (15)	
Former	301 (47.3)	222 (52.4)	79 (37.1)	
Active	239 (37.5)	155 (36.6)	84 (39.4)	
BSA (DuBois)	1.97 (0.212) [76]	1.96 (0.201) [11]	1.98 (0.238) [65]	.2188
Pulse pressure	61.82 (16.812) [81]	62.75 (15.756) [7]	59.06 (19.444) [74]	.0203
Medication				
Platelet inhibitors	373 (58.6) [13]	281 (66.3) [0]	79 (37.1) [13]	<.001
Statin	375 (58.9) [14]	292 (68.9) [0]	83 (39) [14]	<.001
Comorbidities				
Hypertension	414 (65) [3]	288 (67.9) [2]	126 (59.2) [1]	.0011
CT measurements				
Maximal transverse diameter, cm	6.73 (1.7) [31]	6.13 (1.2) [22]	7.9 (1.9) [9]	<.001
Luminal area, cm^2	15.77 (13.2) [96]	12.36 (8.4) [52]	23.27 (18) [44]	<.001
Distance between iliac bifurcatures, cm	7.31 (1.3) [69]	7.43 (1.3) [41]	7.07 (1.3) [28]	.0004
Distance between lowest renal artery to the aortic bifurcature, cm	12.72 (2.6) [71]	12.64 (2.8) [41]	12.89 (1.9) [30]	.0037
UniCa score at maximum size over 15 mm	175.03 (286.6) [50]	190.7 (309.4) [34]	144.02 (232.8) [16]	.0083
Right iliac artery maximal diameter, cm	18.59 (9.1) [112]	18.37 (8.4) [57]	19.12 (10.6) [55]	.7715
Anterior wall thickness, mm	1.17 (0.2) [42]	1.14 (0.2) [29]	1.23 (0.2) [13]	<.001
Distance between the right iliac bifurcature and the aortic bifurcature, cm	6.15 (1.4) [73]	6.14 (1.4) [41]	6.17 (1.4) [32]	.4039
Distance between the aortic bifurcature and os sacrum, cm	6.1 (1.6) [42]	5.89 (1.6) [29]	6.53 (1.7) [13]	<.001
Transverse outer-to-outer diameter of L3, cm	4.5 (0.4) [31]	4.48 (0.4) [24]	4.54 (0.5) [7]	.235
Distance between the left iliac bifurcature and the aortic bifurcature, cm	6.41 (1.6) [72]	6.4 (1.6) [42]	6.43 (1.5) [30]	.4905
UniCa score in thrombus	5.75 (29.2) [120]	6.73 (32.9) [73]	3.68 (18.9) [47]	.105
Suprarenal mural thrombus	367 (57.6) [60]	250 (59) [45]	117 (54.9) [15]	.1033

AI, Artificial intelligence; *BSA*, body surface area; *CT*, computed tomography; *UniCa-score*, calcification score.
All dichotomous variables are reported with numbers and percentage of group population in parentheses. All continuous variables are reported with mean and standard deviation or median and interquartile range, dependent on the normality of data. For each variable, the number of missing data is reported in square brackets. To compare between groups, nonparametric testing was used; χ^2 for dichotomous and Wilcoxon rank-sum for continuous. All measurements of the abdominal aorta were inner-to-inner.

finding could support. However, as platelet inhibitors were also determined as a predictor of rAAA, we propose that statin use along with platelet inhibitors and a hypertension diagnosis may serve as a proxy for individuals diagnosed with underlying vascular conditions.[17-19] Thus, the findings could potentially be interpreted as either underdiagnosis of vascular disease in the rupture cohort, which is certainly plausible, or that preventive treatments may indeed reduce the likelihood of rupture. Interestingly, diabetes was not identified as a feature of importance even though some research indicates a protective quality regarding the presence and expansion rate of

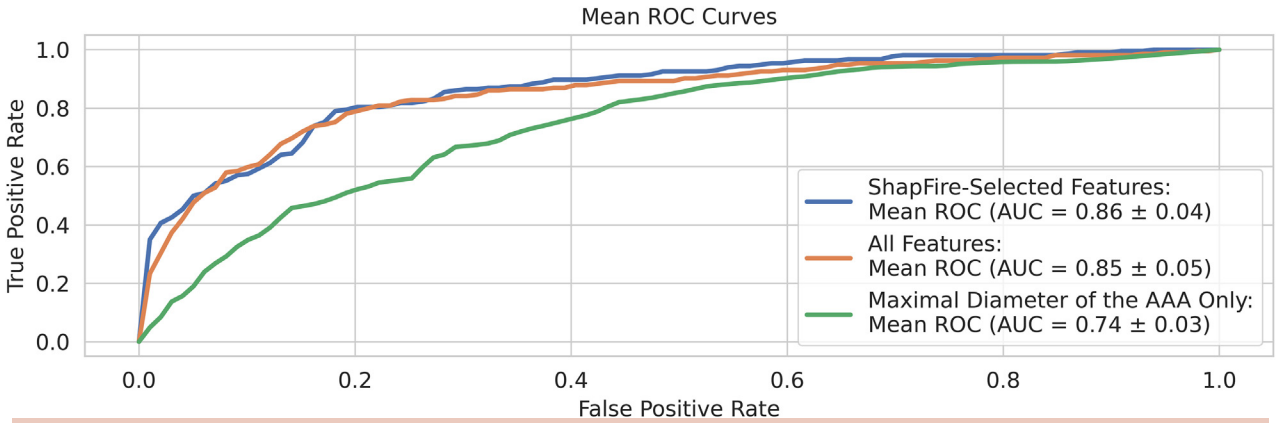


Fig 3. Receiver operating characteristic (*ROC*) curves of three AI models in an internal validation study for predicting ruptured abdominal aortic aneurysm (*rAAA*) using artificial intelligence (AI). The blue curve represents the SHapley Additive exPlanations Feature Importance Rank Ensembling (SHAPFire) AI model, the orange curve represents all features, and the green curve is for using the maximal diameter alone derived from a total population of 637 patients with AAA from a case-control study. *AUC*, Area under the curve.

AAA.[1,18,20] Third, smoking was recognized as an influential feature but of a lesser degree. This is surprising considering that cessation of smoking is the most effective nonsurgical intervention to prevent expansion of the aneurysm.[18,20] Fourth, blood pressure, specifically pulse pressure, and a diagnosis of hypertension were both flagged, which agrees with current literature.[1,20] Fifth and lastly, the presence of an intramural thrombus was identified as a feature of importance decreasing the risk of rupture, which conversely previously has been reported as a mediator of AAA expansion and risk of rupture in observational studies.[21,22]

Implications. Identifying patients at increased risk of rupture of AAA is challenging. Currently, absolute size thresholds are used to determine when to perform a prophylactic surgery. However, a significant number of patients experience rupture before the size limit is reached. This study lays the groundwork for a predictive tool prioritizing patients for elective surgery based on imaging and clinical data, beyond the commonly used AP diameter. Although the SHAPFire model was specifically developed to identify the presence of aortic rupture in patients actively undergoing surgery for AAA, its internal validation revealed a significantly higher accuracy than diameter measurements alone. Although we recognize that applying this model to a different patient population with aortic aneurysms requires further external validation, it opens avenues for future research into rupture risk stratification across varied patient cohorts. Nonetheless, it should be noted that generalizing the model to other cohorts should be approached with caution. The similarity of the SHAPFire model to the model using all available features indicates that the other features may be either repetitive or unimportant. One would furthermore expect the SHAPFire model to be continually more refined given an increasingly larger dataset, which could change the chosen included features or their affinities for the outcome.

When externally validated, this model may assist surgeons in the decision-making process regarding timely prophylactic surgery on patients with AAA.

Strengths and limitations. First, our primary strength lies in the high-quality data from the Danish Vascular Registry, providing a comprehensive view of the relevant population in the Region of Southern Denmark. Secondly, by analyzing data derived from CTA scans rather than applying ML directly to the images, we were able to identify factors associated with rAAA. This approach also enhances the AI's longevity, as minor changes in imaging equipment or software will not impact the outcome. Thirdly, the cohort was manually reviewed and validated to ensure accurate diagnoses.

The study's limitations stem from its observational nature, susceptible to biases related to selection, information, and confounding factors. Although efforts were made to mitigate selection bias by including as many cases as possible, some had to be excluded due to missing data. Notably, only one-third of rupture cases undergo repair, potentially differing from those not reaching the operating theatre.[23] However, obtaining such data is inherently challenging.

Because the dimensions of the abdominal aorta were measured after rupture, the study may be biased due to changes in size occurring during rupture. This could introduce a systematic error, causing the AI models to recommend surgery more aggressively than necessary, or conversely recommend surveillance scan programs when prophylactic surgery would be optimal. However, the pre- and postsizes would be expected to be correlated. To the best of our knowledge, no study has been performed examining the size changes acutely after rupture.

Information bias may have occurred because data were not consistently recorded with the strict quality standards typical of prospective studies, both in medical records and the Danish Vascular Registry. Although the Danish Vascular Registry adopts a prospective data collection approach with stricter criteria, some information was still sourced from medical records. Notably, certain data points were not recorded at the time of the event, such as blood pressure at admission in the case of rupture, as they were deemed invalid. Smoking information was often missing during emergency admissions, so data were selected from either before the rupture or up to 3 months postoperatively. Despite these challenges, the data obtained from CTA scans are considered reliable and accounting for a significant portion of the included features. It is important to acknowledge that, apart from inevitable observer measurements, the robustness of data from CTA scans contributes to the overall validity of the study.

Despite incorporating a multitude of features in the model to minimize confounding, the possibility of residual confounding persists in such studies. The necessary exclusion rate introduces a risk of selection bias. It is worth noting that inherent differences between the groups may exist, further complicating the interpretation of findings. Furthermore, it could be argued that the two groups inherently differ, as rAAA cases are more likely to involve individuals who were previously undiagnosed with AAA. In addition, distinctions may exist between cases of rAAA that undergo surgical intervention and those that do not reach the operating theater. Finally, ethnicity was not factored into the analysis. However, during the study period, the Danish population was predominantly Caucasian, with immigrants and non-Western descendants making up less than 8% of the population.[24]

CONCLUSIONS

We successfully developed an AI decision-making tool to identify patients at increased risk of rupture, which far exceeded the simplistic method relying solely on the anterior-posterior diameter of the abdominal aorta. Notably, the SHAPFire AI tool, based on only 20 features, demonstrated comparable performance to using all 68 features. External validation of the model is warranted before clinical implementation.

AUTHOR CONTRIBUTIONS

Conception and design: NA, KH, AD, JL
Analysis and interpretation: JS, NA, LO, KH, JL
Data collection: ML, AR, AD, JL
Writing the article: JS, NA, LO, JL
Critical revision of the article: JS, NA, LO, ML, AR, KH, AD, JL
Final approval of the article: JS, NA, LO, ML, AR, KH, AD, JL
Statistical analysis: JS, NA
Obtained funding: AD, JL
Overall responsibility: JL

FUNDING

This work was supported by the Independent Research Fund Denmark: Individualized risk assessment of abdominal aortic aneurysms (1030-00339B).

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used OpenAI's AI language model, GPT-4 (Version 4), to check for grammar and spelling mistakes. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

DISCLOSURES

None.

REFERENCES

1. Wanhainen A, Van Herzeele I, Bastos Goncalves F, 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.
2. Wanhainen A, Hultgren R, Linné A, et al. Outcome of the Swedish nationwide abdominal aortic aneurysm screening program. *Circulation*. 2016;134:1141–1148.
3. Siika A, Lindquist Liljeqvist M, Zommorodi S, et al. A large proportion of patients with small ruptured abdominal aortic aneurysms are women and have chronic obstructive pulmonary disease. *PLoS One*. 2019;14:e0216558.
4. Laine MT, Vänttinen T, Kantonen I, et al. Rupture of abdominal aortic aneurysms in patients under screening age and elective repair threshold. *Eur J Vasc Endovasc Surg*. 2016;51:511–516.
5. Li B, Aljabri B, Verma R, et al. Using machine learning to predict outcomes following open abdominal aortic aneurysm repair. *J Vasc Surg*. 2023;78:1426–1438.e6.
6. Spinella G, Fantazzini A, Finotello A, et al. Artificial intelligence application to screen abdominal aortic aneurysm using computed tomography angiography. *J Digit Imaging*. 2023;36:2125–2137.
7. Lundberg S, Lee SI. A unified approach to interpreting model predictions. Accessed December 15, 2024. https://arxiv.org/abs/1705.07874.
8. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj*. 2015;350:g7594.
9. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
10. Eldrup N, Cerqueira C, de la Motte L, Rathenborg LK, Hansen AK. The Danish vascular registry, Karbase. *Clin Epidemiol*. 2016;8:713–718.
11. Altreuther M, Menyhei G. International validation of the Danish vascular registry Karbase: a vascunet report. *Eur J Vasc Endovasc Surg*. 2019;58:609–613.
12. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
13. Ke G, Meng Q, Finley T, et al. *Lightgbm: a highly efficient gradient boosting decision tree*; 2017. Accessed December 15, 2024. https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.
14. Nyrønning L, Skoog P, Videm V, Mattsson E. Is the aortic size index relevant as a predictor of abdominal aortic aneurysm? A population-based prospective study: the tromsø study. *Scand Cardiovasc J*. 2020;54:130–137.
15. Jones GT, Sandiford P, Hill GB, et al. Correcting for body surface area identifies the true prevalence of abdominal aortic aneurysm in screened women. *Eur J Vasc Endovasc Surg*. 2019;57:221–228.
16. Kristensen JSS, Obel LM, Dahl M, Høgh A, Lindholt JS. Gender-specific predicted normal aortic size and its consequences of the

population-based prevalence of abdominal aortic aneurysms. *Ann Vasc Surg*. 2023;91:127–134.
17. Skovbo Kristensen JS, Krasniqi L, Obel LM, Kavaliunaite E, Liisberg M, Lindholt JS. Exploring drug repurposing for treatment of abdominal aortic aneurysms: a systematic review and meta-analysis. *Eur J Vasc Endovasc Surg*. 2023;67:570–582.
18. Obel LM, Diederichsen AC, Steffensen FH, et al. Population-based risk factors for ascending, arch, descending, and abdominal aortic dilations for 60-74–year-old individuals. *J Am Coll Cardiol*. 2021;78:201–211.
19. Obel LM, Lindholt JS, Diederichsen AC, et al. Platelet aggregation is not altered in men with aortic aneurysms. *Thromb Haemost*. 2023;124:277–279.
20. Bhak RH, Wininger M, Johnson GR, et al. Factors associated with small abdominal aortic aneurysm expansion rate. *JAMA Surg*. 2015;150:44–50.
21. Zhu C, Leach JR, Wang Y, Gasper W, Saloner D, Hope MD. Intraluminal thrombus predicts rapid growth of abdominal aortic aneurysms. *Radiology*. 2020;294:707–713.
22. Brunner-Ziegler S, Hammer A, Seidinger D, Willfort-Ehringer A, Koppensteiner R, Steiner S. The role of intraluminal thrombus formation for expansion of abdominal aortic aneurysms. *Wien Klin Wochenschr*. 2015;127:549–554.
23. Lindholt JS. Abdominal aortic aneurysms. *Dan Med Bull*. 2010;57:B4219.
24. Statistics Denmark. IEPCT: the groups of ancestry share of the total population 1. January by ancestry statistics Denmark: Statistics Denmark; 2024. https://www.statistikbanken.dk/statbank5a/SelectVarVal/Define.asp?Maintable=IEPCT&PLanguage=1.

Submitted Aug 17, 2024; accepted Nov 14, 2024.

Additional material for this article may be found online at www.jvascsurg.org.

APPENDIX (online only).

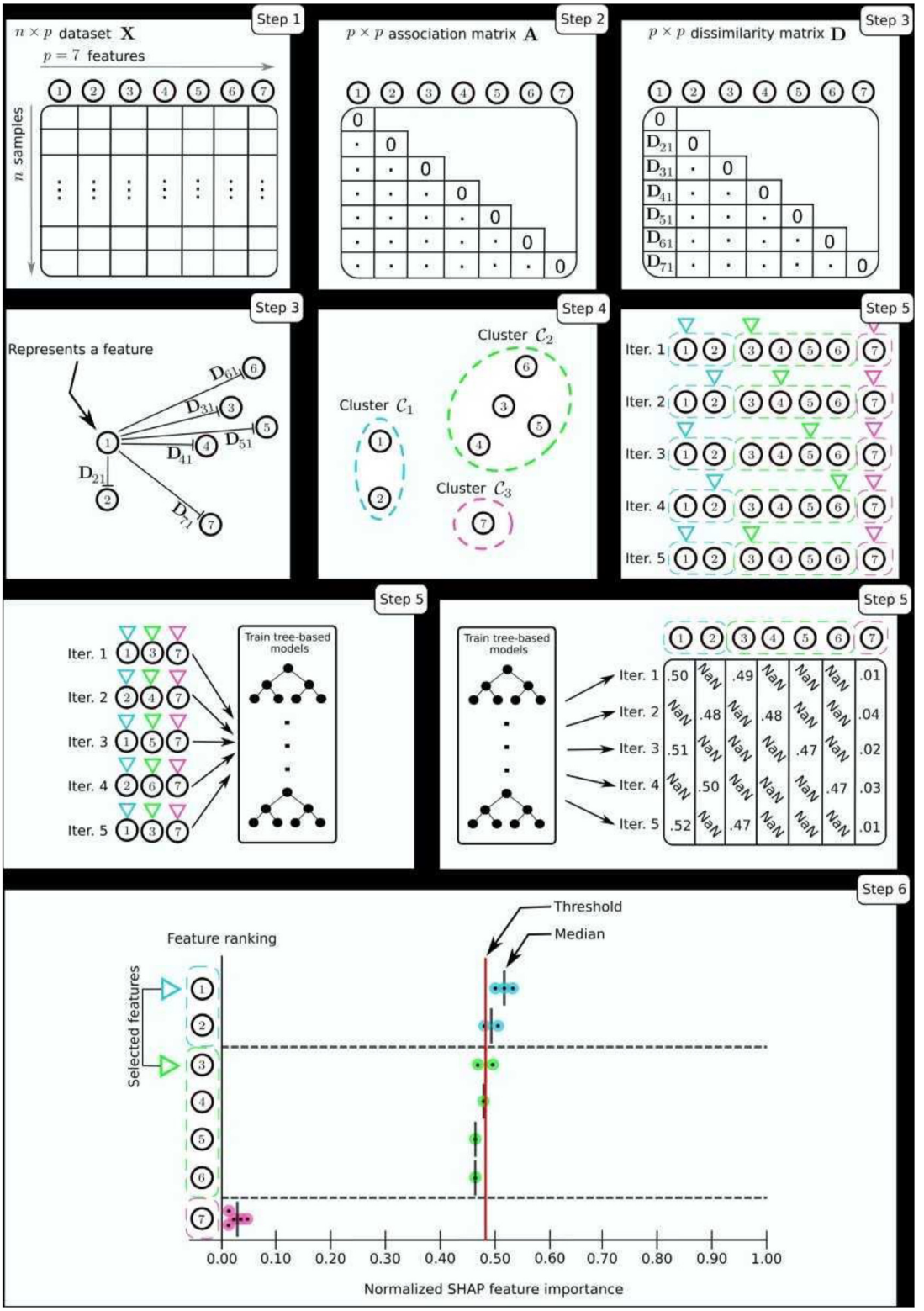
Supplementary Methods (online only)

In the following, we adopt machine learning (ML) terminology and will refer to variables as features; in this context, SHapley Additive exPlanations Feature Importance Rank Ensembling (SHAPFire) is an automated, wrapper-based feature importance ranking and selection tool that leverages the principles of SHAPFire.[7]

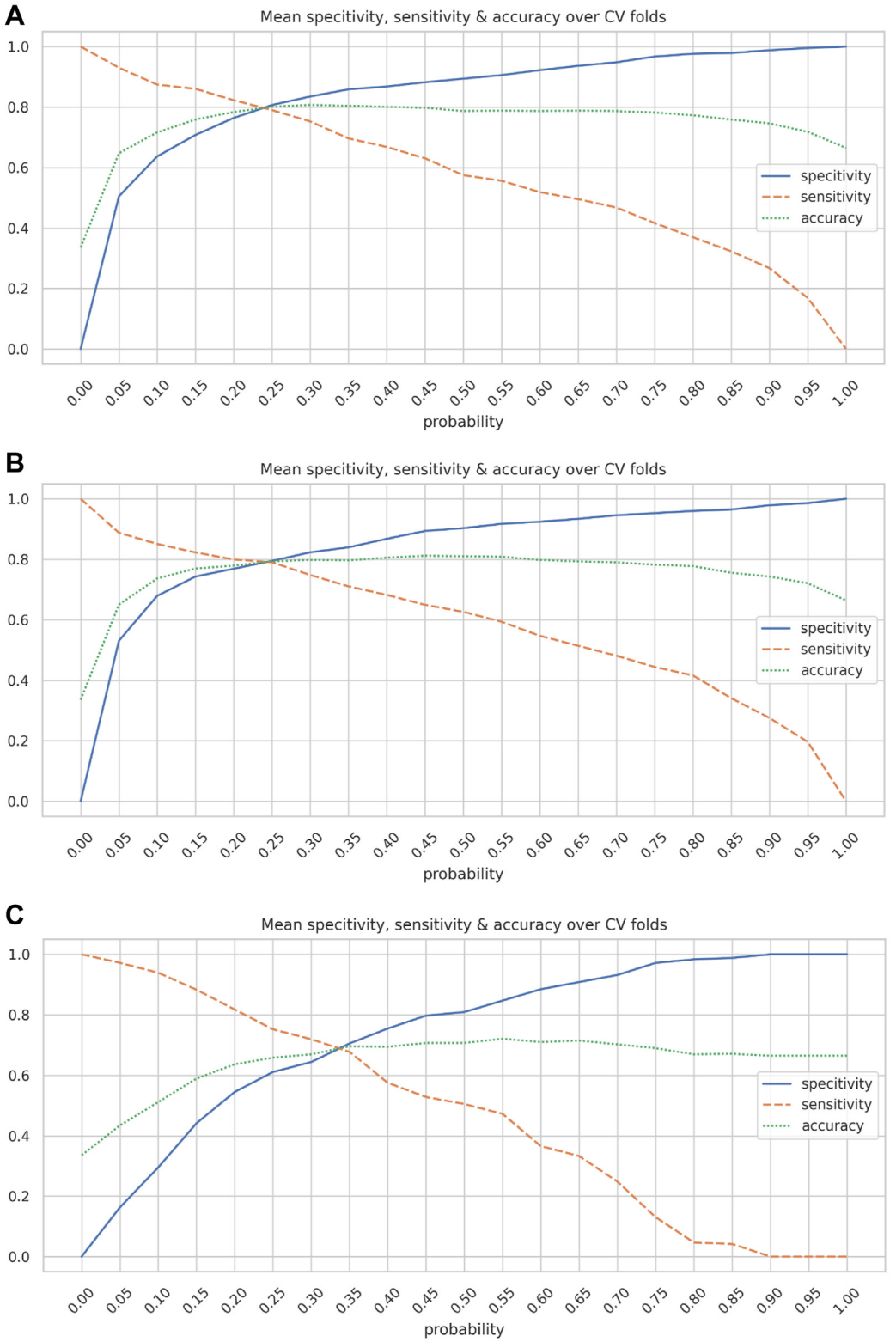
Atop Microsoft's LightGBM gradient boosting decision tree framework[13] and the SHAP Python library,[7] SHAPFire is essentially a machine model interpretation tool. The approach is mainly motivated by the challenge of handling highly associated features in an input dataset, as such highly associated features in an input dataset can affect ML model interpretability, making it hard to obtain accurate feature importance rankings. A concise step-by-step overview of the SHAPFire procedure is given as well as an illustration in Supplementary Fig 1 (online only).

- Step 1. The procedure starts with an input dataset, "X," presented as an "$n \times p$" table. This dataset contains "n" samples and "p" features.
- Step 2: A "$p \times p$" association matrix "A" is formed by assessing the relationship between each pair of features. Various statistical methods are employed based on feature types. For numerical features, Kendall's Tau-B measures the association. For categorical features, Cramer's V is used. For mixed types (numerical and categorical), the correlation ratio serves as the measure. The resulting matrix "A" has values ranging between 0 and 1.
- Step 3: Matrix "A" is transformed into a dissimilarity matrix "D" using the formula "$D = 1 - A$." In this matrix, highly correlated features can be considered closer to each other in terms of distance, whereas those with weaker associations have greater distances from one another.
- Step 4: Agglomerative hierarchical clustering is executed using the dissimilarity matrix "D." This technique groups similar features into clusters with a correlation factor above 0.5, as these likely provide redundant information and could hinder model interpretability in subsequent steps.
- Step 5: For a user-defined number of iterations (preferably no fewer than the largest cluster size to ensure that all features are used at some point in a model): feature subsets are sampled based on the clusters identified. Tree-based ML models are trained, and cross-validated normalized SHAP feature importance scores are extracted. These scores indicate each feature's contribution to predictive accuracy on a scale from 0 to 1.
- Step 6: Features are ranked by their median normalized SHAP feature importance scores, and the highest-ranking feature from each cluster is included in a preliminary feature subset. This candidate feature set undergoes pruning to eliminate less impactful features. The pruning threshold is determined based on a user-provided parameter (a value between 1 and 0, defaulting to 0.1) that signifies the minimum incremental contribution a feature must make to be included.

The process culminates in a final feature subset containing the most informative features for predictive modeling. The SHAPFire artificial intelligence tool subsequently allows the training of a gradient boosting decision tree model using the SHAPFire-selected features.[13]



Supplementary Fig 1 (online only). Step by step process of developing the SHapley Additive exPlanations Feature Importance Rank Ensembling (*SHAPFire*) artificial intelligence (AI) tool. Step 1. Input data: start with a dataset "*X*" containing "*n*" samples and "*p*" features. Step 2. Association matrix: create a matrix "*A*" showing the relationship between each pair of features. Step 3. Dissimilarity matrix: transform "*A*" into a dissimilarity matrix "*D*" to measure how different features are from each other. Step 4. Clustering: group similar features into clusters using hierarchical clustering, focusing on those with a correlation above 0.7. Step 5. Feature sampling: sample feature subsets from clusters, train models, and get SHAP importance scores for each feature. Step 6. Feature selection: rank features by their importance scores, prune less impactful ones based on a threshold (default 0.1), and select the final subset for the model. This process results in selecting the most informative features for the SHAPFire AI tool, which then trains a gradient boosting decision tree model using these features. *NaN,* Not A Number.



Supplementary Fig 2 (online only). **A,** The SHapley Additive exPlanations Feature Importance Rank Ensembling (*SHAPFire*) artificial intelligence (AI). **B,** The AI using all features. **C,** The AI using only the anterior-posterior diameter. *CV,* Cross-validation.

Supplementary Table I (online only). Description of every feature available for the SHapley Additive exPlanations Feature Importance Rank Ensembling (*SHAPFire*) artificial intelligence (AI) model from an internal validation study on AI prediction of ruptured abdominal aortic aneurysm (rAAA)

Variable name	Description	Variable type	Unique	Observations	missing	Mean	SD	minimum	p25	Median	p75	maximum
age	Age, years	Continuous	43	637	0	73	7	35	68	73	77	95
sex_calc	Sex	Binary	2	637	0	1	0	0	1	1	1	1
hyperten	Hypertension	Categorical	4	634	3	1	1	0	0	1	1	9
pseudoa	Pseudoaneurysm	Binary	2	635	2	0	0	0	0	0	0	0
burst_ia	Ruptured iliac aneurysm	Categorical	4	632	5	0	1	0	0	0	0	9
smoking	Smoking	Categorical	4	615	22	1	1	0	1	1	2	2
dm	Diabetes mellitus	Categorical	4	634	3	0	1	0	0	0	0	9
cad	Coronary artery disease	Categorical	4	636	1	0	1	0	0	0	1	9
stroke	Former stroke	Categorical	4	628	9	0	1	0	0	0	0	2
copd	Chronic obstructive pulmonary disease	Categorical	4	635	2	0	1	0	0	0	0	9
fam_aaa	Familial disposition to abdominal aortic aneurysm	Categorical	4	634	3	0	1	0	0	0	0	9
a_diss	Former aortic dissection	Categorical	4	633	4	0	1	0	0	0	0	9
diverticu	Diverticula disease	Categorical	4	633	4	0	1	0	0	0	0	9
auto_im	Autoimmune disease	Categorical	4	628	9	0	1	0	0	0	0	9
sirs	SIRS within the last 3 months	Categorical	4	604	33	0	1	0	0	0	0	9
med_trom	Platelet inhibitors	Categorical	3	624	13	1	0	0	0	1	1	1
med_mag	Acetylsalicylic acid	Categorical	3	624	13	1	0	0	0	1	1	1
med_clo	Clopidogrel	Categorical	3	624	13	0	0	0	0	0	0	1
med_tic	Ticagrelor	Categorical	3	624	13	0	0	0	0	0	0	1
med_per	Persantin	Categorical	3	624	13	0	0	0	0	0	0	1
med_nsaid	NSAID	Categorical	3	623	14	0	0	0	0	0	0	1
med_statin	Statins	Categorical	3	623	14	1	0	0	0	1	1	1
med_ak	Anticoagulants	Categorical	3	624	13	0	0	0	0	0	0	1
med_noak	Direct oral anticoagulants	Categorical	3	624	13	0	0	0	0	0	0	1
med_beta	β-Blockers	Categorical	3	623	14	0	0	0	0	0	1	1
med_thia	Thiazides	Categorical	3	623	14	0	0	0	0	0	0	1
med_ace	ACE inhibitors	Categorical	3	623	14	0	0	0	0	0	1	1
med_calc	Calcium antagonists	Categorical	3	623	14	0	0	0	0	0	1	1
med_bron	Bronchodilators	Categorical	3	624	13	0	0	0	0	0	0	1
med_kort	Oral corticosteroids	Categorical	3	623	14	0	0	0	0	0	0	1
med_chemo	Chemotherapy within 3 months	Categorical	3	623	14	0	0	0	0	0	0	1
weight	Weight	Continuous	81	571	66	83	16	41	72	81	92	162
height	Height	Continuous	47	564	73	174	8	150	170	175	180	206
bmi	Body mass index	Continuous	424	561	76	27	4	15	24	26	30	56
bsa_dubois	Body surface area calculated using the DuBois formula	Continuous	416	561	76	2	0	1	2	2	2	3
bsa_mosteller	Body surface area calculated using the Mosteller formula	Continuous	406	561	76	2	0	1	2	2	2	3
bt_sys	Systolic blood pressure	Continuous	97	556	81	145	21	83	130	145	159	200
bt_dia	Diastolic blood pressure	Continuous	69	556	81	83	14	39	74	83	92	123
map	Mean arterial pressure	Continuous	171	556	81	103	14	56	94	103	113	143
puls_pres	Pulse pressure, mm Hg	Continuous	84	556	81	62	17	19	50	60	71	130
suprarenal_throm	Suprarenal mural thrombus	Categorical	3	577	60	1	0	0	0	1	1	1
aaa_shape	Shape of abdominal aortic aneurysm (fusiform or saccular)	Categorical	3	604	33	1	0	1	1	1	1	2
aaa_bleb	Aortic bleb	Categorical	3	605	32	0	0	0	0	0	0	1
fibrose_ct	Retroperitoneal fibrosis	Categorical	3	601	36	0	0	0	0	0	0	1
ap_dia_max_ct	Maximal anterior-posterior diameter, cm	Continuous	344	607	30	7	1	3	5	6	7	14
trans_dia_ct	Maximal transverse diameter, cm	Continuous	371	606	31	7	2	3	6	6	8	15
circumference_ct	Maximal circumference, cm	Continuous	499	606	31	21	5	9	17	20	24	45
area_ct	Maximal area, cm^2	Continuous	570	606	31	36	18	7	24	30	44	154
lum_area_ct	Maximal luminal area, cm^2	Continuous	479	541	96	16	13	1	7	12	20	94
ap_dia_max_ct_90	Anterior-posterior diameter perpendicular to maximum anterior-posterior diameter	Continuous	353	607	30	6	1	2	5	6	7	13
trans_dia_ct_90	Transverse diameter perpendicular to maximum anterior-posterior diameter, cm	Continuous	356	606	31	7	2	2	5	6	7	15
circumference_ct_90	Circumference perpendicular to maximum anterior-posterior diameter, cm	Continuous	499	606	31	21	5	9	17	19	23	51

Supplementary Table I (online only). Continued.

Variable name	Description	Variable type	Unique	Observations	missing	Mean	SD	minimum	p25	Median	p75	maximum
area_ct_90	Area perpendicular to maximum anterior-posterior diameter, cm^2	Continuous	558	606	31	35	18	4	23	29	42	154
lum_area_ct_90	Luminal area perpendicular to maximum anterior-posterior diameter, cm^2	Continuous	483	541	96	15	13	1	7	11	19	87
trans_dia_l3_ct	Transverse outer-to-outer diameter of L3, cm	Continuous	168	606	31	5	0	1	4	4	5	9
dist_aort_sacrum	Distance between the aortic bifurcature to os sacrum, cm	Continuous	368	595	42	6	2	3	5	6	7	12
illiaca_bifur	Distance between iliac bifurcatures, cm	Continuous	335	568	69	7	1	3	6	7	8	11
ilia_aorta_sin	Distance between the left iliac bifurcature and the aortic bifurcature, cm	Continuous	342	565	72	6	2	3	5	6	7	20
ilia_aorta_dxt	Distance between the right iliac bifurcature and the aortic bifurcature, cm	Continuous	349	564	73	6	1	2	5	6	7	13
art_renis_aorta	Distance between the lowest renal artery and the aortic bifurcature, cm	Continuous	369	566	71	13	3	8	11	12	14	46
ilia_com_max_sin	Left iliac artery maximal diameter, cm	Continuous	199	527	110	18	9	5	13	16	20	73
ilia_com_max_dxt	Right iliac artery maximal diameter, cm	Continuous	208	525	112	19	9	7	13	16	21	75
ca_score_ant_thick	Anterior wall thickness, mm	Continuous	108	595	42	1	0	1	1	1	1	2
ca_unica_max	UniCa score at maximum size over 15 mm	Continuous	448	587	50	175	287	0	12	73	222	2553
ca_unica_thromb	UniCa score in thrombus	Continuous	49	517	120	6	29	0	0	0	0	348
v_ct_quality_yes	Bad quality CT scan	Binary	2	637	0	0	0	0	0	0	0	1
subgroup	Subgroup (training or validation)	Binary	2	637	0	1	1	0	0	1	1	1
burst_aaa	Ruptured abdominal aortic aneurysm	Binary	2	637	0	0	0	0	0	0	1	1

ACE, Angiotensin-converting enzyme inhibitor; *BSA*, body surface area; *CT*, computed tomography; *NSAID*, nonsteroidal anti-inflammatory drug; *UniCa score*, calcification score.

Supplementary Table II (online only). Accuracy of three artificial intelligence (AI) models using maximal Youden's index in an internal validation study for predicting ruptured abdominal aortic aneurysm (rAAA)

	SHAPFire	All features	AP diameter only
Youden's index	0.597	0.585	0.383
Accuracy	0.801	0.793	0.696
Specificity	0.807	0.795	0.705
Sensitivity	0.79	0.79	0.677

AP, Anterior posterior; *SHAPFire*, SHapley additive exPlanations feature importance rank ensembling.