EJVES Short Reports (2018) ■, 1–5

ORIGINAL RESEARCH

Applied Machine Learning for the Prediction of Growth of Abdominal Aortic Aneurysm in Humans

R. Lee [a,*,†], D. Jarchi [b,†], R. Perera [c], A. Jones [a], I. Cassimjee [a], A. Handa [a,‡], D. Clifton [c,‡], on behalf of the Oxford Abdominal Aortic Aneurysm Study, the Oxford Regional Vascular Service Unit

[a] Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK
[b] Department of Engineering Science, University of Oxford, Oxford, UK
[c] Nuffield Department of Primary Care Health, University of Oxford, Oxford, UK

Objective: Accurate prediction of abdominal aortic aneurysm (AAA) growth in an individual can allow personalised stratification of surveillance intervals and better inform the timing for surgery. The authors recently described the novel significant association between flow mediated dilatation (FMD) and future AAA growth. The feasibility of predicting future AAA growth was explored in individual patients using a benchmark machine learning technique.

Methods: The Oxford Abdominal Aortic Aneurysm Study (OxAAA) prospectively recruited AAA patients undergoing the routine NHS management pathway. In addition to the AAA diameter, FMD was systemically measured in these patients. A benchmark machine learning technique (non-linear Kernel support vector regression) was applied to predict future AAA growth in *individual* patients, using their baseline FMD and AAA diameter as input variables.

Results: Prospective growth data were recorded at 12 months (360 ± 49 days) in 94 patients. Of these, growth data were further recorded at 24 months (718 ± 81 days) in 79 patients. The average growth in AAA diameter was 3.4% at 12 months, and 2.8% per year at 24 months. The algorithm predicted the individual's AAA diameter to within 2 mm error in 85% and 71% of patients at 12 and 24 months.

Conclusions: The data highlight the utility of FMD as a biomarker for AAA and the value of machine learning techniques for AAA research in the new era of precision medicine.

Article history: Received 22 January 2018, Revised 15 March 2018, Accepted 25 March 2018,

Keywords: Abdominal aortic aneurysm, Aneurysm progression, Machine learning, Biomarker, Flow mediated dilatation

INTRODUCTION

The clinical management of abdominal aortic aneurysms (AAAs) is defined by three key domains: screening/diagnosis, surveillance, and surgical intervention. With regard to the surveillance of AAAs, it is important to develop tools for the assessment of the likelihood of AAA rupture or for the prediction of future AAA growth. In the setting of clinical research, the true risk of rupture can only be established by allowing AAA rupture without intervention. In comparison, the growth rate of individual AAAs can be ascertained by repeat measurements of the AAA size during surveillance. The average growth rates of AAAs can be observed from cohort studies. However, there is currently no means of predicting the growth of an AAA in individual patients. In a recent survey of international vascular surgery colleagues, "discovering new tests for the prediction of AAA growth" was identified as the top priority for research in AAAs.[1]

In this regard, the novel observation that flow mediated dilatation (FMD, a marker of endothelial function) of the brachial artery is inversely correlated with the rate of future AAA growth, has recently been described. In the study, FMD of the participants was measured at baseline, and they were followed over a 12 month period. There was a significant inverse correlation between baseline FMD and the growth rate recorded over the subsequent 12 months.[2] This highlights the potential utility of FMD as a novel biomarker of AAA progression.

Machine learning techniques are gaining mainstream interest in biomedical research. They are non-linear extensions of standard linear tools from medical statistics. For example, logistic regression maps a vector of values (as

† These authors contributed equally.
‡ These authors are joint senior author.
* Corresponding author.
E-mail address: regent.lee@nds.ox.ac.uk (R. Lee).

https://doi.org/10.1016/j.ejvssr.2018.03.004

Please cite this article in press as: Lee R, et al., Applied Machine Learning for the Prediction of Growth of Abdominal Aortic Aneurysm in Humans, EJVES Short Reports (2018), https://doi.org/10.1016/j.ejvssr.2018.03.004

input) onto a class membership probability (at the output). Logistic regression finds the relationship between the multi-dimensional input and the output (in the range of [0 1] values) using a linear weighted sum of the inputs. Machine learning methods generalise this, by allowing the relationship between the same inputs and outputs to be a non-linear relationship. This allows interactions between input variables to be modelled such that the output value is closer to that which is required.

The role of machine learning has previously been demonstrated in developing prognostic tools for the prediction of cardiovascular disease using ECG data in the China Kadoorie Biobank[3] and other clinical settings.[4] Here the application of a benchmark machine learning technique (non-linear Kernel support vector regression) to predict future AAA growth in *individual* patients is described.

METHODS

This study is based on a prospectively recruited cohort of patients with AAAs (Oxford Abdominal Aortic Aneurysm Study, OxAAA). Every participant gave written consent to take part in the study. The ethics approval reference for this study is SC/0250/13. AAA size data obtained by the National Health Service (NHS) AAA surveillance programme was used. AAA size was measured by the anteroposterior diameter (APD) (outer to outer) on ultrasound. FMD of the *brachial artery* was measured as an additional research assessment in the study participants. Annual AAA % growth was calculated by (ΔAPD/APD at baseline)/(number of days lapsed/365 days).

Full details of patient recruitment and data acquisition of the OxAAA study cohort are as reported recently.[2] A significant correlation between the baseline AAA diameter and AAA growth rate recorded over the subsequent 12 months was observed, but none of the other clinical demographic parameters correlated with future AAA growth. Therefore baseline FMD and AAA diameter were included as the two variables to construct the prediction algorithm. In addition, longer term (24 months) growth data were included in the latest analysis.

Receiver operating characteristic curves (ROC) were plotted first, using two variables (baseline FMD and AAA diameter) to analyse the performance of the generalised linear logistic regression model for discerning growth against a predefined growth rate threshold.

A set of benchmark machine learning techniques was then applied for the prediction of AAA diameter in *individual* patients at 12 and 24 months from baseline. These included non-linear kernel support vector regression (SVR) using two features (FMD, AAA diameter) and hyper-parameter optimisation using nested fivefold cross validations[5,6] (Matlab, V2016b, Natick, MA, USA).

SVR is a regression technique which can be used in the context of linear or non-linear regression, where the linear relation assumption would not be optimal or sufficient to characterise the dynamics of input feature patterns versus model outcome. A Kernel trick can be used in SVR to learn a non-linear function and map feature input into desired model output. Here, non-linear kernel SVR including a Gaussian kernel has been used to non-linearly map the feature input (FMD, AAA size) into the future AAA growth rate.

Table 1. Summary of participant characteristics at the baseline assessment. Q6

Number (male)	94 (82)
Age at consent, years (SD)	74 (8)
AAA size, mm (IQR)	43 (36–48)
Height, m (SD)	1.72 (0.08)
Weight, kg (SD)	83.5 (14)
BMI median (IQR)	27 (24–31)
Blood pressure SBP/DBP, mmHg (SD)	137/77 (15/11)
Smoking status, n (%)	
Current smoker	14 (15)
Past history of smoking (>1 month)	66 (70)
Never smoked	14 (15)
History of ischaemic heart disease, *n* (%)	38 (40)
MI/ACS	33 (35)
Stable angina	18 (19)
Coronary intervention/bypass	34 (36)
History of peripheral arterial disease, *n* (%)	24 (26)
History of cerebral arterial disease, *n* (%)	12 (13)
History of hypertension, n (%)	62 (66)
History of hypercholesterolemia, n (%)	57 (61)
Total cholesterol, mmol/L (IQR)	4 (3.4–5)
High density lipoprotein, mmol/L (IQR)	1.1 (1–1.4)
Low density lipoprotein, mmol/L (IQR)	2.2 (1.7–3.1)
Triglycerides, mmol/L (IQR)	1.3 (0.9–1.9)
History of diabetes mellitus, *n* (%)	14 (15)
HbA1C%, mean (SD)	41 (8)
Oral anti-hyperglycaemics, *n* (%)	11 (12)
Insulin, *n*	0
Chronic kidney disease (eGFR< 60), *n* (%)	21 (22)
Creatinine μmol/L (IQR)	80 (68–96)
Chronic respiratory disease, *n* (%)	15 (16)
Family history of AAA, *n* (%)	20 (21)
History of treated neoplasms, *n* (%)	14 (15)
Regular medication, *n* (%)	
Aspirin	56 (60)
Thienopyridine/ cyclopentyltriazolopyrimidine	14 (15)
Anticoagulants	11 (12)
Statin	71 (76)
β blocker	35 (37)
ACE inhibitor/ARB	60 (64)
C-reactive protein (mg/L, IQR)	2.9 (1.1–7.3)
Median FMD (%, IQR)	2.0 (0.75–4.02)

Note. For variables which demonstrate Gaussian distribution, mean and standard deviation (SD) are presented. For variables which demonstrate non-Gaussian distribution, median and interquartile range (IQR) are presented. AAA = abdominal aortic aneurysm; IQR = interquartile range; BMI = body mass index; SBP = systolic blood pressure; DBP = diastolic blood pressure; MI = myocardial infarction; ACS = acute coronary syndrome; PAD = peripheral arterial disease; TC = total cholesterol; TG = triglycerides; DM = diabetes mellitus; HbA1C = glycated haemoglobin; CKD = chronic kidney disease; eGFR = estimated glomerular filtration rate; ARB = angiotensin II receptor blocker; CRP = C-reactive protein; FMD = flow mediated dilatation of brachial artery.

Please cite this article in press as: Lee R, et al., Applied Machine Learning for the Prediction of Growth of Abdominal Aortic Aneurysm in Humans, EJVES Short Reports (2018), https://doi.org/10.1016/j.ejvssr.2018.03.004