



Original Article

# A Comparative Classification Analysis of Abdominal Aortic Aneurysms by Machine Learning Algorithms

BALAJI RENGARAJAN,<sup>1</sup> WEI WU,<sup>1</sup> CRYSTAL WIEDNER,<sup>2</sup> DAIJIN KO,<sup>2</sup> SATISH C. MULUK,<sup>3</sup>  
MARK K. ESKANDARI,<sup>4</sup> PRAHLAD G. MENON,<sup>5</sup> and ENDER A. FINOL<sup>1,6</sup>

<sup>1</sup>Department of Mechanical Engineering, University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA; <sup>2</sup>Department of Management Science and Statistics, University of Texas at San Antonio, San Antonio, TX, USA; <sup>3</sup>Department of Thoracic & Cardiovascular Surgery, Allegheny Health Network, Allegheny General Hospital, Pittsburgh, PA, USA; <sup>4</sup>Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; <sup>5</sup>Department of Mathematics and Data Analytics, Carlow University, Pittsburgh, PA, USA; and <sup>6</sup>UTSA/UTHSA Joint Graduate Program in Biomedical Engineering, University of Texas at San Antonio, San Antonio, TX, USA

(Received 22 October 2019; accepted 20 January 2020; published online 24 January 2020)

Associate Editor Estefanía Peña oversaw the review of this article.

**Abstract**—The objective of this work was to perform image-based classification of abdominal aortic aneurysms (AAA) based on their demographic, geometric, and biomechanical attributes. We retrospectively reviewed existing demographics and abdominal computed tomography angiography images of 100 asymptomatic and 50 symptomatic AAA patients who received an elective or emergent repair, respectively, within 1–6 months of their last follow up. An in-house script developed within the MATLAB computational platform was used to segment the clinical images, calculate 53 descriptors of AAA geometry, and generate volume meshes suitable for finite element analysis (FEA). Using a third party FEA solver, four biomechanical markers were calculated from the wall stress distributions. Eight machine learning algorithms (MLA) were used to develop classification models based on the discriminatory potential of the demographic, geometric, and biomechanical variables. The overall classification performance of the algorithms was assessed by the accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and precision of their predictions. The generalized additive model (GAM) was found to have the highest accuracy (87%), AUC (89%), and sensitivity (78%), and the third highest specificity (92%), in classifying the individual AAA as either asymptomatic or symptomatic. The k-nearest neighbor classifier yielded the highest specificity (96%). GAM used seven markers (six geometric and one biomechanical) to develop the classifier. The maximum transverse dimension, the average wall thick-

ness at the maximum diameter, and the spatially averaged wall stress were found to be the most influential markers in the classification analysis. A second classification analysis revealed that using maximum diameter alone results in a lower accuracy (79%) than using GAM with seven geometric and biomechanical markers. We infer from these results that biomechanical and geometric measures by themselves are not sufficient to discriminate adequately between population samples of asymptomatic and symptomatic AAA, whereas MLA offer a statistical approach to stratification of rupture risk by combining demographic, geometric, and biomechanical attributes of patient-specific AAA.

**Keywords**—Abdominal aortic aneurysm, Rupture risk evaluation, Image segmentation, Machine learning, Generalized additive model.

## INTRODUCTION

The current clinical management of an abdominal aortic aneurysm (AAA) is based on assessing its size and growth by measuring the maximum diameter of the AAA sac. Clinical intervention is recommended for AAA with maximum diameter greater than 5.5 cm in men and in the range of 5.0 to 5.4 cm in women.<sup>2</sup> In a study by Mastracci *et al.*,<sup>17</sup> an 85% mortality rate was reported for ruptured AAA, while 66% of these deaths occur prior to an emergent intervention. The use maximum diameter as the sole predictor of rupture risk may underscore the importance of other factors, which may not be evident or easily measured in clinical images. Darling *et al.*<sup>4</sup> found rupture to occur at the

Address correspondence to Ender A. Finol, Department of Mechanical Engineering, University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA. Electronic mail: ender.finol@utsa.edu

Balaji Rengarajan and Wei Wu have contributed equally to this work.

minimum wall thickness. Brown *et al.*<sup>1</sup> found that the AAA rupture risk for maximum diameter in the range of 5.0 to 5.9 cm was low and four times greater in females compared to males. Fillinger *et al.*<sup>7</sup> discovered that peak wall stress (PWS) in ruptured AAA was higher than in unruptured AAA, and differentiated the two aneurysm groups better than maximum diameter, thereby concluding that PWS was a more accurate measure of rupture risk than the clinical standard.

Machine learning can be employed to develop statistical inference based on attributes of a training dataset, without relying on predetermined equations, constitutive models, or other heuristics to make predictions on out-of-sample data that is equivalent to the training set in terms of descriptive attributes. Machine learning algorithms (MLA) have been widely used as classifiers in surgery and biomedical engineering. Cui *et al.*<sup>3</sup> applied a Naive Bayes (NB) algorithm to predict osteonecrosis of the femoral head with cannulated screw fixation. Farag *et al.*<sup>6</sup> used k-Nearest Neighbor (KNN) and Support Vector Machine (SVM) to classify lung nodules. Decision Tree (TREE) was applied by Min *et al.*<sup>18</sup> to predict survival in patients with distal bile duct cancer. Lau *et al.*<sup>11</sup> predicted graft failure after liver transplantation with Random Forest (RF). Zheng *et al.*<sup>29</sup> used Lasso Linear Logistic Regression (LASSO) for a classification analysis of various forms of cancer. Multivariate Adaptive Regression Splines (MARS) was applied by Wijeyesundera *et al.*<sup>27</sup> to establish a predictive index for renal replacement therapy after cardiac surgery. Endo *et al.*<sup>5</sup> applied the Generalized Additive Model (GAM) to analyze the associations between transfusion ratios and outcomes (in-hospital mortality and incidence of adverse events).

We hypothesized that using geometric markers (that describe the AAA size, shape, wall thickness, and curvature), demographics information (gender), and biomechanical markers (related to wall stress) will yield a higher classification accuracy when differentiating asymptomatic from symptomatic AAA compared to maximum diameter alone. In addition to addressing the aforementioned hypothesis, a second objective of this work is to compare the classification performance of eight statistical MLA and select the best classifier for rupture risk assessment.

## MATERIALS AND METHODS

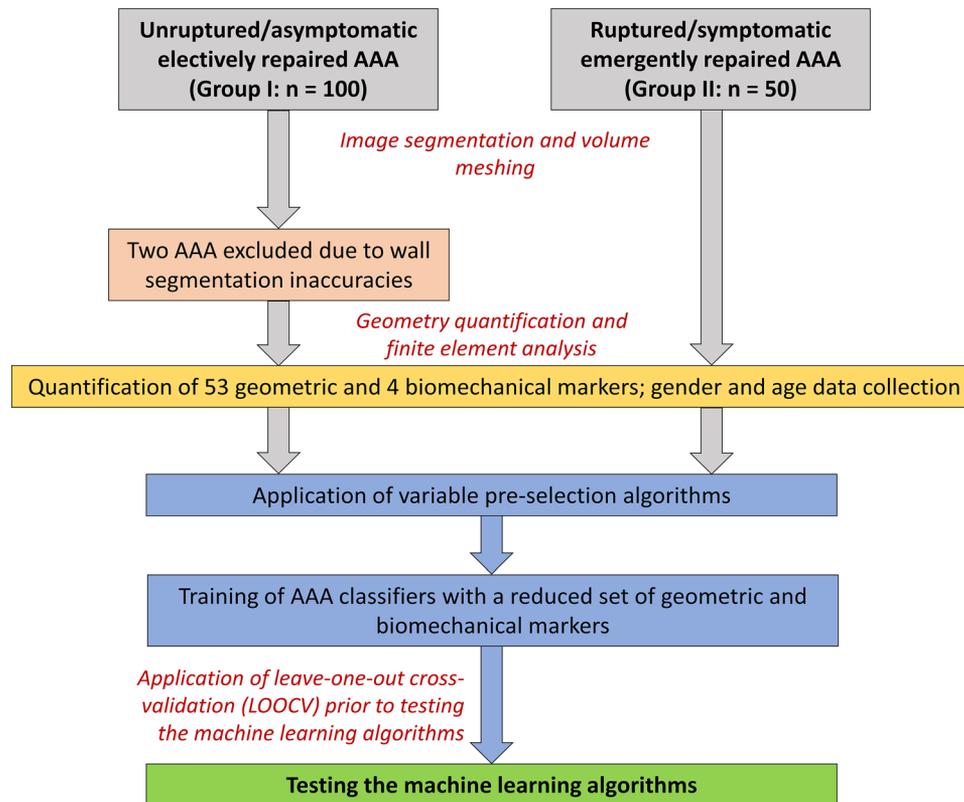
### *Human Subjects Research Study*

The human subjects research protocol for this study was approved by the Institutional Review Boards at Allegheny General Hospital (Pittsburgh, PA) and Northwestern Memorial Hospital (Chicago, IL). As

this was a retrospective review of existing medical records, informed consent was not required. This study consisted of two groups, Group I,  $n = 100$  patients with asymptomatic AAA, who received an elective repair within 6 months of their last imaging follow up, and Group II,  $n = 50$  patients with clinical symptoms or a radiographically confirmed ruptured AAA, who received an emergent repair within 1 month of their last imaging follow up. The flowchart of the study design is shown in Fig. 1. The abdominal computed tomography angiography (CTA) images of the 150 patients were collected after routine follow up, which occurred independent of the execution of the research protocol. These images corresponded to the last CTA scan prior to the (elective or emergent) repair. The inclusion criteria was based on the availability of a CTA exam within 6 months of the repair. We excluded patients who had received unenhanced (non-contrast) CT, those in which the last follow up occurred more than 6 months prior to the repair, and all post-repair CT. Patient age and gender was also registered from existing medical records. Research coordinators at both clinical centers collected the data, de-identified the images, and assigned alphanumeric case numbers to each dataset prior to sharing the data with the investigators.

### *Image Segmentation and 3D Reconstruction*

All segmentations were performed using custom segmentation scripts written in MATLAB (Mathworks Inc., Natick, MA), collectively known as AAASc (v1.03, The University of Texas at San Antonio, San Antonio, TX). Figure 2a illustrates a typical abdominal CTA image, while Fig. 2b shows a common output of AAASc CTA segmentation in which the lumen, inner wall, and outer wall contours are visible. The segmentation algorithms exploit the difference in contrast between the lumen and the surrounding soft tissue regions to identify and segment the lumen boundary by a region-growing method.<sup>24</sup> For the outer wall segmentation, the user selects the best-fitted outer wall boundary from a set of different possible outer wall contours generated by the algorithm.<sup>24</sup> The inner wall segmentation approach employs a trained neural network that infers the inner wall boundary based on a set of computed image features such as regional image texture and intensity.<sup>16</sup> The segmentation protocol generates a point cloud with the three boundaries and outputs volumetric binary masks that contain the lumen, intraluminal thrombus (ILT), and wall. The protocol has been previously validated and implemented to generate patient-specific AAA models for geometric modeling and finite element analysis (FEA).<sup>16,23,25</sup>



**FIGURE 1.** Flowchart of the study design for the first classification analysis. The second classification analysis used all geometric and biomechanical markers, while the third classification analysis used the maximum transverse dimension as the only input variable.

### *Geometric and Finite Element Modeling*

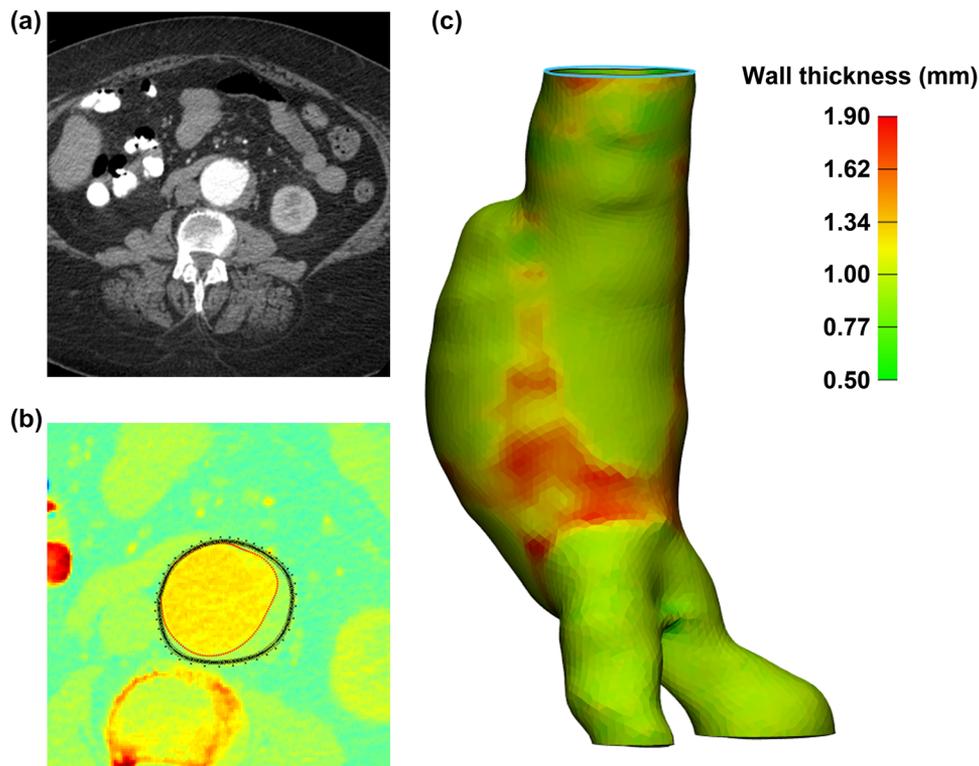
Following image segmentation and binary mask creation, geometric indices representative of the AAA size, shape, and wall thickness were calculated using in-house MATLAB scripts. These geometric markers together constituted fifty-three 1D size, 2D shape, 3D size, 3D shape, second-order curvature, and wall thickness indices, which were calculated for each AAA. The definition and mathematical formulation of the markers are included in Appendix A of the Supplementary Material. Two AAA from Group I were excluded from the geometric modeling protocol due to segmentation inaccuracies. The spatial distribution of wall thickness in an exemplary AAA is illustrated in Fig. 2c. The estimation of AAA wall thickness as an integral component of this protocol was previously verified with post-mortem wall thickness measurements.<sup>16</sup>

The binary masks were used as input to another set of MATLAB scripts, collectively known as AAAMesh,<sup>23</sup> to generate volume and surface meshes that take into account the patient-specific wall thickness distribution. The volume meshes consist of approximately 48,000 to 90,000 quadratic hexahedral

elements and were used with the FEA solver ADINA (Adina R&D, Inc., Watertown, MA) to compute wall stress for each AAA at peak systolic pressure. A Mooney-Rivlin constitutive model was used to represent the AAA wall material properties, as described by Raghavan and Vorp.<sup>22</sup> For such a model, the strain energy density is directly proportional to the first invariant of the left Cauchy-Green deformation tensor, described by Eq. (1),

$$W = \alpha(I_1 - 3) + \beta(I_1 - 3)^2 \quad (1)$$

where  $W$  represents the strain energy density,  $I_1$  the first invariant of the Cauchy-Green tensor, and  $\alpha$  and  $\beta$  are material constants derived from tensile testing of AAA wall specimens. With  $\alpha = 17.4 \text{ N/cm}^2$ ,  $\beta = 188.1 \text{ N/cm}^2$ , and a Poisson's ratio of 0.499, the second order Mooney-Rivlin material was implemented for all FEA models. The results of the FEA simulations were post-processed with Ansys EnSight (Ansys Inc., Canonsburg, PA) to calculate four biomechanical markers: PWS, 99th percentile wall stress (99thWS), 50th percentile wall stress (50thWS), and spatially averaged wall stress (SAWS). These markers are defined in Appendix B of the Supplementary Material and rep-



**FIGURE 2.** AAA segmentation and 3D reconstruction. (a) Exemplary abdominal CTA image; (b) contours of the lumen (red points), inner wall (black crosses), and outer wall (black points) generated during segmentation using AAAVasc; (c) a reconstructed AAA model with non-uniform thickness (in mm).

resent global wall stress metrics for each AAA model. Details on the finite element modeling protocol, including mesh sensitivity analyses, boundary conditions, and simulation set up are reported elsewhere.<sup>23</sup>

#### *Classification Using Machine Learning Algorithms*

We used the eight MLA described in “Introduction” section to test their ability to classify the 148 AAA using the geometric, biomechanical, and demographic markers as input variables to the algorithms. The leave-one-out cross-validation technique (LOOCV) was applied to report the accuracy of prediction for all MLA. We also report on the following performance metrics: area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve, sensitivity, specificity, and precision. Sensitivity is a measure of the true positive cases, i.e. symptomatic AAA that are classified correctly. Specificity is a measure of the true negative cases, i.e. asymptomatic AAA that are classified correctly. Sensitivity and specificity measures are used to generate an ROC curve, from which the AUC is calculated as a measure of overall classifier performance. Precision, which is also known as the positive predictive value, is a measure of the exactness of the prediction. All MLA were

implemented using the statistical computing language R in the RStudio integrated development environment (RStudio, Boston, MA).

## RESULTS

### *Demographic, Geometric and Biomechanical Markers*

AAA patients in Group I had a mean age of  $70 \pm 8$  years, while those in Group II had a mean age of  $72 \pm 11$  years. Since age was unknown for several patients, this variable was not used for subsequent statistical analyses. For Group I, 77% of patients were males and 23% were females, while for Group II, 76% of patients were males and 24% were females. No other demographic variable or clinical data were collected for this study.

The mean and standard deviation of all geometric markers for Groups I and II are summarized in Table 1. Of particular clinical interest,  $D_{\max\text{dir}}$  for Group I was  $56.0 \pm 12.5$  mm, while for Group II it was  $75.6 \pm 19.4$  mm. Symptomatic AAA have a larger vessel volume ( $V$ ) and surface area ( $S$ ) than asymptomatic AAA. ILT volume for Group II ( $139.4 \pm 116.1$  cm<sup>3</sup>) is more than double that of Group I ( $63.9 \pm 58.5$  cm<sup>3</sup>). The length of the abdominal aorta and length of

**TABLE 1. Quantitative summary (mean  $\pm$  standard deviation) of all geometric markers for Groups I (electively repaired AAA) and II (emergently repaired AAA).**

Geometric marker	Group I		Group II	
	Mean	SD	Mean	SD
$D_{\max}$ (mm)	52.83	12.09	69.95	18.18
$D_{\max\text{dir}}$ (mm)	55.96	12.52	75.59	19.43
$D_{\text{ave}}$ (mm)	42.39	11.13	55.43	17.72
$D_{\min}$ (mm)	28.35	8.24	33.13	13.64
$D_{\text{neck.p}}$ (mm)	30.67	9.65	31.98	7.36
$D_{\text{neck.d}}$ (mm)	35.35	20.90	45.76	17.18
$H$ (mm)	100.24	15.14	113.89	17.52
$L$ (mm)	113.57	19.09	133.78	22.69
$H_{\text{neck}}$ (mm)	28.93	21.80	26.01	17.24
$L_{\text{neck}}$ (mm)	35.02	24.40	35.15	21.29
$H_{\text{sac}}$ (mm)	71.31	21.97	87.88	26.44
$L_{\text{sac}}$ (mm)	78.55	25.29	98.63	30.79
$H_b$ (mm)	57.41	18.67	64.52	20.65
$d_c$ (mm)	6.13	5.50	7.39	5.59
$d_{c,\max}$ (mm)	7.65	5.21	9.97	5.94
$\text{TH}_{\min}$ (mm)	0.62	0.38	0.58	0.44
$\text{TH}_{\max}$ (mm)	3.93	2.24	4.54	1.92
$\text{TH}_{\text{ave}}$ (mm)	1.81	0.57	2.01	0.61
$\text{TH}_{D\max}$ (mm)	1.88	0.68	2.13	0.70
$\text{TH}_{\text{mode}}$ (mm)	1.79	0.65	2.01	0.65
$\text{TH}_{\text{median}}$ (mm)	1.80	0.58	2.01	0.62
$\text{TH}_{\text{minvar}}$ (mm)	0.03	0.04	0.04	0.04
$\text{TH}_{\text{maxvar}}$ (mm)	1.14	1.31	1.28	1.02
$\text{TH}_{\text{medianvar}}$ (mm)	0.13	0.18	0.15	0.17
$\text{TH}_{\text{modevar}}$ (mm)	0.04	0.06	0.04	0.04
$\text{TH}_{\text{meanvar}}$ (mm)	0.22	0.24	0.26	0.22
$P_{\text{below}}$	50.24	5.99	49.59	3.33
$P_{\text{above}}$	49.76	5.99	50.41	3.33
$\text{TT}_{\text{ave}}$ (mm)	5.39	3.95	7.57	4.59
$\text{TT}_{\max}$ (mm)	18.90	10.82	27.2	13.66
$\text{TT}_{\min}$ (mm)	9.08e-2	0.30	9.51e-2	0.20
$\text{TT}_{\min\text{Loc}}$ (mm)	0.54	0.38	0.68	0.35
$\text{TT}_{\max\text{Loc}}$ (mm)	0.37	0.22	0.40	0.26
DHr	0.54	0.14	0.61	0.13
DDr	1.80	0.47	2.24	0.61
Hr	0.28	0.21	0.24	0.17
BL	0.57	0.16	0.57	0.17
$\beta$	0.89	0.09	0.90	0.07
$\beta_{\min}$	0.85	0.08	0.84	0.08
$T$	1.11	0.06	1.14	0.08
$C_{\text{ave}}$	1.02	0.02	1.03	0.02
$C_{\max}$	1.10	0.07	1.19	0.25
$C_{\min}$	1.0	0.00	1.00	0.00
$V$ (cm <sup>3</sup> )	166.01	103.81	328.07	211.47
$S$ (cm <sup>2</sup> )	152.65	52.18	234.48	91.17
$V_{\text{ILT}}$ (cm <sup>3</sup> )	63.93	58.51	139.43	116.16
$\gamma$	0.35	0.18	0.40	0.18
IPR	5.30	0.41	5.25	0.37
NFI	1.06	0.04	1.06	0.06
GAA	5.16e-5	1.67e-4	5.64e-5	2.04e-4
MAA	0.03	0.01	0.02	0.01
GLN	2.72	1.07	3.55	1.85
MLN	0.33	0.04	0.38	0.10

The definition and mathematical formulation of these markers is included in Appendix A of the Supplementary Material.

the AAA sac for Group II ( $L = 133.8 \pm 22.7$  mm,  $L_{\text{sac}} = 98.6 \pm 30.8$  mm) were 18 and 25% greater than for Group I ( $L = 113.6 \pm 19.1$  mm,  $L_{\text{sac}} = 78.6 \pm 25.3$  mm), respectively. Symptomatic AAA have thicker walls than asymptomatic AAA: Group II exhibited  $\text{TH}_{\max} = 4.5 \pm 1.9$  mm,  $\text{TH}_{\text{ave}} = 2.0 \pm 0.6$  mm, and  $\text{TH}_{D\max} = 2.1 \pm 0.7$  mm, while Group I yielded  $\text{TH}_{\max} = 3.9 \pm 2.2$  mm,  $\text{TH}_{\text{ave}} = 1.8 \pm 0.6$  mm, and  $\text{TH}_{D\max} = 1.9 \pm 0.7$  mm. IPR and MAA, which are markers that describe the shape and curvature of the AAA wall, were found to be similar for both groups: Group I (IPR =  $5.3 \pm 0.4$  and MAA =  $0.03 \pm 0.01$ ) and Group II (IPR =  $5.3 \pm 0.4$  and mean MAA =  $0.02 \pm 0.01$ ).

The mean and standard deviation of the biomechanical markers for Groups I and II are summarized in Table 2. All global wall stress metrics were found to be higher for Group II compared to Group I. For example, PWS was 43% higher for Group II ( $139.9 \pm 57.7$  N/cm<sup>2</sup>) than Group I ( $97.5 \pm 43.6$  N/cm<sup>2</sup>); Group II had a 35% greater SAWS ( $29.7 \pm 10.8$  N/cm<sup>2</sup>) than Group I ( $22.0 \pm 7.5$  N/cm<sup>2</sup>).

### Machine Learning Classification

In each classification exercise, the MLA resulted in a probability of rupture between 0 and 1. The probability threshold used for statistical assessment of the classification performance was 0.5, namely a probability of rupture  $\geq 0.5$  was considered to be the positive response class (i.e. a ruptured and/or symptomatic AAA), whereas a probability of rupture  $< 0.5$  was identified as the negative response class (i.e. an asymptomatic, unruptured AAA).

### Classification with Variable Pre-selection

The first classification analysis was performed with the markers summarized in Table 3 with the objective of pre-selecting the input variables that achieved the highest classification accuracy for each MLA. TREE used only one variable ( $D_{\max\text{dir}}$ ) to build the classifier. NB, KNN, and RF used all 57 markers as classifiers to differentiate between the two groups (two indices,  $P_{\text{above}}$  and  $P_{\text{below}}$  were found to be highly collinear and thus one was removed from the analysis, reducing the initial set of 58 markers to 57). Variable pre-selection was used by MARS, from which GAM and SVM built their classifier based on the following seven markers (six geometric and one biomechanical):  $D_{\max\text{dir}}$ ,  $\text{TH}_{D\max}$ ,  $\text{TH}_{\text{median}}$ ,  $\text{TH}_{\text{maxvar}}$ ,  $P_{\text{above}}$ ,  $D_{\text{ave}}$ , and SAWS. Appendix C of the Supplementary Material describes the variable pre-selection methods used by each MLA.

Table 4 includes the performance metrics of all MLA listed in ascending order of accuracy. GAM exhibited the highest accuracy (87%), AUC (89%), and sensitivity (78%), and the third highest specificity (92%), with a precision of 83%—the highest amongst the MLA that had sensitivity in excess of 50%; it is considered the best performing classifier of asymptomatic vs. symptomatic AAA for this study. TREE's

**TABLE 2. Quantitative summary (mean  $\pm$  standard deviation) of all biomechanical markers.**

Biomechanical marker (N/cm <sup>2</sup> )	Group I		Group II	
	Mean	SD	Mean	SD
PWS	97.45	43.63	139.92	57.71
99thWS	50.14	16.79	66.94	20.94
50thWS	22.36	8.32	30.60	11.39
SAWS	22.03	7.53	29.71	10.84

The markers are peak wall stress (PWS), 99th percentile wall stress (99thWS), 50th percentile wall stress (50thWS), and spatially averaged wall stress (SAWS). These are global stress metrics calculated from the spatial distributions of wall stress obtained from finite element analysis of each AAA model. The definition of these markers is included in Appendix B of the Supplementary Material.

accuracy was 79%, with a high sensitivity of 74%, although it had a low specificity (82%) and the lowest AUC (60%). One advantage of using TREE as a classifier is the output of an optimal threshold for the variables in the classifier:  $D_{\max\text{dir}} = 6.5$  cm classified the two groups with a maximum accuracy of 79%. MARS' accuracy was 71% and has a higher AUC than TREE (75%); however, its sensitivity was 52% and it had the lowest specificity at 81%. NB, LASSO and RF had similar accuracies (75–79%), AUC (79–83%), specificities (87–94%), and sensitivities (50–54%). KNN and SVM are inefficient in predicting symptomatic AAA as the sensitivities for both MLA are low (KNN = 46%, SVM = 30%). In addition, SVM had the lowest accuracy amongst all MLA. KNN had the highest specificity of all MLA, making it effective in classifying asymptomatic AAA. Figure 3 illustrates the ROC curves for each MLA where it becomes evident that the curve for GAM encloses the other seven MLA at nearly every point, which is verification of its highest AUC.

#### Classification with $D_{\max\text{dir}}$ as the Only Input Variable

The second classification analysis was performed with  $D_{\max\text{dir}}$  as the only input variable for all MLA,

**TABLE 3. Variables used by the machine learning algorithms, which were pre-selected for TREE, LASSO, MARS, SVM, and GAM to yield the highest accuracy of classification (see Appendix C in the Supplementary Material).**

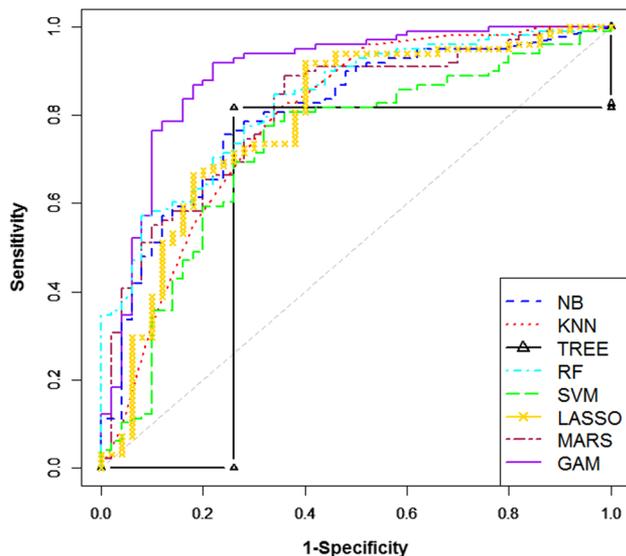
Variable	Machine learning algorithm							
	TREE	NB	KNN	RF	LASSO	MARS	SVM	GAM
$D_{\max\text{dir}}$	X				X	X	X	X
$D_{\text{ave}}$						X		X
$L_{\text{sac}}$					X			
$\text{TH}_{\text{Dmax}}$					X	X	X	X
$\text{TH}_{\text{median}}$						X	X	X
$\text{TH}_{\text{maxvar}}$					X	X	X	X
$\text{TH}_{\text{modvar}}$					X			
$\text{TH}_{\text{min}}$					X			
$P_{\text{above}}$						X	X	X
$\text{TT}_{\text{max}}$					X			
$\text{TT}_{\text{min}}$					X			
$\text{TT}_{\text{minLoc}}$					X			
DDr					X			
$\beta$					X	X		
$C_{\text{max}}$					X			
$C_{\text{min}}$					X			
IPR					X	X		
NFI					X			
MLN					X			
PWS					X			
99thWS					X			
50thWS					X			
SAWS						X	X	X
Gender					X			
All 57 geometric, biomechanical, and demographic markers		X	X	X				

NB, KNN, and RF do not make use of variable pre-selection algorithms.

**TABLE 4. Quantitative performance metrics of the eight machine learning algorithms applied with variable pre-selection, presented in increasing order of overall classification accuracy with a probability threshold of 0.5.**

MLA	Accuracy	AUC	Sensitivity	Specificity	95% CI	$p$ Value [Acc > NIR]	Kappa	Precision
SVM	0.69	0.72	0.30	0.89	(0.61, 0.76)	0.2736	0.2128	0.58
MARS	0.71	0.75	0.52	0.81	(0.63, 0.78)	0.1287	0.3343	0.58
NB	0.75	0.80	0.52	0.87	(0.67, 0.82)	0.0134	0.4094	0.67
RF	0.77	0.83	0.54	0.89	(0.69, 0.84)	0.0028	0.4545	0.71
KNN	0.79	0.79	0.46	0.96	(0.72, 0.85)	0.0004	0.4724	0.85
TREE	0.79	0.60	0.74	0.82	(0.72, 0.85)	0.0004	0.5430	0.67
LASSO	0.79	0.79	0.50	0.94	(0.72, 0.85)	0.0004	0.4838	0.81
GAM	0.87	0.89	0.78	0.92	(0.81, 0.92)	0.0000	0.7088	0.83

AUC represents the area under the curve for the receiver operating characteristic curves illustrated in Fig. 3. NIR stands for no-information-rate, which was 0.66 in our dataset for these LOOCV studies. Except for SVM and MARS, the remaining classifiers resulted in a minimum accuracy (based on their 95% confidence intervals) that surpassed the NIR. GAM had the highest accuracy and the highest Kappa, at 87% and 0.7088, respectively.



**FIGURE 3. Receiver operating characteristic (ROC) curves of the eight machine learning algorithms applied for AAA classification analysis using variable pre-selection (see Appendix C of the Supplementary Material). The methods are Naive Bayes (NB), k-Nearest Neighbor (KNN), Decision Tree (TREE), Random Forest (RF), Support Vector Machine (SVM), Lasso Linear Logistic Regression (LASSO), Multivariate Adaptive Regression Splines (MARS), and Generalized Additive Model (GAM). A quantitative comparison of these algorithms based on global performance metrics is included in Table 4, where GAM yielded the highest accuracy, area under the curve (AUC) and sensitivity, and the second highest specificity.**

given the importance of the clinical standard of care for rupture risk assessment. Table 5 describes the quantitative performance metrics of all MLA, where MARS exhibited the highest accuracy (80%) and AUC (80%). LASSO yielded the highest sensitivity (82%), while SVM and NB produced the highest specificity (90%). The ROC curves for all the MLA are illustrated in Fig. 4, where NB, GAM, and RF appear to out-

perform the other classifiers based on their comparatively high AUC.

#### *Classification with all Geometric, Biomechanical, and Demographic Markers as the Input Variables*

A third classification analysis was performed using all 57 markers as the input variables for each MLA (i.e., no variable pre-selection methods were applied), the results of which are summarized in Appendix D of the Supplementary Material. Table D1 describes the quantitative performance metrics of all MLA, while the ROC curves are shown in Fig. D1.

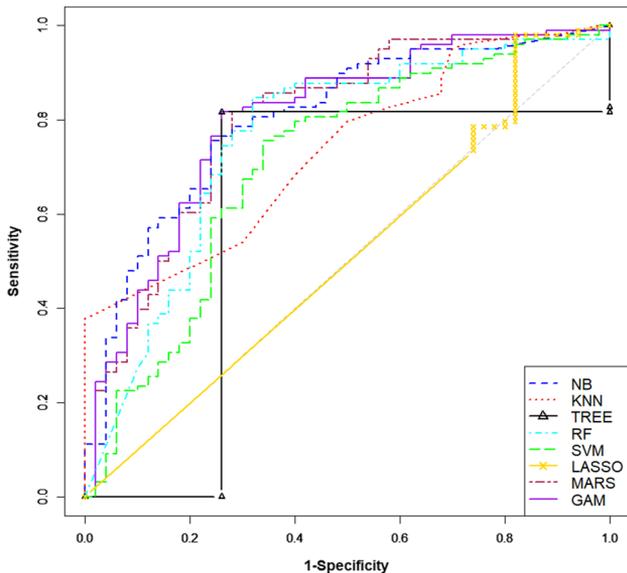
## DISCUSSION

We performed classification analyses of AAA based on the discriminatory potential of their demographic, geometric, and biomechanical characteristics. We tested the hypothesis that using these AAA attributes would yield an improved classification rate compared to the clinical standard of maximum diameter. Eight MLA were applied to a data set of 98 asymptomatic and 50 symptomatic/ruptured aneurysms, and the performance metrics of the algorithms were compared to determine the best overall classifier. To the authors' knowledge, this is the first study on a comprehensive comparison of statistical supervised learning techniques applied to AAA classification using geometric, demographic, and biomechanical markers. We previously used MLA to classify ruptured and unruptured AAA using only geometric markers of size, shape, and wall thickness,<sup>20,25</sup> and surface curvatures.<sup>13</sup> We found that the overall accuracy of the algorithms (GAM, MARS, KNN, RF, NB, TREE, LASSO and SVM) was moderate to high, ranging from 69 to 87%, when using variable pre-selection and a probability threshold of 0.5 [i.e.,  $P(\text{Rupture}) \geq 0.5$ ] considered as the positive

**TABLE 5. Quantitative performance metrics of the eight machine learning algorithms using  $D_{\max\text{dir}}$  as the only input variable.**

MLA	Accuracy	AUC	Sensitivity	Specificity
SVM	0.69	0.72	0.28	0.90
NB	0.72	0.79	0.38	0.90
MARS	0.74	0.74	0.56	0.83
RF	0.74	0.75	0.68	0.77
LASSO	0.41	0.51	0.82	0.20
TREE	0.79	0.60	0.74	0.82
KNN	0.79	0.73	0.68	0.85
GAM	0.78	0.79	0.58	0.88

AUC values correspond to the ROC curves illustrated in Fig. 4.

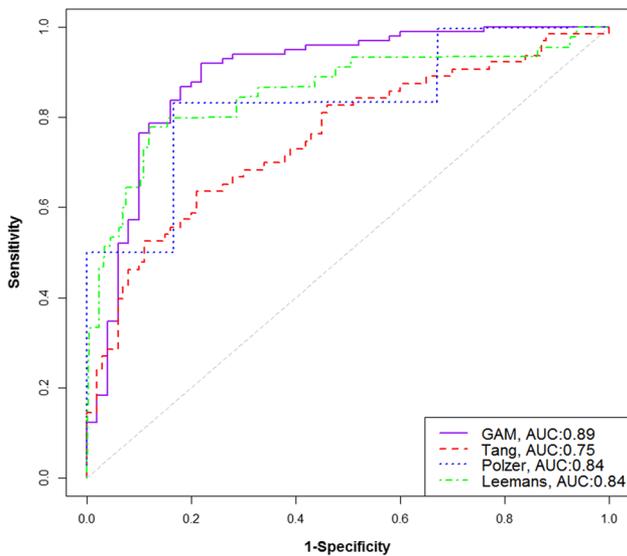
**FIGURE 4. Receiver operating characteristic (ROC) curves of the eight machine learning algorithms using  $D_{\max\text{dir}}$  as the only input variable. A quantitative comparison of these classifiers based on global performance metrics is included in Table 5.**

categorical class. However, the sensitivity, which is the correct prediction rate for symptomatic AAA, was disappointingly low, ranging from 30 to 78%. GAM, which was the best overall classifier, had a high true negative detection probability or specificity (92%), which ensured a high classification rate for asymptomatic AAA. It is anticipated that sensitivity and specificity can be optimized further for any given classifier by virtue of appropriate selection of MLA-specific probability thresholds based on ROC analytics. However, the latter was beyond the scope of this feasibility study, which was focused on comparing the classification performance of MLA and interrogating whether using demographic, geometric, and biomechanical markers would yield a higher classification rate than a maximum diameter based approach.

GAM, which is an extension of generalized linear models (GLM), assumes that the dependent variables are represented as nonlinear additive functions of the independent variables. GLM force linearity on the data, whereas GAM allows for nonlinearity of the data, establishing the underlying detailed data patterns. Two modeling aspects are considered in GAM: which variables should be included in the classifier and how smooth should be the distribution of a variable.<sup>9</sup> The speedy classification of GAM is due largely to its use of an adaptive back-fitting procedure, similar to that used in MARS, to guide the selection of variables and identify their optimal degree of smoothing.<sup>12</sup> The disadvantages of using GAM is the possibility of overfitting and the difficulty of analyzing the response of each predictor independently.

Supervised learning has been used previously to assess AAA rupture risk. Polzer and Gasser<sup>21</sup> quantified rupture risk based on a probabilistic rupture risk index (PRRI) that resulted in an AUC of 84%. They reported on the use of PRRI as the sole predictor of rupture risk and that the directly measured maximum diameter had no contribution to this risk. Conversely, the best performing classifier in the present study, GAM, had an AUC of 89% and used  $D_{\max\text{dir}}$  as one of the primary predictors for the classification. Leemans *et al.*<sup>14</sup> assessed rupture risk using the directly measured maximum diameter and three biomechanical parameters (PWS, peak wall rupture index, and the rupture risk equivalent diameter), yielding an AUC of 85.5%. Tang *et al.*<sup>26</sup> used stepwise multivariate logistic regression to predict rupture and obtained a 64% probability of true prediction, 79% specificity, and an AUC of 75%. Logistic regression is based on the assumption that the predictor and the response variable have a linear relation, which is not true in many clinical applications. By using a nonlinear predictive model, such as GAM, and patient-specific wall thickness, we obtained a comparatively improved classifier of AAA; this comparison is illustrated in Fig. 5 by means of ROC curves and AUC.

GAM uses seven markers to classify the two AAA groups, of which three (SAWS,  $D_{\max\text{dir}}$ , and  $\text{TH}_{D\text{max}}$ ) were identified as the most influential for AAA classification. These are categorized as a biomechanical marker (SAWS), a 1D global size marker ( $D_{\max\text{dir}}$ ), and a 1D local size marker ( $\text{TH}_{D\text{max}}$ ). Wall stress measures such as PWS are used extensively in biomechanics research to assess AAA rupture risk.<sup>7,10,15,28</sup> However, PWS is strongly dependent on the quality of the volume mesh used for FEA and the local wall thickness of the AAA, which can lead to an overestimation of the true maximum stress. SAWS is wall stress averaged over the local surface area of the AAA and is thus a weighted measure of mean wall stress.



**FIGURE 5.** The GAM ROC curve is compared to the ROC curves obtained from the data analyses conducted by Tang *et al.*,<sup>26</sup> Polzer *et al.*,<sup>21</sup> and Leemans *et al.*,<sup>14</sup> who also performed discriminatory analyses of ruptured/symptomatic and unruptured/asymptomatic AAA. These discriminatory analyses<sup>14,21,26</sup> were not based on using the same combination of geometric, biomechanical, and demographic markers as in the present work.

99thWS and 50thWS are not true measures of wall stress but rather percentiles calculated mathematically. Therefore, SAWS may be a more suitable biomechanical discriminator of asymptomatic and symptomatic AAA, since it is related to the surface area of the aneurysm wall. Wall stress is strongly influenced by wall thickness,<sup>19</sup> while thin-walled AAA regions are exposed to high wall stress and are at high risk for rupture.<sup>8</sup> In addition, AAA rupture is known to occur at regions of low wall thickness.<sup>4</sup>  $D_{\max\text{dir}}$  represents the maximum diameter measured as the largest distance between two points on any AAA sac cross section, thereby being the equivalent of the clinical standard of care. Noteworthy is that TREE, which yielded the second highest MLA accuracy (79%), used  $D_{\max\text{dir}}$  as the sole variable for classification. In addition, TREE predicted that the maximum diameter for optimal classification of the two groups is 6.5 cm, in contrast to the clinical standard of 5.0–5.5 cm. This suggests that a maximum diameter of 6.5 cm is an improved discriminator of asymptomatic and symptomatic AAA, compared to the standard of care, for the population sample of 148 AAA in this study.

To address the main hypothesis of this work, the other seven MLA were applied using  $D_{\max\text{dir}}$  as the only variable used to build the classifier. A comparison of the two classification analyses (Tables 4 vs. 5) showed that using  $D_{\max\text{dir}}$  as the sole discriminator is less accurate than using a combination of biomechanical and geometric markers.

For example, GAM, which had the highest accuracy at 87% using seven input variables, was 78% accurate when using  $D_{\max\text{dir}}$ . A further analysis of the GAM  $D_{\max\text{dir}}$  classifier revealed 22 AAA misclassified: 4 false negatives (symptomatic AAA that were misclassified as asymptomatic) and 18 false positives (asymptomatic AAA that were misclassified as symptomatic). The false negatives had a mean  $D_{\max\text{dir}}$  of 55.95 mm, which is comparable to the mean  $D_{\max\text{dir}}$  of Group I (55.96 mm). Similarly, the false positives had a mean  $D_{\max\text{dir}}$  of 67.78 mm, which is comparable to the mean  $D_{\max\text{dir}}$  of Group II (75.59 mm). This yields a reasonable explanation for why the maximum diameter criterion failed to classify correctly these 22 AAA. Conversely, the mean SAWS for the false negatives was 31.32 N/cm<sup>2</sup>, which is comparable to the mean SAWS of Group II (29.71 N/cm<sup>2</sup>), while the mean SAWS for the false positives was 20.56 N/cm<sup>2</sup>, which is comparable to the mean SAWS of Group I (22.03 N/cm<sup>2</sup>). Using SAWS as the only input variable for GAM resulted in the 22 AAA classified correctly, given the similarity of this biomechanical marker with the corresponding means for Groups II (for the false negatives) and I (for the false positives). Therefore, the complex shape of the AAA sac cannot be adequately represented only by  $D_{\max\text{dir}}$ . Rather, it should be quantified by five additional geometric markers that can discriminate between symptomatic and asymptomatic AAA ( $TH_{D\max}$ ,  $TH_{\text{median}}$ ,  $TH_{\text{maxvar}}$ ,  $D_{\text{ave}}$ , and  $P_{\text{above}}$ ) when  $D_{\max\text{dir}}$  alone cannot.

Using variable pre-selection methods resulted in more accurate classification outcomes compared to using all 57 markers in the classifiers (Tables 4 vs. D1). One additional comparative analysis was performed with GAM using its six geometric markers to build the classifier vs. using its sole biomechanical marker for classification. The outcome of this analysis is described in Appendix E of the Supplementary Material. Tables E1 and E2 reveal the nonlinearity of the variables in the model, while Table E3 shows that geometric or biomechanical markers by themselves cannot achieve the same accuracy as its combined use in the classifier. This is further corroborated in Fig. E1, which illustrates the highest AUC is obtained when all 7 markers are used to build the classifier.

The present work is subject to several important limitations. There is suspected inter-observer variability in the segmentation of the clinical images, although its effect on the geometric and biomechanical markers was not quantified. There is also variability in the pixel size of the CTA images amongst the medical records. Since wall thickness estimation is bounded by the pixel size and the intensity gradient across the wall, the larger the pixels, the less precise is the ensuing wall thickness prediction. The exclusion of ILT in the FEA

modeling limits the results of the study to stresses predicted only by wall mechanics. This limitation likely yields an over prediction of PWS, 99thWS, 50thWS, and SAWS, as ILT tends to provide the wall with a mechanical barrier to the normal forces caused by blood pressure. The quantification of geometric indices *a priori* is another limiting aspect of our geometry quantification approach. There could be size and shape measures important for differentiating individual AAA that are not taken into account by our methodology but that could be predicted by using other techniques. It is possible to use additional AAA attributes, such as tissue composition metrics (obtained from immunohistochemistry), which would be subject to the availability of AAA wall specimens for subsequent histological analysis. Finally, the classification analyses presented in this work could be adopted as a rupture risk assessment strategy. This is justified by the fact that all AAA in Group I were asymptomatic, unruptured, and patients received an elective repair within 6 months of the last CTA follow up. Conversely, all AAA in Group II were symptomatic and patients received an emergent repair within 1 month of the last CTA follow up; however, not all had a confirmed rupture visible in the clinical images.

In summary, for AAA classification based on geometric, demographic, and biomechanical markers, GAM showed the highest accuracy (87%), prediction probability for symptomatic AAA (78%), and third highest prediction probability for asymptomatic AAA (92%), amongst eight supervised learning techniques evaluated. Six geometric markers ( $D_{\max\text{dir}}$ ,  $D_{\text{ave}}$ ,  $\text{TH}_{\text{Dmax}}$ ,  $\text{TH}_{\text{median}}$ ,  $\text{TH}_{\text{maxvar}}$ ,  $P_{\text{above}}$ ) and one biomechanical marker (SAWS) yielded the best overall performance of the GAM classifier. Such performance is comparatively superior to using only geometric or biomechanical markers. Using patient-specific geometric and biomechanical attributes, and a nonlinear predictive model for the MLA classifier, yields a classification accuracy (87%) greater than maximum diameter alone (78%). Implementing MLA as classifiers for rupture risk assessment is clinically feasible since the computational times required to execute the algorithms are relatively low: a few minutes on a standard Windows personal desktop for the sample size of 148 AAA used in this work. In addition, the classification accuracy increases concomitantly with the size of the training dataset. Hence, one can envision a large database of AAA demographic, geometric and biomechanical features, created from thousands of unruptured and ruptured AAA. The database would be used to train a MLA classifier and test it on every new diagnosed AAA with the purpose of assessing its rupture risk based on the “likeness” of the new AAA to either the unruptured or the ruptured group.

## ELECTRONIC SUPPLEMENTARY MATERIAL

The online version of this article (<https://doi.org/10.1007/s10439-020-02461-9>) contains supplementary material, which is available to authorized users.

## ACKNOWLEDGMENTS

Research funding was provided in part by National Institutes of Health Award No. R01HL121293. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The use of ANSYS Enight is gratefully acknowledged through an educational licensing agreement with Ansys, Inc.

## CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

## REFERENCES

- <sup>1</sup>Brown, P. M., D. T. Zelt, and B. Sobolev. The risk of rupture in untreated aneurysms: the impact of size, gender, and expansion rate. *J. Vasc. Surg.* 37(2):280–284, 2003.
- <sup>2</sup>Chaikof, E. L., R. L. Dalman, M. K. Eskandari, B. M. Jackson, W. A. Lee, M. A. Mansour, T. M. Mastracci, M. Mell, M. H. Murad, L. L. Nguyen, G. S. Oderich, M. S. Patel, M. L. Schermerhorn, and B. W. Starnes. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J. Vasc. Surg.* 67(1):2–77, 2018.
- <sup>3</sup>Cui, S. S., L. K. Zhao, Y. M. Wang, Q. Dong, J. X. Ma, Y. Wang, W. Zhao, and X. Ma. Using Naive Bayes classifier to predict osteonecrosis of the femoral head with cannulated screw fixation. *Injury.* 49(10):1865–1870, 2018.
- <sup>4</sup>Darling, R. C., C. R. Messina, D. C. Brewster, and L. W. Ottinger. Autopsy study of unoperated abdominal aortic aneurysms. The case for early resection. *Circulation.* 56(3 Suppl):II161–II164, 1977.
- <sup>5</sup>Endo, A., A. Shiraishi, K. Fushimi, K. Murata, and Y. Otomo. Outcomes of patients receiving a massive transfusion for major trauma. *Br. J. Surg.* 105(11):1426–1434, 2018.
- <sup>6</sup>Farag, A. A., A. Ali, and S. Elshazly. Feature fusion for lung nodule classification. *Int. J. CARS.* 12(10):1809–1818, 2017.
- <sup>7</sup>Fillinger, M. F., M. L. Raghavan, S. P. Marra, J. L. Cronenwett, and F. E. Kennedy. In vivo analysis of mechanical wall stress and abdominal aortic aneurysm rupture risk. *J. Vasc. Surg.* 36(3):589–597, 2002.
- <sup>8</sup>Gasser, T. C. Biomechanical rupture risk assessment: a consistent and objective decision-making tool for abdominal aortic aneurysm patients. *Aorta.* 4(2):42–60, 2016.
- <sup>9</sup>Jeong, C., J. H. Min, and M. S. Kim. A tuning method for the architecture of neural network models incorporating

- GAM and GA as applied to bankruptcy prediction. *Expert Syst. Appl.* 39(3):3650–3658, 2012.
- <sup>10</sup>Larsson, E., F. Labruto, T. C. Gasser, J. Swedenborg, and R. Hultgren. Analysis of aortic wall stress and rupture risk in patients with abdominal aortic aneurysm with a gender perspective. *J. Vasc. Surg.* 54(2):295–299, 2011.
- <sup>11</sup>Lau, L., Y. Kankanige, B. Rubinstein, R. Jones, C. Christophi, V. Muralidharan, and J. Bailey. Machine-learning algorithms predict graft failure after liver transplantation. *Transplantation.* 101(4):E125–E132, 2017.
- <sup>12</sup>Leathwick, J. R., J. Elith, and T. Hastie. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Model.* 199(2):188–196, 2006.
- <sup>13</sup>Lee, K., J. Zhu, J. Shum, Y. Zhang, S. C. Muluk, A. Chandra, M. K. Eskandari, and E. A. Finol. Surface curvature as a classifier of abdominal aortic aneurysms: a comparative analysis. *Ann. Biomed. Eng.* 41:562–576, 2013.
- <sup>14</sup>Leemans, E. L., T. P. Willems, C. H. Slump, M. J. van der Laan, and C. J. Zeebregts. Additional value of biomechanical indices based on CTA for rupture risk assessment of abdominal aortic aneurysms. *PLoS ONE.* 13(8):e0202672, 2018.
- <sup>15</sup>Maier, A., M. Gee, C. Reeps, J. Pongratz, H.-H. Eckstein, and W. Wall. A comparison of diameter, wall stress, and rupture potential index for abdominal aortic aneurysm rupture risk prediction. *Ann. Biomed. Eng.* 38:3124–3134, 2010.
- <sup>16</sup>Martufi, G., E. S. Di Martino, C. H. Amon, S. C. Muluk, and E. A. Finol. Three-dimensional geometrical characterization of abdominal aortic aneurysms: image-based wall thickness distribution. *J. Biomech. Eng.* 131(6):061015, 2009.
- <sup>17</sup>Mastracci, T. M., L. Garrido-Olivares, C. S. Cinà, and C. M. Clase. Endovascular repair of ruptured abdominal aortic aneurysms: a systematic review and meta-analysis. *J. Vasc. Surg.* 47(1):214–221, 2008.
- <sup>18</sup>Min, K. W., D. H. Kim, B. K. Son, E. K. Kim, S. B. Ahn, S. H. Kim, Y. J. Jo, Y. S. Park, J. Seo, Y. H. Oh, S. Oh, H. Y. Kim, M. J. Kwon, S. K. Min, H. R. Park, J. Y. Choe, J. Y. Jeon, H. I. Ha, and J. W. Lee. Invasion depth measured in millimeters is a predictor of survival in patients with distal bile duct cancer: decision tree approach. *World J. Surg.* 41(1):232–240, 2017.
- <sup>19</sup>Mower, W. R., L. J. Baraff, and J. Sneyd. Stress distributions in vascular aneurysms: factors affecting risk of aneurysm rupture. *J. Surg. Res.* 55(2):155–161, 1993.
- <sup>20</sup>Parikh, S. A., R. Gomez, M. Thirugnanasambandam, S. S. Chauhan, V. De Oliveira, S. C. Muluk, M. K. Eskandari, and E. A. Finol. Decision tree based classification of abdominal aortic aneurysms using geometry quantification measures. *Ann. Biomed. Eng.* 46:2135–2147, 2018.
- <sup>21</sup>Polzer, S., and T. C. Gasser. Biomechanical rupture risk assessment of abdominal aortic aneurysms based on a novel probabilistic rupture risk index. *J. R. Soc. Interface.* 12(113):20150852, 2015.
- <sup>22</sup>Raghavan, M. L., and D. A. Vorp. 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.* 33:475–482, 2000.
- <sup>23</sup>Raut, S. S., P. Liu, and E. A. Finol. An approach for patient-specific multi-domain vascular mesh generation featuring spatially varying wall thickness modeling. *J. Biomech.* 48(10):1972–1981, 2015.
- <sup>24</sup>Shum, J., E. S. Di Martino, A. Goldhammer, D. H. Goldman, L. C. Acker, G. Patel, J. H. Ng, G. Martufi, and E. A. Finol. Semiautomatic vessel wall detection and quantification of wall thickness in computed tomography images of human abdominal aortic aneurysms. *Med. Phys.* 37(2):638–648, 2010.
- <sup>25</sup>Shum, J., G. Martufi, E. S. Di Martino, C. B. Washington, J. Grisafi, S. C. Muluk, and E. A. Finol. Quantitative assessment of abdominal aortic aneurysm geometry. *Ann. Biomed. Eng.* 39:277–286, 2011.
- <sup>26</sup>Tang, A., C. Kauffmann, S. Tremblay-Paquet, S. Elkouri, O. Steinmetz, F. Morin-Roy, L. Cloutier-Gill, and G. Soulez. Morphologic evaluation of ruptured and symptomatic abdominal aortic aneurysm by three-dimensional modeling. *J. Vasc. Surg.* 59(4):894–902, 2014.
- <sup>27</sup>Wijeyesundera, D. N., K. Karkouti, J. Y. Dupuis, V. Rao, C. T. Chan, J. T. Granton, and W. S. Beattie. Derivation and validation of a simplified predictive index for renal replacement therapy after cardiac surgery. *JAMA.* 297(16):1801–1809, 2007.
- <sup>28</sup>Xenos, M., S. H. Rambhia, Y. Alemu, S. Einav, N. Labropoulos, A. Tassiopoulos, J. J. Ricotta, and D. Bluestein. Patient-based abdominal aortic aneurysm rupture risk prediction with fluid structure interaction modeling. *Ann. Biomed. Eng.* 38(11):3323–3337, 2010.
- <sup>29</sup>Zheng, S. F., and W. X. Liu. An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification. *Comput. Biol. Med.* 41(11):1033–1040, 2011.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.