Published in final edited form as:

Proc Inst Mech Eng H. 2018 September ; 232(9): 922–929. doi:10.1177/0954411918794724.

On the Assessment of Abdominal Aortic Aneurysm Rupture Risk in the Asian Population Based on Geometric Attributes

Tejas Canchi^a, Eddie Y.K Ng^a, Sriram Narayanan^b, and Ender A. Finol^c

^aSchool of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, tejas002@e.ntu.edu.sg

^bDepartment of General Surgery, Tan Tock Seng Hospital, Jalan Tan Tock Seng, Singapore 308433, snarayan@doctors.org.uk

^cDepartment of Mechanical Engineering, University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas, USA, ender.finol@utsa.edu

Abstract

This study aims to review retrospectively the records of Asian patients diagnosed with Abdominal Aortic Aneurysm (AAA) to investigate the potential correlations between clinical and morphological parameters within the context of whether the aneurysms were ruptured or unruptured. A machine learning based approach is proposed to predict the rupture status of Asian AAA by comparing four different classifiers trained with clinical and geometrical parameters obtained from computed tomography images. The classifiers were applied on 312 patient datasets obtained from a regulatory-approved database. The datasets included 17 attributes under three classes: unruptured AAA, ruptured AAA, and normal aorta without aneurysm. Four different classification models, namely Decision Trees, Naïve Bayes, Logistic Regression, and Support Vector Machines were applied to the patient dataset. The models were evaluated by ten-fold cross-validation and the classifier performances were assessed with classification accuracy, area under the curve (AUC) of receiver operator characteristic, and F measures. Data analysis and evaluation were performed using the WEKA machine learning application. The results indicated that Naïve Bayes achieved the best performance amongst the classifiers with a classification accuracy of 95.2%, an AUC of 0.974, and an F measure of 0.952. The clinical implications of this work can be addressed in two ways. The best classifier can be applied to prospectively acquired data to predict the likelihood of aneurysm rupture. Next, it would be necessary to estimate the attributes implicated in rupture risk beyond just maximum aneurysm diameter.

Keywords

Abdominal Aortic Aneurysm; Rupture Risk; Geometric Attributes; Machine Learning

Corresponding Author: Tejas Canchi, School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore, Tel: +65-67904555, Fax: +65-67924062, tejas002@e.ntu.edu.sg.

Conflict of Interest
None to declare.

Introduction

An Abdominal Aortic Aneurysm (AAA) is a dilation of the infra-renal segment of the aorta¹. As a result of this condition, there is progressive weakening of the arterial wall which is a risk factor for the rupture of the aneurysm. However, the expansion of the aorta varies amongst individuals and not all AAA cases will result in a rupture. The U.K. Small Aneurysm Trial Participants² concluded that a cut-off diameter of about 5 – 5.5 cm should be used to recommend surgery for AAA patients. Emergency surgery of the AAA involves high risk and hence it would be useful if a prediction model can be used to assess the risk rupture of an aneurysm without using a one-size fits all approach such as maximum diameter.

Several factors are associated with assessment of AAA rupture risk. Ruptured AAA are seen to have less tortuosity and larger cross sectional diameter asymmetry³. This is in accordance with studies showing higher stresses in asymmetric AAAs^{4,5}. Anatomically accurate AAA models must incorporate non-uniform wall thickness in order to get a realistic assessment of stresses developed on the arterial wall. DiMartino et al⁶ quantified the arterial wall thickness using a laser micrometer with specimens from patients marked for surgical repair. Wall thickness measured in ruptured AAA patients (3.6 ± 0.3 mm) was significantly different from those having electively repaired AAA (2.5 ± 0.1 mm). They also observed an inverse correlation between the wall thickness and local tissue strength. Raghavan et al⁷ analyzed four AAA tissue samples extracted from cadavers (three unruptured, one ruptured) and concluded that there was considerable regional variation of wall thickness with significantly reduced values at the rupture site. Based on analysis by Sacks et al., it has been determined that patient specific AAA have highly tortuous and asymmetric morphologies with regional and local changes in surface curvature and wall thickness values⁸. It is therefore critical that the aneurysm shape and the wall thickness be considered while assessing AAA rupture risk.

Various computational methods have been used to model the AAA mechanical environment, namely finite element modelling (FEM) and machine learning methods. However, FEM has its limitations as many factors cannot be input to a FE model, e.g. race and family history⁹. Machine learning methods are believed to be more efficient as they can be implemented with a large number of parameters over a large patient cohort without the need for executing simulations with an FEM code. For example, using machine learning and data mining methods to develop models that learn from clinical and biomechanical input data, the models would make predictions based on parameters chosen *a priori* that are known to affect AAA risk of rupture. A summary of machine learning methods used in the classification of AAA population groups is described elsewhere¹⁰.

Machine learning methods have been used earlier to predict risk of rupture in AAA as well as to estimate parameters such as wall shear stress (WSS), which is reported as one of the biomechanical factors strongly correlated to rupture risk. To find correlations between WSS, blood viscosity, density, velocity, and geometric parameters of AAA and carotid bifurcation models, Jordanski et al.¹¹ tested three machine learning algorithms: multivariate linear regression, multilayer perceptron neural network and Gaussian conditional random fields. They were able to predict the temporal evolution of WSS in both the AAA and carotid

bifurcation model with high accuracy. The values of WSS correlated in 8 of the 10 time steps of the cardiac cycle and the Gaussian conditional random field performed the best among the methods tested. This was an extension of their previous work on data mining methods to estimate WSS from a number of computational fluid dynamics (CFD) simulations that involve geometric attributes as well¹². They used a neural network model to estimate the systolic stress, achieving a reduction in time from 114 min for CFD to 0.1 min using data mining methods for 85 patients. Soudah et al.¹³ also used artificial neural networks to reduce computational times for estimating WSS. They used 243 idealized AAA models and performed FEA on them. Using the results obtained using the finite element technique, two different neural networks were developed and trained: a Mesh Neural Network (MNN) and a Tension Neural Network (TNN). The MNN created an aneurysm mesh in terms of four geometric factors (asymmetry factor, aneurysm diameter, aneurysm thickness, aneurysm length) and the TNN is coupled with the MNN to calculate the peak WSS on every node of the mesh for a given arterial pressure. This method was able to demonstrate high accuracy in the estimation of WSS when compared to the FEA, although the work was performed with idealized AAA models.

Statistical correlations between biomechanics and shape indices have been carried out by Guillermo et al.¹⁴ where they performed CFD on 13 AAA patient models, and plotted correlations of peak wall shear stress (PWSS) and peak intraluminal pressure (PIP) with defined shape indices in 1D, 2D and 3D. All the AAA had maximum diameters of less than 40 mm and hence maximum diameter and stresses were not highly correlated. Nevertheless, they concluded that length, asymmetry and saccular index significantly influence the PWSS and PIP. Izbicki et al.¹⁵ presented an assessment of aortic aneurysm rupture risk by applying machine learning methods with the use of patient medical records while identifying the significant factors that play a role in rupture risk. Shum¹⁶ proposed to use computed tomography (CT) image data to improve the performance of an AAA rupture prediction model by using three-dimensional image processing techniques to characterize indices of aneurysm size, shape and regional wall thickness. To this end, multivariate analysis of variance was used to determine the significant difference between the indices and the aneurysm condition as it relates to its unruptured/ruptured status. Shum further investigated the quantitative measurement of AAA geometry and wall thickness to assess rupture risk using the Decision Trees method¹⁷. It was shown that geometry indices and regional variations of AAA shape should be taken into consideration to improve the accuracy in predicting rupture risk. Giannoglou et al.¹⁸ reported on the relationship between the peak wall stresses that can cause aneurysm rupture and several parameters characteristic of an aneurysm's geometry. They showed that three-dimensional analysis of curvature is an important factor in rupture risk assessment.

In this work, four machine learning classifiers were tested for their ability to assess rupture risk in a sample of the Asian AAA population. Such testing was based on performance metrics that include classification accuracy of ruptured AAA, sensitivity and specificity of each classifier, and a comparative analysis by means of AUC values. Subsequently, a feature selection algorithm was applied to rank the most significant attributes correlated with AAA rupture risk, yielding a parametric model. The novel contribution of this work is the derivation of geometric attributes associated with AAA rupture risk in a cohort of Asian

patients and a parametric model that can be used in a clinical setting for the assessment of such risk.

Materials and Methods

The input dataset was collected retrospectively from 312 in-patient medical records drawn from an IRB approved database. All patients were treated at Tan Tock Seng Hospital in Singapore. The dataset included 17 attributes under three classes: U, R, and N. The nine attributes for patient characteristics were A, G, R, S, ES, D, Hyp, PDA, and BMI. Geometric indices measured from standard of care CT images constituted the other 8 attributes, which are as follows: D1a, D1c, NL, DAAA, L1, D2, D3, and D4. These indices are illustrated schematically in Figure 1. The data mining application WEKA was used to carry out the comparative machine learning analysis¹⁹. WEKA incorporates an imputation function wherein missing values in a dataset are replaced by means of the distribution; this function was used to impute the missing attributes in the current AAA dataset.

In this work, four different classifiers were tested to assess the risk of rupture in a population sample of Asian AAA consisting of unruptured and ruptured aneurysms, and normal aortas without aneurysm. The attributes used for the classifiers consisted of 9 clinical parameters obtained from existing medical records and 8 geometric indices calculated from the segmented and 3D-reconstructed CT abdominal images. The range, mean and standard deviation values for the quantifiable numerical attributes in the full dataset are shown in Table 1.

Amongst the 26 ruptured AAA patients in the dataset, 24 were men while only two were women. The majority of the ruptured AAA patients (21) were in the age group of 58 to 78 years with the average age being 72.3 years. The average DAAA for this set of patients was 65.7 mm with the highest DAAA being 82.0 mm. All ruptured AAA patients were ex-smokers and hypertensive with only 9 showing occurrence of diabetes as well. The mean, range and standard deviations for the quantifiable numerical attributes in the ruptured group of the dataset are shown in Table 2. A stratified 10-fold cross validation method was used to measure model performance using four different classifiers.

There are two evaluation phases in machine learning. The first phase is to choose the most effective method to measure the performance during training. The second phase is to assess the performance for the final outcome of the trained model. The present dataset is considered an unbalanced distribution model with limited data. Unbalanced datasets are those in which there are large differences between large samples of negative data and small samples of positive data²⁰.

Classifier performance was tested using the cross-validation method. Cross validation measures the expected error in the number of misclassifications made over the total number of examples that used the performance of the classifier. *K*-fold cross validation splits the original dataset into *k* uniformly sized subsamples. Next, the classifier is applied during *k* iterations that involve training and testing. Within each iteration, one subsample is named the test set and the rest are designated for training. The holdout set is used to calculate the

error rate and k error estimations are averaged to obtain an overall error rate. The most common used value of k is 10, i.e. 10-fold cross validation, which splits the data into 10 subsamples: 9 for training and 1 for testing. This process was used to compare the performance of four classifiers: Naïve Bayes (NB), Decision Trees (DT), Support Vector Machines (SVM), and Logistic Regression (LR). The algorithms were evaluated by metrics such as classification accuracy, AUC, and F measures. A description of the algorithms and metrics is provided in Appendix A.

Receiver Operator Characteristic (ROC) Analysis

ROC curves are an effective way to visualize the performance of classifiers. Provost et al.²¹ argued that the application of classification accuracy may not be a fair indicator of performance and recommended the use of ROC curves. The accuracy index provides a simple numerical measurement for the overall efficiency of the classifier. The ROC plot displays the ratio between specificity and sensitivity through a single threshold measurement²². It is a plot in which the x-axis represents (1-specificity) and the y-axis represents the sensitivity of the classifier. The plot measures the performance of the classifier by means of the true positive fraction and false positive fraction at all possible threshold scores.

The perfect ROC curve is the one that arches up to the top left hand corner of the plot. As the slope of the curve is closer to a 45-degree line, the classifier is less accurate. A simple way to represent the ROC performance with a single metric is the AUC, which can be used to analyze multiple classifiers as the performance is measured independently of the chosen threshold. The perfect classifier has an AUC of 1.0²³.

Feature Selection

A wrapper based feature selection algorithm was applied to each of the four classifiers. This method yielded the most significant attributes (i.e., geometric indices) for each classifier.

Results

A stratified 10-fold cross validation method was used to measure model performance using the four different classifiers. As shown in Table 3, NB achieved the best performance amongst the classifiers with a classification accuracy of 95.2%, an AUC of 0.974, and an F-measure of 0.952. The results emphasize the outcomes of classes U and R through the confusion matrix. As the goal was to predict AAA rupture risk, the performance of predicting the true positive instances was the key measurement. LR showed a classification accuracy of 94.2%, F-measure of 0.942, and an AUC of 0.885. The least effective classifier in the investigation was SVM with a classification accuracy of 93.9%, F-measure of 0.941, and an AUC of 0.844. A standard confusion matrix key is presented in Table 4 while detailed output parameters on the individual confusion matrices are described in Table 5. The confusion matrices show that the four classifiers are able to classify the normal aorta patients well with almost no differences in the accuracy between them. As this is a 3-class problem, the accuracy of identifying each class (U, R or N) is important. To this end, the NB classifier is capable of classifying more of the U class even as the R classification is the

same for the NB, LR and SVM classifiers. This is also reflected in the overall accuracy of classification in which the NB outperformed the other classifiers.

The DT algorithm, illustrated in Figure 2, shows the sequence of attributes that must be considered to assess rupture risk with this machine learning algorithm. Neck length is the most significant geometry index, followed by DAAA, D3, G, D, R, and D1c, in decreasing order of importance. The decision tree showed a classification accuracy of 92.0%, F-measure of 0.918, and ROC AUC of 0.831. A comparison of the ROC curves for each classifier is shown in Figure 3. The NB classifier shows the highest AUC among the four classifiers with 0.974. A wrapper based feature selection algorithm was applied using each classifier to evaluate a subset of features that are significant in each method. Table 6 lists the features for each classifier; this algorithm results in the best model inclusive of the most significant attributes (geometric and clinical).

Discussion

The goal of this work was to identify a suitable machine learning classifier that can be used at the bedside to assess AAA rupture risk in the Asian patient population. To this end, four classifiers were tested for performance in the classification of ruptured and unruptured AAA based on existing medical records and abdominal CT images of 312 subjects. As seen from the metrics of each classifier, the NB algorithm is the best performer.

Patients with hypertension and a smoking history, along with being Caucasian, exhibit the highest risk for AAA rupture. Prior investigations into Caucasian patients reported that AAA wall surface curvature is highly correlated with wall stress¹⁸, while AAA sac length is the most significant geometric index contributing to AAA rupture risk assessment¹⁶. These works suggest that aneurysmal sac shape and size play a significant role in the progression of the aneurysm to rupture and must be considered in any assessment of rupture risk.

In the present work, the DT classifier with feature selection showed that AAA neck length is the most significant geometric index correlated to rupture, followed by the maximum transverse diameter and the diameter of the right iliac artery, in order of importance. Noteworthy is that these geometric measures are significantly different in Asians compared to Caucasian patients. While DT would be a suitable tool to aid clinicians at the bedside given the graphical manner in which the testing outcome is obtained, it was the third most effective classifier in this study with a classification accuracy of 92.0%.

Comparing the wrapper based feature selection outcomes across classifiers shows that the maximum transverse diameter is the common marker of AAA rupture risk amongst all classifiers. Nevertheless, features like neck length, the diameter of the right iliac artery, and the distance from the lowest caudal artery to the iliac bifurcation are also important in discriminating between ruptured and unruptured AAA. The models predicted by feature selection are consistent with the morphological differences seen in Asian aortas compared to their Caucasian counterparts^{24,25}.

The NB classifier with wrapper based feature selection was the best performing model consisting of the following attributes as the most significant factors correlated with AAA

rupture risk: neck length, right iliac artery diameter, maximum transverse diameter, the distance between the caudal artery and iliac bifurcation, and previous dissection of the aorta. Among the patient history attributes, smoking was not a significant discriminator amongst any of the classifiers. Clinically, smoking is seen to be highly correlated with AAA incidence in America and Europe. Given the ethnicity of the population studied, we infer that smoking may not have a strong correlation with risk of AAA rupture. The NB and LR classifiers show that previous dissection of the aorta is an important discriminator of ruptured and unruptured AAA.

This work is subject to several important limitations. The attributes of the dataset used in this work are known as 1D geometric indices. However, the geometry of an AAA can be quantified with 2D and 3D indices such as curvature and volume, which may also play a significant role in rupture risk. Future work will incorporate those additional attributes that will likely improve the accuracy of the classifiers and help clinicians in their decision making. The dataset is also subject to imputation, which means that there were missing values that were estimated using the mean of the distribution. A larger dataset is also needed to build a definitive predictive model that can be a standard for patient care.

In this work, four machine learning algorithms were applied to classify a retrospectively acquired patient dataset into normal aorta, unruptured or ruptured AAA on the basis of 17 attributes obtained from existing medical records and measured from segmented standard of care abdominal computed tomography images. Overall, the Naïve Bayes algorithm yielded the best performance with a 95.2% classification accuracy based on the following five attributes as the most statistically significant discriminators of AAA rupture after feature selection: (i) neck length, (ii) right iliac artery diameter, (iii) maximum transverse diameter, (iv) the distance between the caudal artery and iliac bifurcation, and (v) previous dissection of the aorta. The clinical implications of this outcome can be addressed in two ways. The Naïve Bayes classifier can be applied to prospectively acquired data to predict the likelihood of aneurysm rupture during patient follow ups. Next, it would be necessary to estimate the attributes implicated in rupture risk beyond just maximum aneurysm diameter.

Acknowledgments

This work has been funded by a Research Student Scholarship from the Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore to Tejas Canchi, and a U.S. National Institutes of Health award (R01HL121293) to Ender A. Finol. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. National Institutes of Health.

Appendix A

Description of Machine Learning Algorithms used

Decision Trees

Decision/Classification trees are represented as a flowchart that illustrates optional moves from a series of decisions. They can be used to represent classification as well as regression analysis. A classification tree is used to label the class that depends on the value of the attribute. Each internal node denotes an attribute test which controls node division. Terminal

nodes are assigned the status of the class variable being studied. The growing structure stops when all the attributes belong to the same target class or when maximum tree depth is reached.

Naïve Bayes

Naïve Bayes is a classification method based on the concept of conditional probabilities. It assumes that all factors that affect the value of the target variable are independent of each other. For each attribute, it calculates the probability, conditional on the target variable. This is done for all the attributes in the training set. Thereafter, a joint conditional probability is estimated and eventually, using the Bayes' rule, the conditional probability of the target variable is calculated. This is repeated for all the factors/attributes. The Naïve Bayes is a suitable classifier for use when there are several nominal factors in the problem.

Support Vector Machines

Support vector machines are a type of machine learning method where the algorithm tries to find the vectors in other words, the distance between the boundary separating two classes and the points depicting values of each factor on either side. The distance between the boundary and the closest point is called the margin. The most optimal boundary is the line (in a binary class problem) that has the highest margin.

Logistic Regression

Logistic regression is a machine learning classifier which is used when the target variables are binary. It can be used to describe the relationship between the class variable and one or many independent variables that may be numerical or nominal in nature. The output is generally a 0 or a 1. Logistic regression generates a probabilistic function that describes whether each variable belongs to one or the other binary state of the class variable. The process can be described in graphically in terms of a sigmoid function. The method can also be extended to cases where there are more than 2 categories. This is called multinomial logistic regression.

Performance Metrics

Confusion Matrix

A confusion matrix is a table containing information that assesses the performance of a classifier on the test set. In binary classification, data is separated into two sets; positive (P) and negative (N). Therefore, the confusion matrix consists of four types of results: two groups of correct classifications as true positive (TP) and true negative (TN), and two incorrect classification groups as false positive (FP) and false negative (FN). The total instances of the four groups equals the number of test examples. The groups are defined as follows:

- TP: the target was predicted yes and the actual class is yes.
- TN: the target was predicted no and the actual class is no.
- FP: when the system predicted yes and the actual class is no.

- FN: when the system predicted no and the actual class is yes.

Accuracy, Sensitivity, Specificity, and Precision

These evaluation metrics provide a simple method for comparing the performance of various classifiers. Accuracy calculates the percentage of accurately classified instances and can be expressed in terms of the previously defined groups according to Eq. (1).

$$Accuracy = \frac{TN + TP}{FN + TN + TP + FP} \quad (1)$$

Sensitivity (known as the true positive rate or recall) calculates the percentage of positive instances being accurately recognized as positive, given by Eq. (2).

$$Sensitivity = \frac{TP}{FN + TP} \quad (2)$$

Specificity (known as the true negative rate) calculates the percentage of negative instances being accurately recognized as negative, according to Eq. (3).

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

Precision calculates the extracted instances that are related to each other from the dataset, given by Eq. (4).

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

F measure

F measure is used in the evaluation of classifiers in unbalanced datasets; it incorporates precision and recall into a single metric. It can be used effectively to represent the true positive to the arithmetic mean of predicted positives and real positives, according to Eq. (5).

$$F\ measure = \frac{2*Precision*Sensitivity}{Precision + Sensitivity} \quad (5)$$

Nomenclature and Units (if applicable)

U	Unruptured AAA
R	Ruptured AAA

N	Normal aorta without aneurysm
A	Age (years)
G	Gender (M or F)
R	Race
S	Smoker
ES	ex-smoker
D	Diabetes (Y/N)
Hyp	Hypertension (Y/N)
PDA	Previous dissection of aorta (Y/N)
BMI	Body Mass Index (kg/m ²)
D1a	Proximal neck diameter aortic diameter at level of lowest renal artery (mm)
D1c	Aortic diameter at start of aneurysm (mm)
NL	Neck Length (mm)
DAAA	AAA maximum diameter (mm)
L1	Aortic length from lowest renal artery to bifurcation (mm)
D2	Maximum diameter of left common iliac artery (mm)
D3	Maximum diameter of right common iliac artery (mm)
D4	Diameter at the aortic bifurcation (mm)
DT	Decision Tree
LR	Logistic Regression
NB	Naïve Bayes
SVM	Support Vector Machines

References

- [1]. Vorp DA. Biomechanics of abdominal aortic aneurysm. J Biomech. 2007 12 31;40(9):1887–902. [PubMed: 17254589]
- [2]. Participants TU. Mortality results for randomised controlled trial of early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms. The Lancet. 1998 11 21;352(9141):1649–55.
- [3]. Fillinger MF, Racusin J, Baker RK, et al. Anatomic characteristics of ruptured abdominal aortic aneurysm on conventional CT scans: implications for rupture risk. J Vasc Surg. 2004 6 30;39(6): 1243–52. [PubMed: 15192565]

- [4]. Vorp DA, Raghavan ML, Webster MW. Mechanical wall stress in abdominal aortic aneurysm: influence of diameter and asymmetry. *J Vasc Surg.* 1998 4 30;27(4):632–9. [PubMed: 9576075]
- [5]. Fillinger MF, Marra SP, Raghavan ML, et al. Prediction of rupture risk in abdominal aortic aneurysm during observation: wall stress versus diameter. *J Vasc Surg.* 2003 4 30;37(4):724–32. [PubMed: 12663969]
- [6]. Di Martino ES, Bohra A, Geest JP, et al. Biomechanical properties of ruptured versus electively repaired abdominal aortic aneurysm wall tissue. *J Vasc Surg.* 2006 3 31;43(3):570–6. [PubMed: 16520175]
- [7]. Raghavan ML, Kratzberg J, de Tolosa EM, et al. Regional distribution of wall thickness and failure properties of human abdominal aortic aneurysm. *J Biomech.* 2006 12 31;39(16):3010–6. [PubMed: 16337949]
- [8]. Sacks MS, Vorp DA, Raghavan ML, et al. In vivo three-dimensional surface geometry of abdominal aortic aneurysms. *Ann Biomed Eng.* 1999 7 1;27(4):469–79. [PubMed: 10468231]
- [9]. Howard DP, Banerjee A, Fairhead JF, et al. Population-Based Study of Incidence of Acute Abdominal Aortic Aneurysms with Projected Impact of Screening Strategy. *J Am Heart Assoc.* 2015 8 25;4(8): e001926. [PubMed: 26289347]
- [10]. Canchi T, Kumar SD, Ng EY, et al. A review of computational methods to predict the risk of rupture of abdominal aortic aneurysms. *BioMed Res. Int.* 2015 10 5;2015.
- [11]. Jordanski M, Radovic M, Milosevic Z, et al. Machine Learning Approach for Predicting Wall Shear Distribution for Abdominal Aortic Aneurysm and Carotid Bifurcation Models. *IEEE J. Biomed. Health. Inform.* 2016 12 14.
- [12]. Filipovic N, Ivanovic M, Krstajic D, et al. Hemodynamic flow modeling through an abdominal aorta aneurysm using data mining tools. *IEEE Trans. Inf. Technol. Biomed* 2011 3;15(2):189–94. [PubMed: 21134818]
- [13]. Soudah E, Rodriguez JF, Lopez R. Mechanical stress in abdominal aortic aneurysms using artificial neural networks. *J Mech. Med. Biol* 2015 6;15(03):1550029.
- [14]. Va Guillermo, José Alberto Va, Eduardo S, et al. Statistical Analysis for Rupture Risk Prediction of Abdominal Aortic Aneurysms (AAA) Based on Its Morphometry. *J Mech. Med. Biol* 2017 3 2:1750065.
- [15]. Izbicki R, Lee AB, Finol EA. Assessment of Aortic Aneurysm Rupture Risk. *arXiv preprint arXiv:1106.0545.* 2011 6 3.
- [16]. Shum J Risk assessment of abdominal aortic aneurysms by geometry quantification measures, Ph. D. thesis, Carnegie Mellon University, 2011.
- [17]. Shum J, Martufi G, Di Martino E, et al. Quantitative assessment of abdominal aortic aneurysm geometry. *Ann. Biomed. Eng* 2011 1 1;39(1):277–86. [PubMed: 20890661]
- [18]. Giannoglou G, Giannakoulas G, Soulis J, et al. Predicting the risk of rupture of abdominal aortic aneurysms by utilizing various geometrical parameters: revisiting the diameter criterion. *Angiology.* 2006 Aug;57(4):487–94. [PubMed: 17022385]
- [19]. Hall M, Frank E, Holmes G, et al. The WEKA data mining software: an update. *SIGKDD Explor.* 2009 11 16;11(1):10–8.
- [20]. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One.* 2015 3 4;10(3) : e0118432. [PubMed: 25738806]
- [21]. Provost FJ, Fawcett T, Kohavi R. The case against accuracy estimation for comparing induction algorithms. In: *International Conference on Machine Learning*, Madison, WI, Jul 22-Jul 24, 1998 Vol. 98, pp. 445–453.
- [22]. Fawcett T An introduction to ROC analysis. *Pattern Recognit. Lett.* 2006 6 30;27(8):861–74.
- [23]. Metz CE. Basic principles of ROC analysis. In *Semin. Nucl. Med* 1978 10 1 (Vol. 8, No. 4, pp. 283–298). WB Saunders.
- [24]. Cheng SW, Ting AC, Ho P, et al. Aortic aneurysm morphology in Asians: features affecting stent-graft application and design. *J Endovasc Ther.* 2004 12;11(6):605–12. [PubMed: 15615550]
- [25]. Mladenovic AS, Markovic ZZ, Hyodoh HH. Anatomic differences of the distal aorta with dilatation or aneurysm between patients from Asia and Europe as seen on CT imaging. *Eur. J Radiol* 2012 9 30;81(9):1990–7. [PubMed: 21658872]

Figure 1:
Schematic of the eight geometric indices measured in each AAA after image analysis of the standard of care CT.

Figure 2:
Decision Tree schematic ensuing from the application of this machine learning method to the testing subset.

Figure 3:
ROC curves for the four classifiers.

Table 1:

Mean, standard deviation, and range of the numerical attributes in the full dataset.

Attribute	Range	Mean $\pm$ Standard Deviation
Age (Yrs.)	23.0 – 99.0	68.6 $\pm$ 13.8
D1a (mm)	17.2 – 26.1	21.6 $\pm$ 1.6
D1c (mm)	15.8 – 26.6	22.4 $\pm$ 2.0
NL (mm)	0.0 – 40.0	11.9 $\pm$ 12.5
DAAA (mm)	17.2 – 82.0	37.3 $\pm$ 17.2
L1 (mm)	96.0 – 146.0	120.2 $\pm$ 11.3
D2 (mm)	8.0 – 29.0	15.1 $\pm$ 3.7
D3 (mm)	8.0 – 28.0	15.0 $\pm$ 3.7
D4 (mm)	14.9 – 27.1	22.3 $\pm$ 2.0
BMI (kg/m ²)	11.7 – 49.6	23.5 $\pm$ 4.3

Table 2:

Mean, standard deviation, and range of the numerical attributes in the ruptured group of the dataset.

Attribute	Range	Mean $\pm$ Standard Deviation
Age (yrs.)	58 – 88	72.3 $\pm$ 8.2
D1a (mm)	19.8 – 23.6	21.6 $\pm$ 1.3
D1c (mm)	19.9 – 26.1	23.3 $\pm$ 1.6
NL (mm)	12 – 28	20.0 $\pm$ 3.9
DAAA (mm)	53 – 82	65.7 $\pm$ 8.2
L1 (mm)	103 – 143	122.5 $\pm$ 12.1
D2 (mm)	16 – 29	20.7 $\pm$ 3.5
D3 (mm)	15 – 28	21.1 $\pm$ 3.4
D4 (mm)	20.9 – 27.1	24.6 $\pm$ 1.3
BMI (kg/m ²)	16.2 – 37.8	25.4 $\pm$ 4.4

Table 3:

Comparison of classifier performance metrics.

Classifier/Performance Metrics	Classification Accuracy	F Measure	ROC AUC
DT	92.0%	0.918	0.831
NB	95.2%	0.952	0.974
LR	94.2%	0.942	0.885
SVM	93.9%	0.941	0.844

Table 4:

Standard confusion matrix for a machine learning classifier.

	Negative (Correct Class)	Positive (Correct Class)
Negative (Prediction)	True Negative	False Negative
Positive (Prediction)	False Positive	True Positive

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5:

Confusion matrices for each classifier.

	N	U	R
N	158	0	0
U	0	120	8
R	0	7	19
(a) NB classifier confusion matrix			

	N	U	R
N	158	0	0
U	0	117	11
R	0	14	12
(b) DT classifier confusion matrix			

	N	U	R
N	157	0	1
U	4	118	6
R	0	7	19
(c) LR classifier confusion matrix			

	N	U	R
N	158	0	0
U	0	116	12
R	0	7	19
(d) SVM classifier confusion matrix			

Table 6:

Models generated after wrapper based feature selection for each classifier, yielding the most statistically significant attributes in order of importance.

NB	SVM	LR	DT
PDA	A	R	NL
NL	DAAA	PDA	DAAA
DAAA	L1	NL	D3
L1	-	DAAA	-
D3	-	D3	-