CLASSIFICATION OF AORTIC STIFFNESS FROM EIGENDECOMPOSITION OF THE DIGITAL VOLUME PULSE WAVEFORM

Natalia Angarita-Jaimes[1], *Stephen R. Alty*[2⋆], *Sandrine C. Millasseau*[3] *and Philip J. Chowienczyk*[3]

[1]Heriot-Watt University, Department of Mechanical Engineering, Edinburgh, U.K.
[2]King's College London, Centre for Digital Signal Processing Research, WC2R 2LS, U.K.
[3]GKT School of Biomedical Sciences, St. Thomas' Hospital, London SE1 7EH, U.K.

ABSTRACT

Aortic stiffness as measured by aortic pulse wave velocity (PWV) has been shown to be an independent predictor of Cardiovascular Disease (CVD), however, the measurement of PWV is time consuming. Recent studies have shown that pulse contour characteristics depend on arterial properties such as arterial stiffness. This paper presents a method for estimating PWV from the digital volume pulse (DVP), a waveform that can be rapidly and simply acquired by measuring transmission of infra-red light through the finger pulp. PWV and DVP were measured on 461 subjects attending a cardiovascular prevention clinic at St Thomas' Hospital, London. Using a non-linear Kernel based Support Vector Machine (SVM) classifier, it is possible to achieve results of up to 88% sensitivity and 82% specificity on unseen data. Further, we show that this approach outperforms traditional Artificial Neural Network (ANN) methods. This technique could be employed by health professionals to rapidly diagnose patients' cardiovascular fitness in general practice clinics.

1. INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of mortality in the developed world. An estimated 17 million people die of CVDs every year, in particular heart attacks and strokes (according to the World Health Organization). Factors such as advancing age, heredity or family history, gender and ethnicity are independent causes for cardiovascular disease that science can do little about. However extensive clinical and statistical studies have identified that a substantial number of these deaths (at least one third) can be attributed to preventable major risk factors such as tobacco use, high blood pressure, high cholesterol, diabetes and obesity. Physical inactivity and unhealthy diet are other main contributory factors which increase individual's risk to CVDs. Despite the knowledge of all these factors, many of the deaths and most of the heart attacks are the result of *under-diagnosis* and/or *under-treatment*.

Currently, several methods exist to evaluate CVD risk. One of the most well-known and still widely used is the Framingham risk calculator [1]. The Framingham Heart Study based on a white population from North America has been operational since the 1950s and includes a cohort of 5573 patients. The study found a quantitative relationship between multiple risk factors through a parametric model that significantly predicted the occurrence of several cardiovascular diseases. However, because this model (and others) rely mainly on the measure of these factors (which are time consuming and in some cases invasive) they are impractical for *casual screening*. Therefore, an assessment of total (global) risk based on the summation of all major risk factors would be clinically useful, hence the development of an appropriate and simple method for the evaluation of individual risk of CVD, based on a single measure, needs to be established. In recent studies aimed at achieving this goal, it has been shown that an increase of the stiffness in large arteries is strongly associated with increased CVD risk. A number of indirect and direct measures of large artery stiffness have been proposed and among these various indices, a recognized and reliable method is Pulse Wave Velocity (PWV). PWV is the measurement of the average speed of propagation of the arterial pulse wave in the aorta (the main artery taking blood from the heart to the rest of the body). Ageing, accompanied by an increase in arterial stiffness, leads to an increase in aortic PWV and results [2] have indicated that this measurement taken alone appeared as a strong predictor of cardiovascular mortality with high performance values as assessed by the standard Framingham equations. In a further study [3] on a French population of predominantly hypertensive patients, arterial stiffness measured through PWV predicted the occurrence of CVD more accurately than those provided by classical risk factors assessed through the Framingham or multivariate Cox models [4].

Traditionally, the *carotid-femoral pressure pulse* has been used to evaluate aortic stiffness because the pressure wave (PWV) can be more easily recorded at these two sites and its distance is great enough to allow an accurate calculation of the time interval between the two waves. From this measurement of the PWV, an index has been defined to characterize the shape of the arterial pulse that is known to change with age and certain pathologies. This index, called the Augmentation Index (AIx) measures the interaction of the forward pressure wave with the reflected wave from the distal circulation. Such interaction is measured as the increase in the peak pressure

⋆Please address all correspondence to steve.alty@kcl.ac.uk

Fig. 1. DVP recorded by measuring the transmission of IR light through the finger pulp.

and is used as a surrogate measure of arterial stiffness. In addition to the changes related to age, the AIx varies with the presence of cardiovascular disease and has been shown [5] to be related to classical risk factors; especially the pressure pulse measured at the top of the aorta seems to correlate well. Two major drawbacks of this technique, however, are the expensive equipment needed and the high level of technical expertise required for obtaining an adequate waveform for analysis. The technique used to measure the carotid-to-femoral PWV is *applanation tonometry* which gives a direct measurement of aortic PWV but this is costly. For this reason, and in order to generalize the evaluation of arterial stiffness among primary care physicians, (to make it suitable for use in large clinical studies), a technique that requires no specialized technical skill would be most desirable.

1.1. Digital Volume Pulse Approach

A promising alternative is the volume pulse waveform measured at the finger tip, the so-called Digital Volume Pulse (DVP). The DVP can be rapidly and simply obtained by measuring infrared light absorption of arterial blood in the finger pulp (technically referred to as *photoplethysmography*). The pioneering work of Takazawa et al [6] has shown that the contour of DVP is similar to that of a carotid pressure pulse. Unlike tonometry (used to estimate PWV), photoplethysmography is inexpensive and operator independent (see Fig. 1). Furthermore, Millasseau et al [7, 8] have substantially documented that the pressure pulse and the DVP contain the same information and that both are determined by a direct wave and a reflected wave. For these reasons using features extracted from the DVP waveform to estimate arterial stiffness and hence cardiovascular disease risk is very attractive.

Signal subspace analysis is performed on the DVP waveform data to extract suitable features for classification. Support Vector Machine (SVM) [9, 10] supervised learning technique (introduced in section 3) was then applied to find the best set of features to give good prediction of high and low PWV. These features were also tested using the more traditional Artifical Neural Network (ANN) approach to provide a benchmark for comparison with the SVM method.

2. FEATURE EXTRACTION

Previous work in this field [11] was based on features extracted from the DVP waveform which were, however, *physiologically* motivated. That is to say, they were based on indicators that were related to the physical properties of the aorta and arteries in general. These included the peak-to-peak time of the waveform (generally having two peaks) as this is affected by the stiffness of the blood vessels, also the crest-time of the first peak and the relative amplitudes of the two peaks. In this paper, by contrast, we make no assumptions about the physical nature of the waveform and rely instead on signal processing and decomposition. Exhaustive tests were made on a number of different methods of extracting suitable features from the DVP waveform (and these are presented here [12]). They included Kernel Principal Component Analysis (PCA) [13], Wavelet Packet (WP) decomposition and signal subspace analysis. Ultimately, it was found that a certain range of the eigenvalues of the covariance matrix (formed by the autocorrelation of the DVP waveform with its mean removed) outperformed all the other features and methods by some margin. Briefly, the covariance matrix $\mathbf{A} = \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ can be decomposed using Singular Value Decomposition (SVD) into its *eigenvectors*, $\mathbf{U}$ and $\mathbf{V}$ and *eigenvalues*, $\mathbf{\Sigma}$ where $\tilde{\mathbf{x}}$ is the DVP amplitude with its mean removed. Specifically, the range of eigenvalues $\sigma_3, \cdots, \sigma_9$ inclusive were found empirically to give the best results. Furthermore, this study is based on a much larger data set (by three times) and therefore implies a much greater statistical significance.

3. SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) [9] have received a great deal of attention recently proving themselves to be very effective in a variety of pattern classification tasks. They have been applied to a number of problems ranging from hand-written character recognition, bioinformatics to automatic speech recognition (amongst many others) with a great deal of success. A brief summary of the mathematical theory of SVMs follows, for a complete treatment please see [10]. Consider a binary classification task with a set of linearly separable training samples

$$S = \{ \ (\mathbf{x}_1, y_1) \quad \cdots \quad (\mathbf{x}_m, y_m) \ \}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^d$, i.e., $\mathbf{x}$ lies in a d-dimensional input space, and y_i is the class label such that $y_i \in \{-1, 1\}$. The label indicates the class to which the data belongs. A suitable discriminating function could then be defined as:

$$f(\mathbf{x}) = sgn(\langle \mathbf{w}, \mathbf{x} \rangle + b) \,. \tag{2}$$

Where vector $\mathbf{w}$ determines the orientation of a discriminant plane (or hyperplane), $\langle \mathbf{w}, \mathbf{x} \rangle$ is the inner product of the vectors, $\mathbf{w}$ and $\mathbf{x}$ and b is the *bias* or offset. Clearly, there are an infinite number of possible planes that could correctly classify the training data. Intuitively one would expect the choice of a line drawn through the "middle", between the two classes, to be a reasonable choice. This is because small perturbations of each data point would then not affect the resulting classification. This therefore implies that a good separating plane is one that is more general, in that it is also more likely to accurately classify a new set of, as yet unseen, test data. It is thus the object of an optimal classifier to find the best *generalising hyperplane* that is equidistant or furthest from each set of points. The set of input vectors is said to be *optimally separated* by the hyperplane if they are separated without error and the distance between the closest vector and the hyperplane is maximal. This approach leads to the determination of just one hyperplane.

3.1. Soft-Margin Classifier

Typically, real-world data sets are in fact linearly inseparable in input space, this means that the maximum margin classifier approach is no longer valid and a new model must be introduced. This means that the constraints need to be relaxed somewhat to allow for the minimum amount of misclassification. Therefore the points that subsequently fall on the wrong side of the margin are considered to be errors. They are, as such, apportioned a lower influence (according to a preset *slack variable*) on the location of the hyperplane. In order to optimise the soft-margin classifier, we must try to maximise the margin whilst allowing the margin constraints to be violated according to the preset slack variable ξ_i. This leads to the minimisation of: $\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\xi_i$ subject to $y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$ for $i = 1, \ldots, m$. The minimisation of linear inequalities is typically solved by the application of Lagrangian duality theory [10]. Hence, forming the primal Lagrangian,

$$L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\beta_i\xi_i - \sum_{i=1}^{m}\alpha_i\left[y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) - 1 + \xi_i\right], \quad (3)$$

where α_i and β_i are independent *Lagrangian multipliers*. The *dual-form* can be found by setting each of the derivatives of the primal to zero thus, $\mathbf{w} = \sum_{i=1}^{m} y_i\alpha_i\mathbf{x}_i$ and $\sum_{i=1}^{m} y_i\alpha_i = 0$, then re-substituting into the primal thus,

$$L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle \ . \quad (4)$$

Interestingly, this is the same result as for the maximum margin classifier. The only difference is the constraint $\boldsymbol{\alpha} + \boldsymbol{\beta} = C$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta} \geq 0$, hence $0 \leq \boldsymbol{\alpha} \leq C$. This implies that the value C, sets an upper limit on the Lagrangian optimisation variables α_i, this is sometimes referred to as the *box constraint*. The value of C offers a trade-off between accuracy of data fit and regularisation, the optimum choice of C will depend on the underlying nature of the data and is usually determined by *cross-validation* (whereby the classifier is tested on a section of *unseen* data). These equations can be solved mathematically using Quadratic Programming (QP) algorithms. There are many online resources of such algorithms available for download, see website referred to in [10] for an up to date listing.

3.2. Kernel Functions

It is quite often the case with real-world data that not only is it linearly non-separable but it also exhibits an underlying non-linear characteristic nature. Kernel mappings offer an efficient solution by non-linearly projecting the data into a higher dimensional feature space to allow the successful separation of such cases. The key to the success of Kernel functions is that special types of mapping, that obey Mercer's Theorem, offer an *implicit* mapping into feature space. This means that the explicit mapping need not be known or calculated, rather the inner-product itself is sufficient to provide the mapping. This reduces the computational burden dramatically and in combination with SVM's inherent generality largely mitigates the so-called "*curse of dimensionality*". Further, this means that the input feature inner-product can simply be substituted with the appropriate Kernel function to obtain the mapping whilst having no effect on the Lagrangian optimisation theory. Hence, the relevant classifier function then becomes:

$$f(\mathbf{x}) = sgn\left[\sum_{i=1}^{nSVs} y_i\alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right] \quad (5)$$

where $nSVs$ denotes the number of support vectors, y_i are the labels, α_i are the Lagrangian multipliers, b the bias, $\mathbf{x}_i$ the *Support Vectors* previously identified through the training process, and $\mathbf{x}$ the test data vector. The use of Kernel functions transforms a simple linear classifier into a powerful and general non-linear classifier. There are a number of different Kernel functions available [10], however, one of the most consistently useful is the *Gaussian Radial Basis Function* (RBF) Kernel, given by

$$K(\mathbf{x}_i, \mathbf{x}) = exp(-\|\mathbf{x}_i - \mathbf{x}\|^2/2\sigma^2) \ . \quad (6)$$

It was found that using this Kernel gave the best performance for the classifier.

4. RESULTS

Here a binary classifier based upon the OSU SVM toolbox for MATLAB® [14] was employed, using a Gaussian RBF Kernel, in combination with a soft margin classifier. After much experimentation a constraint factor of $C = 1000$ and Gaussian RBF Kernel with $\sigma = 1.2$ was found to be optimum and used to train and test the classifier to obtain the results below. A cohort of 461 subjects recruited from a south-east London hypertension clinic, with complete DVP waveform data and PWV measurements were used in this study. The PWV values were grouped into low and high values. Studies [15] have shown that values of $< 9\text{ms}^{-1}$ are low risk and values $> 11\text{ms}^{-1}$ indicate a high CVD risk category. Moreover, the mean PWV value of our cohort was found to be

	SVM Classifier		ANN Classifier	
	High	**Low**	**High**	**Low**
Target High	88%	12%	79%	21%
Target Low	18%	82%	24%	76%

Table 1. SVM vs. ANN classification of PWV

around 10ms^{-1}, hence, a binary target label was determined according to this threshold. The cohort was gapped to remove those subjects with PWV of between 9 and 11ms^{-1} to avoid ambiguity of classification. The remaining 315 records underwent multi-folded cross-validation whereby 90% were used for training and 10% for testing in any given fold. As shown in Table 1, the SVM method yields a high degree of classification accuracy, with a significantly high proportion, 88%, of true positives achieved (i.e. the *sensitivity*). There was a slightly lower result of only 82% true negatives (i.e. the *specificity*). However, this is of less importance as clinicians are more concerned by *false negatives*, (patients who were incorrectly classified as having a low CVD risk), of which there were only 12%. Hence, the overall average becomes 85% successful classification. By comparison the ANN approach achieved at best only 79% sensitivity, 76% specificity and 78% overall. Hence the SVM method outperformed the ANN method by quite some margin. Moreover, the use of eigenvalue features improved the robustness and reliability of the classifier significantly over that presented in our previous work [11].

5. CONCLUSIONS

A method to accurately classify patients into high and low PWV (equivalent to high and low CVD risk) using features extracted from their DVP waveform is presented. Support Vector Machine classification is shown to provide superior results when compared with the popular Artificial Neural Network approach. Measuring DVP is very simple and rapid, hence this method offers the very exciting property of being suitable for use by health professionals, such as GPs, as a casual screening facility for the prevention of CVD related injury and mortality for minimal cost to hospitals and health authorities.

6. FUTURE WORK

Future work is currently under way to investigate new feature extraction methods motivated by new signal processing techniques to further improve the classification results. The statistical significance of this study grows as we continue to gather new data on subjects from different geographical, ethnical and pathological backgrounds.

7. ACKNOWLEDGEMENT

We would like to express our gratitude to the patients of St. Thomas' Hospital involved in the study for allowing us to collect and analyse their data.

8. REFERENCES

[1] K. M. Anderson, P. M. Odell, P. W. F. Wilson, W. B. Kannel, "Cardiovascular disease risk profiles," *American Heart Journal*, 121:293–298, 1991.

[2] J. Blacher, R. Asmar, S. Djane and M. Gerard, "Aortic Pulse Wave Velocity as a Marker of Cardiovascular Risk in Hypertensive Patients", *Hypertension*, 33:1111–1117, 1999.

[3] P. Boutouyrie et al, "Aortic stiffness is an independent predictor of primary coronary events in hypertensive patients: a longitudinal study," *Hypertension*, 39:10–15, 2002.

[4] S. M. Grundy, R. Pasternak, P. Greenland, S. Smith Jr. and V. Fuster. "Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations", *Journal of the American College of Cardiology*, Vol. 34, Issue 4, pp 1348–1359, October 1999.

[5] J. Nurnberger et al, "Augmentation index is associated with cardiovascular risk," *Journal of Hypertension*, 20(12):2407–2414, 2002.

[6] K. Takazawa, N. Tanaka, M. Fujita, O. Matsuoka, T. Saiki, M. Aikawa, S. Tamura and C. Ibukiyama, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmograph waveform", *Hypertension*, 32:365-370, 1998.

[7] S. C. Millasseau, F. G. Guigui, R. P. Kelly et al., "Noninvasive assessment of the digital volume pulse. Comparison with the peripheral pressure pulse", *Hypertension*, 36:952–956, 2000.

[8] S. C. Millasseau, R. P. Kelly, J. M. Ritter, P. J. Chowienczyk, "Determination of age-related increases in large artery stiffness by digital pulse contour analysis," *Clinical Science (London)* 103:371–377, 2002.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[10] N. Christianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, http://www.support-vector.net, 2000.

[11] S. R. Alty, S. C. Millasseau, P. J. Chowienczyk and A. Jakobsson, "Cardiovascular Disease Prediction Using Support Vector Machines", 46th IEEE Mid-West Symposium on Circuits and Systems, Egypt, December 2003.

[12] N. Angarita-Jaimes, "Support Vector Machines for Improved Cardiovascular Disease Risk Prediction", M.Sc. Thesis, Department of Electronic Engineering, King's College London, 2005.

[13] J. Shawe-Taylor and N. Christianini, *Kernel Methods for Pattern Recognition*, Cambridge University Press, http://www.kernel-methods.net, 2004.

[14] J. Ma, Y. Zhao and S. Ahalt, *OSU SVM Classifier Matlab Toolbox (version 3.00)*, http://www.eleceng.ohio-state.edu/maj/osu_svm/, 2002.

[15] J. McLaughlin, M. McNeill, B. Braun and P. D. McCormack, "Piezoelectric sensor determination of arterial pulse wave velocity", *Physiological Measurement*, 24:693–702, 2003.