Detecting vascular age using the analysis of peripheral pulse

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2814630, IEEE Transactions on Biomedical Engineering

Detecting vascular age using the analysis of peripheral pulse

Michele Sorelli*, Antonia Perrella, and Leonardo Bocchi, *Member*, *IEEE*

***Abstract*—Vascular ageing is known to be accompanied by arterial stiffening and vascular endothelial dysfunction, and represents an independent factor contributing to the development of cardiovascular disease. The microvascular pulse is affected by the biomechanical alterations of the circulatory system, and has been the focus of studies aiming at the development of non-invasive methods able to extract physiologically relevant features. *Objective*: proposing an approach for the assessment of vascular ageing based on a support vector machine (SVM) learning from features of the pulse contour. *Methods*: the supervised classifier was trained and validated over 20935 models of pulse wave, obtained with a multi-Gaussian decomposition algorithm, applied to laser Doppler flowmetry signals of 54 healthy, non-smoker subjects. *Results*: the multi-Gaussian model showed a mean R^2 of 0.98 and an average normalized root mean square error of 0.90, demonstrating the ability to reconstruct the pulse shape. Over 30 training and validation experiments, the SVM showed a mean Pearson's r of 0.808 between the rate of waves classified as *old* and the age of the subjects, along with an average area under the ROC curve of 0.953. *Conclusion*: the SVM showed the capability to discriminate differently aged individuals. *Significance*: the proposed method might detect the ageing-related modifications of the vascular tree; furthermore, since diabetes promotes vascular alterations comparable to ageing, this approach may be also suitable for the screening of diabetic angiopathy.**

***Index Terms*—Laser Doppler Flowmetry, Vascular Ageing, Microcirculation, Pulse Decomposition Analysis, Support Vector Machine.**

I. Introduction

AGEING is deemed a major non-reversible risk factor for cardiovascular disease, contributing per se to the progression of vascular dysfunction and to the incidence of hypertension, stroke, coronary heart disease, and heart failure [1]. Vascular ageing is characterized by large artery stiffening and has been independently correlated with endothelial dysfunction [2]. The gradual accumulation of reactive oxygen species (ROS) has been identified as an important underlying mechanism, since it functionally inactivates the endothelial production of nitric oxide (NO), hence impairing the physiological NO-dependent regulation of the vascular muscle cells tone [3]. Moreover, the ageing-related endothelial dysfunction is fuelled by the upregulation of inflammatory cytokines, which promote chronic low-grade inflammation, thus favouring endothelial leakage, and further contributing to ROS production [4]. Several ageing processes are involved in the reduction of arterial compliance in the elderly, namely the thickening of the tunica media, the decrease in the elastin-to-collagen ratio, and the increased collagen cross-linking due to non-enzymatic glycosylation [4]. In accordance with the Moens-Korteweg equation [5], vascular stiffening manifests itself through an increase in pulse wave velocity (PWV); this is usually evaluated transcutaneously by measuring the pulse *foot-to-foot* transit time between the common carotid artery and the common femoral artery, where the *foot* is identified with the instant immediately preceding the steep pressure rise during early systole. PWV is generally considered as a robust, and reproducible method for assessing arterial stiffness and has been related with adverse CVD prognosis [6]. Nevertheless, the biomechanical properties of the circulatory system, in addition to being reflected by PWV, exert in general also a complex effect on the shape of the arterial pulse. Accordingly, theoretical models have been devised in an attempt to quantify the total arterial compliance and resistance, and derive the local visco-elastic properties of the vascular tree from arterial pressure or Doppler flow waveforms recorded *in vivo* [7]. Physiological information on the circulatory system can also be obtained from the microvascular pulse contour. The analysis of the second derivative of the photoplethysmographic digital volume pulse (DVP), often termed *acceleration photoplethysmogram*, has been proposed as a straightforward approach to derive indices of vascular ageing and tone, based on the ratio of characteristic peaks [8], [9], [10], [11]. However, despite its simplicity, this method does not offer a clear physiological interpretation, and the reliance on the waveform second derivative makes it considerably vulnerable to noise and artefacts. Other physiologically-motivated parameters, derived from the analysis of the DVP first derivative, have been proposed by Millasseau et al. [12], namely: the *reflection index* (RI) and the *stiffness index* (SI). Specifically, the RI is defined as the ratio of the diastolic DVP maximum or inflection point to the amplitude of the systolic peak, and has been linked to peripheral pressure wave reflection and small artery tone; the SI, a marker of arterial stiffness, is instead obtained from the ratio of subject height to the relative delay between the systolic and diastolic DVP peaks, and has been shown to correlate positively with age. Pulse decomposition analysis (PDA) is another sophisticated technique for characterizing the microvascular waveform. This approach is based on the concept that the arterial pulse contour is mainly affected by two major reflection sites (i.e. the juncture between the thoracic and abdominal aorta, and the aortic bifurcation of the common iliac arteries), which give rise to waves that counter-propagate with respect to the forward-travelling systolic pulse

This study was supported by Ente Cassa di Risparmio di Firenze, Florence, Italy (grant number 2015.0914).

M. Sorelli, A. Perrella, and L. Bocchi are with the Department of Information Engineering, University of Florence, Florence, Italy (*correspondence e-mail: michele.sorelli@unifi.it).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2814630, IEEE Transactions on Biomedical Engineering

[13]. In PDA, the pulse waveform is thus decomposed into a varying number of basic components (generally from 3 to 5), generated from Gaussian, log-normal or Rayleigh-type models [14], so as to separately reconstruct and characterize the primary systolic wave and the secondary reflected (or re-reflected) pulses. Recently, some authors have applied machine learning algorithms which were able to classify differently aged individuals [15] or detect high values of PWV [16], based on features derived from the microvascular pulse waveform. In this work, we first present a four-Gaussian model for the extraction of physiological pulse features from the peripheral perfusion waveform. We then report on the implementation of a *support vector machine* supervised learning model which analyzes the pulse feature data extracted with the PDA algorithm, in order to detect vascular ageing patterns and discriminate between subjects of different ages.

II. Materials and Methods

A. Subjects and experimental setup

54 healthy, non-smoking volunteers with no known history of cardiovascular disease were enrolled for this study. Their age distribution is shown in Fig. 1. Research activities were carried out in accordance with the guidelines of the Declaration of Helsinki of the World Medical Association. All the included participants received detailed information on the research protocol and its purpose, and signed an informed consent form. Data were collected and treated anonymously in accordance with privacy regulations.

Microvascular perfusion was recorded on the pulp of the right hallux with a Periflux 5000 Laser Doppler Flowmetry (LDF) system (Perimed, Sweden). The LDF measurements were performed at rest in a temperature-controlled environment ($T \approx 22\,^{\circ}\mathrm{C}$), following a 10-minute acclimatization period, with the subjects lying supine in a comfortable position. Perfusion signals were acquired with a $32\,\mathrm{Hz}$ sampling frequency; a $0.03\,\mathrm{s}$ time constant was selected for the instrument output low-pass filter, so as not to dampen the heartbeat-related pulsatile components. The LDF system was calibrated conforming to the manufacturer's specifications before the initiation of the study, using the Motility Standard and the Zeroing Disc provided with the instrument. Calibration was checked approximately every two weeks, as specified in the instrument manual: no further intervention was required, as the obtained values were always within the reference tolerance levels (250 ± 15 AU in Motility Standard, and 0 ± 1 AU on the Zeroing Disc).

B. Multi-Gaussian decomposition of the LDF pulse wave

The developed PDA algorithm is comprised of four main sub-blocks, namely: LDF signal pre-processing, pulse detection, pulse multi-Gaussian modeling, and feature extraction. The following paragraphs separately present the implementation details of these stages.

1) Pre-processing: A few preprocessing operations were needed to improve the performance of the proposed multi-Gaussian modeling algorithm. In detail, the LDF perfusion signals were upsampled from the original $32\,\mathrm{Hz}$ to $128\,\mathrm{Hz}$, using 8 original values to generate each sample of the expanded signal; next, a low-pass FIR interpolation filter (length: 33 samples; normalized cut-off frequency: 0.5) was contextually applied to minimize spectral distortion.

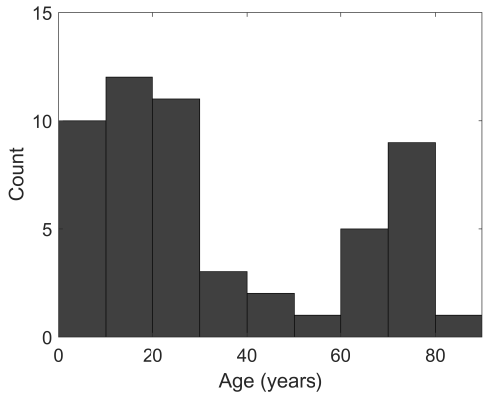


Figure 1. Age distribution of the subjects.

2) Detection: Pulse segmentation represents the first major stage of the multi-Gaussian decomposition algorithm. A local maxima detector is initially applied to find the systolic perfusion peaks, imposing a maximum theoretical heart rate of $150\,\mathrm{bpm}$ in order to limit the amount of false positives, associated with secondary ripples. Next, the algorithm detects the end-diastolic perfusion troughs as the absolute minima between consecutive peaks. These preliminary time references are employed for the elimination of *false* cardiac cycles, exploiting the observation that the amplitude of the pulse wave is not subject to rapid cycle-to-cycle changes; specifically, only candidate pulses whose valley-to-peak excursion exceeds 50% of the previous ones' are preserved at this stage. A further refinement step is needed to improve the accuracy of the original detections, as *true* primary systolic peaks and end-diastolic depressions do not always correspond to the absolute maxima and minima of the separate cardiac cycle time windows. Regarding the systolic references, the algorithm searches for any maxima preceding the original solution, whose prominence with respect to the corresponding valley does not fall below 80% of the original peak amplitude; on the other hand, the existing end-diastolic references are changed for possible local minima within the original trough-to-peak time interval, if their amplitude is lower than 20% of the local peak excursion. Besides the aforementioned feature points, the proposed pulse modeling method employs the identification of an *incisura* on the descending limb of the pulse, with the purpose of separating the forward-travelling systolic wave from the diastolic phase of the pulse, generally associated with delayed components that are reflected in the arterial tree due to vascular impedance mismatch. As a preliminary step, LDF perfusion signals are detrended by subtracting a cubic spline interpolant fitted to the end-diastolic detections, thus isolating the cardiac pulse wave. The identification of the pulse incisura is based on the analysis of the contour's first derivative, $p'(t)$, obtained by means of a 3-point differentiator:

$$p'(t) = \frac{p(t+1) - p(t-1)}{2} \tag{1}$$

where $p(t)$ represents the LDF perfusion signal at discrete time t; a 7-point moving average filter is subsequently applied to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2814630, IEEE Transactions on Biomedical Engineering

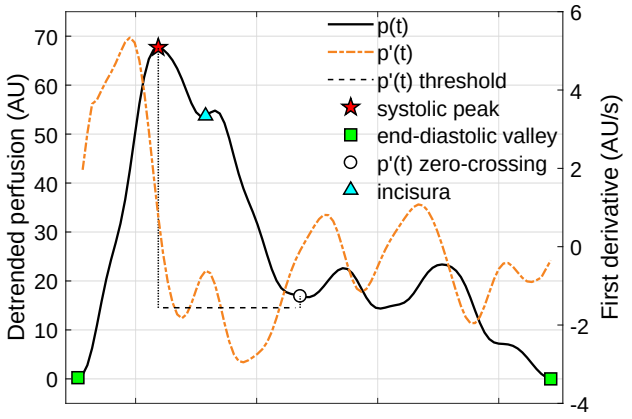


Figure 2. LDF pulse detection and reference points identification.

$p'(t)$, so as to attenuate the undesired effect of local spurious fluctuations. Fig. 2 schematizes the approach followed for the determination of the incisura: within each cardiac cycle, the algorithm first searches for the earliest negative-to-positive zero crossing of $p'(t)$ following the systolic reference (if none are detected, the end-diastolic valley is selected instead); then, the time span between this point and the systolic peak is analyzed for the presence of local $p'(t)$ maxima exceeding the average pulse slope in the same interval: if detected, the earliest of them is adopted as the incisura reference, otherwise the original $p'(t)$ zero crossing is chosen. This method proved to be effective for the distinction of the systolic primary wave from the other diastolic components, even when their partial temporal superposition produces only a slight notch in the pulse contour, but still without introducing an excessive sensitivity to minor $p'(t)$ variations.

3) Modeling: Following the identification of these feature points, a four-Gaussian model is fitted to each detected pulse:

$$G_{pulse}(t) = \sum_{i=1}^{4} g_i = \sum_{i=1}^{4} \alpha_i \cdot e^{-\frac{(t-\mu_i)^2}{2\sigma_i{}^2}} \qquad (2)$$

where the g_1 component is related to the systolic ventricular ejection, while g_2, g_3, and g_4 are introduced to represent the pulse reflections within the circulatory tree (and hence constrained to either side of the incisura, accordingly). Model identification was performed with the Levenberg-Marquardt optimization algorithm for non-linear least squares curve fitting problems [17]. In general, upon optimization, the model parameters are properly tuned with respect to specific properties of each separate pulse, in order to move the optimization starting point close to the ideal solution, and limit the search hypervolume, thus improving the stability of the algorithm. As detailed in Table I, the amplitude of the four Gaussians is initialized and constrained on the basis of the maximum excursion of the pulse, while their standard deviations are configured by bounding the corresponding full width at half maximum (FWHM) to the span of the systolic and diastolic phases, according to the following relation:

$$\mathrm{FWHM} = 2\sigma\sqrt{2\log(2)} \quad \Rightarrow \quad \sigma = \frac{\mathrm{FWHM}}{2\sqrt{2\log(2)}} \qquad (3)$$

It must be noted that, despite the means of the three *reflection* Gaussians being initialized so as to be progressively distributed along the diastolic portion of the pulse, their original order is not necessarily retained after optimization. Therefore, the algorithm rearranges them so as to keep the inequality constraint $\mu_2<\mu_3<\mu_4$ always satisfied. After model identification is completed, a series of exclusion criteria is imposed in order to reduce the impact of misfitted pulses. In detail, the algorithm discards potential movement artefacts by setting a 3-σ upper limit on the systolic α_1 amplitude. Furthermore, spuriously merged cardiac cycles are removed on the basis of an analogous upper boundary enforced on the pulse duration, T_p. On the other hand, the simultaneous detection of small 3-σ outlier values of T_p and α_1 is associated with the presence of spurious split pulses. Finally, the conjoint detection of large σ_1 and R^2 outliers is adopted to identify the potential mis-reconstruction of the systolic pulse component. On the whole, N = 20935 waveform models were thus extracted from the dataset of LDF perfusion signals. The qualitative performance of the proposed multi-Gaussian modeling algorithm is displayed in Fig. 3.

4) Feature extraction: A set of waveform descriptors was defined in order to characterize each LDF pulse model and generate the input dataset of the SVM classifier. To this end, we referred to recent literature on the analysis of digital plethysmographic pulse, and took the *stiffness* and *reflection* indexes into consideration. However, in the present study the original definitions of SI and RI were adapted to the framework of the LDF pulse multi-Gaussian decomposition algorithm. A modified SI was thus derived from the time span between the forward wave and the centroid of the three diastolic Gaussians (Fig. 4a), and then normalized to the overall pulse duration ($\mathrm{SI_{norm}}$). The original concept of RI was instead modified by computing the RI as the percentage ratio of the areas beneath the diastolic (A_d) and systolic (A_s) pulse models (Fig. 4a). As performed in [16], the *crest time* (CT), i.e. the duration of systolic pulse ascent, was also evaluated (Fig. 4a) since increased CT values have been observed in subjects with hypertension or atherosclerosis [18]; as for the SI, the CT was normalized to the pulse duration ($\mathrm{CT_{norm}}$). Additionally, these three physiological waveform properties were integrated with the percentage amplitude ratios α_2/α_1, α_3/α_1, α_4/α_1 and the delays Δt_{1-2}, Δt_{1-3}, Δt_{1-4} of the diastolic components relative to the systolic forward wave (Fig. 4b). In summary, the study dataset included N input vectors $\mathbf{x}_i$ of the form:

$$\mathbf{x}_i = \Big(\mathrm{SI_{norm,i}},\ \mathrm{RI_i},\ \mathrm{CT_{norm,i}},\ \frac{\alpha_{2,\mathrm{i}}}{\alpha_{1,\mathrm{i}}},\ \frac{\alpha_{3,\mathrm{i}}}{\alpha_{1,\mathrm{i}}},\ \frac{\alpha_{4,\mathrm{i}}}{\alpha_{1,\mathrm{i}}},\ ...$$
$$...\Delta t_{1-2,i},\ \Delta t_{1-3,i},\ \Delta t_{1-4,i}\Big) \qquad i = 1, ..., N.$$

C. Support Vector Machines for binary classification

Support vector machines (SVM) are a category of efficient linear supervised learning systems for binary classification problems, first introduced by Vapnik [19]. The following section covers a brief overview of SVM mathematical theory, before moving on to deal with the implementation details of the vascular ageing SVM classifier. For an exhaustive treatment, please refer to [20], [21].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2814630, IEEE Transactions on Biomedical Engineering

Table I
CONFIGURATION OF THE LEVENBERG-MARQUARDT OPTIMIZATION ALGORITHM.

COMPONENT		LOWER BOUND	STARTING POINT	UPPER BOUND
	a_1	$0.5 \cdot \Delta p$	$0.8 \cdot \Delta p$	Δp
g_1	μ_1	$t_{P\ start}$	$t_{P\ start} + 0.5 \cdot \Delta t_{sys}$	t_{inc}
	σ_1	$0.5 \cdot \Delta t_{sys}/2\sqrt{2\log(2)}$	$\Delta t_{sys}/2\sqrt{2\log(2)}$	$1.5 \cdot \Delta t_{sys}/2\sqrt{2\log(2)}$
	a_2	0	$0.4 \cdot \Delta p$	$0.6 \cdot \Delta p$
g_2	μ_2	t_{inc}	t_{inc}	$t_{P\ end}$
	σ_2	$0.1 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$	$0.25 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$	$0.33 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$
	a_3	0	$0.4 \cdot \Delta p$	$0.6 \cdot \Delta p$
g_3	μ_3	t_{inc}	$t_{inc} + 0.33 \cdot \Delta t_{dia}$	$t_{P\ end}$
	σ_3	$0.1 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$	$0.25 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$	$0.33 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$
	a_4	0	$0.4 \cdot \Delta p$	$0.6 \cdot \Delta p$
g_4	μ_4	t_{inc}	$t_{inc} + 0.67 \cdot \Delta t_{dia}$	$t_{P\ end}$
	σ_4	$0.1 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$	$0.25 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$	$0.33 \cdot \Delta t_{dia}/2\sqrt{2\log(2)}$

$t_{P\ start}, t_{P\ end}$: left and right time boundaries of the pulse (corresponding to the end-diastolic valleys);
t_{inc}: incisura;
$\Delta t_{sys}, \Delta t_{dia}$: systolic and diastolic time spans;
Δp: maximum amplitude excursion of the pulse.

In a two-class learning problem, the training dataset consists in general of a number of *example* vectors $\mathbf{x}_i$, each associated to a label $y_i \in (-1, +1)$ which identifies the corresponding class. A linear classifier is based on the definition of a linear discriminant function of form:

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b, \tag{4}$$

where $\mathbf{x} : f(\mathbf{x}) = 0$ identify a *decision* hyperplane, which virtually divides the input space I into two separate parts. In these regions $f(\mathbf{x})$, which is proportional to the signed distance to the hyperplane, is either positive or negative; therefore, a classification rule can be directly derived from:

$$R(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b). \tag{5}$$

The simplest formulation of a SVM model is the so-called *maximal margin classifier*, i.e. the classifier which maximizes the geometric margin between two linearly separable classes. Consider the points $\mathbf{x}_+$ and $\mathbf{x}_-$, within the positive and negative regions, which are closest to a given hyperplane $f(\mathbf{x}) = 0$. By assuming they are equidistant from the decision boundary, the following relations hold:

$$\begin{aligned} f(\mathbf{x}_+) &= \mathbf{w}^{\mathrm{T}}\mathbf{x}_+ + b = c, \\ f(\mathbf{x}_-) &= \mathbf{w}^{\mathrm{T}}\mathbf{x}_- + b = -c. \end{aligned} \tag{6}$$

The expression of the geometric margin M_f can then be derived from the equations:

$$\begin{aligned} \mathbf{w}^{\mathrm{T}}\mathbf{x}_+ + b - c = 0, \\ \mathbf{w}^{\mathrm{T}}\mathbf{x}_- + b + c = 0, \end{aligned} \tag{7}$$

which define the corresponding hyperplanes parallel to the decision boundary, whose reciprocal distance (i.e. the margin) can then be expressed as:

$$M_f = \frac{2c}{\|\mathbf{w}\|}. \tag{8}$$

Based on this formulation, the hyperplane which determines the widest separation between the classes can be obtained by maximizing $1/\|\mathbf{w}\|$ or, equivalently, by minimizing $\frac{1}{2}\|\mathbf{w}\|^2$, thus leading to the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{w},b} \ \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to} \ \ y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b) \geq 1 \quad i = 1, ..., N. \end{aligned} \tag{9}$$

However, in real-world classification problems, datasets cannot often be separated by a linear decision boundary, as classes may significantly overlap in the input space domain. This observation is the basis for the formulation of the *soft margin classifier*, in which the constraint of the maximal *hard* margin classifier (Eq.9) is relaxed so as to allow the presence of misclassified instances:

$$\begin{aligned} &\min_{\mathbf{w},b} \ \frac{1}{2}\|\mathbf{w}\|^2 + \mathrm{C}\sum_{i=1}^{N}\xi_i \\ &\text{subject to} \ \ y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b) \geq 1 - \xi_i \quad i = 1, ..., N \end{aligned} \tag{10}$$

where $\xi \in {R_0}^+$ are *slack variables* which allow a number of observations to fall either within the margin ($0 \leq \xi \leq 1$) or on the wrong side of the decision boundary ($\xi > 1$). The cost term C ($\mathrm{C} \in {\mathrm{R}_0}^+$) represents a key *hyperparameter* of the soft margin SVM, in that it determines the relative importance between margin maximization and training accuracy. As explained later on, C must be properly tuned.
The above optimization problem is generally solved using the dual formulation obtained through the method of Lagrange multipliers:

$$\max_{\alpha} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j {\mathbf{x}_i}^{\mathrm{T}}\mathbf{x}_j$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2814630, IEEE Transactions on Biomedical Engineering

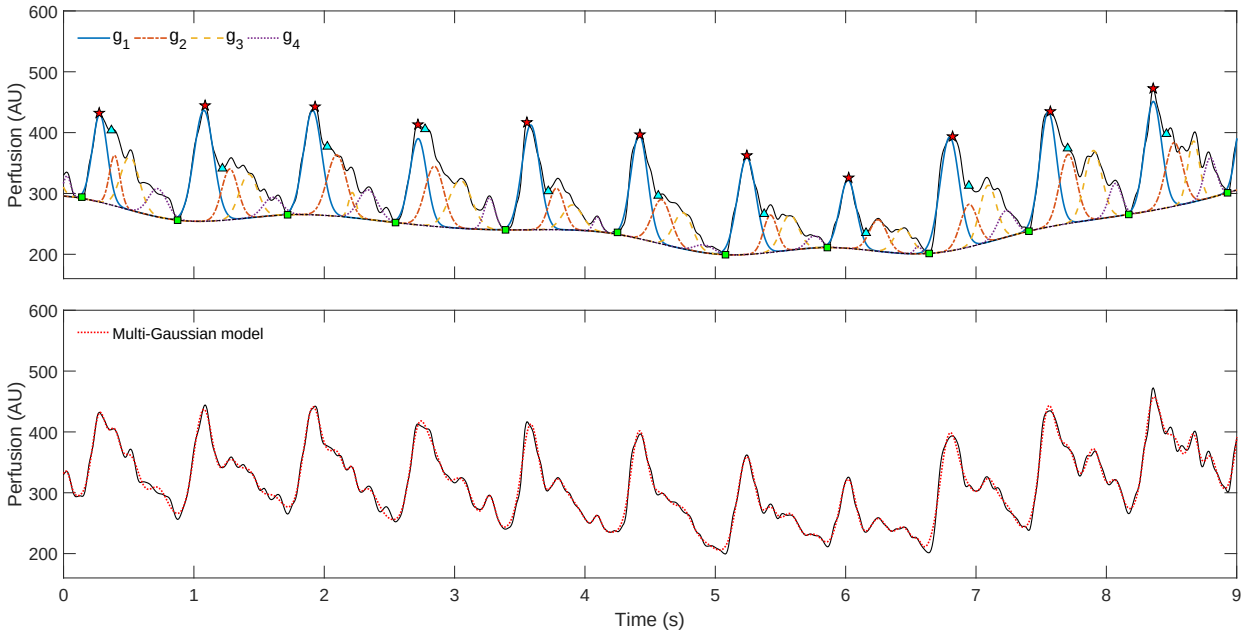


Figure 3. Multi-Gaussian decomposition of the LDF pulse wave: single components (top) and general model (bottom).

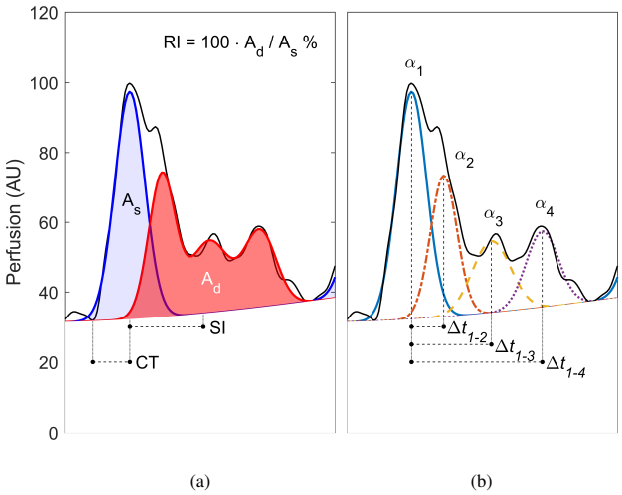


Figure 4. LDF pulse feature extraction for vascular ageing classification.

$$\text{subject to } \begin{cases} \sum_{i=1}^{N} y_i \alpha_i = 0 & i = 1, ..., N \\ 0 \leq \alpha_i \leq C \end{cases} \tag{11}$$

The dual optimization problem leads to the following form for the hyperplane weight vectors:

$$\hat{\mathbf{w}} = \sum_{i=1}^{N} \hat{\alpha}_i y_i \mathbf{x}_i. \tag{12}$$

where only those input examples $\mathbf{x}_i$ associated to non-zero $\hat{\alpha}_i$ coefficients lie on or within the soft margin. These instances are the so-called *support vectors*, as the solution $\hat{\mathbf{w}}$ depends only on them. In the present work, the above optimization problem was solved with Platt's Sequential Minimal Optimization algorithm for fast SVM training [22].

In general, SVM classifiers can achieve higher flexibility and training accuracy if the original input space I is mapped into a new *feature* space $F = \{\phi(\mathbf{x}) | \mathbf{x} \in I\}$ according to a non-linear transformation, so as to allow the linear learning system to generate complex non-linear decision boundaries. As a result, the discriminant hyperplane becomes:

$$f(\mathbf{x}) = \hat{\mathbf{w}}^{\mathrm{T}} \phi(\mathbf{x}) + b = \sum_{i=1}^{N} \hat{\alpha}_i y_i \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}) \rangle + b, \tag{13}$$

while the Lagrange dual function gets the form:

$$L = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle. \tag{14}$$

In the above formulation, the feature map $\phi : I \rightarrow F$ appears only through inner products; as a consequence, the knowledge of the *kernel* function:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle \tag{15}$$

is sufficient to train the SVM in F, without an explicit calculation of the underlying feature map $\phi(\mathbf{x})$. The widely used *radial basis function* kernel was selected in the present study:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}, \tag{16}$$

where the *kernel scale* γ represents, along with the soft margin constant C, a hyperparameter which directly affects the flexibility of the resulting SVM and, accordingly, needs careful configuration. More specifically, when the kernel scale is small, a given example $\mathbf{x}_i$ will be associated to a non-zero $K(\mathbf{x}_i, \mathbf{x}_j)$ with respect to any other input data, which will thus affect the $f(\mathbf{x})$ solution, favouring the generation of a smooth classification boundary. On the other hand, high γ values will make the kernel function decay faster to a value of 0 for increasing $\|\mathbf{x}_i - \mathbf{x}_j\|$ distances; therefore, the discriminant function $f(\mathbf{x})$ will be influenced only by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2814630, IEEE Transactions on Biomedical Engineering

the closest training examples, leading to a higher degree of curvature and boundary flexibility.

Analogous considerations can be made for the C hyperparameter, whose role is to weigh the penalty assigned to misclassified examples: high values will increasingly discourage the admission of any misclassified training instances, thus inducing narrow, sinuous margins, whereas a small constant will in turn lead to lower $1/\|\mathbf{w}\|$ values (Eq. 10) and then to a wider separation.

D. Vascular ageing classification

In the present study, LDF pulse feature data were standardized (by subtracting their mean and dividing by their standard deviation) so as to scale them to a similar range. The binary classes of the SVM model were defined on the basis of a cut-off age of 40 years: each property vector $\mathbf{x}_i$ extracted from the LDF signal of a given subject was labelled accordingly. Fig. 5 shows the waveform feature distributions for the *young* (i.e. age < 40 years) and *old* (i.e. age $\geq$ 40 years) groups. In the remainder of the article, the latter will be referred to as the *positive* class. K-fold cross-validation was used to assess the SVM classifier performance, combining the results of K rounds of separate training, and validation splits of the original set of input data [21]. For every fold, dataset partition was conducted so as not to place part of the $\mathbf{x}_i$ vectors of a given subject into both training and validation sets, thus avoiding any potential bias due to similarity within the same subject. In detail, a rounded 10% of the subjects (and their associated $\mathbf{x}_i$ examples) was used to make up the validation set of each fold; this scheme resulted in 11 folds comprehending mutually exclusive validation sets of 5 subjects (reduced to 4 in the last fold). The following set of scores was adopted for characterizing the classification capability of the SVM model: the classifier *sensitivity* or true positive rate (TPR); the *specificity* or true negative rate (TNR); the F_1 measure (i.e. the harmonic mean of sensitivity and precision, ranging from 0 to 1) [23]; and the Matthews Correlation Coefficient (MCC) [24], which assumes a value of 0 when the performance of the classifier is equivalent to a random guess, reaching a maximum value of +1 for a perfect prediction, while falling to -1 in case of a total disagreement between predictions and *true* classes. These metrics were averaged over the 11 iterations, so as to derive a global evaluation of the classifier.

For the sake of completeness, their definitions are reported below:

$$\begin{aligned} TPR &= \frac{TP}{TP+FN}, \\ TNR &= \frac{TN}{TN+FP}, \\ F_1 &= \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \\ MCC &= \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}}, \end{aligned} \tag{17}$$

where TP and FP namely indicate true and false positive classifications, while TN and FN represent true and false negatives, respectively.

As stated in the previous section, the soft margin constant C and the kernel scale γ carry a significant impact on the behaviour of the SVM classifier, and hence require proper tuning. This *model selection* was performed separately within each of the 11 folds, through a grid search over the values $\mathrm{C}_{grid} \in (10^1,\ 10^{1.5},\ 10^2)$ and $\gamma_{grid} \in (10^0,\ 10^{0.5},\ 10^1)$; the optimal pair of hyperparameters was identified on the basis of a nested 3-fold validation over the training set, by maximizing the average geometric mean of TPR and TNR over the resulting three validation subsets. This parameter was adopted so as to take the dataset unbalance into account and avoid producing SVM models that behaved poorly on the contextual *minority* class, specifically the age $\geq$ 40 group. Higher ($\mathrm{C}_{\mathrm{grid}}$,$\gamma_{grid}$) solutions were not systematically searched, as preliminary tests associated them with significant over-fitting regimes.

III. Results

The proposed model demonstrated the ability to fit and characterize the microvascular LDF pulse, showing a coefficient of determination R^2 of 0.98 ± 0.03 (mean $\pm\, \sigma$) over the N pulse waves. Goodness of fit was additionally evaluated on the basis of a normalized root mean square error (NRMSE), defined as:

$$\mathrm{NRMSE} = 1 - \frac{\|p(t) - G(t)\|}{\|p(t) - \overline{p}\|} \tag{18}$$

where $p(t)$ and $G(t)$ are respectively the LDF signal and the multi-Gaussian model, while $\overline{p}$ represents the mean perfusion. Based on this formulation, a $\mathrm{NRMSE} = 1$ would theoretically correspond to a perfect data fit, whereas a value of 0 would imply that the multi-Gaussian model does not perform better than a straight line. An average NRMSE score of 0.90 ± 0.03 was obtained on the set of 54 LDF perfusion signals, markedly exceeding the performance shown by a previous version of the modeling algorithm, presented in [25].

For each of the N pulse waves composing the study dataset, the fitting procedure was repeated 100 times, applying a random initialization of the multi-Gaussian model; at each iteration, the starting parameter values were extracted from a uniform probability distribution, within the respective lower and upper boundaries listed in Table I. Then, the normalized interquartile range of the nine morphological features, $\mathrm{IQR}_{\mathrm{norm}}$ (defined as the interquartile range, divided by the median), was estimated for each wave so as to assess their sensitivity to the initial fitting conditions. The results (median, IQR), obtained for each waveform descriptor, are the following: $\mathrm{SI}_{\mathrm{norm}}$: 0.21% (0.00%, 7.81%); RI: 0.05% (0.00%, 3.15%); $\mathrm{CT}_{\mathrm{norm}}$: 0.01% (0.00%, 0.84%); α_2/α_1: 0.58% (0.00%, 17.64%); α_3/α_1: 1.16% (0.00%, 31.30%); α_4/α_1: 0.73% (0.01%, 35.28%); Δt_{1-2}: 0.14% (0.00%, 13.74%); Δt_{1-3}: 0.29% (0.00%, 22.60%); Δt_{1-4}: 0.17% (0.00%, 10.73%). These outcome values highlight the stability of the proposed multi-Gaussian modeling algorithm. Regarding the presence of relatively wide IQR ranges, it is relevant to recall how the original reconstruction method finely tunes the initial conditions of the Levenberg-Marquardt optimizer on the basis of specific characteristics of each detected pulse, thus contributing to its stability. However, in the present test, parameter initialization was random and not fully driven by a priori information on

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2814630, IEEE Transactions on Biomedical Engineering

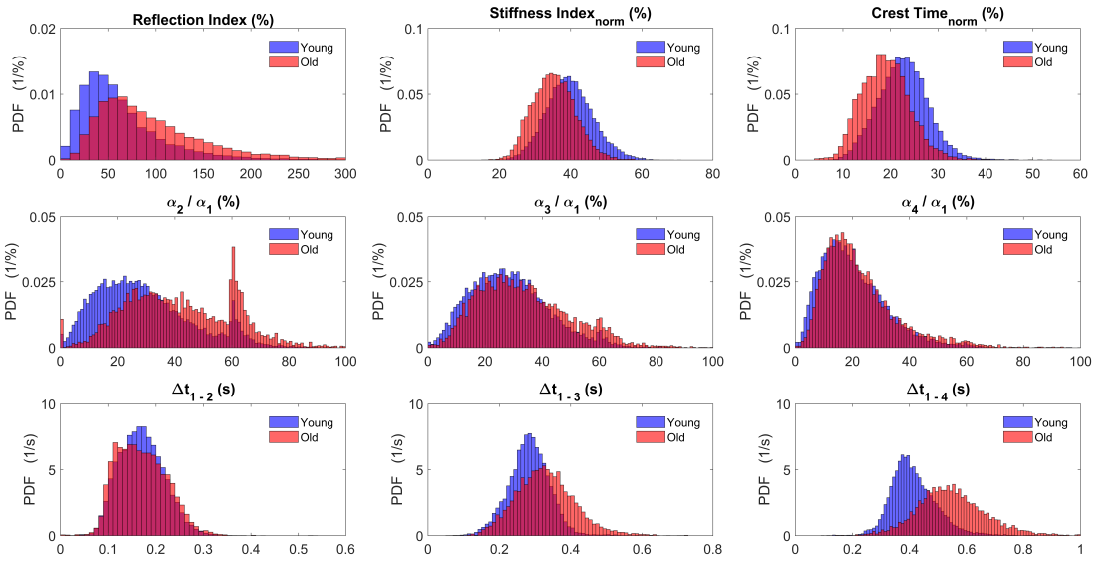


Figure 5. Microvascular feature probability density functions (PDF) for the *young* (i.e. age<40 years) and *old* (i.e. age≥40 years) classes.

the pulse profile, and this might have had a detrimental effect on the overall performance.

The evaluation of the SVM pulse classifier, based on the approach detailed in the previous paragraph, was repeated over 30 successive experiments, so as to consider the variability linked to the different partition of the input data for K-fold cross-validation. Concerning the classification of the single modeled pulse waves, the resulting average TNR and TPR (95% CI) were respectively 0.902 (0.896, 0.907) and 0.646 (0.636, 0.657); a mean F_1 score of 0.678 (0.668, 0.688) was obtained, along with a MCC of 0.550 (0.539, 0.561). The box plots of the fold-wise mean performance scores are shown in Fig. 6.

Although these results indicate the presence of a relevant fraction of false negatives, given the predefined age threshold of 40 years, a different perspective can be obtained if we shift the focus from pulse to subject classification (Figs. 7 and 8). A high correlation between the actual age of the subjects (from the 11 validation sets) and the rate of waves classified as *Old-Pulse* is indeed observed, reflected by a mean Pearson's correlation coefficient of 0.808 (0.803, 0.812). Therefore, this *Old-Pulse Rate* (OPR) was adopted as a criterion to discriminate between young and old individuals. For each experiment, a receiver operating characteristic (ROC) curve was thus generated by varying the OPR threshold, and evaluating the subject true/false positive predictions on all folds. The overall average area under the curve (AUC), hence obtained from the entire set of experiments, was 0.953 (0.951, 0.956), which is indicative of a very good discrimination performance. Table II summarizes these global results.

Table II
PERFORMANCE METRICS FOR VASCULAR AGEING CLASSIFICATION.

CLASSIFICATION	METRIC	MEAN	95% CI
LDF Pulse	TNR	0.902	(0.896, 0.907)
	TPR	0.646	(0.636, 0.657)
	F_1	0.678	(0.668, 0.688)
	MCC	0.550	(0.539, 0.561)
Subjects	Pearson's r	0.808	(0.803, 0.812)
	ROC AUC	0.953	(0.951, 0.956)

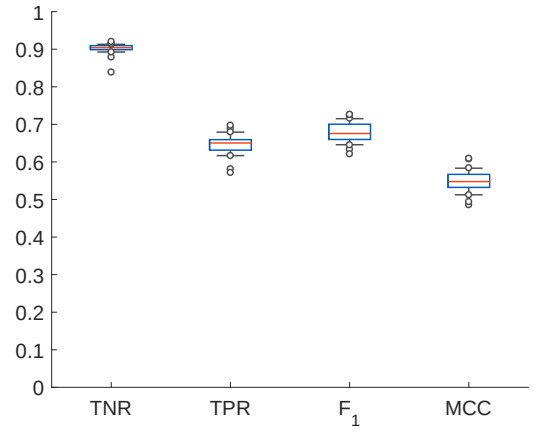
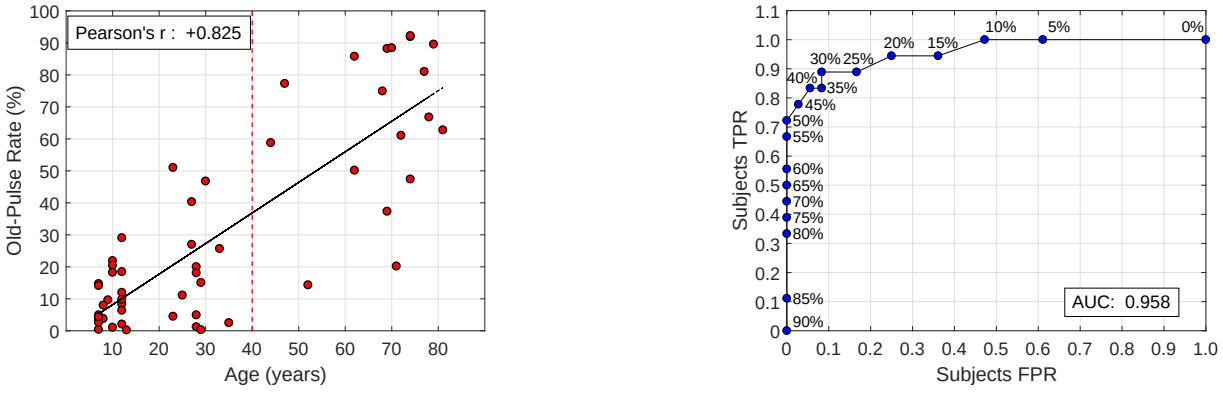


Figure 6. Pulse ageing classification: box plots of the fold-wise mean performance scores achieved over the 30 cross-validation experiments.

IV. Discussion

In order to derive an indication on the qualitative aspect of the *old* and *young* LDF pulses, as predicted by the SVM classifier, the input feature vectors corresponding to the centroids of the *true positive* and *true negative* clusters were identified. These samples were then adopted to reconstruct two representative pulse models (normalized to the amplitude of the systolic component, α_1, for the sake of comparison). Their shape, shown in Fig. 9, is in accordance with the observations made by Millasseau et al. [26], which report a reduction in the DVP *incisura* with ageing. The authors contextually referred to the findings of Dawber and Lax, which state that DVP waveforms with no inflection along the descent of the systolic wave or without any *dicrotic* local minima are more prevalent in older healthy individuals or in subjects with CVD [27], or arteriosclerosis [28]. The general physiological interpretation is that the ageing-related stiffening of the large

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2814630, IEEE Transactions on Biomedical Engineering



(a) (b)

Figure 7. Subject ageing classification: (a) sample correlation plot between the age of the subjects and the *Old-Pulse Rate* (OPR), corresponding to the maximum Pearson's r achieved in the experiments; (b) sample ROC curve obtained by shifting the OPR threshold (displayed next to the points).

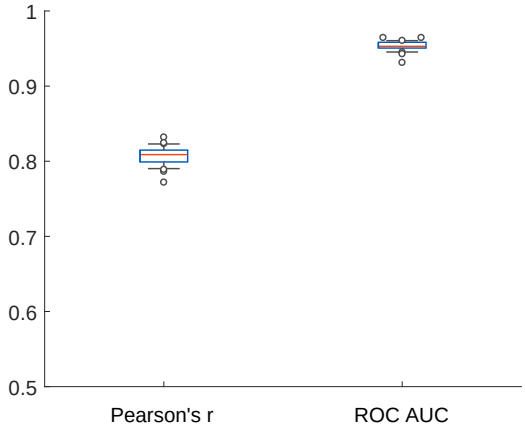


Figure 8. Subject ageing classification: box plots of the Pearson's correlation coefficients and the ROC AUC estimates obtained from the 30 experimental iterations.

conduit arteries determines a faster back-propagation of the pressure waves reflected by the peripheral circulation, and thus a higher degree of temporal superposition with the forward-travelling systolic pulse. This hypothesis is in fact reflected by the lower SI_{norm} index of the *Old-Pulse* model, and by its higher systolic peak value. Other salient evidence is the larger RI (107.7%), compared to the *Young-Pulse* model (65.1%): this may be indicative of an elevated impedance mismatch in the peripheral resistance vessels, due to an altered level of smooth muscle cells tone. Nevertheless, caution should be used while interpreting the LDF waveform and its quantitative descriptors through the same theoretical framework established for the analysis of the plethysmographic pulse, as these techniques have different underlying measurands. Indeed, LDF provides a non-absolute measurement of tissue perfusion which, under a series of assumptions, scales with the local concentration of red blood cells and their average velocity [29]. On the other hand, photoplethysmography is sensible to local blood volumetric changes; under pulsatile conditions these are indirectly related to perfusion, as the local variation in blood concentration is determined by the difference between the inflow and outflow of blood from the measurement volume. Mizeva et al. have recently demonstrated a significant correlation between the oscillations of digital plethysmographic and LDF signals within the low frequency intervals associated with endothelial, myogenic and

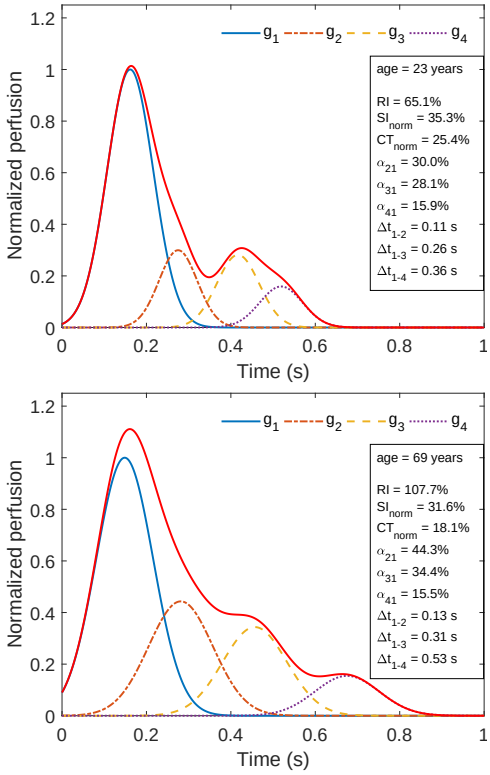


Figure 9. LDF pulse models generated from the centroids of the *true negative* (top) and *true positive* (bottom) feature clusters.

neurogenic vasomotion mechanisms [30]. However, to the authors' knowledge a thorough comparison of the DVP and LDF microvascular pulse contours is lacking and further work is needed for the characterization of the LDF waveform changes in response to physiological and pharmacological stimuli, or to specific pathological conditions. This would be highly relevant for the identification of specific LDF pulse descriptors which are suitable for physiological interpretation. Furthermore, the performance of the vascular ageing classifier may be generally improved through the preliminary removal of redundant input variables with limited discriminant capability in this specific classification task. A feature subset selection approach, such as the minimum-redundancy-maximum-relevance algorithm proposed by Peng [31], may therefore be adopted in future implementation of the SVM classifier.

V. Conclusion

This paper presented an approach for the automated assessment of vascular ageing based on the modeling of the microvascular cardiac waveform. The multi-Gaussian decomposition algorithm developed for feature extraction demonstrated to be suitable for reconstructing the shape of the LDF pulses. Moreover, the SVM supervised classifier achieved overall good performance in the discrimination of *Young* and *Old-Pulse* models, with respect to the arbitrary 40 years threshold adopted for the generation of the binary classes. More interestingly, a high correlation between the rate of *Old-Pulse* predictions and the underlying age of the test subjects was observed, and very high ROC AUC values were obtained, by adopting this rate as a criterion value for discriminating differently aged individuals. Therefore, the presented methodology may have the underlying capability to detect the progressive ageing of the vascular system. Since diabetic angiopathy is known to be characterized by vascular alterations comparable to ageing [32], this approach may also be suitable for the screening of diabetic patients. We thus aim to investigate the possibility of implementing this supervised learning methodology for discriminating between diabetic and control perfusion pulse waves.

References

[1] S. S. Mahmood, D. Levy, R. S. Vasan, and T. J. Wang, "The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective," *The Lancet*, vol. 383, no. 9921, pp. 999–1008, 2004.

[2] G. F. Mitchell, H. Parise, E. J. Benjamin, M. G. Larson, M. J. Keyes, J. A. Vita, R. S. Vasan, and D. Levy, "Changes in arterial stiffness and wave reflection with advancing age in healthy men and women," *Hypertension*, vol. 43, no. 6, pp. 1239–1245, 2004.

[3] M. Tesauro, A. Mauriello, V. Rovella, M. Annicchiarico-Petruzzelli, C. Cardillo, G. Melino, and N. Di Daniele, "Arterial ageing: from endothelial dysfunction to vascular calcification," *Journal of Internal Medicine*, vol. 281, no. 5, pp. 471–482, 2017.

[4] P. Mistriotis and S. T. Andreadis, "Vascular aging: molecular mechanisms and potential treatments for vascular rejuvenation," *Ageing Research Reviews*, vol. 37, pp. 94–116, 2017.

[5] W. W. Nichols, M. F. O'Rourke, and C. Vlachopoulos, *McDonald's Blood Flow in Arteries, Sixth Edition: Theoretical, Experimental and Clinical Principles*. CRC Press, 2011.

[6] S. Laurent, L. Marais, and P. Boutouyrie, "The noninvasive assessment of vascular ageing," *Canadian Journal of Cardiology*, vol. 32, pp. 669–679, 2016.

[7] J. Alastruey, T. Passerini, L. Formaggia, and J. Peiró, "Physical determining factors of the arterial pulse waveform: theoretical analysis and calculation using the 1-D formulation," *Journal of Engineering Mathematics*, vol. 77, no. 1, pp. 19–37, 2012.

[8] K. Takazawa, M. Fujita, K. Yabe, T. Sakai, T. Kobayashi, K. Maeda, Y. Yamashita, M. Hase, and C. Ibukiyama, "Clinical usefulness of the second derivative of a plethysmogram (acceleration plethysmogram)," *Journal of Cardiology*, vol. 23, pp. 207–217, 1993.

[9] H. Takada, K. Washino, J. Harrell, and H. Iwata, "Acceleration plethysmography to evaluate aging effect in cardiovascular system," *Medical Progress Through Technology*, vol. 21, no. 4, pp. 205–210, 1996.

[10] K. Takazawa, N. Tanaka, M. Fujita, O. Matsuoka, T. Saiki, and M. Aikawa, "Assessment of vasoactive agents and vascular aging by second derivative of the photoplethysmograph waveform," *Hypertension*, vol. 32, pp. 365–370, 1998.

[11] I. Imanaga, H. Hara, S. Koyanagi, and K. Tanaka, "Correlation between wave components of the second derivative of plethysmogram and arterial distensibility," *Japanese Heart Journal*, vol. 39, no. 6, pp. 775–784, 1998.

[12] S. C. Millasseau, R. Kelly, J. M. Ritter, and P. J. Chowienczyk, "The vascular impact of aging and vasoactive drugs: comparison of two digital volume pulse measurements," *American Journal of Hypertension*, vol. 16, pp. 467–472, 2003.

[13] M. C. Baruch, D. E. R. Warburton, S. S. D. Bredin, A. Cote, D. W. Gerdt, and C. M. Adkins, "Pulse Decomposition Analysis of the digital arterial pulse during hemorrhage simulation," *Nonlinear Biomedical Physics*, vol. 5, no. 1, 2011.

[14] R. Couceiro, P. Carvalho, R. P. Paiva, J. Henriques, I. Quintal, M. Antunes, J. Muehlsteff, C. Eickholt, C. Brinkmeyer, M. Kelm, and C. Meyer, "Assessment of cardiovascular function from multi-Gaussian fitting of a finger photoplethysmogram," *Physiological Measurement*, vol. 36, pp. 1801–1825, 2015.

[15] G. Straface, L. Landini, M. Barrella, M. Bevilacqua, A. Evangelisti, and L. Bocchi, "Analysis of the microcirculatory pulse wave: age-related alterations," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 7362–7365.

[16] S. R. Alty, N. Angarita-Jaimes, S. C. Millasseau, and P. J. Chowienczyk, "Predicting arterial stiffness from the digital volume pulse waveform," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2268–2275, 2007.

[17] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical analysis*. Springer, 1978, pp. 105–116.

[18] J. B. Dillon and A. B. Hertzman, "The form of the volume pulse in the finger pad in health, arteriosclerosis, and hypertension," *American Heart Journal*, vol. 21, no. 2, pp. 172–190, 1941.

[19] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*. ACM press, 1992, pp. 144–152.

[20] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[21] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer Series in Statistics, 2001, vol. 1.

[22] J. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods – Support Vector Learning*, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. MIT Press, 1999, pp. 185–208.

[23] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[24] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica et Biophysica Acta*, vol. 405, no. 2, pp. 442–451, 1975.

[25] M. Sorelli, A. Perrella, and L. Bocchi, "Cardiac pulse waves modeling and analysis in laser Doppler perfusion signals of the skin microcirculation," in *CMBEBIH 2017: Proceedings of the International Conference on Medical and Biological Engineering 2017*. Springer Singapore, 2017, vol. 62, pp. 20–25.

[26] S. C. Millasseau, R. Kelly, J. M. Ritter, and P. J. Chowienczyk, "Determination of age-related increases in large artery stiffness by digital pulse contour analysis," *Clinical Science*, vol. 103, no. 4, pp. 371–377, 2002.

[27] T. R. Dawber, H. E. Thomas Jr, and P. M. McNamara, "Characteristics of the dicrotic notch of the arterial pulse wave in coronary heart disease," *Angiology*, vol. 24, no. 4, pp. 244–255, 1973.

[28] H. Lax, A. W. Feinberg, and B. M. Cohen, "Studies of the arterial pulse wave: the normal pulse wave and its modification in the presence of human arteriosclerosis," *Journal of Chronic Diseases*, vol. 3, no. 6, pp. 618–631, 1956.

[29] A. Humeau, W. Steenbergen, H. Nilsson, and T. Strömberg, "Laser Doppler perfusion monitoring and imaging: novel approaches," *Medical & Biological Engineering & Computing*, vol. 45, no. 5, pp. 421–435, 2007.

[30] I. Mizeva, C. Di Maria, P. Frick, S. Podtaev, and J. Allen, "Quantifying the correlation between photoplethysmography and laser Doppler flowmetry microvascular low-frequency oscillations," *Journal of Biomedical Optics*, vol. 20, no. 3, p. 037007, 2015.

[31] H. Peng and C. Ding, "Minimum redundancy and maximum relevance feature selection and recent advances in cancer classification," *Feature Selection for Data Mining*, vol. 3, no. 2, pp. 185–205, 2005.

[32] S. S. DeLoach and R. R. Townsend, "Vascular stiffness: its measurement and significance for epidemiologic and outcome studies," *Clinical Journal of the American Society of Nephrology*, vol. 3, no. 1, pp. 184–192, 2008.