Artificial Intelligence in Medicine 53 (2011) 127–138

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques

Enric Monte-Moreno*

Department Signal Theory and Communications, Universitat Politècnica de Catalunya, Campus Nord, Edifici D5, C/Jordi Girona, 1-3, 08034, Barcelona, Spain

ARTICLE INFO

Article history:
Received 14 May 2010
Received in revised form 2 May 2011
Accepted 8 May 2011

Keywords:
Machine learning
Photoplethysmography
Noninvasive measurement
Blood glucose estimate
Blood pressure estimate

ABSTRACT

Objective: This work presents a system for a simultaneous non-invasive estimate of the blood glucose level (BGL) and the systolic (SBP) and diastolic (DBP) blood pressure, using a photoplethysmograph (PPG) and machine learning techniques. The method is independent of the person whose values are being measured and does not need calibration over time or subjects.

Methodology: The architecture of the system consists of a photoplethysmograph sensor, an activity detection module, a signal processing module that extracts features from the PPG waveform, and a machine learning algorithm that estimates the SBP, DBP and BGL values. The idea that underlies the system is that there is functional relationship between the shape of the PPG waveform and the blood pressure and glucose levels.

Results: As described in this paper we tested this method on 410 individuals without performing any personalized calibration. The results were computed after cross validation. The machine learning techniques tested were: ridge linear regression, a multilayer perceptron neural network, support vector machines and random forests. The best results were obtained with the random forest technique. In the case of blood pressure, the resulting coefficients of determination for reference vs. prediction were $R^2_{SBP} = 0.91$, $R^2_{DBP} = 0.89$, and $R^2_{BGL} = 0.90$. For the glucose estimation, distribution of the points on a Clarke error grid placed 87.7% of points in zone A, 10.3% in zone B, and 1.9% in zone D. Blood pressure values complied with the grade B protocol of the British Hypertension society.

Conclusion: An effective system for estimate of blood glucose and blood pressure from a photoplethysmograph is presented. The main advantage of the system is that for clinical use it complies with the grade B protocol of the British Hypertension society for the blood pressure and only in 1.9% of the cases did not detect hypoglycemia or hyperglycemia.

1. Introduction

Measurements of blood glucose levels (BGL) and systolic (SBP) and diastolic (DBP) pressures are currently made by means that are either invasive or require mechanical intervention. To measure glucose levels current systems require blood sampling and use of certain reagents. Blood pressure is assessed by mechanically measuring the pressure difference between an external agent (the cuff squeezing the arm) and the pressure of the blood vessels. An estimate is made by comparing the pressures. An alternative to these methods is the use of a non invasive device such as a photoplethysmograph (PPG) that can assess the distal heart beat (i.e. signal shape as measured on the finger's capillaries) by means of differences in light absorption. This can be done for instance, with a pulse oximeter.

The need for a non invasive method for estimating both values is justified by the size of the affected population. According to Wild et al. [1] at least 171 million people worldwide suffer from diabetes. Monitoring glucose levels requires the use of glucometers, disposable needles, and blood test strips, which can be expensive and cumbersome to use. Monitoring blood pressure can help to protect against complications associated with diabetes, such as retinal, renal and cardiovascular disease. Moreover, according to Wolf-Maier et al. [2], the age and sex adjusted prevalence of hypertension (defined as >140/90 mmHg) was 28% in North American countries and 44% in European countries. Because hypertension increases the risk of heart attack, heart failure, stroke and kidney disease, a cuffless and easy-to-use method for monitoring blood pressure might improve daily control of blood pressure levels. Moreover both variables are routinely measured in intensive care units. Specifically, a non-invasive device that provides real-time, continuous arterial pressure and glucose measurements is of interest because of its lower risk of infection.

Additionally, our interest in simultaneous measurement of BP and BGL is particularly strong because in 2004, 75% of adults with

* Corresponding author. Tel.: +34 93 4016435; fax: +34 93 4016447.
E-mail address: enric.monte@upc.edu

0933-3657/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2011.05.001

self-reported diabetes had blood pressures greater than or equal to 130/80 mmHg, or used prescription medications for hypertension [3], this represents an additional argument in favor of a noninvasive monitoring system for both quantities.

There are precedents for measurements of BGL, SBP and DBP from the PPG signal. A PPG signal is obtained by illuminating the skin and measuring changes in light absorption. This signal is obtained from a photoplethysmograph, which is based on small light-emitting diodes facing a photodiode. In our case the light consisted on a red beam and a beam in the near infrared band. The light absorption depends on the heart beat because the blood vessels in the finger expand and contract with each heart beat. The difference between the minimum absorption and the peak absorption of the light intensity is proportional to the distal cardiac pulse. For non-invasive glucose measurements, most of the systems that were previously developed are based on near infrared spectroscopy and the majority have proved to be unsuccessful [4]. The difficulty lies in the fact that the light absorption of glucose is two orders of magnitude below that of water, making reliable measurements difficult and introducing the need for repeated system recalibration either between individuals or over time. Haaland et al. [5] describe some of the difficulties associated with this approach. For BP, the PPG signal was used by both Cattivelli and Garudadri [6] and Chen et al. [7], both are based on the pulse arrival time, i.e. by the time interval between the QRS apex on an electrocardiogram and on a beat-by-beat finger-tip photoplethysmogram generated by an oximetric sensor that requires complex equipment and frequent calibration. Other methods, summarized in Allen [8], are difficult to use.

An approach to non invasive systolic blood pressure estimation which specifically uses the shape of the PPG waveform [9], is based on the fact that a pressure cuff on the right arm makes the pressure pulses in the arteries distal to the cuff disappear due to the collapse of the artery under the cuff. In the paper the authors record the PPG signal from the two index fingers, and by analyzing the derivative of the signal and by correlating both PPG signals with the cuff pressure estimate the actual SBP. Note that the method does not estimate the DBP level. Another approach that makes use of the shape of the PPG [10], found that increases on the SBP and DBP were reflected on the DC and AC components of the PPG waveform. We do not follow the approach proposed in this paper because of the need of calibration of the system to each individual. Finally, in Ref. [11], the authors measure the shape of the PPG signal, and compute features from the acceleration pulse wave derived from the PPG signal, which are used as input of a decision tree followed by a multiple regression analysis. A limitation of this method is that it estimates only systolic blood pressures and does not use information about the heart rate.

In contrast to these methods we propose an approach that does not measure differences in time intervals or in light absorption at different wavelengths, but rather measures the effect of physiological alterations on the *shape* of the PPG waveform and on the *heart rate*. These measurements are related to the individual's hemodynamics and the glucose levels.

In this paper we propose a device that is structured into three modules. A summary of the system itself is shown in Fig. 1. From the signals entering the PPG, the *activity detection* (AD) module selects a window of 1-min duration containing a clean signal (comprising 4500 consecutive samples). This module eliminates artifacts and avoids loss of signal due to movements of the finger. A *signal processing* (SP) module processes the signal in this window and extracts a vector of features. The output of this module is a fixed length vector. Finally, a *machine learning* (ML) module infers the function that relates the output of the SP module to the variables, BGL, SBP and DBP.

The system was trained and tested with cross validation in a population of 410 individuals. The measurements used as input and reference for training the system were obtained as follows:

- PPG input measurements were made with the fingertip pulse oximeter device *iPod Digital Oximeter* [12]. For details on the accuracy of the pulse oximeter see Appendix A.
- SBP and DBP target measurements were made with an aneroid sphygmomanometer (conventional sphygmomanometer cuff).
- BGL target measurements were made with the self-monitoring blood glucometer *Accu-Chek Aviva* [13].

It should be stressed that the database consisted of 410 different individuals, with one measurement per individual and that the design criterion was that the system should not require calibration either between individuals or over time.

Our underlying assumption was that a functional relationship exists between the PPG signal and blood pressure and glucose levels. This functional relationship can be explained by several facts, such as the fact that the pulse oximeter yields a waveform that is the result of the propagation of the ECG signal through the arteries, veins and capillaries, also by the fact that the PPG signal reflects the hemodynamics of the individual. In addition the state of the autonomic nervous system is partly reflected in the measured waveform and also because the shape depends partially on the relationship between distal capillary bed and blood viscosity. We also assumed that a machine learning system would be able to infer this functional relationship from a labeled database. Our results bear out these assumptions.

The paper is structured as follows: in Section 2 we summarize the physiological facts that support our selection of the features used to estimate pressures and glucose levels. In Section 3.1 we describe an AD module for filtering out artifacts and extracting a useful window of signal; this window was then used as input to the SP module. In Section 3.2 we briefy describe the SP module that computed the feature vector used by the ML algorithm. In Section 3.3 we present a description of the ML module and the ML techniques. In Section 3.4 we describe the database. In Section 4 we present our results and finally Section 5 is devoted to conclusions and future work.

2. Physiological factors relevant to the proposed system

In this section we summarize the physiological properties that support a theoretical estimation of blood pressure and glucose level from the PPG signal. We also relate these properties to the signal processing techniques described in Sections 3.1 (the AD module) and 3.2 (the SP module). This section lists physiological facts that are briefly described, along with the signal processing techniques used to capture their characteristics. Afterwards, in Section 3.2, we provide a detailed description of each signal processing technique.

Blood viscosity and vessel compliance. Blood viscosity increases under conditions of insulin resistance in fat cells due to the presence of elevated plasma levels of free fatty acids. Blood viscosity also depends on both blood pressure and glucose level [14]. Thus the viscosity of the blood alters the flux of blood in the capillaries and therefore the shape of the PPG pulse. Vessel compliance is related to the vessel wall response to increases in pressure that are in turn determined by the state of the cardiovascular system.

- Both properties are reflected in the degree to which the PPG waveform is damped; this information can be obtained via spectral analysis and an autoregressive model of the waveform.

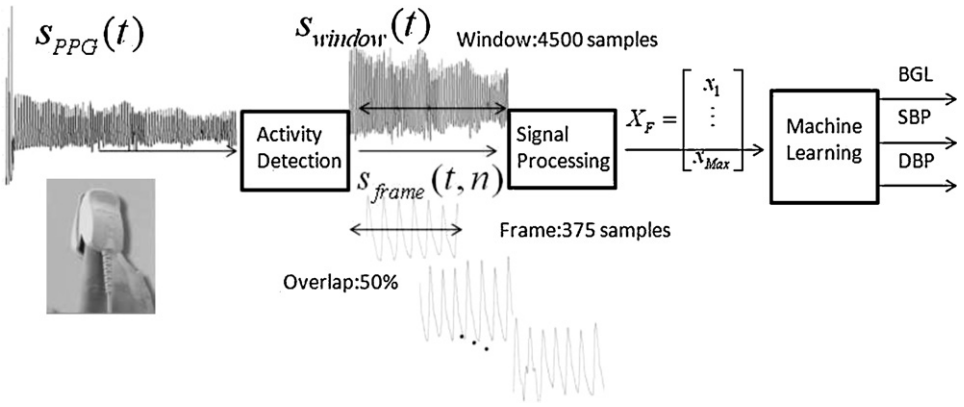

Fig. 1. Diagram of the system.

Hemodynamics. The baroreflex[1] relates the heart rate to the blood pressure [15,16]. Several studies [17] have explored the spectral properties of the HR and its relationship to blood pressure levels. These studies identified a correlation and a transfer function between different frequencies and the mean arterial pressure. Previous investigators have described a relationship between spectral bands of the pulse interval and systolic arterial pressure [18], and between HR variability and the baroreflex in humans [19]; the non-linearity of the latter relationship has also been reported [20,21].

- This justifies the use of ML techniques flexible enough to infer the nonlinear relationship between the spectral characteristics of the HR signal and the blood pressure, as well as the use of spectral analysis features.

Metabolic syndrome. The metabolic syndrome [22] consists of hypertension, obesity, and insulin resistance and other pathologies. The syndrome causes changes in glucose and blood pressure levels as well as the state of the autonomic system and general hemodynamics, and is related to each subject's individual demographic and morphometric characteristics.

- This requires that, we use variables such as age, weight and body mass index, in addition to describing the shape of the pulse.

Relationship between diabetes and the HR variability. There is evidence that diabetes alters the relationship between HR and blood pressure [23]. Diabetic neuropathies associated with diabetes mellitus result from diabetic microvascular injury to the small blood vessels that supply the nerves. These injuries are reflected in the power spectrum of the HR [24] and manifest as a much lower HR variability. Moreover, they are indirectly reflected in blood pressure dynamics, due to the fact that the parasympathetic system is more affected than the sympathetic.

- Since there is a functional relationship between altered glucose levels due to diabetes and HR variability, the features that will capture this relationship are the power spectrum of the distance between pulses and statistics about the HR and the HR variability.

Emotional states. It is known that emotional states [25], such as stress, anger, happiness, surprise, etc. change the values of the blood pressure and glucose by altering the activity of the sympathetic and parasympathetic nervous systems. These emotional states have been correlated with HR variability and the spectral shape of the cardiac tracing. McCraty et al. [26] showed that different emotional states are reflected in the shape of the power spectrum of the HR variability.

- This also justifies the use of features related to the shape of the PPG pulse and the power spectrum and statistics of the HR (i.e mean value, standard deviation, skewness, etc.).

Breathing and the autonomic nervous system. Several studies have shown that the respiratory frequency can regulate the blood pressure [27,28] and that in diabetic patients, the two are closely related [29]. It is also known that the respiratory rate can be extracted from the PPG signal[30].

- This justifies the use of the low frequency components of the PPG energy, which are a measure of the respiratory rate.

3. Materials and methods

This section gives a detailed account of the three parts that constitute the system, shown together in Fig. 1. The recorded signal from the PPG is digitalized into the sequence $S_{PPG}(t)$ at a sample rate of 75 samples/s and is the input to the activity detection module. The AD module selects a window of the $S_{PPG}(t)$ containing 4500 consecutive samples of clean signal. The window is then segmented into frames. The output of the AD module is:

- A vector $S_{window}(t)$ that consists of 1 min of clean signal (window length, $L_{window} = 4500$ consecutive samples).
- An array $S_{frame}(\tau, n)$, that consists of frames of 5 s (375 samples) with an overlap of 50% that were extracted from $S_{window}(t)$. The index τ indicates the sample number in a frame and n is the frame number. The number of samples in a frame is $L_{frame} = 375$, and the number of frames in a window is $N_{frame} = 24$.

Both $S_{window}(t)$ and $S_{frame}(\tau, n)$ are input to the signal processing module, whose output is the vector X_F of dimension 33. This vector contains the set of features used in the machine learning module. Finally we present the ML module and justify the technique selected for inferring the function relating X_F and the desired targets, i.e. the reference values of BGL, SBP and DBP.

The criterion for deciding $L_{frame} = 375$ was that a frame of 5 s was long enough to record several heart beats, and at the same time, it was short enough to consider that the change in amplitude due to the breathing rate was negligible. A frame duration lower than 5 s might capture too few heart beats, and therefore, the detection of activity might be unreliable. However, a frame duration signifi-

[1] A negative feedback system that controls short-term changes in blood pressure.

cantly longer than 5 s might yield unreliable endpoints, i.e., frames with a transient. The τ index is the local time in the n th frame.

3.1. Activity detection module

In this section we summarize the structure of the AD module. The aim of the AD module is to remove the corrupted parts of the signal (i.e. initial transient, spurious clicks, loss of signal, noise, saturations, etc.), giving as output the vector $S_{\text{window}}(t)$ and the array $S_{\text{frame}}(\tau, n)$. The AD module was designed similarly to the techniques used for detecting voice activity in speech technologies [31]. We were interested in a robust and computationally light module. This was done by computing sets of features that used complementary information. Also some features had a dual use, as they were used in both the AD and SP modules.

The activity detection module consists of the following parts:

1. A submodule that computes the feature vector X^n_{AD} for each frame.
2. A decision function based on a linear classifier that for each frame outputs one of two classes, either "signal", or "not signal".
3. A finite state automaton (FSA).

This module takes as input the PPG signal, $S_{\text{PPG}}(t)$, and sequentially computes $S_{\text{frame}}(\tau, n)$. For each frame we compute a feature vector X^n_{AD}, that is the input to a decision function. The decision function for each X^n_{AD} outputs the class of the frame. The class of each frame is the input to a FSA that determines the endpoints of the clean signal, i.e., determines the 1 min window of consecutive samples of clean signal. Once these endpoints are found, the system outputs the set of frames $S_{\text{frame}}(\tau, n)$ that fall between the endpoints along with the signal $S_{\text{window}}(t)$. Next we describe each of these parts in detail.

3.1.1. Set of features of the vector X^n_{AD}

In this subsection we present the set of features that are computed from each frame in order to form the vector X^n_{AD}. The vector X^n_{AD} is used as the input to a classifier (decision function) that classifies the frame as either "signal" or "not signal".

Below we present a summary of the features used in the vector X^n_{AD}. When a given feature is also used in X_F, we give a detailed description in Section 3.2. *Kaiser–Teager energy of the PPG signal.* The Kaiser–Teager energy is a standard tool used in speech processing for end-point detection, and for finding energy profiles of signals with periodic signal components. The Kaiser–Teager energy (KTE) was computed at the frame level. The KTE_n is based on a harmonic generation model of a sinusoid [32]. The KTE_n model assumes that the signal is a sinusoid x_t of the form $x(t) = A\cos(\Omega t + \phi)$ and the energy is computed by the left part of Eq. 1. After simple algebra [33] it can be shown that with the sinusoid model, $\text{KTE}_n(t)$ is approximately proportional to the square of the amplitude times the angular frequency $A^2\Omega^2$.

$$\text{KTE}(t) = x(t)^2 - x(t+1)x(t-1) \approx A^2\Omega^2 \tag{1}$$

for $t = 1, \ldots, L_{\text{frame}}$.

This measure can be extended to signals with a richer spectral content, by computing the value for each bin of the FFT of a sequence at frame level [34]. The Teager energy computed by Eq. 1 in the case of periodic signals works as if the Teager was computed for each bin of the FFT and then added. Therefore we have a measure sum of the energy for all bins.

Thus the $\text{KTE}_n(t)$ has the following properties

- for a periodic waveform of high amplitude the mean value of the KTE_n in a frame is high.
- for background noise, transients and clicks, the mean value of the KTE_n in a frame is low.

Therefore it is a good candidate for a feature in a classifier that separates clean signal from noise, transients or artifacts.

The $\text{KTE}_n(t)$ sequence for the n th frame was computed as,

$$\text{KTE}_n(t) = S^2_{\text{frame}}(t, n) - S_{\text{frame}}(t+1, n)S_{\text{frame}}(t-1, n) \tag{2}$$

for $t = 2, \ldots, L_{\text{frame}} - 1$. Then from $\text{KTE}_n(t)$ we computed, the mean KTE^{μ}_n, variance KTE^{σ}_n, interquartile range $\text{KTE}^{\text{iqr}}_n$ and skewness $\text{KTE}^{\text{skew}}_n$.

Qi-Zheng energy. The Qi-Zheng energy (QZE_n) is the result of filtering the energy sequence at each frame, using a filter designed to be optimal for detecting beginning or ending edges of a given signal. The coefficients of the filter are presented elsewhere [35]. This model for detecting the edges of a signal assumes exponential growth of the edge to be detected, and the model is best for detecting beginning or ending transients of periodic signals [35]. It filters out clicks, pops, sudden falls in amplitude, or noise with growth-shapes that differ from an exponential. This feature is complementary to the others in the sense that it takes high values at the endpoints.

Spectral entropy. The spectral entropy sequence H^S_n is a scalar associated to the n th frame (see detailed description in Section 3.2, spectral entropy).

The entropy H^S_n has the property that, for a flat or tilted power spectrum of a frame (i.e. absence of signal or presence of noise or clicks) it has a high value, while for a tone or harmonic signal it has a low value. Therefore, one can use the spectral entropy to characterize the PPG signal at frame level, which will have a low entropy for clean signal zones, while transients and artifacts will have high entropy (i.e. flat or irregular spectrum).

Zero crossing rate. The Zero crossing rate Z^c_n was computed for the n th frame. First, we filtered $S_{\text{frame}}(\tau, n)$ with a high-pass filter in order to remove the mean. Given that the filter was FIR of order 4 and that the transient was negligible, there was no need to apply a window. Then, the Z^c_n was computed from the number of crossings divided by the number of samples in the frame L_{frame}. In the presence of a clean PPG signal, this variable took values between $5/L_{\text{frame}}$ and $10/L_{\text{frame}}$, whereas in cases of sudden losses of signal, clicks or transients, the value was either too low or too high.

The X^n_{AD} vector. Finally the vector X^n_{AD} was constructed by aggregating the features described above, i.e.:

$$X^n_{\text{AD}} = \left[\text{KTE}^{\mu}_n, \text{KTE}^{\sigma}_n, \text{KTE}^{\text{iqr}}_n, \ \text{KTE}^{\text{skew}}_n, \text{QZE}_n, H^S_n, Z^c_n\right]^T$$

3.1.2. Decision function

The last block was a decision function that was implemented as a classifier that takes X^n_{AD} as input and decides between two classes, i.e. class "signal" (PPG signal) and class "no signal" (corruption/loss of signal, background noise). The decision function consisted of a linear classifier [36], that performs the classification by means of a hyperplane on the space of the features listed above. This classifier was trained from a hand-labeled subset of the database. The selection of the training subset and the labeling was done by the author. The signal-to-noise ratio of the useful segments of the PPG signal was higher than 30 dB and had clear harmonic components, whereas the transients and segments with a loss of signal had irregular shapes and a much lower energy. Therefore a linear classifier had a high performance in terms of the detection rate, and the set of examples used for training the classifier was selected to have a representation of the possible transients and artifacts of the signal. Note that the fraction of the recording time that corresponded to the "no signal" class was much smaller than the one corresponding

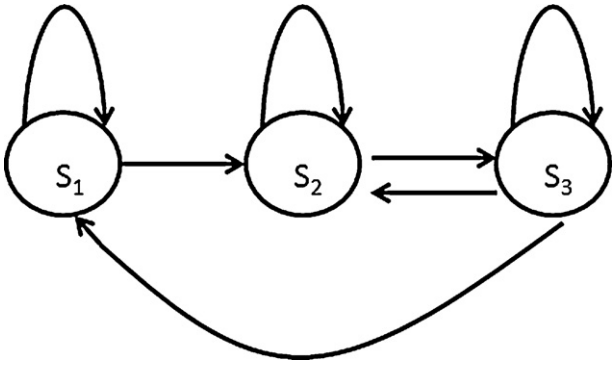

Fig. 2. Transitions between states that were allowed by the finite state automaton.

to the "signal" class. Therefore the training was biased, increasing the prior probability of the "no signal" case.

3.1.3. Finite state automaton

We used a FSA to segment the PPG signal. The decisions on the transitions were done based on counts on the number of consecutive frames classified by the decision function as "signal" or "no signal" (see Section 3.1.2). Three states were defined,

- s_1=Spurious/absent signal,
- s_2=In-ppg,
- s_3 =Leaving-ppg,

in order to model the ambiguous zones between the two classes. The time spent in each state was used to filter false positives/negatives and allow for a time margin at each endpoint of activity. The transitions were decided from counts on the number of frames classified by a decision function, and the allowed transition between states was specified by the directed graph shown in Fig. 2.

Each state consisted of a set of counters and had as input the class given by the decision function. The counters, were defined as follows:

- C_S^i was the number of consecutive frames at state i classified as "signal".
- C_{NS}^i was the number of consecutive frames at state i classified as "no signal".

The rules for the transition between states (see Fig. 2) were the following:

- $s_1 \rightarrow s_2$ if $C_S^1 \geq 4$, otherwise, stay at s_1
- $s_2 \rightarrow s_3$ if $C_{NS}^2 \geq 2$, otherwise, stay at s_2
- $s_3 \rightarrow s_2$ if $C_S^3 \geq 2$, otherwise, stay at s_3
- $s_3 \rightarrow s_1$ if $C_{NS}^3 \geq 4$, otherwise, stay at s_3

The thresholds were decided by inspection from a set of examples (see Section 4.1) and adjusted by hand on a set of 20 examples; this, was sufficient to capture a clean segment of the PPG signal. An alternative might have been deciding the thresholds by means of an integer optimization technique or even by enumeration as the number of possible configurations is low enough using the whole database. Nevertheless as the base classifier had a low error rate, we selected the recordings with false negatives and false positives so that the thresholds were reliable in adverse conditions, otherwise the thresholds would be biased to an optimistic situation.

The clean signal was considered to be in the states s_2 and s_3. Once a 1 min segment of consecutive samples was found, this clean signal was assigned to the outputs $S_{\text{window}}(t)$ and $S_{\text{frame}}(\tau, n)$.

3.1.4. Output of the AD module

Once the FSA has spent 1 min (i.e. 24 consecutive frames) in states s_2 and/or s_3, the module copies the samples of $S_{\text{PPG}}(t)$ in this interval into the vector $S_{\text{window}}(t)$ and the frames in this interval into the variable $S_{\text{frame}}(\tau, n)$. Both of these were the output of the module. Note that the Kaiser–Teager energy and the spectral entropy were also used in the SP module.

3.1.5. Robustness to changes in the PPG sensor

Another important aspect of the design is its robustness to changes in the photoplethysmographic instrument itself. A change in the sensor affects the frequency response and/or gain in the amplifier. In the future in order to compensate for the differences between photoplethysmographs, a cepstral subtraction [31] approach should be used to compensate for changes in the transfer function of each photoplethysmographs. This aspect of the preprocessing is left for future work.

3.2. Signal processing module

In this section we summarize the features computed from the PPG signal. The physiological properties mentioned in Section 2 were used as a guide to select the set of features in X_F. The output of this module was a vector X_F that, consisted of the aggregated features that partially characterize the subject's hemodynamic and autonomic status. A duration of 1 min for $S_{\text{window}}(t)$ was selected in order to have sufficient samples for computing the heart rate and respiratory rate. For the frames, the time length and overlap were selected so that each frame would have at least four or five heart pulses, and also most of the time lasted less than the respiratory rate (i.e. most frames would be shorter than the time required for one breath).

The feature vector X_F generated in this module consists of two kinds of features:

- global features computed from $S_{\text{window}}(t)$ and
- features aggregated from traits computed on $S_{\text{frame}}(\tau, n)$. The number of frames ($N_{\text{frame}}$) was 24. These features model the time variation of certain magnitudes across the window being measured.

Autoregressive model of the PPG waveform. The autoregressive (AR) coefficients model the spectral envelope of the PPG signal; this technique was used because it assumes an underlying model of a signal generated by an impulse that goes through a series of resonant tubes [31]. This can act as a coarse model for propagation of the ECG signal through the arteries, veins and capillaries, and reflects the changes in the shape of the signal due to the propagation through tubes of different diameters, as well as to the effects of viscosity. In addition, this method has the interesting feature that the AR coefficients also model the shape of the basic pulse. When the order of the model was selected, we required that the shape of the AR power spectrum of a given pulse follow correctly the periodogram of the PPG signal. In our experiments we used an AR model of order 5. The coefficients were computed from the Yule-Walker equations [31] and were derived from sample covariances. This feature was considered to be a global feature and consisted of a vector AR_{PPG} of length 5.

Kaiser–Teager energy of the PPG signal We computed the Kaiser–Teager energy features via two approaches, with one at the window level and another at the frame level.

- At the window level we first computed the sequence of instantaneous energy KTE(t) as

$$\mathrm{KTE}(t) = S_{\mathrm{window}}(t)^2 - S_{\mathrm{window}}(t+1)S_{\mathrm{window}}(t-1).$$

for $t = 2, \ldots, L_{\mathrm{window}} - 1$. From the sequence KTE($t$), we estimated an AR model of order 5. The coefficients of the model were assigned to a vector $\mathrm{KTE}_{\mathrm{AR}}$ of length 5.
- At the frame level, we computed the $\mathrm{KTE}_n(t)$ sequence from Eq. 2, and we characterized each frame by the mean KTE_n^{μ}, variance KTE_n^{σ}, interquartile range $\mathrm{KTE}_n^{\mathrm{iqr}}$ and skewness $\mathrm{KTE}_n^{\mathrm{skew}}$. These features are also used at the AD module at the frame level (see Section 3.1.1). From these sequences, we summarized the behavior over the whole window by computing averages for each frame. These averages were the mean KTE^{μ}, variance KTE^{σ}, interquartile range $\mathrm{KTE}^{\mathrm{iqr}}$ and skewness $\mathrm{KTE}^{\mathrm{skew}}$.

Heart rate statistics. The heart rate (HR) statistics were considered to be a global feature and were computed from the $S_{\mathrm{window}}(t)$. First, the signal was filtered with a band pass filter in order to remove the mean and attenuate the effect of the dicrotic notch. The specifications of the filter can be found at Appendix B. Afterwards the HR was computed from the zero crossings of the signal (i.e. the time interval between alternate crossings). This method was found to be more robust than computing the HR from signal peaks. Using the sequence of time intervals between heart beats, we computed the mean value HR^{μ}, the variance HR^{σ}, the interquartile range $\mathrm{HR}^{\mathrm{iqr}}$ and the skewness $\mathrm{HR}^{\mathrm{skew}}$ for the whole window.

Oxygen saturation range. The feature $\mathrm{OSR}_{\mathrm{PPG}}$ is given by the pulse oximeter (i.e. the PPG sensor) and provides the mean value for peripheral oxygen saturation over the entire window.

Information about the individual. The selected variables were age *AG*, weight *WGT* and body mass index *BMI*. Whether an individual was taking beta blockers was found to be irrelevant from a performance point of view, and was therefore discarded.

Spectral entropy statistics. This feature was computed for each frame $S_{\mathrm{frame}}(\tau, n)$ and provides an index of confidence in the presence of spectral lines and/or the tilt of the spectrum. The spectral entropy H_n^S of a frame is a scalar obtained by computing the entropy function (see Eq. 4) from the normalized power spectrum [37,38]. Note that as the power spectrum is normalized so that the area is equal to one, it can be interpreted as a probability distribution, and the use of the entropy function therefore makes sense.

This measure has the property that for tones (i.e. narrow spectral lines), the entropy will take low values, whereas when the spectrum is flat or has a low tilt, the entropy will take high values. In other words, this feature measures the damping of the pulses, the spectral shape, harmonic components, and presence of noise. The power spectrum is computed by means of the short time fast Fourier transform (FFT) of n th frame $S_{\mathrm{frame}}(\tau, n)$. The duration of the frame assures that at least several heart pulses are present in the analysis frame.

This feature was computed for each frame as follows:

First, we computed the FFT of $S_{\mathrm{frame}}(\tau, n)$ of length L_{frame}, which was zero paddled, in order to increase the length to the nearest power of two. The final length of the zero paddled frame (L_{FFT}) was 512.

$$X^n \leftarrow \mathrm{FFT}(S_{\mathrm{frame}}(\tau, n), L_{\mathrm{FFT}}) \tag{3}$$

Then we normalized the square of the absolute value of each bin of the FFT:

$$P_X^n[k] \leftarrow \frac{|X^n[k]|^2}{\sum_{j=1}^{L_{\mathrm{FFT}}} |X^n[j]|^2} \quad k = 1 \ldots L_{\mathrm{FFT}}$$

Finally, we computed the entropy H_S^n from P_X^n:

$$H_n^S \leftarrow \sum_{k=1}^{L_{\mathrm{FFT}}} P_X^n[k] Log\left(P_X^n[k]\right) \tag{4}$$

This gave us a sequence of estimates of the entropy at the frame level $S_{\mathrm{frame}}(\tau, n)$. From the sequence H_n^S, we computed for each frame, its mean value H_S^{μ}, variance H_S^{σ}, interquartile range H_S^{iqr} and skewness H_S^{skew}.

Energy profile of the PPG signal. The log-energy profile $\mathrm{Log}E_n$ of the PPG signal was computed at frame level from $S_{\mathrm{frame}}(\tau, n)$ in order to estimate the respiratory rate. Unlike other methods based on the wavelet transform [39], we decided to use the log-energy of the sequence of frames, aiming for a low complexity estimate, despite the poorer precision. The $\mathrm{Log}E_n$ was computed as follows:

$$\mathrm{Log}E_n \leftarrow Log\left(\sum_{\tau=1}^{L_{\mathrm{frame}}} S_{\mathrm{frame}}^2(\tau, n)\right)$$

The frame duration of 5 s allowed for computation of the respiratory rate of a healthy adult at rest. The feature used for summarizing the spectral variability of the energy profile were the AR parameters of order 5 computed after subtracting the mean. The vector of AR coefficients $\mathrm{Log}E^{\mathrm{AR}}$ was computed from the sequence $\mathrm{Log}E_n$ and summarizes the spectral content of the sequence of log energy over all of the frames. The AR vector $\mathrm{Log}E^{\mathrm{AR}}$ was a model of the fluctuations of the energy of the PPG relative to the respiratory rate. Note that in our case, the objective was not to generate an exact estimate of the respiratory rate, but to find a rough estimate of the rate and its power spectrum. In addition, we computed as descriptive statistics the log energy profile of $S_{\mathrm{window}}(t)$, the variance $\mathrm{Log}E^{\sigma}$ and the interquartile range $\mathrm{Log}E^{\mathrm{iqr}}$.

The X_F vector. Finally, the vector X_F was constructed by aggregating the features mentioned above. This feature vector was used as the input to a machine learning module.

$$\begin{aligned} X_F = \big[&\mathrm{KTE}_{\mathrm{AR}}, \mathrm{KTE}^{\mu}, \mathrm{KTE}^{\sigma}, \mathrm{KTE}^{\mathrm{iqr}}, \mathrm{KTE}^{\mathrm{skew}}, \ \mathrm{HR}^{\mu}, \\ &\times \mathrm{HR}^{\sigma}, \mathrm{HR}^{\mathrm{iqr}}, \mathrm{HR}^{\mathrm{skew}}, \mathrm{AR}_{\mathrm{PPG}}, \mathrm{OSR}_{\mathrm{PPG}}, AG, WGT, BMI, H_S^{\mu}, H_S^{\sigma}, \\ &\times H_S^{\mathrm{iqr}}, \ H_S^{\mathrm{skew}}, \mathrm{Log}E^{\mathrm{AR}}, \mathrm{Log}E^{\sigma}, \mathrm{Log}E^{\mathrm{iqr}}\big]^T \end{aligned}$$

3.3. Machine learning module

In this section we describe the module that infers from labeled examples a function that relates the feature vector X_F and the desired targets (DBP, SBP and BGL). We tested four different ML techniques, each with different structures and properties.

A requirement for the ML techniques selected was that each should be flexible enough to deal with the peculiarities of the data, have ways to control the over-generalization phenomenon, and be robust enough to deal with noisy measurements. It should be noted that the reference values have some inherent uncertainty that can be interpreted as noise (see Section 3.4). This error originates from the error tolerances of the measuring devices. On the other hand

the effect of a small amount of additive noise in the references has been shown to improve accuracy when making predictions [40].

Additional requirements for the module were that it should generalize from the training examples to unseen inputs, should not need calibration, should not require too much computational resources to generate the output (note that this was not required during training), and finally, should be stable, meaning that small changes in the training database composition, should not give rise to large differences in performance. The stability of the algorithm gives us an indication of the reliability of the trained system and, together with accuracy, served as the criterion for selecting the appropriate ML technique for this module. Stability was checked by cross-validation, and in Section 4.4, we rank the algorithms by the standard deviation of the residual sum of squares (RSS) between different rotations of the training and validation database.

In this subsection we describe the structure of each ML technique, how it was trained, and how the structure and parameters were optimized so that results could be compared.

To select the best structure and parameters we used 80% of the database as a training and validation database (TrnValDB) and the remaining 20% of the database (TstDB) for testing and comparison purposes. Note that each sample of a database consisted of a unique recording. Selection of the best structure and testing were done in two phases.

- In the first phase, different architectures and sets of parameters were tried for each algorithm, and selection was based on the best performance after 10-fold cross validation on TrnValDB, i.e., 90% of the TrnValDB database was used for training and the remainder for computing the validation results. This was done 10 times by rotating the databases.
- In the second phase, once the best configuration was found, the performance was evaluated on the TstDB database.

The machine learning techniques tested were the following:

- *Linear regression* (LR). For the linear regression method [41], we trained a univariate model $y^j = \beta_0^j + \sum_{i=1}^{n} \beta_i^j X_F(i)$, where $j = \{\text{SPB,DPB,GBL}\}$, using the ridge regression algorithm [41]. Ridge regression was selected because, we found that estimates of model parameters by direct minimization of the RSS were unreliable. This happened both in the multivariate and univariate cases, as the covariance matrix was badly conditioned. This occurred because the vector X_F has features that are highly correlated. The result was a high RSS on the validation and test databases and high values for some of the β_i^j. The ridge regression method minimizes a penalized residual sum of squares and reduces the variability due to the high correlation between features. The penalization parameter λ of the ridge regression imposed a size constraint on the coefficients, thus correcting the high values of the estimated coefficients. This phenomenon is explained by Hastie et al. [41] on page 59. Weights were computed from the input feature matrix $\mathbf{X} = [X_F^1, \ldots, X_F^M]$, where M was the number of observation vectors used for computing the parameters β of the linear regression model. The vector of the reference values (targets) for each observation was denoted as y and the regularization parameter as λ. The ridged regression estimate was; $\hat{\beta}^{\text{ridge}} = \left(X^T X + \lambda I\right)^{-1} X^T y$. The value of λ was the mean value of the optimal values computed in each of a 10-fold cross validation on the TrnValDB database. The quality criterion was the value of the coefficient of determination R^2 on the testing database, TstDB. The algorithm was programmed in Matlab by the author.
- *Neural networks* (NN). For the neural networks method we used a multilayer perceptron trained by means of a conjugate gradient algorithm (Fletcher Powell) [42] minimizing the RSS. The multilayer perceptron was selected because it is able to approximate non-linear relationships. The structure consisted of a hidden layer and an output layer for each value of interest. The outputs shared the information from the hidden units in order to make use of the common traits of the functions to be estimated; this is known to improve performance [43]. The conjugate gradient algorithm was selected to enable fast convergence and the capacity for dealing with high dimension inputs. The Levenberg–Marquart algorithm was rejected because it was too slow. In order to deal with the problem of convergence to local minima, we used a multistarting scheme (i.e., train several times and select the network with the best validation performance). In addition, the input data were standardized. The mean and variance values of the standardization process were estimated from TrnValDB and were used to standardize both TrnValDB and TstDB. The best configuration (number of hidden units and type of nonlinear function) was determined from a 10-fold cross validation on the TrnValDB database. This configuration had 10 neurons in the hidden layer, three at the output, and a hyperbolic tangent representing the non-linearity. The quality criterion was the coefficient of determination R^2 value on the test database TstDB. The software was the neural network toolbox for Matlab.
- *Support vector machines* (SVM). For SVM regression, we used the ϵ-insensitive loss function [44], and the kernels were the Gaussian and the polynomial. As the Gaussian kernel assumes a common sigma for all dimensions, we standardized the data as explained above. The values of sigma were varied from 0.1 to 5, and for the polynomial kernel, orders from 1 to 5 were tested. In all cases different values of the soft margin C and ϵ tube were tried. It should be noted that the algorithm was sensitive to outliers and that numerical problems were encountered for some combinations of parameters. The best configuration (kernel type, soft margin, ϵ and kernel parameters) was determined from a 10-fold cross validation on the TrnValDB database. The best kernel was a Gaussian kernel. The quality criterion was the R^2 value on the testing database TstDB. The software was the SVM toolbox [45] for Matlab.
- *Classification and regression trees* (CART). Classification and regression trees are ML techniques that infer logical trees composed of disjunctions of conjunctions [46], and at each node an element of the input vector is compared to a threshold. This ML technique was not tested with cross validation because preliminary testing yielded a rather poor performance. It was also discarded because we tested the random forest technique, which is a generalization of the CART algorithm.
- *Random forest* (RF). Random forests are a ML technique based on a set (forest) of classification and regression trees [47], each trained in a different way. The output of the system is an aggregation of the outputs of the trees of the forest. The methodology is justified by the fact that the mathematical expression of the error of a classifier depends on the bias and variance of each of the trees of the forest. In order to take advantage of the fact that averages reduce the bias and variance, the algorithm introduces a systematic controlled variability and bias [47]. This is done by bootstrapping the training data and introducing randomness at the selection of the testing feature at each node. The result is that the mean of the outputs reduces the variance and bias of the estimate; therefore the system has a lower global error rate. The execution time of this method is low because the base classifier is a decision tree, and the only computations are comparisons at each node, in other words, there are no multiplications. Another advantage is related to the fact that each node compares only one feature, making the system robust with respect to the scale of the inputs and the correlations between them. Consequently there is no need to standardize or whiten the input data. The best configuration (number of trees in the forest, number of features tested at each node and maxi-

Table 1
Characteristics for the reference values.

Measurement	Glucose (mg/dl)	SBP (mmHg)	DBP (mmHg)
Min	49	90	60
Max	393	180	120
Mean	139	123	78
SD	66	21	16
Range	343	90	60

mum number of samples at a given node) was determined from a 10-fold cross-validation on the TrnValDB database. The quality criterion was the R^2 value on the testing database TstDB. The performance of the RF was insensitive to variations of the parameters in a certain margin, i.e., for a number of trees higher than 30 there was no improvement, and the size of the set at each node could vary from 1 to 10 without worsening the performance. The algorithm was programmed in Matlab by the author.

3.4. Description of the database

The data base was collected from 213 male and 197 female individuals, aged from 9 to 80 years (mean±SD=37.97±13.32), (see Table 2). The participants were selected from personnel of the Universitat Politecnica de Catalunya (# 71) and from an ambulatory primary care center (# 339). The database was recorded on a successive recruitment base without selection. The sampling biases come from the fact that a part of the database consisted of healthy young individuals (university personnel) and persons from the ambulatory medical assistance staff. Another bias of the data was that pediatric and geriatric populations were under-represented. A total of 79 people reported suffering from diabetes. The measurements were taken by personnel trained in the technique. University measurements were taken by the author, and ambulatory measurements were taken by a medical doctor. Each subject's PPG was measured with the fingertip pulse oximeter device *iPod Digital Oximeter* [12], blood pressure measurements were made an aneroid sphygmomanometer and the glucose levels were measured with a self-monitoring of blood glucose (SMBG) meter *Accu-Chek Aviva* [13]. Table 1 summarizes the measured blood pressure and glucose levels.

The reference values in the database are not perfect. There are variations due to estimating error of the SMBG meter, as well as uncertainty due to the fact that BP values were measured manually. We found that the manual estimates of BP values were sometimes rounded off, which resulted in slight clustering that can be seen in Figs. 4 and 5. This can be considered to be a rounding error, and has a maximum value of 5 mmHg, which is within the grading criteria of the BHS [48] (Table 6) Table 2.

4. Training and results

The experimental part consisted first of selecting the best machine learning technique and then assessing its performance. In this section we first review the training of the AD module, present the ML technique selected and finally discuss the performance of the definitive system. The final system was implemented in a Matlab program entitled SIMPRESGLUC.

Table 2
Characteristics for the study population.

Measurement	Weight (kg)	Height (cm)	Age (years)
Min	35	138	9
Max	111	192	80
Mean	65.5	164.9	37.9
SD	14.7	9.6	13.3

Note that the experiments were done in two phases, in the first phase we compared the ML algorithms, and in the second phase we tested the best algorithm. The cross validation methodology was done at each phase with different proportions of elements assigned the training, validation and test databases.

4.1. Training the activity detection module

The activity detection module was trained from labeled examples. Preliminary tests showed that performance on endpoint detection did not depend on the subset used. We trained the classifier with one-fifth of the database manually labeled, and left the parameters fixed for all other experiments. Examples were selected so that the training database of the AD module had enough examples of possible variations, including clicks, loss of signal, transients, etc. The thresholds of the FSA were determined from 20 examples. Inspection of the results from the database showed that the rest of the database was well segmented. Note that the final performance of the whole system was tested by cross validation on the segmented data, but the segmentation algorithm was not retrained.

4.2. Feature selection

Several methods exist for selecting a subset of features that maximizes the performance of a ML system. Some are based on enumeration, such as the sequential search techniques [42], while other are based on projections, such as the use of PCA or factor analysis.

We did not attempt feature selection using sequential search techniques, due to the computational requirements. On the other hand, two of the ML techniques that we used (LR-ridge, RF) rank the usefulness of the features, while another technique (SVM), is robust to the presence of irrelevant features. In the case of LR, the regularization parameter λ in the ridge regression algorithm shrank the regression coefficients by their contribution to the final results (see Ref. [41], p. 61). In our case, the coefficients shrank in a similar way as λ increased, and therefore from the set of weights β_i^j of the ridge regression, we could not conclude that given features were either irrelevant or highly predictive. On the other hand, the RF has a method for ranking the features based on how the performance degrades when a given input feature is randomly permuted in the test [47]. In this case we observed a similar phenomenon: i.e., the contribution of the features to performance was more or less similar. The LR-ridge and RF both indicated that the features contributed in a similar manner to the final result. As such, it was not necessary to do a feature selection process.

The age, weight and BMI variables were not predictive when used alone (i.e., $R^2 < 0.1$), but nevertheless improved the accuracy of the results. It was found that with the information related to the PPG signal alone, the predicted values were reasonable, but improved when used simultaneously with the information about the individual.

4.3. Selection of the ML algorithm

The criterion for selecting the ML algorithm was the value of the coefficient of determination R^2 on the test database. The TstDB database consisted of 20% of the original database (82 individuals) sampled randomly. The TrnValDB database contained the remaining 80% (328 individuals). The ML algorithms were tuned on the TrnValDB in order to maximize their performance on a 10-fold cross validation. Once the best structure and set of parameters were found, the algorithms were trained again with the TrnValDB database and tested with the TstDB database. As shown in Table 3 the RF method give the best performance.

Table 3
Performance of the ML algorithms.

	LR	NN	SVM	RF
R^2_{BGL}	0.52	0.54	0.64	0.88
R^2_{SBP}	0.59	0.65	0.72	0.90
R^2_{DBP}	0.53	0.63	0.68	0.86

Table 4
Final performance of the system.

	BGL	SBP	DBP
R^2_{RF}	0.90	0.91	0.89

4.4. Stability of the ML algorithms

In order to have a reliable idea of how the algorithms performed on novel data, we computed the standard deviation of the R^2 between the ten rotations of the train and validation database (TrnValDB). The variability of the performance gives an indication of the stability of the predictive method. For each of the three measures (BGL, SBP and DBP), Random Forest ranked first, as it displayed less variability on the cross-validation for all three measures. This method was followed by the linear regression algorithm, the SVM and finally the NN. In the case of the NN, the performance displayed significant variance despite multistarting.

4.5. Final results

The system's performance was assessed using the random forest method and was tested on all the 410 individuals. The random forest technique was selected as the predictive method because it had both the best R^2 value and the best stability.

The performance was assessed for the whole database by means of a 10-fold cross validation; i.e., we used 90% of the database for training, leaving 10% as testing database, and then rotated ten times. The performance in terms of the R^2 is summarized in Table 4.

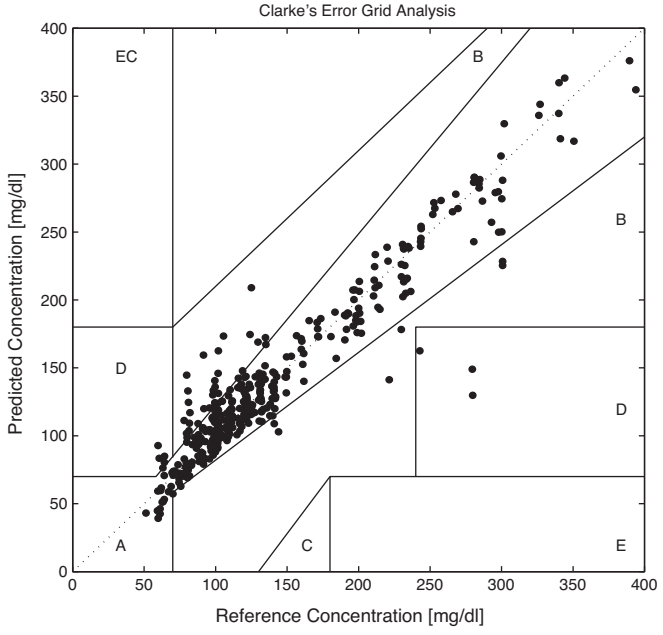

Fig. 3. Results of the Clarke error grid.

Table 5
Results of the Clarke error grid.

Zone	A	B	C	D	E
Percentage (%)	87.71	10.32	0	1.96	0
Total	357	42	0	8	0

4.5.1. Glucose results

In Fig. 3 we present the Clarke error grid, which divides the scatter plot into five regions. These regions quantify the accuracy of the blood glucose reference values compared to the predicted values for different types of errors.

The grid is divided into five regions, each of which is associated with a different interpretation of the prediction error. Details about the grid have been described previously [49]. In Table 5 we present the number and percentage of points in each region. Our results show that 87.7% of points were in region A, which covers predictions within 20% of the reference sensor. Another 10.3% of the data went to region B, which contains predictions more than 20% away from the reference sensor but would not lead to inappropriate treatment, in the sense that although the measurement is inaccurate by at most 20% of the value, it does not miss cases of hypoglycemia or hyperglycemia or give false positives for either hypoglycemia or hyperglycemia. No points fell into region C, which represents the false positives either of hyperglucemia or hypoglucemia. Only 2% of the points fell into region D, which typically indicates that the prediction would not be able to detect hypoglycemia or hyperglycemia. Finally no points fell into region E, which consists of prediction errors that confuse hypoglycemia for hyperglycemia or viceversa.

The resulting value of the coefficient of determination R^2_{BGL} was 0.90; therefore the RSS was 10% of the sample variance.

The Clarke grid scatterplot was generated via a Matlab function [50].

4.5.2. Blood pressure results

In Fig. 4 we present the scatter-plot for systolic blood pressures and in Fig. 5 the plot for diastolic blood pressures. The values of the resulting coefficients of determination were $R^2_{SBP} = 0.91$ and $R^2_{DBP} = 0.89$, meaning that in both cases the RSS was approximately 10% of the sample variance.

Table 6 displays the results and the error thresholds for grading the system according to the British Hypertension Society (BHS) criteria. The BHS protocol [48] grades devices by cumulative

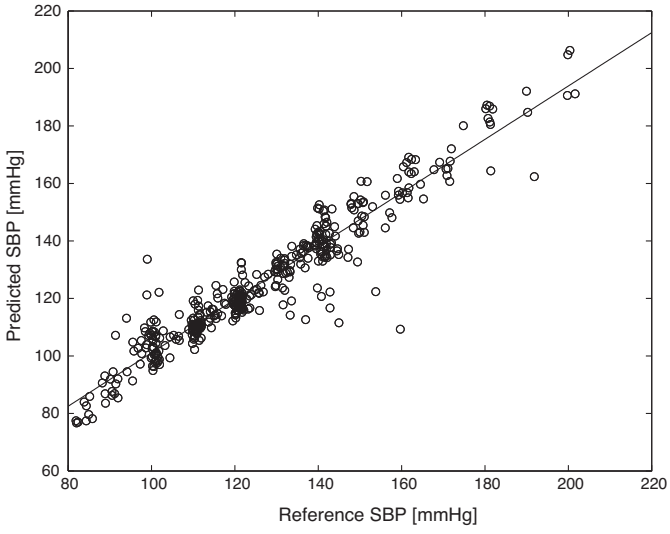

Fig. 4. Scatter plot of systolic blood pressures.

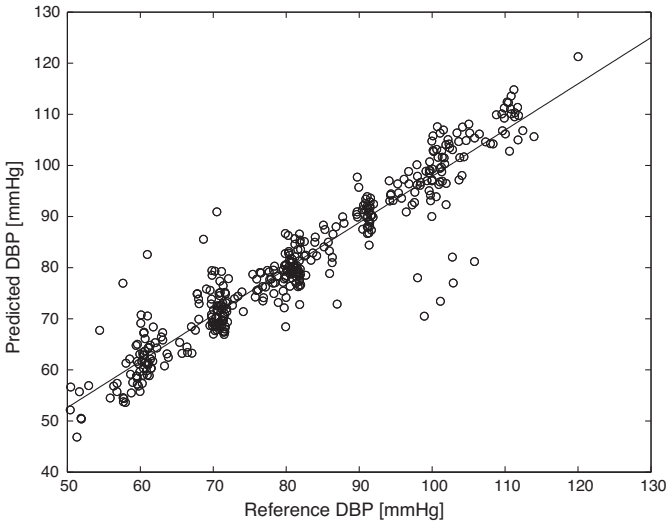

Fig. 5. Scatter plot of diastolic blood pressures.

Table 6
Comparison with the grading criteria of the BHS [48].

	≤5 *mmHg*	≤10 *mmHg*	≤15 *mmHg*
SBP Totals	256	354	373
DBP Totals	324	370	385
SBP (%)	62.43%	86.34%	90.97%
DBP (%)	79.02%	90.24%	93.90%
Grade A BHS(%)	60%	85%	95%
Grade B BHS(%)	50%	75%	90%
Grade C BHS(%)	40%	65%	85%

Table 7
Random permutation of the references.

Measurement	BGL	SBP	DBP
R^2_{Train}	0.0174	3.7100e-004	0.0070

percentages of readings, falling within 5, 10, and 15 mmHg of the standard sphygnmanometer measurement. Our predictions agreed with grade B of the BHS protocol. Note that the system did not meet class A criteria due to the presence of 17 outliers with an error greater than 15 mmHg, but it did meet criteria for cumulative percentage of readings less than 5 mmHg and 10 mmHg.

4.6. A test for overfitting

It is known that machine learning techniques can overfit. We performed a test for overfitting as suggested by Smith [4], which consisted of training with a random permutation of the reference values. In a case of overfit the result should be an R^2 close to one in the training database. We obtained a random permutation of the reference values from the entire database and trained the system with a training database that consisted of 369 individuals. The results (presented in Table 7) show that the system was not able to learn the function with the permuted references. These results indicate that there is a functional dependence between the input and output, and that this system can capture this relationship.

5. Conclusions and future work

This study shows that it is possible to simultaneously estimate the BGL, SBP and DBP levels from a PPG signal. In Section 2 we presented possible physiological explanations for why these variables might be estimated from a PPG signal. Our results show that the system has an accuracy similar to that of the standard measurement devices, i.e., $R^2_{BGL} = 0.90$ for the glucose levels and grade B of the BHS standard for blood pressures. An interesting property of the system is that it is subject independent and does not require calibration; therefore a, portable and easy-to-use device could be fabricated that implements the system proposed in this paper [51]. Another property is that the response of the system was linear with the values to be estimated, that is, there were no saturation levels, as shown in Figs. 3–5. This means that the system is reliable even at extreme degrees of hypo/hypertension or glucose levels. We tested the system on a database with reference values that were not ideal; nevertheless the results indicated that the system worked well for the ranges of variables shown in Table 1. Also we have devised an activity detection module, which filters out the initial transients, and the drops in the signal level. Although the pulse oximeter normally produced an immediately good quality signal, but the initial transient was not of a standard duration. There were also drops in the signal level due to movements of the hand where the pulse oximeter was attached. Although there was high variability in the length of the recording to obtain a 1 min signal window, in most of the cases it was little more than a minute. The activity detection is of importance in the continuous estimation, where the drops in signal energy can be rather frequent. In addition the computational resources needed to implement the system proposed in this paper are low, and the processing time for the 1 min window was less than 10 s, using a standard PC and software programmed in Matlab. Note that the results might be biased by the fact that the database used for selecting the best algorithm, is also used for the validation of the system.

The clinical use of the system we propose would be an easy and portable device for continuous measurement of blood pressure and glucose level for 24 h, with the advantage of a device that does not have to be calibrated between individuals. The system complies with Grade B of the British Hypertension Society's protocol, which is considered the minimum to pass the accepted criteria of the International Protocol or the AAMI Protocol [52]. Another clinical use would be ambulatory blood pressure monitoring, allowing for the monitoring of the variability of the blood over a day either at home or at the hospital. This clinical application is important for essential and refractory hypertension, which requires strict control of the treatment and medication.

As the device measures blood glucose levels, we evaluated the clinical performance by using the Clark error grid that quantifies the clinical accuracy of blood glucose estimates generated by meters as compared to a reference value. More than 98% of the test points were in Regions A and B of the Clark error grid, which are the estimates within a 20% margin from the reference value, whereas only 1.96% (8) were in Region D, which are estimates with an error that fails to detect hypoglycemia or hyperglycemia. In addition, no points were in Regions C (false positives of hyperglycemia or hypoglycemia) or E (wrongly classifies hyperglycemia as hypoglycemia or vice versa). One clinical application could be glucose control with continuous recording of diabetes II patients. This system would be of great usefulness in intensive care units because it allows a 24-h period of continuous monitoring of blood glucose levels and blood pressure, without interventions that may have a risk of infection.

As future work we propose to use as reference the glucose measurements from standard blood tests and to use a more accurate measurement of the blood pressure. In addition we propose to recruit a larger population with a greater range of BGL, SBP and DBP. Additionally, the system will be tested on severe cases of hypotension, and if good results can be obtained it may be useful for early detection of sepsis in intensive care units. Other improvements include making the system robust to changes in the sensor by using cepstral subtraction [31] (see Section 3.1.5) or estimating other physiological measurements related to the state of the

autonomic nervous system that are usually obtained from blood analysis. Finally, a real-time version of the SIMPRESGLUC software could be devised that would output the SPB, DBP and BGL with a delay of seconds, instead of the current version's minute.

Acknowledgments

The author would like to thank Jose B. Marin´o, Dr. Jose Antonio Fiz, Dr. Pep Roca, Dr. Hu Yue Xhian and specially Pau Bofill for their helpful comments that greatly improved the manuscript. The author also thanks the two anonymous referees for many helpful comments and suggestion that have greatly improved this paper. This project was financed by TEC2009-14094-C04-01.

Appendix A.

Accuracy of the pulseoximeter.

SpO2 (70-100%) (1 SD)		
	No motion	Adults ±2 digits
	Motion	Adults ±3 digits
	Low perfusion	Adults ±3 digits
Heart Rate		
	No motion (18-300 BPM)	Adults ±3 digits
	Motion (40-240 BPM)	Adults ±5 digits
	Low perfusion (40–240 BPM)	Adults ±3 digits

Appendix B.

Specifications of the band-pass filter used for computing the heart rate statistics were as follows:

- Chebyshev Type I filter of order 6.
- Lower stop frequency 0.01 Hz at −10 dB.
- Lower pass frequency 0.1 Hz.
- Upper pass frequency 0.2 Hz.
- Upper stop frequency 0.4 Hz at −30 dB.

The lower stop frequency was selected to eliminate the DC component and low frequency components, such as the respiratory rate. The band pass was selected to capture the margin of the heart rate,, and the upper stop frequency was selected in order to attenuate off band components. The attenuations were selected as a trade off between the need for eliminating interferences and the need for having a low-order filter.

References

[1] Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes. Diabetes Care 2004;27(5):1047–53.
[2] Wolf-Maier K, Cooper RS, Banegas JR, Giampaoli S, Hense H-W. Hypertension prevalence and blood pressure levels in 6 European Countries, Canada, and the United States. JAMA 2003;289(18):2363–9.
[3] Diabetes statistics. American Diabetes Association. Last accessed May 2, 2011. URL http://www.diabetes.org/diabetes-basics/diabetes-statistics.
[4] Smith JL. The pursuit of noninvasive glucose: hunting the deceitful turkey. Last accessed May 2, 2011. URL http://www.mendosa.com/noninvasive_glucose.pdf.
[5] Haaland M, Ries M, Koepp VTW, Eaton R. Reagentless near-infrared determination of glucose in whole blood using multivariate calibration. Appl Spectrosc 1992;46(10):1575–8.
[6] Cattivelli FS, Garudadri H. Noninvasive cuffless estimation of blood pressure from pulse arrival time and heart rate with adaptive calibration. In: International workshop on wearable and implantable body sensor networks. 2009. p. 114–9.
[7] Chen W, Kobayashi T, Ichikawa S, Takeuchi Y, Togawa T. Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration. Med Biol Eng Comput 2000;5:569–74.
[8] Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007;28:R1–39.
[9] Nitzan M, Patron A, Glik Z, Weiss A. Automatic noninvasive measurement of systolic blood pressure using photoplethysmography. BioMed Eng OnLine 2009;8(1):28.
[10] Jeong I, Jun S, Um D, Oh K, Yoon H. Non-invasive estimation of systolic blood pressure and diastolic blood pressure using photoplethysmograph components. Yonsei Med J 2010;51(3):345–53.
[11] Suzuki S, Oguri K. Cuffless and non-invasive systolic blood pressure estimation for aged class by using a photoplethysmograph. In: Engineering in medicine and biology society. 30th annual international conference of the IEEE. 2008. p. 1327–30.
[12] ipod (r) digital oximeter. Last accessed May 2, 2011. URL http://www.nonin.com/OEMSolutions/iPod.
[13] Accu-chek aviva. Last accessed May 2, 2011. URL https://www.accu-chek.com/us/glucose-meters/aviva.html.
[14] Yildirim C, Mete S, Kamber D. Blood viscosity and blood pressure: role of temperature and hyperglycemia. Am J Hypertens 2001;14(5):433–8.
[15] Lanfranchi PA, Somers VK. Arterial baroreflex function and cardiovascular variability: interactions and implications. Am J Physiol Regul Integr Comp Physiol 2002;283(4):815–26.
[16] Cavalcanti S, Belardinelli E. Modeling of cardiovascular variability using a differential delay equation. IEEE Trans Biomed Eng 1996;43(10): 982–9.
[17] Cerutti C, Barres C, Paultre C. Baroreflex modulation of blood pressure and heart rate variabilities in rats: assessment by spectral analysis. Am J Physiol Heart Circ Physiol 1994;266(5):H1993–2000.
[18] Fazan R, de Oliveira M, Dias da Silva VJ, Joaquim LF, Montano N, Porta A, et al. Frequency-dependent baroreflex modulation of blood pressure and heart rate variability in conscious mice. Am J Physiol Heart Circ Physiol 2005;289(5):H1968–75.
[19] deBoer RW, Karemaker JM, Strackee J. Hemodynamic fluctuations and baroreflex sensitivity in humans: a beat-to-beat model. Am J Physiol Heart Circ Physiol 1987;253(3):680–9.
[20] Keener JP, Sneyd J. Mathematical physiology. Interdisciplinary applied mathematics, vol. 8. New York: Springer Verlag; 1998.
[21] Di Rienzo M, Parati G, Radaelli A, Castiglioni P. Baroreflex contribution to blood pressure and heart rate oscillations: time scales, time-variant characteristics and nonlinearities. Philos Trans R Soc A: Math Phys Eng Sci 2009;367(1892):1301–18.
[22] Cornier M-A, Dabelea D, Hernandez TL, Lindstrom RC, Steig AJ, Stob NR, et al. The metabolic syndrome. Endocr Rev 2008;29(7):777–822.
[23] Ducher M, Cerutti C, Gustin MP, Abou-Amara S, Thivolet C, Laville M, et al. Noninvasive exploration of cardiac autonomic neuropathy. Four reliable methods for diabetes? Diabetes Care 1999;22(3):388–93.
[24] van Ravenswaaij-Arts CMA, Kollee LAA, Hopman JCW, Stoelinga GBA, van HP, Geijn. Heart rate variability. Ann Intern Med 1993;118(6):436–47.
[25] Sapolsky RM, Romero LM, Munck AU. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. Endocr Rev 2000;21(1):55–89.
[26] McCraty R, Atkinson M, Tiller WA, Rein G, Watkins AD. The effects of emotions on short-term power spectrum analysis of heart rate variability. Am J Cardiol 1995;76(14):1089–93.
[27] Grossman E, Grossman A, Schein M, Zimlichman R, Gavish B. Breathing-control lowers blood pressure. J Hum Hypertens 2001;15(5):263–9.
[28] Elliot W, Izzo J, White W, Rosing D, Snyder C, Alter A, et al. Graded blood pressure reduction in hypertensive outpatients associated with use of a device to assist with slow breathing. J Clin Hypertens 2004;10(6):553–9.
[29] Schein M, Alter A, Levine S, Baevsky T, Nessing A, Gavish B. High blood pressure reduction in diabetics with interactive device-guided paced breathing: final results of a randomized controlled study. J Hypertens 2007;25(2): 292.
[30] Leonard P, Beattie TF, Addison PS, Watson JN. Standard pulse oximeters can be used to monitor respiratory rate. Emerg Med J 2003;20(6):524–5.
[31] Rabiner LR, Juang B-H. Fundamentals of speech recognition. Englewood Cliffs, NJ, USA: Prentice Hall; 1993.
[32] Dunn RB, Quatieri TF, Kaiser JF. Detection of transient signals using the energy operator. In: IEEE International conference on Acoustics, Speech, and Signal Processing (ICASSP93). 1993. p. 145–8.
[33] Kaiser JF. Some useful properties of teager's energy operators. In: IEEE international conference on acoustics, speech, and signal processing (ICASSP93). 1993. p. 149–52.
[34] Ying G, Mitchell CD, Jamieson LH. Endpoint detection of isolated utterances based on a modified teager energy measurement. In: IEEE international conference on acoustics, speech, and signal processing (ICASSP93). 1993. p. 732–5.
[35] Li Q, Zheng J, Tsai A, Zhou Q. Robust endpoint detection and energy normalization for real-time speech and speaker recognition. IEEE Trans Speech Audio Process 2002;10(3):146–57.
[36] Duda RO, Hart PE, Stork DG. Pattern classification. New York: Wiley-Interscience Publication; 2000.
[37] Renevey P, Drygajlo A. Entropy based voice activity detection in very noisy conditions. In: Dalsgaard P, Lindberg B, Benner H, Tan Z-H, editors. In: EUROSPEECH-2001. Aalborg, Denmark; 2001:1887–1890.
[38] Shen J-L, Hung J-W, Lee L-S. Robust entropy-based endpoint detection for speech recognition in noisy environments. In: Proc. ICSLP98. Australasian Speech Science and Technology Association; 1998. p. 1015–8.
[39] Leonard P, Douglas J, Grubb N, Clifton D, Addison P, Watson J. A fully automated algorithm for the determination of respiratory rate from the photoplethysmogram. J Clin Monit Comput 2006;20(February 4):33–6.
[40] Breiman L. Randomizing outputs to increase prediction accuracy. Machine Learn 2000;40(3):229–42.

[41] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning springer series in statistics. New York: Springer; 2001.
[42] Bishop C. Neural networks for pattern recognition. Oxford: Clarendon Press; 1995.
[43] Abu-Mostafa YS. Learning from hints in neural networks. J Complex 1990;6:192–8.
[44] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge, U.K: Cambridge University Press; 2000.
[45] Canu S, Grandvalet Y, Guigue V, Rakotomamonjy A, SVM and kernel methods matlab toolbox. Perception Systmes et Information. INSA de Rouen. Rouen, France; 2005. Last accessed May 2, 2011. URL http://asi.insa-rouen.fr/enseignants/arakotom/toolbox/index.html.
[46] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Belmont, CA, USA: Wadsworth and Brooks; 1984.
[47] Breiman L. Random forests. Machine Learn 2001;45(1):5–32.
[48] O'Brien E, Waeber B, Parati G, Staessen J, Myers MG. Blood pressure measuring devices: recommendations of the European Society of Hypertension. Br Med J 2001;322(7285):531–6.
[49] Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987;10(5):622–8.
[50] Codina EG, Clarke error grid analysis; 2008. matlabcentral fileexchange. Last accessed May 2, 2011. URL http://www.mathworks.com/matlabcentral/fileexchange/20545-clarke-error-grid-analysis.
[51] Monte E. Descripción de un sistema para estimar el nivel de glucosa y de presión sanguínea a partir de las medidas de un pulsioximetro e información auxiliar. Internal Report 252 Mon. Universitat Politècnica de Catalunya. Barcelona, Spain; 2008 (June).
[52] Validated blood pressure monitors. Last accessed May 2, 2011. URL http://www.bhsoc.org/blood_pressure_list.stm.