Predicting Arterial Stiffness From the Digital Volume Pulse Waveform

Stephen R. Alty*, *Member, IEEE*, Natalia Angarita-Jaimes, Sandrine C. Millasseau, and Philip J. Chowienczyk

***Abstract*—Cardiovascular disease (CVD) is currently the biggest single cause of mortality in the developed world, hence, the early detection of its onset is vital for effective prevention therapies. Aortic stiffness as measured by aortic pulse wave velocity (PWV) has been shown to be an independent predictor of CVD, however, the measurement of PWV is complex and 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 the transmission of infra-red light through the finger pulp. PWV and DVP were measured on 461 subjects attending a clinic in South East London. Techniques for extracting features from the DVP contour based on physiology and information theory were compared. Low and high stiffness were defined according to a threshold level of PWV chosen to be 10 m/s^{-1}. Using a support vector machine-based classifier, it is possible to achieve high overall classification rates on unseen data. Further, the use of support vector regression techniques lead to a direct real-valued estimate of PWV which outperforms previous methods based on multilinear regression. We, therefore, conclude that support vector machine-based classification and regression techniques provide effective prediction of arterial stiffness from the simple measurement of the digital volume pulse. This technique could be usefully employed as a cheap and effective CVD screening technique for use in general practice clinics.**

***Index Terms*—Cardiovascular disease (CVD), digital volume pulse (DVP), photoplethysmography, pulse wave velocity (PWV), support vector machines (SVMs).**

I. Introduction

CARDIOVASCULAR disease (CVD) is the leading cause of mortality in the developed world. An estimated 17 million people die every year from CVD (mainly from myocardial infarction and stroke; source: World Health Organization). Established risk factors for CVD include age, sex, cigarette smoking, high blood pressure (hypertension), serum cholesterol and the presence or absence of diabetes mellitus. The current approach for estimating the risk of a CVD event (such as a myocardial infarction or stroke) within an individual involves the use of these factors in a "risk calculator" derived from regression equations relating levels of individual risk factors to CVD events in prospective follow-up studies such as the Framingham Heart study [1] (or the Cox model study [2]). However, such risk calculators do not identify a substantial minority of subjects who subsequently develop CVD. CVD occurs principally as a result of atherosclerosis and arteriosclerosis, inflammatory and degenerative conditions of the arterial wall. Initially, changes occur at the cellular level but then lead to changes in the mechanical properties of the wall. The possibility that biophysical measures of the properties of the arterial wall may provide a measure of CVD risk has thus received attention recently [3], [4]. One of the most promising measurements is arterial stiffness. Arteries stiffen with advancing age and *premature* stiffening may result from a combination of arteriosclerosis and atherosclerosis. In addition, arterial stiffening leads to systolic hypertension and increased load on the heart. Arterial stiffness may thus not only provide a marker of the effects of atherosclerosis/arteriosclerosis on the arterial wall, but in itself lead to adverse haemodynamic effects that increase CVD risk. In studies to date [3], [4] large artery stiffness, as measured by pulse wave velocity (PWV) (shown in the following) has proven to be a powerful independent predictor of CVD events, more closely related to CVD risk than traditional risk factors.

A. Estimation of Arterial Stiffness by PWV

Direct measurement of arterial stiffness by simultaneous measurement of change in arterial pressure and diameter is technically challenging and in practice the technique most widely used to measure stiffness is the determination of arterial PWV. PWV is the velocity with which the pressure pulse propagates through the arterial tree and is related to arterial stiffness by the Moens–Korteweg [5] equation

$$\mathrm{PWV} = \sqrt{\frac{E_{\mathrm{inc}} h}{2r\rho}} \tag{1}$$

where E_{inc} is the incremental Young's elastic modulus of the arterial wall, h is the wall thickness, r is the radius of the artery, and ρ is the density of blood. Carotid-femoral PWV is the measurement that has been used in most outcome studies to date. It includes the aorta and large elastic arteries that are most susceptible to age-related stiffening and which determine dynamic loading on the heart. Carotid-femoral PWV can be determined

Manuscript received January 18, 2006; revised January 7, 2007. The work of S. C. Millasseau was supported by Micro Medical Ltd. *Asterisk indicates corresponding author.*

*S. R. Alty is with the Centre for Digital Signal Processing Research, Division of Engineering, King's College London, Strand, London WC2R 2LS U.K. (e-mail: steve.alty@kcl.ac.uk).

N. Angarita-Jaimes was with King's College London, Centre for Digital Signal Processing Research, London WC2R 2LS U.K. She is now with the School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT U.K. (e-mail: menncaj@leeds.ac.uk).

S. C. Millasseau was with the King's College, School of Medicine (Cardiovascular Division), St. Thomas' Hospital, London SE1 7EH, U.K. (e-mail: sandrine_millasseau@yahoo.fr).

P. J. Chowienczyk is with the King's College St. Thomas' Hospital, School of Medicine (Cardiovascular Division), London SE1 7EH U.K. (e-mail: phil.chowienczyk@kcl.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2007.897805

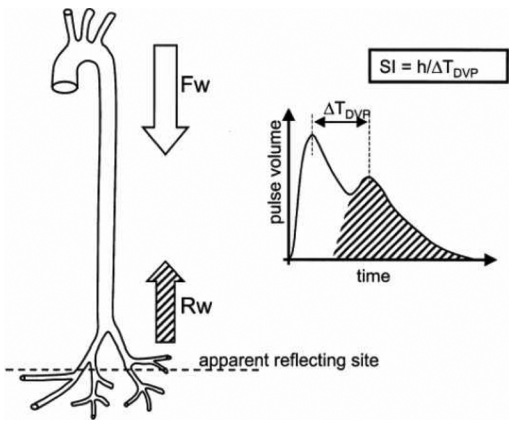

Fig. 1. Stiffness index is related to the time delay (here $\mathrm{PPT} = \Delta T_{\mathrm{DVP}}$) between the systolic and diastolic components of the waveform and the subject's height h. Hence, $\mathrm{SI}_{\mathrm{DVP}} = h/\Delta T_{\mathrm{DVP}}$, (from [8]).

by application of a noninvasive pressure sensor over carotid and femoral arteries. The time delay ΔT between the foot of the pressure pulse arriving at the carotid and femoral arteries is measured and the path length L between the arteries estimated from distance between the site of application of the sensors. PWV then equals $L/\Delta T$ measured in meters per second^{-1}. The sensors may be applied simultaneously to both arterial sites or to each site sequentially using the "R-wave" of the ECG as a time reference.

B. Estimating Arterial Stiffness by Pulse Contour Analysis

Determination of PWV as described before requires a skilled technician, involves the subject disrobing to expose the femoral artery, and requires specialized equipment. An alternative, simpler more widely applicable technique would be of great advantage in screening for CVD. We have previously suggested that the contour of the digital volume pulse (DVP) may be used to estimate arterial stiffness [6]–[8]. The DVP waveform can be rapidly acquired by measuring absorption of infrared light across the finger pulp (technically referred to as *photoplethysmography*). This varies with red blood cell density and hence in variation of blood vessel diameter during the cardiac cycle. Although the absolute amplitude of the DVP depends upon local factors such as temperature and perfusion of the hand, its contour is mainly determined by characteristics of the heart and large arteries. In many healthy subjects, especially in younger subjects with low arterial stiffness, the DVP exhibits an initial systolic peak followed by a diastolic peak, thought to be formed—at least in part—by pressure waves reflected from the periphery of the circulatory system. The time between the systolic and diastolic peaks, the *peak-to-peak time* (PPT) is related to the time taken for the pressure wave to propagate from the heart to periphery and back and is thus related to PWV in the large arteries. A so-called *stiffness index* (SI) derived from subject height (as a measure of path length) and PPT can be used to estimate PWV (see Fig. 1). In older subjects and in subjects with premature arterial stiffening and in some healthy subjects

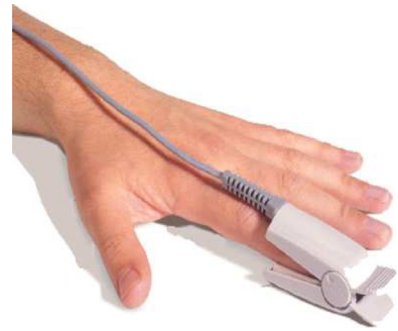

Fig. 2. DVP recorded by measuring the transmission of IR light through the finger pulp.

that presumably have differing anatomy, however, the systolic and diastolic peaks in the DVP become difficult to identify and SI becomes unreliable as an estimate of stiffness.

This paper sets out to show that it is possible to get a good estimate of a given patient's PWV (and by inference, arterial stiffness and thereby Cardiovascular health) simply and quickly from their DVP waveform alone. Previously published works on related topics have tended to rely on a very limited cohort of subjects. For example, Bhattacharya *et al.* [9] present an interesting paper on an initial study into the diagnostic potential of the photo-plethysmographic waveform based on the results obtained from four subjects. In our paper, we had access to data collated from a cohort of 461 subjects. Moreover, the aim of our study was to determine which features of the DVP waveform could be used *most effectively* to estimate PWV. A number of different *physiologically motivated* and *signal subspace based* features (see Section III) were extracted from the DVP waveform and tested using machine learning techniques to determine the optimum set or combination thereof. Support vector machine (SVM) [10]–[12] supervised learning theory (introduced in Section IV) was applied to find the best features extracted from the DVP waveform to give good prediction of high and low PWV. These features were also tested using the more traditional artificial neural network (ANN) [13] machine learning approach to provide a benchmark for comparison with the SVM method.

II. Photoplethysmography

The data acquisition system consisted of a handheld photoplethysmography data capture device (Micro Medical Ltd., Chatham, U.K.). The equipment transmitted infrared light at 940 nm and was placed on the index finger of the right hand (see Fig. 2). The signal from the plethysmograph was digitized using a 12-bit analog-to-digital converter with a sampling frequency of 100 Hz. The DVP signals were recorded over a 10-s period resulting in the capture of a number of waveforms. These sequences of waveforms were subsequently ensemble averaged to obtain a single waveform per subject. The final waveforms were subject to baseline wander removal and normalized to 100 samples per waveform and the amplitude was arbitrarily scaled ranging from 0 to 1000.

A. Study Population

A cohort of 461 subjects were recruited from the local area of South East London. No subject had a previous history of CVD or was receiving vasoactive drugs. The subjects were in a range of 16 to 81 years of age and the average age was 50 with a standard deviation of 13.6 years. A number of basic physiological indicators were measured from each subject including systolic and diastolic blood pressure along with their DVP waveform and their PWV. Specifically, the Carotid-femoral PWV was measured by ECG-referenced carotid and femoral tonometry (SphygmoCor, Atcor, Australia). The mean value of the stiffness index in all subjects (9.4 m/s^{-1}) was of similar magnitude to that of PWV (10.4 m/s^{-1}).

III. Feature Extraction

Numerous measurements (or *features*) can be derived from the DVP waveform. Previous work in this field [8] has led to the selection of what we now term *physiologically motivated features*. That is to say, parameters associated with the physiological properties of the aorta and arterial characteristics in general. Since then, we have focused our attention on exploiting information theoretic approaches to feature extraction and we refer to these as *signal subspace-based features*. The details of the two extraction processes are elucidated in Sections III-A and III-B, respectively.

A. Physiologically Motivated Features

Several different combinations of features were compared and after experimentation it was found that, in fact, a specific set of four of these features give the best classification of high or low PWV. Interestingly, two of these features have been independently cited in the literature [6], [8], and [15] as having a bearing on cardiovascular pathology. When the DVP was first measured in 1941, Dillon and Hertzman [15] observed that subjects with hypertension and/or arteriosclerosis had an "increase in the crest time" compared to healthy subjects. The *crest time* (CT), i.e., the time from the foot of the waveform to its peak [as shown in Fig. 3(c) and (d)] has proved to be a useful feature for the classifier. Also, *peak-to-peak time* (PPT), defined as the time between the first peak and the second peak or inflection point of the DVP waveform [see Fig. 3(c) and (d)]. As mentioned previously in the Introduction, the second peak/inflection point on the DVP is generally accepted to be due to reflected waves. So its timing would be related to arterial stiffness and PWV. This has indeed been confirmed in two separate studies where PPT was found to correlate well with aortic transit time [6] and the so-called *stiffness index* (SI) (subject height divided by PPT) was shown to correlate well to aortic PWV [8].

The definition of PPT depends on the DVP waveform as its contour varies with subjects. When there is a second peak as is the case with "Waveform A" in Fig. 3(a), PPT is defined as time between the two maxima. In other words, PPT is the time between the two positive to negative zero-crossings of the derivative. However, in some DVP waveforms, there is no clear second peak as in "Waveform B" in Fig. 3(b). In this case, PPT is defined as the time between the peak of the waveform and the inflection point on the downslope of the waveform [which is a local maximum of the first derivative, as shown in Fig. 3(d)]. These three features then: PPT, CT, and SI were empirically found to be among the best features for accurate classification of PWV.

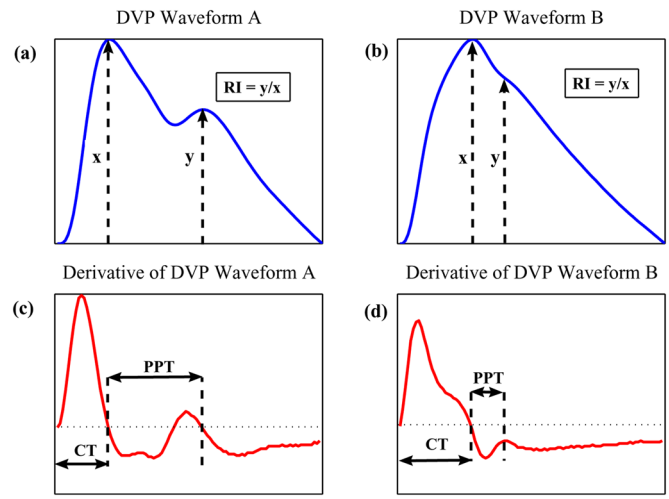

Fig. 3. (a)–(d) Two DVP waveforms and their respective derivatives.

B. Signal Subspace-Based Features

This method of feature extraction, by contrast, makes no assumptions about the physical generation of the waveform and relies 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 [16]). They included kernel principal component analysis [17], wavelet packet decomposition, and signal subspace analysis. In fact, it was found that certain ranges of the eigenvalues of the covariance matrix (formed by the autocorrelation of the DVP waveform with its mean removed) outperformed all the other information theory motivated features and methods. These subspace-based features can be related to the Fourier components in the power spectrum of the DVP waveform. In order to calculate these features, the autocorrelation sequence (ACS) of the DVP data (with mean removed, denoted by $\tilde{\mathbf{d}}$) is determined up to 100 lags. Then, a 100×100 covariance matrix, $\mathbf{A}$, is formed by constructing a Toeplitz matrix [18] from half of this sequence (the ACS is symmetrical and real). Hence, the elements of $\mathbf{A}$ can be formed as follows:

$$a(i,j) \equiv a(j,i) = \sum_{n=1}^{N} \tilde{d}(n-i)\tilde{d}(n-j), \quad 0 \le i, \ j < 100 \tag{2}$$

where $\tilde{d}(n)$ is the nth element of the DVP amplitude data vector with its mean removed, i.e., $\tilde{\mathbf{d}} = \{\tilde{d}(1), \ldots, \tilde{d}(N)\}$ with the data vector normalized to length $N = 100$. These covariance matrices (one per subject) can be then be decomposed using Eigenvalue decomposition [18] such that

$$\mathbf{A} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^{-1}. \tag{3}$$

Here $\mathbf{V}$ is the matrix of *eigenvectors* of $\mathbf{A}$ and $\mathbf{\Sigma}$ its *eigenvalues*, where $\mathbf{\Sigma} = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_N\}$. During experimentation, it was found that some ranges of eigenvalues, for in-

stance, $\hat{\mathbf{\Sigma}} = \{\sigma_3, \ldots, \sigma_9\}$, gave good results. It is hypothesized that the first one or two eigenvalues σ_1 and σ_2 primarily represent the fundamental signal subspace data, which is common to all waveforms in the database and the smallest eigenvalues constitute the noise subspace, hence, the removal of these improve the orthogonality of the data and its subsequent classification.

C. Combinations of Features

While it was impractical to attempt all the possible combinations of feature sets, it was found that certain combinations of physiologically motivated and signal subspace-based features showed the most consistently reliable results overall. Hence, in this paper, we present the findings based primarily on experiments with just physiologically motivated features and then just signal subspace-based features and, finally, combinations of the two types and show the best of these types of data sets in Section V. Further, Table V contained in the Appendix shows the Pearson correlation coefficients determined for all 11 features used in the results of this study.

IV. Support Vector Machines

SVMs [10]–[12] 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 handwritten character recognition, bioinformatics to automatic speech recognition (among many others) with a great deal of success. A brief summary of the mathematical theory of SVMs follows, for a complete treatment please see [12].

A. Hard-Margin Classifier

Consider a binary classification task with a set of linearly separable training samples

$$S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\} \tag{4}$$

where $\mathbf{x} \in \Re^d$, i.e., $\mathbf{x}_i$ 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

$$f(\mathbf{x}) = \mathrm{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + b) \tag{5}$$

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 from the origin. 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 *generalizing 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 by maximizing the margin we effectively need to minimize $(1/2)\|\mathbf{w}\|^2$ subject to the following constraint: $y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1$.

B. Soft-Margin Classifier

Typically, real-world data sets are in fact linearly inseparable in input space, this implies that the maximum margin classifier approach is no longer valid and a new model must be introduced. This means that the constraints of the maximum or *hard* margin classifier (which requires that there are no errors of classification) 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 optimize the soft-margin classifier, we must try to maximize the margin while allowing the margin constraints to be violated according to the preset slack variable ξ_i. This leads to the minimization of: $(1/2)\|\mathbf{w}\|^2 + C\sum_{i=1}^m \xi_i$, 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 minimization of linear inequalities is typically solved by the application of Lagrangian duality theory [12]. Hence, forming the primal Lagrangian

$$L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^m \xi_i - \sum_{i=1}^m \beta_i \xi_i - \sum_{i=1}^m \alpha_i \left[y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1 + \xi_i \right] \tag{6}$$

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 resubstituting into the primal thus

$$L(\boldsymbol{\alpha}) = \sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i,j=1}^m y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \tag{7}$$

Interestingly, this is the same result as for the maximum margin classifier. The only difference is the constraint $\alpha_i + \beta_i = C \,\forall\, i$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta} \geq \mathbf{0}$, hence, $0 \leq \alpha_i \leq C \,\forall\, i$. This implies that the value C, sets an upper limit on the Lagrangian optimization variables α_i and β_i, sometimes referred to as the *box constraint*. The value of C offers a tradeoff between accuracy of data fit and regularization, 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). The dual criterion $L(\boldsymbol{\alpha})$ is maximized and this is typically solved using quadratic programming (QP) algorithms. There are many online resources of such algorithms available for download (see the website referred to in [12] for an up to date listing).

Another possible realization of the soft-margin classifier, is termed ν-SVM and uses the so-called ν-parameterization approach [14]. In this case, the parameter C is replaced by a parameter $\nu \in (0, 1]$ which is asymptotically a lower bound on the

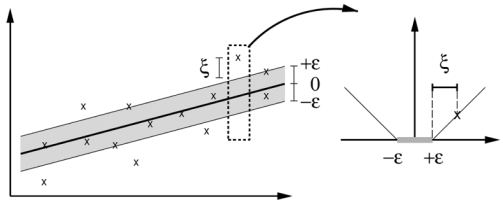

Fig. 4. Diagram depicting support vector regression in linear input space with ε-insensitive loss function "tube" [19].

number of support vectors and an upper bound on the number of margin errors (i.e., data points that lie on the wrong side of the hyperplane). This has the overall effect of providing a quasi-linear range on the regularization constraint (the actual relationship will depend on the data set) and provides a useful framework from which to perform a grid search for hyper-parameter optimization. This method has been used in the classification results of this paper.

C. Support Vector Machines for Regression

SVMs lend themselves as easily to the task of regression with only a small extension in theory. Essentially they allow for real-valued targets to be estimated by modeling a linear function (see Fig. 4) in *feature space* (see Section IV). The same maximum margin concept is maintained but it is augmented by a so-called *ε-insensitive loss-function*; so long as the training data points lie within the range of the loss function then no error is deemed to have occurred. In a manner analogous to the soft-margin classifier, errors are accounted for by the inclusion of slack variables that allow data points to violate this constraint in a limited fashion. This leads to a slightly modified set of constraints requiring the minimization of: $(1/2)\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*)$, subject to $y_i - \langle \mathbf{w}, \mathbf{x}_i\rangle - b \leq \varepsilon + \xi_i$, and $\langle \mathbf{w}, \mathbf{x}_i\rangle + b - y_i \leq \varepsilon + \xi_i^*$ with $\xi_i, \xi_i^* \geq 0$ for $i = 1, \ldots, m$. The solution for the previous QP problem is provided once again by the use of the Lagrangian duality theory, however, we have omitted a full derivation for the sake of brevity (please see [19] and the references therein).

D. Kernel Functions

It is quite often the case with real-world data that it is linearly inseparable, it may, however, exhibit a relatively simple underlying nonlinear characteristic nature (such as quadratic). Kernel mappings offer an efficient solution to this problem by nonlinearly projecting the input 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 calculation of a Kernel inner-product itself is sufficient to provide the mapping. This simplifies 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 while having no effect on the Lagrangian optimization theory. Hence, the relevant classifier function then becomes

TABLE I
KEY FOR PRESENTED DATA SETS

Dataset	Features
P1	{CT, PPT}
P2	{CT, PPT, SI}
Σ1	$\{\sigma_3, \ldots, \sigma_9\}$
Σ2	$\{\sigma_2, \ldots, \sigma_9\}$

$$f(\mathbf{x}) = \operatorname{sgn}\left[\sum_{i=1}^{\mathrm{nSVs}} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right] \tag{8}$$

and for regression

$$f(\mathbf{x}) = \sum_{i=1}^{\mathrm{nSVs}} (\alpha_i - \alpha_i^*)\, K(\mathbf{x}_i, \mathbf{x}) + b \tag{9}$$

where $nSVs$ denotes the number of support vectors, y_i are the labels, α_i and 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 nonlinear classifier (or regressor). There are a number of different Kernel functions available [12], however, the most commonly used are the *gaussian radial basis function* (GRBF) kernel, given by

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2\right) \tag{10}$$

where γ defines the size or width of the radial basis function, and the *polynomial* kernel, given by

$$K(\mathbf{x}_i, \mathbf{x}) = (\langle \mathbf{x}_i, \mathbf{x}\rangle + c)^p \tag{11}$$

where c is an offset constant and p is the polynomial degree or order.

V. RESULTS

This section contains the best results obtained after thorough experimentation for both support vector classification and support vector regression. Both these methods were compared to the baseline technique of a "standard" three-layer ANN. Results for classification are presented in percentages with overall (or "total" rate) then the *sensitivity* or "true positive" rate followed by the *specificity* or "true negative" rate. The regression results are compared in terms of the standard deviation of their errors from the actual values for the entire data set. In actual fact, a wide range of possible feature sets and combinations exist that performed the required tasks satisfactorily, however, for ease of readability and conciseness, we have selected those which we found to perform the best overall. The results presented here are based on the best two sets of each of the physiologically motivated features, two sets of signal subspace-based features and finally the two best combinations of the each type of feature set. During experimentation, the two sets of physiologically-based features that performed the best were found to be: $P1 = \{\mathrm{CT}, \mathrm{PPT}\}$ and $P2 = \{\mathrm{CT}, \mathrm{PPT}, \mathrm{SI}\}$. The two sets of signal

TABLE II
ANN CLASSIFICATION RATES % AND (SD) FOR VARIOUS DATA SETS (SEE TABLE I)

	Datasets					
	Physiological		Eigenvalues		Combinations	
	P1	P2	Σ1	Σ2	Σ1+P1	Σ1+P2
Total	80.2 (1.8)	78.1 (3.1)	79.2 (1.8)	81.3 (0.0)	80.5 (1.3)	81.3 (0.1)
Sens	83.3 (3.2)	82.8 (6.8)	81.0 (5.5)	84.4 (1.9)	86.0 (1.4)	85.8 (1.2)
Spec	78.0 (4.8)	76.7 (6.0)	78.5 (5.7)	77.9 (1.2)	74.9 (1.6)	77.1 (0.8)

subspace based features that performed the best were also found accordingly, $\Sigma 1 = \{\sigma_3, \ldots, \sigma_9\}$ and $\Sigma 2 = \{\sigma_2, \ldots, \sigma_9\}$. The key to these feature sets is shown in Table I. Finally, the execution time taken for the feature extraction process for all feature types was very short, typically of the order of less than one second and significantly shorter than the training process for either SVM or ANN classifier on a standard PC, (the training times being of the order of a minute or two).

The entire cohort of 461 subjects, with complete DVP waveform data and PWV measurements was used in this study. The PWV values were grouped into low and high values. Studies [21] have shown that values of <9 m/s^{-1} are low risk and values >11 m/s^{-1} indicate a high CVD risk category. Moreover, the mean PWV value of our cohort was found to be around 10 m/s^{-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 11 m/s^{-1} to avoid ambiguity of the target classes. The remaining 315 records underwent three fold cross-validation whereby 90% were used for training and 10% for testing in any given fold.

A. ANN Benchmark

An ANN [13] benchmark method was used to provide baseline results for both classification and regression with which to compare those of the SVM techniques. There are many different architectures available for an ANN, but perhaps the most commonly used arrangement is the three-layer feed-forward network with resilient back propagation which is favored for its resistance to local-minima and fast convergence properties. The hidden layer neurons used the sigmoid transfer function and the output node used a signum transfer function for classification (a linear function was used for regression). Only the number of hidden nodes then remained to be specified. This was chosen by cross-validation on separate unseen data to optimize generalization performance. Networks with between 4 and 7 hidden nodes were considered and the cross-validation errors always showed a clear optimum within this range; its precise location depended on the dataset but was always between 4 and 7 hidden nodes. The MATLAB neural network toolbox was used to perform these experiments and the results for ANN classifier are shown in Table II. It can be seen that many data sets give a reasonably high overall classification rate around 80%–81%. In general, however, the classifier tends to favor sensitivity over specificity by around five to ten percentage points. This implies a minor tendency toward false positives, which is generally preferable to the opposite bias. Notwithstanding, it can be seen that for the ANN, combinations of physiological and eigenvalue-based features provide little improvement over the eigenvalue-based sets alone. We shall compare and contrast these results with those obtained from the support vector classifier in Section V-B.

B. Support Vector Classification

Here a binary classifier based upon the Ohio State University SVM toolbox for MATLAB [20] was employed using a number of different kernel functions in combination with a soft margin classifier. It was found during experimentation that the GRBF kernel performed as well or better than the others and, hence, the results in Table III are based on this kernel (the performance of the other kernels is discussed below in Table IV). After performing a thorough model parameter grid searches (see Fig. 5), results were averaged from a "block" of nine individual results obtained from a range of both constraint factor ν and GRBF width γ. This technique was applied to all the results to mitigate against over-training of the model parameters, ensuring a more general classifier and unbiased results. As shown in Table III, the SVM method using *physiologically motivated* feature set P1 yields a fairly high degree of classification accuracy, with a significantly high proportion 90.3% of true positives achieved. There was a slightly lower result of only 77.4% true negatives. Hence, the overall average successful classification rate becomes 84.0%. By comparison the ANN approach achieved at best only 83.3% sensitivity, 78.0% specificity, and 80.2% overall. Hence, the SVM method outperformed the ANN method (with the exact same data) by quite some margin. Perhaps of greater significance, however, are the results for the *signal subspace-based features*, which show a distinct improvement in the specificity when compared with those of the physiologically motivated features, especially for the SVM classifier. Here, it is readily possible to achieve sensitivities in the region of 90%, with specificities of over 80%, bringing overall classification to 85.3% (this rate can be achieved with a range of model parameters and represents an overall mean rate). Again, Tables II and III indicate that the ANN results are inferior to those of the SVM, moreover, they show that the eigenvalue based features with SVM gave better results than those based on physiological features alone. Finally, combinations of both physiological and signal subspace feature sets were tested and the best two pairs are also included in the table. The other two combinations came quite close, but none were as effective as the two presented. It can be seen that the combinations of $\Sigma 1 + P1$ and $\Sigma 1 + P2$ gave the best results. The latter set gave an overall classification rate of 87.5%, with a balanced sensitivity and specificity of the same value. This is significantly the best figure achieved thus far in this study and represents a very high classification rate for PWV based on features extracted solely from the DVP waveform. Tests with different kernel functions were performed to compare with those of the GRBF kernel;

TABLE III
SVM CLASSIFICATION RATES % AND (SD) FOR VARIOUS DATA SETS USING GRBF KERNEL (SEE TABLE I)

	Datasets					
	Physiological		Eigenvalues		Combinations	
	P1	P2	Σ1	Σ2	Σ1+P1	Σ1+P2
Total	84.0 (0.5)	84.0 (0.5)	85.1 (1.4)	85.3 (1.6)	86.1 (0.9)	87.5 (0.0)
Sens	90.3 (2.0)	88.2 (2.1)	90.3 (3.3)	90.3 (4.0)	86.7 (1.4)	87.5 (0.0)
Spec	77.4 (0.5)	79.9 (1.0)	80.2 (2.8)	80.5 (3.4)	85.3 (2.5)	87.5 (0.0)

TABLE IV
OVERALL SVM CLASSIFICATION RATES % AND (SD) FOR VARIOUS KERNEL FUNCTIONS FOR EACH DATA SET (SEE TABLE I)

	Datasets					
	Physiological		Eigenvalues		Combinations	
Kernel	P1	P2	Σ1	Σ2	Σ1+P1	Σ1+P2
Linear	84.0 (0.5)	83.3 (0.0)	82.3 (0.9)	80.9 (0.5)	84.4 (0.0)	85.8 (0.5)
Poly2	83.9 (0.8)	84.0 (0.5)	83.9 (1.4)	84.5 (1.1)	86.0 (1.2)	86.8 (0.5)
Poly3	83.7 (0.9)	83.9 (1.1)	85.0 (0.9)	84.6 (1.4)	86.1 (0.9)	86.5 (0.0)
GRBF	84.0 (0.5)	84.0 (0.5)	85.1 (1.4)	85.3 (1.6)	86.1 (0.9)	87.5 (0.0)

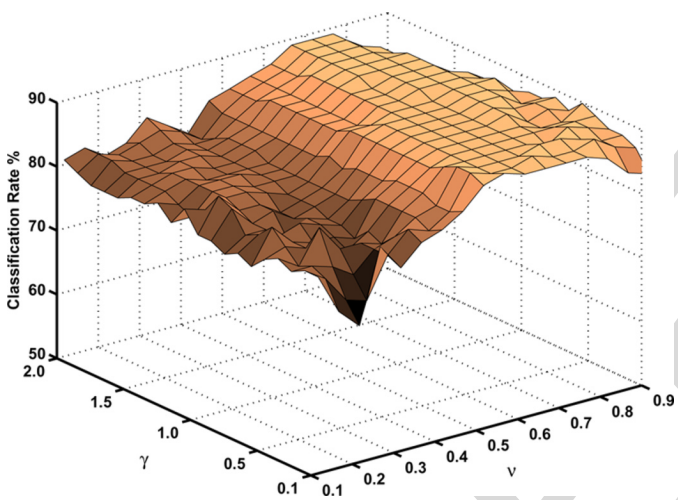

Fig. 5. Grid search showing classification rate is high for a wide range of γ and ν for data set $(\Sigma 1 + P2)$.

linear and polynomial (second and third degree) were tested but the GRBF kernel consistently gave the best overall results of the four kernel functions for all of the data sets (see Table IV).

C. Support Vector Regression

The same SVM toolbox was capable of providing regression figures and was used in this study also. Again, after performing careful grid searches, a GRBF Kernel, using a typical constraint factor in the region of $C = 1000$ and $\sigma = 1.0$, appeared to give the best results (though these model values were not critical). The PWV data from all subjects was used (with no gapping), multifolded cross-validation was employed whereby 90% of the population was used for training and the remaining 10% for testing. The test results were ten-fold cross-validated and subsequently aggregated and compared to the actual target PWV. The standard deviation (SD) of the difference (error) between the actual and predicted PWV was found to be 2.25 m/s^{-1} with the physiologically motivated features alone and just 2.13 m/s^{-1} for the signal subspace-based features. Though the improvement with combinations of features was marginal, it was possible to achieve a SD of just 2.03 m/s^{-1}. Nevertheless, these results compare very favorably to the traditional multiregression techniques whereby a best SD of 2.60 m/s^{-1} was achieved. Finally, the best ANN regressor achieved a SD of 2.07 m/s^{-1}, thus, the difference in performance of the SVR and ANN based regressor was much less significant here. Hence, it has been demonstrated that it is possible to directly predict (to a degree of accuracy approaching that of the measuring equipment i.e., ± 1 m/s^{-1}) the PWV from features extracted solely from the DVP waveform.

VI. CONCLUSION

A method to accurately classify subjects into high and low PWV (equivalent to high and low CVD risk) using features extracted from their DVP waveform has been presented. Both SVM methods of *classification* and *regression* are shown to provide superior results when compared with either traditional multilinear classifiers or the popular ANN approaches in this specific application. Many of the methods and feature sets within the study were capable of achieving better than 80% successful classification, with the overall best classification of PWV reaching 87.5%. The best results were obtained while using a combination of different features extracted from the DVP waveform (i.e., both physiologically motivated and signal subspace-based features). Measuring DVP and extracting these features is very simple and rapid, we, therefore, conclude that this method offers the very exciting property of being suitable for use by health professionals as a screening facility for the assessment of CVD risk, such as in a General Practice clinic.

VII. FUTURE WORK

Work is currently under way to investigate new feature extraction methods to further improve the classification and regression results. The statistical significance of this study grows as we continuously gather new data on subjects from different geographical, ethnical, and pathological backgrounds. Furthermore, as subject follow-up outcome data becomes available, we intend to investigate and disseminate a study into the ability of the classifier to predict actual CVD events.

TABLE V
CORRELATION AMONGST PAIRS OF FEATURES USED IN THE STUDY

	CT	PPT	SI	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9
CT	1.000	-0.547	-0.147	-0.633	0.303	-0.175	0.126	0.066	0.512	0.513	0.591
PPT	-0.547	1.000	0.189	0.449	-0.339	-0.092	-0.450	-0.419	-0.678	-0.642	-0.743
SI	-0.147	0.189	1.000	0.206	-0.176	0.031	-0.073	-0.040	-0.161	-0.135	-0.237
σ_2	-0.633	0.449	0.206	1.000	-0.183	0.524	0.164	0.309	-0.445	-0.485	-0.429
σ_3	0.303	-0.339	-0.176	-0.183	1.000	0.470	0.290	0.340	0.330	0.248	0.366
σ_4	-0.175	-0.092	0.031	0.524	0.470	1.000	0.775	0.805	0.149	0.041	0.145
σ_5	0.126	-0.450	-0.073	0.164	0.290	0.775	1.000	0.943	0.497	0.406	0.490
σ_6	0.066	-0.419	-0.040	0.309	0.340	0.805	0.943	1.000	0.525	0.416	0.456
σ_7	0.512	-0.678	-0.161	-0.445	0.330	0.149	0.497	0.525	1.000	0.976	0.806
σ_8	0.513	-0.642	-0.135	-0.485	0.248	0.041	0.406	0.416	0.976	1.000	0.807
σ_9	0.591	-0.743	-0.237	-0.429	0.366	0.145	0.490	0.456	0.806	0.807	1.000

APPENDIX

Table V shows the Pearson correlation coefficients determined for all 11 features used in the results of this study, i.e., $\{\mathrm{CT}, \mathrm{PPT}, \mathrm{SI}, \sigma_2, \ldots, \sigma_9\}$.

ACKNOWLEDGMENT

The authors would like to thank the patients of St. Thomas' Hospital who were involved in this study for allowing them to collect and analyze their data. Also, they would like to thank Prof. P. Sollich for his helpful comments during the preparation and revision of this paper.

REFERENCES

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

[2] 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," *J. American College Cardiology*, vol. 34, no. 4, pp. 1348–1359, 1999.

[3] J. Blacher, R. Asmar, S. Djane, and M. Gerard, "Aortic pulse wave velocity as a marker of cardiovascular risk in hypertensive patients," *Hypertension*, vol. 33, pp. 1111–1117, 1999.

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

[5] W. W. Nichols and M. F. O'Rourke, *McDonald's Blood Flow in Arteries. Theoretical, Experimental and Clinical Principles*. New York: Oxford Univ. Press, 1999.

[6] P. J. Chowienczyk *et al.*, "Photoplethysmographic assessment of pulse wave reflection: Blunted response to endothelium-dependent beta2-adrenergic vasodilation in type II diabetes mellitus," *J. American College Cardiology*, vol. 34, pp. 2007–2014, 1999.

[7] S. C. Millasseau *et al.*, "Noninvasive assessment of the digital volume pulse. Comparison with the peripheral pressure pulse," *Hypertension*, vol. 36, pp. 952–956, 2000.

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

[9] J. Bhattacharya, P. P. Kanjila, and V. Muralidhar, "Analysis and characterisation of photo-plethysmographic signal," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 1, pp. 5–11, Jan. 2001.

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

[11] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, pp. 121–167, 1998.

[12] N. Christianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, MA: Cambridge Univ. Press, 2000.

[13] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[14] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Computation*, vol. 12, pp. 1207–1245, 2000.

[15] J. B. Dillon and A. B. Hertzman, "The form of the volume pulse in the finger pad in health, arteriosclerosis, and hypertension," *American Heart J.*, vol. 21, pp. 172–190, 1941.

[16] N. Angarita-Jaimes, "Support vector machines for improved cardiovascular disease risk prediction," M.Sc. dissertation, Dept. Electron. Eng., King's College London, London, U.K., 2005.

[17] J. Shawe-Taylor and N. Christianini, *Kernel Methods for Pattern Recognition*. Cambridge, MA: Cambridge Univ. Press, 2004.

[18] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins Univ. Press, 1996.

[19] A. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, 2004.

[20] J. M. Y. Zhao and S. Ahalt, "Ohio State University (OSU) SVM classifier MATLAB toolbox (Version 3.00)," 2002.

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

Stephen R. Alty (M'00) received the B.Eng. degree (Hons.) in electronic engineering from the University of Liverpool, Liverpool, U.K., in 1992 and the Ph.D. degree in signal processing from Liverpool John Moores University, Liverpool, U.K., in 1998.

Since 2002, he has been an Assistant Professor with the Centre for Digital Signal Processing Research, King's College London, London, U.K. He was an Assistant Professor with Liverpool John Moores University. His research interests include adaptive signal processing and machine learning with application to spectral analysis, speech, and biomedical pattern recognition.

He is a member of the IEEE Signal Processing and the IEEE Engineering in Medicine and Biology Societies.

Natalia Angarita-Jaimes was born in Bucaramanga, Colombia, in 1980. She received her degree in electronic engineering from Los Andes University of Bogotá, Bogotá, Colombia, in 2003, and the M.Sc. in digital signal processing from King's College London, London, U.K., in 2005. She is currently pursuing the Ph.D. degree in the development of new imaging systems that enable the full 3-D location of a particle in fluid flows to be applied to flow measurement in complex biological systems from the School of Mechanical Engineering, University of Leeds, Leeds, U.K.

She worked as a Research Assistant with the École des Mines de Nantes, Nantes, France, where she developed an interest in biomedical engineering.

Ms. Angarita-Jaimes is a member of the IET.

Sandrine C. Millasseau received the B.S. degree in electronic engineering from École Nationale Supérieure de l'Électronique et de ses Applications (ENSEA), Paris, France, and the Ph.D. degree in biomedical engineering from King's College London, London, U.K., in 2003.

She has subsequently worked as a Support and Development Engineer on a wide range of biomedical instrumentation projects.

Philip J. Chowienczyk is Professor of Cardiovascular Clinical Pharmacology with School of Medicine, King's College, London, U.K., and an Honorary Consultant Physician with Guy's and St. Thomas' Foundation Trust Hospital, London, U.K. He trained in physical science, biomedical instrumentation, and clinical medicine and his current research interests include technological development of noninvasive tests of arterial health.