Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2020 February ; 11317: . doi:10.1117/12.2549277.

A machine learning approach for abdominal aortic aneurysm severity assessment using geometric, biomechanical, and patient-specific historical clinical features

Golnaz Jalalahmadi¹, María Helguera^{1,2}, Cristian A. Linte^{1,3}

¹Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, USA

²Instituto Tecnológico José Mario Molina Pasquel y Henríquez - Unidad Lagos de Moreno, Jalisco, México

³Biomedical Engineering Department, Rochester Institute of Technology, Rochester, USA

Abstract

Recent studies monitoring severity of abdominal aortic aneurysm (AAA) suggested that reliance on only the maximum transverse diameter (D_{max}) may be insufficient to predict AAA rupture risk. Moreover, geometric indices, biomechanical parameters, material properties, and patient-specific historical data affect AAA morphology, indicating the need for an integrative approach that incorporates all factors for more accurate estimation of AAA severity. We implemented a machine learning algorithm using 45 features extracted from 66 patients. The model was generated using the J48 decision tree algorithm with the aim of maximizing model accuracy. Three different feature sets were used to assess the prediction rate: i) using D_{max} as a single-feature set, ii) using a set of all features, and, lastly iii) using a feature set selected via the BestFirst feature selection algorithm. Our results indicate that BestFirst feature selection yielded the highest prediction accuracy. These results indicate that a combination of several specific parameters that comprehensively capture AAA behavior may enable a suitable assessment of AAA severity, suggesting the potential benefit of machine learning for this application.

Keywords

Abdominal aortic aneurysm; machine learning; feature selection; BestFirst; geometrical indices; biomechanical parameters; clinical data; accuracy; sensitivity; specificity

1. INTRODUCTION

Abdominal aortic aneurysm (AAA) is an irreversible expansion of the infra-renal aorta at least 1.5 times of a healthy aortic diameter. In light of its high mortality rate, this condition is becoming one of the leading causes of death worldwide [1–3]. The current clinical guidelines consider only the maximum transverse diameter (D_{max}) of AAA (5.5 cm for men, and 5 cm for women) to monitor abdominal aortic aneurysm (AAA) severity, risk of rupture, and need for repair [4, 5]. Nevertheless, it has been reported that patients featuring a smaller

diameter than the established D_{max} criterion have ruptured, while patients with $D_{max} > 9$ cm have survived without need for surgery or repair [6].

Moreover, it has been shown that in addition to geometric indices such as D_{max} , a variety of biomechanical parameters and material properties affect AAA morphology and behavior [3, 7–9]. Wall stress distribution and mainly the peak wall stress - which is directly related to the failure site on the AAA surface - computed using finite element analysis (FEA) have been helpful to comprehensively consider the effects of all mentioned parameters [10–12]. Additionally, it has been illustrated that patient-specific historical data, such as the presence of clinical pre-existing conditions or risk factors, such as smoking, hypertension, or chronic kidney disease (CKD) may increase AAA severity and risk of rupture [6, 13].

Therefore, an all-inclusive predictive model approach that incorporates not only geometric and biomechanical factors, but also patient-specific historical clinical data, may provide a more accurate appraisal of a patient's AAA severity.

To this extent, machine learning (ML) algorithms can play a strong role in management of AAA in a patient-specific manner utilizing all various data sources mentioned above. However, ML methods have rarely been used to study AAA, and even more so to predict in-hospital mortality [13–15] of patients or to help select the best set of geometrical indices to study AAA [16–19].

In the current study, our goal is to merge patient-specific data from three different sources - geometrical indices, biomechanical parameters and clinical factors - into a ML model, with the aim of maximizing its accuracy to predict the severity of individual AAA. To the best of our knowledge, this is one of the few studies to integrate geometric, biomechanical, and patient-specific historical data into a machine learning model for AAA severity assessment, therefore suggesting the potential benefit of employing machine learning techniques to better assess and characterize AAA severity.

2. METHODOLOGY

2.1 Datasets

Contrast enhanced computed tomography angiography (CTA) images of 66 patients were retrospectively obtained from the Division of Vascular Surgery at the University of Rochester following approval by the Human Subject Review Board. The images were used to calculate 26 geometrical indices and 4 biomechanical parameters, details of which are available in [11, 20, 21]. Moreover, 15 patient-specific clinical factors, consisting of patient demographics, habitual information such as risk factors, and other clinical pre-existing conditions, were also extracted from the clinical dataset.

We grouped patients according to their rupture status obtained from the clinical inventory into two groups: **Group 1** (the “at risk of rupture” group, labeled as 1), consisted of patients who underwent interventional AAA repair due to a high risk of rupture; and **Group 2** (the “not at risk of rupture” group, labeled as 0), consisting of patients who did not receive an elective repair and were not in danger of rupture.

2.2 Methods

We devised and implemented a machine learning algorithm using the 45 geometric, biomechanical and clinical features (Figure 1). We used the WEKA (University of Waikato, Hamilton, New Zealand), an open source machine learning software, to perform feature scaling, feature selection, model training, testing and cross validation, and to calculate the accuracy of the model [22] (Figure 1).

Because of their nature, different features have a different range of values that, in return, can affect the speed and performance of the ML algorithm. To increase the speed and accuracy of the model, feature scaling and mean normalization were performed in a pre-processing step. As such, we first identified the variance of each feature sample relative to its mean across the population, then normalized its deviation by the standard deviation of the feature across the dataset (Eq. 1). This process ensured that all normalized features were in the range of $-1 \leq x_{i-new} \leq 1$.

$$x_{i-new} = \frac{x_i - \bar{x}}{s_i} \quad (\text{Eq. 1})$$

where x_{i-new} is the value of feature i , $\bar{x}$ is the mean of all the values of feature i , and s_i is the standard deviation of feature i .

Three different feature sets were used in the proposed algorithm. First we used all 45 mentioned features. Then we resumed to performing a feature selection using BestFirst - a greedy stepwise backward feature selection method [6, 13, 22]. The purpose of feature selection was to identify the features that discriminate the AAA population into the classified groups more accurately, to reduce overfitting, and improve the overall classifier performance [13, 18]. Finally, we used the D_{max} criterion alone, which has been used as the standard clinical criterion, as the only feeding feature to the ML model.

Our proposed model was generated using the J48 decision tree algorithm whose output consists of a patient's classification as belonging to one of two groups: Group 1 - patients at risk of rupture, labeled as 1, or Group 2 - patients not in danger of rupture, labeled as 0. J48 is a decision tree algorithm that works based on the information gain. It utilizes all the available features and in each iteration, the algorithm selects the feature with the highest information gain, which, in turn, will help to classify the data more precisely than the features with lower information gain. This iterative process continues until all the samples in one subset are classified as part of the same class or until all available features have been exhausted [16, 17, 22].

We evaluated the performance of the training model using a five-fold cross validation study on 55 patients. We then assessed the performance of the model using the remaining 11 patient datasets, none of which were used for either training or cross-validation. The 5-fold cross validation method, which is the standard method to assess the training performance and robustness of machine learning algorithms, divides the whole training set into 5 equal sets of 11 patients, keeping one set out for testing, while the algorithm is trained on the remaining 4 sets. This procedure was therefore repeated 5 times, leading to five accuracy

estimations, which were averaged to represent the final accuracy of the ML model for the cross-validation set [6, 16, 17]. The performance rate of the model was then computed by applying the trained model on the testing set.

The evaluation of the model was based on sensitivity (true positive prediction rate), specificity (true negative prediction rate), and accuracy (true positive + true negative prediction rate). The ground truth patient classification was established based on the available clinical data that clearly indicated whether the patient underwent repair (labeled as “1”) or not (labeled as “0”) [13, 22]. Moreover, Kappa statistics - a measure of how randomly the results are occurring, which should always be less than or equal to 1- was also used to compare the performance of ML model for different feature sets. If Kappa is 0, it indicates that the results are highly unreliable, while a K value closer to 1 implies that the results are reliable, and the predicted accuracy can be trusted [13,22].

3. RESULTS

First, we implemented the machine learning algorithm using all features and achieved a cross-validation accuracy of 56% and a test set accuracy of 45%. In the effort to further improve prediction error, we first performed a feature selection pre-processing step prior to feeding the features into the machine learning algorithm. Following BestFirst feature selection, the model yielded a prediction accuracy of 76% on the cross-validation set and 82% on the test set, Table 1 lists the summary of the results. Next, we assessed the prediction rate of the D_{max} criterion alone, by using D_{max} as the sole feature in the proposed algorithm. This single feature approach yielded a 71% accuracy for the cross-validation and 64% accuracy for the test set.

Using BestFirst, thirteen features were selected: D_{max} (maximum transverse diameter), L_{neck} (length of the AAA neck), BL (Bulge location), γ (ratio of volume of intraluminal thrombus to the AAA volume), MLN (a measure of irregularities on the AAA surface), peak wall stress location (of AAA wall where the peak wall stress happens [23]), presence of HTN (hypertension), presence of Afib (Atrial fibrillation), presence of COPD (Chronic Obstructive Pulmonary Disease), presence of neoplasm (abdominal tumor), amount of ACE (angiotensin converting enzyme), smoking status, and gender. More details about these features can be found in [16, 24].

To analyze the data, we used the area under (AUC) the Receiver operating characteristic (ROC) analysis. The ROC is a plot of the true positive rate (sensitivity) against the false positive rate, and AUC's value closer to one indicates optimal performance (Figure 2).

4. DISCUSSION

To the best of our knowledge, this study is one of the very few to incorporate information from several classes of features that may affect AAA severity and rupture risk using ML methods [6]. Moreover, this is the first study to assess model prediction accuracy on a separate dedicated test set.

The J48 decision tree machine learning algorithm was used to investigate which parameters can better predict the AAA rupture severity more accurately. Three different sets of features were used: i) a set of all the 45 features including geometrical, biomechanical, and clinical factors, ii) a set of the most representative and impactful features selected using BestFirst feature selection method, which resulted in the selection of 13 most relevant features, and iii) only D_{max} which has been used as the standard criterion in the monitoring process of AAA.

Table 1 presents a summary of our results. It can be seen that the feature selection method provides higher accuracy for both the cross-validation and test set compared to the other two methods. Using all or only a single feature might lead to model over-fitting or under-fitting. For instance, under-fitting occurred when predicting the patient's at risk of rupture (sensitivity) using only a single feature - D_{max} . While the sensitivity of the BestFirst feature selection model was the highest across all three models, its specificity was lower compared to the specificity of the D_{max} single-feature model. However, from Table 1, it is obvious that using a single feature D_{max} - is not reliable based on the Kappa statistics ($K = 0$, hence highly unreliable results), and low AUC value (on the order of 0.5 for both the cross-validation and test set). An AUC of 0.5 is indicating that D_{max} has a 50% chance to predict risk of the rupture which is as accurate as making the decision by flipping of a coin. While the BestFirst feature selection model has 71% and 75% chance to predict the risk of the rupture for the cross-validation and test set, respectively, which in both cases is much higher than using D_{max} only.

To further validate our results, we ran the machine learning model using peak wall stress (PWS) as the sole feature. The results for both the cross validation and test set were similar to the resulting accuracy, sensitivity, and specificity, when using D_{max} . It has been shown that PWS and D_{max} change concomitantly with one another, i.e., PWS increases with increasing D_{max} [10, 16, 21]. Therefore, using either of these properties as the single feature for the machine learning model would result in a linear fit between the classes (0: un-ruptured or 1: ruptured AAA) and the feature.

In this model, the number of ruptured and un-ruptured AAAs above and below of the median is important. In our database, for both D_{max} and PWS, the number of ruptured and un-ruptured AAAs above and below of their corresponding median stays the same, thus leading to similar results when choosing D_{max} or PWS as the sole feature in the model.

While there has been no work done on a separate test set and therefore there is no statistical data reported, our study showed higher accuracy for all 3 feature sets when we compare our cross-validation results with previous studies [6]. Moreover, the highest value that has been reported in the literature for Kappa is $k = 0.37$ for cross validations [16]. In this study, the Kappa value for cross validation using the BestFirst model was $k = 0.42$, which represents an increase from $k = 0$ for D_{max} , and $k = -0.06$ for the all-features model.

5. CONCLUSION

The main objective of this work was to present an improved and potentially more accurate machine learning model to predict AAA severity. Our results indicate that a combination of several specific parameters that comprehensively capture the AAA behavior is more suitable to assess AAA severity rather than D_{max} or PWS alone. Thirteen parameters, including 5 geometrical indices, 1 biomechanical parameter, and 7 clinical factors, were selected using the BestFirst feature selection model in this work. The combination of these parameters demonstrates the effect of different aspects of geometry, the biology of the tissue, and the health and life-style of the patients on the growth and severity of AAA. In order to have a more accurate monitoring and decision-making procedure for AAA surveillance, consideration of such information is essential.

To the best of our knowledge, this is one of the very few studies to integrate geometric, biomechanical, and patient-specific historical data into a machine learning model for AAA severity assessment, and our results suggest the potential benefit of employing machine learning techniques to better assess and characterize AAA severity supporting better clinical-decision making.

6. REFERENCES

- [1]. Niestrawska JA, Viertler C, Regitnig P et al., "Microstructure and mechanics of healthy and aneurysmatic abdominal aortas: experimental analysis and modelling," *J R Soc Interface*, 13(124), (2016).
- [2]. Ramadan A, Al-Omran M, and Verma S, "The putative role of autophagy in the pathogenesis of abdominal aortic aneurysms," *Atherosclerosis*, 257, 288–296 (2017). [PubMed: 28139205]
- [3]. Wittek A, Derwich W, Fritzen C-P et al., "Towards non-invasive in vivo characterization of the pathophysiological state and mechanical wall strength of the individual human AAA wall based on 4D ultrasound measurements," *ZAngewMathMech*, 1–20 (2018).
- [4]. Vu KN, Y. Kaitoukov F, et al., "Rupture signs on computed tomography, treatment, and outcome of the abdominal aortic aneurysms," 5, 281–93 (2014).
- [5]. Leemans EL, Willems TP, van der Laan MJ et al., "Biomechanical Indices for Rupture Risk Estimation in Abdominal Aortic Aneurysms," *J Endovasc Ther*, 24(2), 254–261 (2017). [PubMed: 27872318]
- [6]. Fernando García-García Eleni Metaxa, Christodoulidis Stergios et al., "Prognosis of abdominal aortic aneurysms: A machine learning-enabled approach merging clinical, morphometric, biomechanical and texture information," *IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, 463–8 (2017).
- [7]. Wilson JS, Bersi MR, Li G et al., "Correlation of Wall Microstructure and Heterogeneous Distributions of Strain in Evolving Murine Abdominal Aortic Aneurysms," *Cardiovasc Eng Technol*, 8(2), 193–204 (2017). [PubMed: 28378165]
- [8]. Stevens RRF, Grytsan A, Biasetti J et al., "Biomechanical changes during abdominal aortic aneurysm growth," *PLoS One*, 12(11), e0187421(2017). [PubMed: 29112945]
- [9]. Reeps C, Maier A, Pelisek J et al., "Measuring and modeling patient-specific distributions of material properties in abdominal aortic aneurysm wall," *Biomech Model Mechanobiol*, 12(4), 717–33 (2013). [PubMed: 22955570]
- [10]. Heng MS, Fagan MJ, Collier JW et al., "Peak wall stress measurement in elective and acute abdominal aortic aneurysms," *J Vasc Surg*, 47(1), 17–22; discussion 22 (2008). [PubMed: 18060730]

- [11]. Jalalahmadi G, Helguera M, Mix DS et al., "Toward modeling the effects of regional material properties on the wall stress distribution of abdominal aortic aneurysms," *Proceeding of Medical Imaging SPIE*, 10578, (2018).
- [12]. Wu W, Rengarajan B, Thirugnanasambandam M et al., "Wall Stress and Geometry Measures in Electively Repaired Abdominal Aortic Aneurysms," *Ann Biomed Eng*, (2019).
- [13]. Monsalve-Torra A, Ruiz-Fernandez D, Marin-Alonso O et al., "Using machine learning methods for predicting inhospital mortality in patients undergoing open repair of abdominal aortic aneurysm," *J Biomed Inform*, 62, 195–201 (2016). [PubMed: 27395372]
- [14]. Karthikesalingam A, Attallah O, Ma X et al., "An Artificial Neural Network Stratifies the Risks of Reintervention and Mortality after Endovascular Aneurysm Repair; a Retrospective Observational study," *PLoS One*, 10(7), e0129024(2015). [PubMed: 26176943]
- [15]. García G, Tapia A, and De Blas M, "Computer-supported diagnosis for endotension cases in endovascular aortic aneurysm repair evolution," *Comput Methods Programs Biomed*, 115(1), 11–9 (2014). [PubMed: 24721658]
- [16]. Shum J, Martufi G, Di Martino E et al., "Quantitative assessment of abdominal aortic aneurysm geometry," *Ann Biomed Eng*, 39(1), 277–86 (2011). [PubMed: 20890661]
- [17]. Lee K, Zhu J, Shum J et al., "Surface curvature as a classifier of abdominal aortic aneurysms: a comparative analysis," *Ann Biomed Eng*, 41(3), 562–76 (2013). [PubMed: 23180028]
- [18]. Urrutia J, Roy A, Raut SS et al., "Geometric surrogates of abdominal aortic aneurysm wall mechanics," *Med Eng Phys*, 59, 43–49 (2018) [PubMed: 30006003]
- [19]. Parikh Shalin A., Gomez Raymond, Thirugnanasambandam Mirunalini et al., "Decision Tree Based Classification of Abdominal Aortic Aneurysms Using Geometry Quantification Measures," *Annals of Biomedical Engineering*, 46(12), 2135–2147 (2018). [PubMed: 30132212]
- [20]. Jalalahmadi G, Helguera M, and Linte CA, "A numerical framework for studying the biomechanical behavior of abdominal aortic aneurysm," *SPIE Proc Medical Imaging*, 101372A, (2017).
- [21]. Jalalahmadi G, Helguera M, Mix DS et al., "(Peak) Wall Stress as An Indicator of Abdominal Aortic Aneurysm Severity," *IEEE Xplore*(18311472), (2018).
- [22]. Witten IH, Frank E, Hall MA et al., [Data Mining: Practical Machine Learning Tools and Techniques] Elsevier, (2017).
- [23]. Mcgloughlin TM, Doyle B, Callanan A et al., "A Finite Element Analysis Rupture Index (FEARI) as an additional tool for abdominal aortic aneurysm rupture," *Vascular Disease Prevention* 6, 114–121 (2009).
- [24]. Tang A, Kauffmann C, Tremblay-Paquet S et al., "Morphologic evaluation of ruptured and symptomatic abdominal aortic aneurysm by three-dimensional modeling," *J Vasc Surg*, 59(4), 894–902.e3 (2014). [PubMed: 24439318]

Fig 1:

Outline of analysis protocol: CTA images of 66 patients were used to construct the 3-D model of individual AAAs and calculate 26 geometrical and 4 biomechanical patient-specific parameters. Clinical datasets of patients were utilized to include 15 patient-specific clinical factors. Using these features, ML model with different sets of features were applied to predict the rupture severity.

Fig 2. ROC plots for all features, BestFirst and Only D_{max} sets for a) cross-validation sets, and b) test sets. It can be seen that the BestFirst has the largest AUC for both cross-validation and test-set, indicating better performance and higher accuracy.

Table 1:

Model performance using different feature sets

Features type	Kappa statistics	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)
Cross Validation Results					
All Features	-0.06	0.55	25	69	56
D_{max}	0	0.50	0	100	71
BestFirst Feature Selection	0.42	0.71	56	85	76
Test Set Results					
All Features	-0.18	0.55	25	57	45
D_{max}	0	0.50	0	100	64
BestFirst Feature Selection	0.21	0.75	75	86	82