

1 2

3 4

# Type 2 diabetes screening test by means of a pulse oximeter

Enrique Monte Moreno\*, Maria José Anyo Luján, Montse Torres Rusiñol, Paqui Juárez Fernández, Pilar Núñez Manrique, Cristina Aragón Triviño, Magda Pedrosa Miquel, Marife Alvarez Rodríguez, M. José González Burguillos

**Abstract**—In this paper, we propose a method for screening for the presence of type 2 diabetes by means of the signal obtained from a pulse oximeter. The screening system consists of two parts; the first analyses the signal obtained from the pulse oximeter, and the second consists of a machine-learning module.

The system consists of a front end that extracts a set of features from the pulse oximeter signal. These features are based on physiological considerations. The set of features were the input of a machine-learning algorithm that determined the class of the input sample, i.e. whether the subject had diabetes or not. The machine-learning algorithms were random forests, gradient boosting, and linear discriminant analysis as benchmark. The system was tested on a database of 1,157 subjects (two samples per subject) collected from five community health centres.

The mean receiver operating characteristic (ROC) area found was 69.4% (median value 71.9% and range [75.4%–61.1%]), with a specificity=64% for a threshold that gave a sensitivity=65%.

We present a screening method for detecting diabetes that has a performance comparable to the glycated haemoglobin (haemoglobin A1c HbA1c) test, does not require blood extraction, and yields results in less than five minutes.

**Index Terms**—Type 2 Diabetes; Screening; Machine learning; Statistical learning, Boosting; Photoplethysmography; Signal processing; Noninvasive treatment.

## I. INTRODUCTION

The current techniques used to diagnose diabetes are based on urine, glycated haemoglobin (haemoglobin A1c HbA1c) and oral glucose tolerance (OGTT) tests [1], [2]. In these cases, the test depends on variations in the glucose level in the blood or on the measurement of the alteration of the structure of haemoglobin cause by long exposure to high levels of glucose. In this paper, we propose a method for detecting whether a person has diabetes by means of the signal obtained from a pulse oximeter. This screening method has several advantages, in that the diagnosis is done with a non-invasive measure obtained from pulse oximetry and does not require

This work has been supported by the Spanish Ministerio de Economía y Competitividad, contract TEC2015-69266-P.

\*Enrique Monte Moreno is with the Department Signal Theory and Communications, Universitat Politècnica de Catalunya, Campus Nord, Edifici D5, C/Jordi Girona, 1-3, 08034, Barcelona, Spain. (email:enric.monte@upc.edu)Corresponding author. Tel.:+34 93 4016435;fax:+34 93 4016447.

Maria José Anyo Luján is at the Engineering department manager, of Synmed. Avda. Príncipe De Asturias 43, 08012 Barcelona, España

Montse Torres Rusiñol, Paqui Juárez Fernández, Pilar Núñez Manrique, Cristina Aragón Triviño, Magda Pedrosa Miquel, Marife Alvarez Rodríguez, M. José González Burguillos are at the SAP Delta del Llobregat. Institut Català de la Salut (ICS).c. Rosell, 4-6. 08902 L' Hospitalet de Llobregat. Spain

laboratory analysis such as in the case of a urine, glycated haemoglobin, or glucose tolerance test or a self-test HbA1c kit that requires a blood sample. Pulse oximetry is done by means of a photoplethysmogram (PPG), which gives a signal that is proportional to the perfusion of blood to the dermis. In this paper, we show that the information given by the PPG signal allows for a screening method that detects the presence of diabetes.

This paper is structured as follows: in section II, we discuss the physiological properties that are reflected in the PPG signal. In section III, we discuss the signal processing techniques applied to extract the relevant features for the machine learning module that does the classification. In section IV, we justify the selection of the machine-learning algorithms. In section V, we describe the database on which the system was tested. In section VII, we present the results, and finally, in section VIII, we discuss the methodology and results of this work.

Although the diagnosis of diabetes is established by the measurement of plasma glucose [1], an easy to use screening method might help to identify patients who are either pre-diabetics or undiagnosed. According to a survey done from 1988 to 1994 [3], the prevalence of diagnosed diabetes was estimated to be 5.1% of the U.S. adult population, and the prevalence of undiagnosed diabetes was 2.7%, which means that one in every three persons with diabetes is undiagnosed. However, the trend of prevalence of diabetes is increasing, and according to [4], the prevalence of diagnosed diabetes in the US by 2050 will be 7.2% (29 million persons). If the proportion of diagnosed to undiagnosed diabetes is maintained, the number of undiagnosed diabetics might be about ten million. In England [5], the prevalence has grown from 10% in 2003 to 35% in 2011. Therefore, the actual estimates are unreliable. In the case of the Spanish population, a national study [6] to examine the prevalence of diabetes and impaired glucose regulation conducted at 100 health centres from 2009–10 with 5,000 participants, found that the prevalence of diagnosed diabetic patients was 7.8% and that the percentage of undiagnosed diabetics was 6.0% (in both cases, the CI was  $\pm 0.7\%$ ). This indicates a total of 2.8 million persons. In the specific case of the region where the study was conducted, [7] found that the prevalence rates of known diabetes, unknown diabetes, and impaired glucose tolerance were 6.4%, 3.9%, and 11.9% in men and 6.9%, 3.4%, and 11.9% in women, respectively. The use of a screening device is justified by the increasing prevalence of type 2 diabetes in the general population. The diagnosis criteria according to [1], are as

follows:

- Symptoms of diabetes and causal glucose level greater than  $200\text{mg/dL}$
- Fasting plasma glucose greater than  $126\text{mg/dL}$
- Plasma glucose level after 2 h in a glucose tolerance test greater than  $200\text{mg/dL}$

If any of these three criteria is met after several repeated test, the diagnostic is positive. Early diagnosis of diabetes is important; according to [8] and [9], half of the people who meet the diagnostic criteria for diabetes are undiagnosed, and it might take up to 12 years before clinical recognition, and when the patients are diagnosed, up to 25% have established retinopathy, and 50% have clinical evidence of diabetic tissue damage and additional cardiovascular risk factors.

One of the advantages of the method proposed in this paper is the reduction in the cost of wide spread screening that might alert from undiagnosed diabetes or prediabetic states. [10] found that the cost per identified case between \$176 and \$236 (US dollars in 2003, comparing different screening methods), corresponding to medical costs that include laboratory tests, personnel time, and other material costs but excluding non-medical costs such as transportation, and time spent traveling. Interestingly, the authors defined the cost as the total cost of detection divided by the total number of cases identified.

The test that we propose might improve this ratio because the test is cost effective and easy to use. The costs consist of the price of a pulse oximeter, and the software can be run on a standard PC, which is usually available at the test point for other purposes. It does not need reagents, and because the measurement mechanism is simple and involves measuring one minute of signal, there is no need for a specialised technician. The clinical implication of the method is that it allows for a fast and inexpensive way of screening for diabetes. As stated in [11], shifting from glucose diagnostic criteria to HbA1c, would have a profound impact on the diagnosis of diabetes because of the resources, time, and costs needed for the 2-h plasma glucose and/or fasting plasma glucose diagnostic test. The method proposed in this paper has slightly lower performance than the HbA1c test but does not need reagents, allowing for testing at the point of care (e.g., a doctor's office or pharmacy), also does not need blood extraction, the results is known immediately and the cost of the device is extremely low. This allows the testing of patients who otherwise would not be tested, either because of the need for blood extraction, the costs of the test, or the fact that routine medical checks tend to be short. Another clinical benefit comes from the fact that the alternative, based on glycated haemoglobin, gives an indirect indication through changes in the structure of the haemoglobin, which depends on the diet that the patient has followed in the previous three months. Therefore, as the method we propose does not depend on the diet, it will not give false negatives as a result of the diet that the patient has followed. In addition, as explained in section II, the underlying measurement is related to alterations in the cardiac and arterial system, so the result of the test that we propose in this paper can give an indirect indication of the extent to which the diabetic disease has harmed the cardiac system of the patient.

## II. PHYSIOLOGICAL FACTORS THAT EXPLAIN THE EXTRACTED FEATURES

In this section, we will justify the use of the signal obtained from a photoplethysmogram (PPG) to measure indirectly the hemodynamics and the presence of either hyperglycaemia or damage produced by diabetes in a given subject. A PPG gives a signal that is proportional to the pressure pulse of the cardiac cycle, and the shape of the pulse is influenced by the different parameters of the arteries and the hemodynamical state of the person.

One of the physiological parameters that is affected by diabetes is heart rate variability (HRV) [12]. Appendix A gives a summary of HRV and of how it is measured. The HRV statistics change with the presence and duration of diabetes and also allows for the characterisation of the severity of the disease. [12] found a linear relationship between HRV, the duration of diabetes, and age; i.e., the HRV decreases with each additional year, and in the case of diabetes, the decrease is not only larger but also proportional to the duration of the disease. Figure 1 of [12], listed various examples of HRV and compared HRV between control individuals and patients with diabetes, showing the sharp contrast in the variability between the healthy and the diabetic patients.

One of the assumptions of this work is that HRV can be derived from the PPG waveform. This assumption is backed by previous results in the literature. [13] tested the relationship between HRV measured from electrocardiogram recordings and the signal obtained with PPG. They also checked the agreement using various statistical tools and concluded that the measurement obtained by means of the PPG signal is accurate enough. For this paper, it is relevant that their tests showed that the RR interval from the ECG is aligned with the PP interval. In addition, [14] the authors conducted an extensive review of the accuracy of the estimate of HRV from the pulse rate variability (PRV), which is measured from the PPG. They compared with detail 32 publications, taking into account various factors, and found that the PRV is accurate for healthy subjects at rest, with discrepancies appearing as a result of physical or mental stress, and that the variability is somewhat overestimated by the PRV, reflecting coupling effects between respiration and the cardiovascular system.

Another assumption of this work is that features from the ECG can be inferred from the PPG waveform through processing of the signal. For instance, [15], simultaneously measured the ECG and the PPG signals and showed that there was correspondence between waveforms. As mentioned above, [14] provided a thorough summary of the literature referring to the use of the comparison between the PPG technology and the ECG. Although most of the studies they considered are related to the reliability of computing the HRV with PPG as a surrogate for the ECG, they showed the relationship between the systolic peaks and the diastolic minima with the features of the pulse wave signal, which allows for the different pulse intervals to be substituted for the RR interval.

The strong assumption that we are making in this paper, which is plausible but not supported by the literature, is that, from the PPG signal, through processing, we can obtain surrogates

for the  $Q$ ,  $R$ , and  $T$  points of the  $PQRST$  complex of the ECG. If these surrogates provide information related to the different points of the ECG, we can make use of the fact that the intervals between these points reflect the alterations caused by the diabetic disease. On the other hand, empirically we have found that the parameters we propose as surrogates improve the classification rate.

The usefulness of these parameters are justified by the following facts. For instance [16], showed that episodes of hypoglycaemia induced abnormalities in cardiac repolarisation, including lengthening of the  $QT$  interval, and  $QT$  dispersion and the normalised  $QT_c$ , which is defined as the ratio  $QT_c = QT/RR$ , so it is the dispersion with respect to the fraction of the  $RR$  interval. In addition [17], showed that Hypoglycaemia creates electrophysiologic alterations causing  $PR$  interval shortening,  $ST$ -segment depression,  $T$ -wave flattening, a reduction of the  $T$ -wave area, and  $QT_c$ -interval prolongation. [18], found that, in children and adolescents with type 1 diabetes, the  $QT_c$  is lengthened (note that, in this paper, the corrected  $QT$  interval is defined as  $QT_c = QT/\sqrt{RR}$ ), and an interesting aspect is that, besides the increase in the length of the  $QT_c$ , the range of this interval increased and the maximum value was 4 times the mean value. In contrast with healthy individuals [19], the  $QT_c$  is not prolonged under normal conditions, although, after eating a carbohydrate-rich meal the effect of hyperglycaemia temporarily increases the  $QT_c$  interval. Other ECG parameters are altered such as a flattening of the  $T$  wave,  $ST$ -depression,  $PQT$  intervals, the duration of the QRS interval, which justifies the search for surrogate estimators of the  $PQRST$  points, and the characteristic shapes of wave forms in amplitude and durations. The above findings coincide with those of the Framingham heart study [20] in that, from the epidemiological point of view there was an association of hyperglycaemia with reduced heart rate variability, and this effect was especially marked in diabetics. In the paper the authors suggested that the alteration of the heart rate variability is caused by the presence of abnormalities in neural regulatory mechanisms. This mechanism is explained, for instance by [21], who stated that cardiovascular autonomic diabetic neuropathy is one of the most common diabetes-associated complications and that changes in heart rate variability are often the earliest symptom, even before the symptoms of diabetes appear. The two studies coincide in finding that heart rate variability decreases as a function of time in the case of patients with diabetes. Interestingly, the neural regulatory mechanisms affected by the presence of diabetes also control other parameters that affect the shape of the PPG signal, such as the blood pressure and respiration ([21] and [22]).

### III. SIGNAL PROCESSING OF THE PPG WAVEFORM

The recordings were made by means of a CONTEC CMS50D+ pulsi oximeter with a USB connection to a standard desktop computer (for more details, see Appendix B). The total duration of the recording was one minute following the stabilisation of the signal.

The signal processing module consisted of two submodules: the first was the activity detection module, the purpose of

which is to ensure the quality of the signal, in the sense that it rejects recordings where the signal has a low amplitude or bad quality, caused either by the movement of the subject or by incorrect placement of the pulse oximeter. The second processes the signal to extract the features that are relevant for discrimination between diabetics and non-diabetics subjects.

The initial version of the system had a high number of candidate feature sets (about 110), which were reduced after an exploratory phase. To avoid introduce significant biases, the author decided to do the feature selection by training the machine-learning module (see section IV) with the first three community health centres, and testing with the fourth and fifth. For a description of the demography of the community health centres see section V. As shown in the results section VII, this selection of the features did not introduce a strong bias in the results. Among the various possible validation strategies, the author considered that the natural unit partitioning the data is the community health centre, because the strategy does not mix unknown confounding variables and is more similar to a real-world situation. The criterion for selecting the relevant features was the ranking given by the random forest algorithm [23] and the backward feature selection technique [24] using as a classifier the random forest algorithm and the gradient boosting algorithm (see section IV for the justification of the selection of these classification algorithms). Note that, although the random forest is common to both methods, the criterion for the ranking is different; the internal ranking in the random forest algorithm is based on the permutation of values of the input features, while the backward feature selection is a greedy search that tries combinations of features. The results and rankings were consistent, and the set of selected subset of features is presented in subsection III-B. The selected sets of features are described in section III-B, and consisted of nine sets, which added up to a total of 69 features for the input vector for the machine-learning algorithm. As there had already been a filtering phase for the candidate features, the relative ranking of the selected features was more or less flat, which resulted in the fact that, when doing a ranking by taking each feature by itself, the members of each set of features appeared mixed and did not coincide with the ranking done by the class of features. In addition, the ranking changed by CHC when a cross validation was done.

#### A. Activity detection algorithm

An activity detection module (ADM) based on a finite state automaton was implemented. The activity detection algorithm was used to guarantee that the recorded signal was of sufficient quality. The details of the activity detection module are described in [22]. The criterion for selecting a one-minute analysis window is a trade-off between the fact that the short-term HRV can be roughly estimated from a recording of a few minutes (see Appendix A) and the operational restrictions (the time slots for the recordings) on the normal functioning of the community health centres where the samples were collected. The total number of samples in the analysis window was 3,600, corresponding to a minute at a sampling rate of 60 samples/second. The ADM

consisted of a local classifier that operated on frames of five seconds (300 samples) that made a local decision on the classes 'signal'/'not signal'. This classifier made the decision using as features the Kaiser Teager energy, Qi-Zheng energy, spectral entropy, and zero crossing rate. This local decision used finite state-automaton with three states and simple rules for deciding whether the signal was correct or the recording should be discarded.

### B. Feature extraction

The features extracted from the PPG signal obtained after the ADM, which was a vector 3,600 in length. We will denote this signal  $x[n]$ . The extracted features are related to the general shape of the of the PPG signal and to the heart rate variability statistics.

Before processing to extract the features, the PPG signal was first filtered by means of a median 1D filter of length 5 to remove spurious peaks. Then it was low-pass filtered by means of a Chebyshev filter of type 2 of the order 7, which had a cut-off normalised frequency of 1/8, with 40dB in the stopband, and subsequently the signal was filtered by a simple pre-emphasis high-pass filter with impulsional response:

$$h_{preemp}[n] = \delta[n] - 0.98\delta[n - 1] \quad (1)$$

This pre-processing not only gave a cleaner signal but also made clearer the systolic peaks, the diastolic minima (which denote the onset of the systole, the slope between the two points) and the presence of the dicrotic notch [14]. These features will be useful for the processing done in subsection III-B2.

1) *Cepstral analysis*: In this section, we describe the features obtained from the cepstral analysis of the general shape of the PPG signal. The cepstral analysis has been used to characterise the HRV [25] and [26], specifically for the case of diabetics. The use of the cepstral analysis allows us to better distinguish the LF and HF components in a representation that has fewer parameters. Another advantage of using cepstral analysis is that it transforms a convolution into an addition [27], which allows for a rough separation of some of the information concerning the state of the arteries from the cardiac excitation. Note that we emphasise that it is partial information, because the PPG signal gives an incomplete representation of the state of the circulatory system.

**Short-term cepstrum.** The use of the short-term cepstrum is justified by the fact that it will capture the specific traits of the shape of the PPG signal by means of a few coefficients. To compute the short-term cepstrum of the signal [27], and the signal  $x[n]$  was divided into frames  $\mathbf{x}_i$  of length  $L_{frame} = 240$ , with an overlap of 50%, giving a sequence of  $N_f$  frames, where the index  $i$  indicates the frame number

$$\mathbf{x}_i = \begin{bmatrix} x[(i-1) \times L_{frame}], x[(i-1) \times L_{frame} + 1], \\ \dots, x[i \times L_{frame}] \end{bmatrix} \quad (2)$$

For each frame, the cepstrum was computed as follows:

$$c_i^{st} = \text{ifft}(\log(\text{fft}(\mathbf{x}_i))) \quad (3)$$

As the channel component of the cepstrum corresponded to the first cepstral coefficients, we retained the first  $N_{ceps}^{st} = 10$  coefficients of the mean cepstrum, which we will denote the reduced cepstral vector  $c_i^r \leftarrow c_i[1 : N_{ceps}^{st}]$ . Finally, the input feature corresponding to the cepstrum was defined as the mean value of the frame cepstrum vector.

$$c_{\text{Mean}}^{st} = \frac{1}{N_f} \sum_{i=1}^{N_f} c_i^r \quad (4)$$

The computation of the mean cepstrum is justified by the fact that we can characterise the state of the arteries of the patient by means of an approximation of the impulsional response of the system that relates the heart signal to the final observation. Given that each  $x_i$  can be understood as a varying excitation convolved by a channel and the cepstrum transforms a convolutional mix into an addition, the mean value will be the unvarying part, which corresponds to the channel. This idea has been used in speech processing to determine the spectral characteristics of channels [28].

**Long-term cepstrum.** The use of the long-term cepstrum is justified by the fact that it will capture information about the evolution of the shape of the PPG signal and the excitation, which reflects the periodicities in the repetition of the basic shape of the signal. Note that, in contrast to the short-term cepstrum, for the long term we do not average the cepstrum values. The cepstrum of a given time series consists of the addition of the convolutive terms that generate the observed signal  $x[n]$ . Because the excitation and channel correspond to different cepstral coefficients, we used as a feature the first  $N_{ceps}^{lt} = 20$  coefficients of the cepstrum computed on all the observation vectors  $x[n]$ .

$$c_{\text{glob}}^{lt} = \text{ifft}(\log(\text{fft}(x))) [1 : N_{ceps}^{lt}] \quad (5)$$

The number of coefficients retained in both cases was decided by the author after the inspection of a few examples. **Zero crossing rate cepstrum.** Another way of capturing HRV is to compute the cepstral coefficients of the zero crossing points of  $x[n]$ . These coefficients will give the general shape of the evolution of the periodical component of the signal. We will denote this feature  $c_{zc}$ , it corresponds to the first  $N_{ceps}^{zc} = 10$  cepstral coefficients of the distances of the zero crossing points of  $x - \text{mean}(x)$ .

**General comments about the cepstrum analysis.** In the initial experiments, to determine which features were relevant, we found that the use of LPC-cepstrum gave worse performance than the fast Fourier transform (FFT) based cepstrum, so it was discarded. Something similar occurred with the raw LPC coefficients and the line spectrum pairs (LSP) coefficients, which were also discarded. Note also that

the selection of the number of coefficients corresponds to an implicit cepstral liftering, i.e., to multiplying the cepstral sequence by a pulse of shorter duration. The use of an additional liftering function, such as a ramp, did not improve performance, which we explain by the fact that the base classifiers are decision trees (see section IV), which make local decisions at each node by means of a comparison, so the scaling of an individual coefficient does not affect the result.

2) *Surrogate parameters* : In section II, we justified the usefulness of the parameters of the ECG for determining whether a subject had diabetes. In this section, we present features (indirectly related to the ECG signal) that allow us to distinguish between subjects with and without diabetes. It is important to remember that we posit that some features of the ECG are reflected in the PPG signal.

The first feature of the ECG is the RR distance, which in the PPG signal  $x[n]$  corresponds to the distance between systolic peaks. From the RR distance between the peaks of  $x[n]$ , we created a new vector that we will denote  $x_{RR}[n]$ . The elements of  $x_{RR}[n]$  are the distances (in samples) between the peaks of  $x[n]$ . In addition, to determine other features of interest from the PPG signal, we recorded the positions of the peaks, which we will denote  $x_{pos}[n]$ .

#### Ratio of low-frequency to high-frequency (HF) energies.

To find the ratio of low-frequency to high-frequency energies of the HRV, we computed the FFT of length  $Nfft$  (greater than the length of the signal  $x_{RR}[n]$ ) as follows;

$$X_{RR} = FFT_{Nfft}(x_{RR}[n]) \quad (6)$$

From  $X_{RR}$ , we computed the energy in the HF and LF bands. These bands in our case were defined in the normalised frequency (i.e., a sampling frequency corresponding to 1.0) <sup>1</sup> as follows:  $LF = [0.0, 0.1]$  and  $HF = [0.1, 0.2]$ . As feature, we took the ratio between the sum of the absolute values of the bins at LH and HF intervals of  $X_{RR}$ ,

$$R_{atioLH_2HF} = \frac{\sum_{k=0}^{\lfloor 0.1 \times Nfft \rfloor} |X_{RR}[k]|}{\sum_{k=\lceil 0.1 \times Nfft \rceil}^{\lfloor 0.2 \times Nfft \rfloor} |X_{RR}[k]|} \quad (7)$$

**Position of the highest peak of the RR and cepstrum of the RR** Two other features were extracted from the vector  $x_{RR}[n]$ , the normalised frequency corresponding to the highest peak of  $X_{RR}$  and the first five coefficients of the cepstrum of the  $x_{RR}[n]$  signal. These features will be denoted as follows:  $Pos_{RR}^{Peak}$  and  $Ceps_{RR}$ .

**Surrogate features** Next, to find the surrogates of the  $PQRST$  points of the ECG signal, we created a vector that contained information about the critical points of the PPG signal. This vector was generated from the pre-processed signal  $x[n]$  by two filtering steps. First,  $x[n]$  was filtered by means of a high-pass filter that emphasised the points' maximal slope and the maxima and minima of the signal.

<sup>1</sup>Note that there is ambiguity in the definition, we have followed the definition given by [29], which differs, for instance, from that of Matlab, where the normalised frequency is defined as half the sampling frequency. In contrast [27] uses normalised radian frequencies.

The filter was  $h_{der}[n] = \delta[n] - \delta[n - 2]$ , which yielded  $x_{hpass}[n] = x[n] * h_{der}[n]$ . Next, the sequence  $x_{hpass}[n]$  was filtered by a zero-phase [29] low-pass filter with a cut-off frequency at 0.05, (Chebyshev filter of type 2 of the order 8). This low-pass filter was designed to preserve the maxima/minima of  $x_{hpass}[n]$  which are the critical/turning points of  $x[n]$ . This new signal will be denoted as  $x_{CrtPnts}[n]$ . As the low-pass filter was zero phase, the relative delay between  $x[n]$  and  $x_{CrtPnts}[n]$  is of two samples. Therefore, from the vector  $x_{pos}[n]$ , one can infer the position of the  $R$  peak in the vector  $x_{CrtPnts}[n]$ . This alignment with the  $R$  points allows consideration of the neighbouring peaks. We will take these maxima/minima as surrogates for the real set of points  $PQRST$ .

Note that, although the distances of the peaks obtained are roughly consistent with the distances of the ECG, we are not claiming that they correspond to these points. Our claim is that they are either plausible explanations or surrogates that allow for the discrimination between subjects with and without diabetes. We have found that the statistics derived from these points are informative, in the sense that they contribute to the discrimination between classes. As discussed in section II, it is known that the statistics related to the intervals  $QT$  and  $QT_c$  are relevant to the problem at hand.

We determined two categories of critical points in  $x_{CrtPnts}[n]$ . The first was obtained from the immediately preceding minimum peak of  $x_{CrtPnts}[n]$  aligned with the  $R$  points given by  $x_{pos}[n]$ , which we posit corresponds to the surrogate for the  $Q$  point, and the second was obtained from the immediate following maximum which we posit corresponds to the surrogate for the  $T$  point. The statistics related to the peak immediately following the  $R$  peak, which would be the surrogate for the  $S$ , also give discriminative information.

The set of features elaborated from these distances were the statistics of the surrogate intervals  $QT$ ,  $QT_c = QT/\sqrt{RR}$ , and  $QS$ , which each consisted of a vector with the mean, median, inter-quantile range, absolute mean deviation, standard deviation, and range. These two features will be denoted, respectively as  $QT_c\_Stats$ , and  $QS\_Stats$ .

**Demographic features.** We generated a vector with the set of several demographic features, such as weight (W), height (H), age (A), gender (G), body mass index (BMI), and body fat percentage (BF). The use of the body fat percentage is justified by the fact that the body fat is a predictor of non-insulin dependent diabetes mellitus (see [30]). The BF variable was computed from the formula for adults given in [31]. The formula is  $BF\% = 1.20 \cdot BMI + 0.23 \cdot age - 10.8 \cdot sex - 54$ . The information on whether the subject smoked or not was not used because, in the database, this information was missing in about the 20% of records. The vector with the demographic information was constructed as follows:

$$X_{Demo} = [W, H, A, G, BMI, BF] \quad (8)$$

3) *Final feature vector* : The final feature vector consisted of the aggregation of the partial features, which we define as

follows:

$$Global_{Fea} = [c_{Mean}^{st}, c_{glob}^{lt}, c_{zc}, Ceps_{RR}, R_{ratioLH_2HF}, Pos_{RR}^{Peak}, QT, QT_{c\_Stats}, QS_{Stats}, X_{Demo}] \quad (9)$$

This vector is the input of the machine-learning module.

### C. Feature importance

As stated at the beginning of this section, the initial set of variables was much higher than the one that was finally used. Note that as there are 9 different categories of features, the ranking of the 69 individual features was partially mixed, i.e. the ranking did not cluster the variables by specific categories. On the other hand, the information about the real impact of each category of features is important for the reproducing or improving the method.

A side product of the Random Forest method is a score associated to each feature, that is proportional to the degradation of the performance when the data of a given input is permuted, maintaining the other features intact. This degradation of the performance is computed from the 'out of bag errors'. This allows for ranking the variables, and also for obtaining a ranking score per category. We decided to use as performance index per category the mean value of this score of each variable of a given category. The ranking per category was the following:

- 1) Short term cepstrum
- 2) Long term cepstrum
- 3) Zero crossing rate cepstrum
- 4) Ratio Low frequency to High frequency (HF) ratio of energies
- 5) Position of the highest peak of the RR and cepstrum of the RR
- 6) Surrogate features
- 7) Demographic features

The feature importance was computed for each of the initial 69 features, which means that we are not discounting the total effect of category of variables.

The fact that the demographic features were the least important at the level of each single feature, was confirmed empirically by an experiment where we did not use this category of features. The degradation of the performance was in a range of 2% to 4% of the ROC, depending on the CHC.

Also, we believe that a better estimate of the 'Surrogate features' by increasing the duration of the recordings, should improve the performance of the system and the ranking of this category of features. The reason is justified in section II.

## IV. MACHINE LEARNING MODULE

The system was tested with two different machine-learning algorithms: random forests (*RF*) [32] and gradient boosting (*GB*) [33]. As a benchmark, we took the linear discriminant analysis classifier (LDA) (see [23] chapter 4). The reason for selecting as classifiers the *RF* and *GB*, is that the base classifier of each technique is a classification and regression tree (CART), which allows us to use the raw features without

specific rescaling or transformation of the features [23]. This robustness to the heterogeneity of the type of feature is due to the way in which the CART algorithm decides the output class. The algorithm makes local decisions by comparing the value of a given feature to a threshold, and each comparison is done independently of the other nodes or features. As the feature vector we are using is composed of features of different origin, with different statistical properties, the use of the CART technique as a base classifier is justified. In contrast, linear discriminant analysis, logistic regression, neural nets, and support vector machines are sensitive to the coding/scaling of the features.

Another common characteristic of *RF* and *GB* is their robustness to the collinearity of the input features. In contrast with *RF* and *GB* for other machine-learning algorithms, such as those mentioned above, the input data must be either normalized or whitened because of the sensitivity of these learning algorithms to correlated data or unbalanced ranges. This can give rise to ever-increasing values of the parameters unless regularisation steps are taken (see discussions in [24] and [23]). Extreme values of the parameters give poor performance in generalisation. In addition, numerical algorithms are prone to instabilities when inputs vary in a correlated way.

- In the case of neural networks (multilayer perceptron) or LDA, collinearity can give rise to extreme values of the weights unless regularisation techniques are applied.
- In the case of radial basis functions (neural networks) or support vector machines based on radial basis functions, collinearity gives rise to the need for a high number of units, as there is the need to model the dispersion of the data along certain directions.

Related to the collinearity effect is the fact that some of the features might have information in common, being in some cases nearly repetitions. The decision trees are fairly immune to degradation caused by both effects, simply because the algorithm in this case selects at each node the variable that locally gives the best improvement on the Gini factor. By creating a scatter plot of all the input features, it was detected empirically that this effect was present in our database.

### A. Random forest comments

The algorithm was programmed by the author in Matlab. This algorithm has different hyper-parameters, such as: 1) the number of samples per leaf, 2) the number of features tested per node, and 3) the number of trees in the forest.

Different combinations of these parameters were explored by cross validation on the training database for each testing CHC, and we found that the best results coincided with the recommendations of Breiman [32], which were a few samples per leaf, in our case one, and the number of features tested per node, equal to the square root of the total number of features (using the base 10 logarithm gave very similar results). The performance stabilised at 600 trees, and increasing the number of trees did not improve the performance of the algorithm.

### B. Gradient boosting comments

The algorithm was programmed by the author in Matlab by taking as a base classifier a decision tree algorithm and following the description in [33]. The free parameters of the algorithm were set by cross validation. As there was little difference between the values of the hyperparameters between different experiments, they were set as follows. The cost function was the MSE, and the sub-sampling of the training database was done with a fraction of 0.75, which improved the performance in comparison to not doing the sub-sampling. The regularisation parameter (shrinkage factor) was set at  $v = 0.005$ . The size of the trees was selected to be  $J = 10$  and the number of iterations was  $M = 1500$ .

## V. DESCRIPTION OF THE DATA

The data were collected in five Community Health centres (CHCs) around Barcelona during the spring of 2013, when the flu season had ended [34]. At each CHC, the samples were collected from approximately 250 patients, who gave informed consent. For most of the patients, two recordings (i.e., two PPG measurements) were done. The participants in our study were patients who had routine blood extractions at a CHC. Although there were no explicit exclusion criteria, in very busy moments, some samples for the study were not taken.

### Uniformity of the CHC

We will describe the composition of the CHC regarding the demographics and the possible confounding variables. This is justified by the fact that, to evaluate the system (see section VII), we decided to take the CHC as the unit of validation. By taking the CHC as the unit for partitioning training, validation, and testing, the possible biases and confounding factors specific to each CHC are not mixed. To ensure that the partitioning is valid, we will show in this section that the demographic distribution between CHCs is uniform.

Total number of patients was 1,170, of whom 340, (29.06%) were diabetic and 830 (70.94%) were not diabetic. The proportion of diabetics to non-diabetics was very different between CHCs, as shown in Table II, with two groups of similar CHCs i.e., (1,2, 3, and 4) and an outlier with a higher proportion of diabetics, i.e., CHC 5. With regard to gender differences, the occurrence of diabetes was higher among men; see Table I. The fact that a subject was a type II diabetic was established from the medical record. Note that, in the sample, there are a number of cases of non-diagnosed diabetic and pre-diabetic patients. This will introduce bias to the rate of false negatives because a certain number of samples will have the wrong label. The fact that there are wrongly labelled 'not diabetic' patients means that the sensitivity and specificity of our test, when trained with a perfectly labelled database, should perform better. The design of the database guaranteed that the diagnosed diabetics were truly diabetics and that the proportion of diabetics to non-diabetics would be representative of a real situation in a CHC. In contrast, as the number of recruited patients was high and the diagnosis is expensive, we accepted a number of non-diagnosed diabetics. The uncertainty about the number

of wrongly labelled cases will be in the range of 49 out of 830 (6% prevalence in the general population). However as the data were collected in community health centres from patients who had routine blood extractions, it is expected that the rate of undiagnosed diabetics and pre-diabetics should be low because undiagnosed diabetics correspond to patients who had the blood extraction because they had previously visited the doctor for other medical problems. Therefore, if the diabetic symptomatology were evident, the doctor would have taken action.

The body mass index for diabetics and non-diabetics was also similar between CHCs as shown in Table IV. The histograms of the BMI are shown in Figure 2, which shows that the mean value and inter-quantile range are higher in the case of diabetic patients. The general weight histogram is shown in Figure 3, which shows that both distributions are similar, with a difference caused by a few exceptional outliers in the case of the non-diabetics.

In Table III, we show the percentage of smokers. The global mean is 16.18% and the proportion of diabetic subjects who smoke was lower than the case of non-diabetics. The proportions are not broken down by centres because this information was lacking in one of the CHCs. The distribution of smokers (diabetics or non-diabetic) was almost equal in all CHCs with about one point of difference.

The fasting distribution of the glucose level in *mg/dl* is shown in Figure 1, which shows that, although the mean of the histogram of the diabetic subjects was a 25% higher, interestingly, the mode (i.e., the most likely value) was almost the same in both populations, about 90 *mg/dl*<sup>2</sup>. The difference in the mean comes from the fact that the tails of the distribution in the case of diabetic subjects is much longer. The inter-quantile range (IQR, i.e., the values between the lower 25% and the upper 75%), in the case of the histogram of diabetic subjects, is double compared to the case of non-diabetics.

The age distribution between the diabetic population and the non-diabetic population was similar; as shown in Figure 4, the mean and the inter-quantile ranges are almost the same. Perhaps the only difference is that the mode is higher in the case of the non-diabetic histogram.

In this section, we have shown that the only significant demographic difference between the diabetic and non-diabetic groups is related to the fasting level of glucose and that the CHCs have similar demographic structures.

## VI. METHODOLOGICAL ISSUES

In designing experiments (see, for instance, [35]), a key point for obtaining reliable statistics is dealing with confounding factors that are not controlled in the experiment. If the confounding factors are common to the different natural

<sup>2</sup>Note that the abscissa is not in the same scale.



divisions of the experiment (in this case the CHC), then a 10-fold cross-validation is correct, in the sense that it reduces the variability caused by non-controlled confounders, so the means and confidence margins are more accurate. In our study, if we aggregate all the individuals regardless of the CHC and test in a set of individuals from a mix of the CHCs the results with the 10-fold cross-validation are much better, yielding an upward bias of the ROC by several points, i.e., a mean ROC area of 77.3%, with a standard deviation of 3.4%, which is 7 points better than the results obtained when factoring by CHC (for comparison purposes, see Table V). In our case, the confounding factors that we think that were not uniform between CHCs were the profession, age distribution and national origin of the patient. This information was not available and therefore could not be used for computing the performance taking into account the different factors.

Therefore, instead of a random partition in a 10-fold cross-validation, we decided to take each CHC as the unit for computing the mean and variability, i.e., doing a rotation of training with four CHCs, and testing with the remainder. This partition of the database allowed us to determine the range of the mean caused by specific confounding factors in the test database that were not present in the training database. The results are shown in Table V and in Figures 5, 6 and 7.

The adjustment of the parameters of the classifier was done with a 10-fold cross-validation on the four CHCs used as the training database (i.e., the validation database was extracted from the training data). Note that the use of cross validation for determining the parameters of the classifiers does not mix the confounding factors that might be present in the test database. This partition of the database allowed us to determine the range of the mean due to specific confounding factors of the test database that were not present in the training database. Note also that as there were two samples from most of the subjects, the allocation of samples between the train and validation databases was based on the subject index, i.e., we did not mix samples from the same individual in different databases.

We explain these findings by the fact that although the CHCs are very similar in terms of the different aspects discussed in section V, there might be other conditions that are specific to each CHC that biased the results. We believe that the testing on different CHCs gives a better (more realistic) estimate of the performance on an unseen population, and a better estimate of the variability.

## VII. RESULTS

The data were processed as explained in section III, and tested with three different classifiers: RF, GB, and LDA. The criterion for determining the quality of the results was the receiver operating characteristic (ROC) area [23], which allows for assessing the trade-off between sensitivity and specificity. The ROC area is defined as the lower area of the curve that relates the sensitivity (true positive rate) to the false positive rate (true negative rate or 1.0-specificity), as we vary the decision threshold of the classifier. That is, a given threshold determines the sensitivity and specificity of

the classifier.<sup>3</sup>

The results corresponding to the ROC area obtained by each classifier are presented in Table V, and the ROC figures for the RF and GB algorithms are shown in Figures 6, 5 and 7. The algorithms RF and GB yield similar results; i.e., ROC areas and the shapes of the ROC figures are similar, and the performance for each CHC vary within a few percentage points. Note also that the LDA classifier gave a systematically lower value for the classification results. The results for CHC 3 and 5 were lower due to the fact that the demographics were different from the other CHCs.

## VIII. CONCLUSION

In this paper, we have presented a method of screening for the diagnosis of diabetes based on the signal obtained from a PPG. One of the advantages of this method is that it is a fast test, the results can be obtained in two or three minutes, and the price is low because it needs only a PPG sensor and the computation can be done using either a low-cost computer or a smart phone. Note that the procedure consists of obtaining a sample of the pulse oximeter signal of one minute in duration, which does not require of qualified personnel, and the whole process (including the computation of the results) takes less than five minutes (including possible repetitions of the measure). This is in contrast to plasma glucose measurements or glycated haemoglobin tests, which require the extraction of a blood sample, and laboratory measurements. Although the measurement of glycated haemoglobin is faster than the oral glucose tolerance test (OGTT), the cost is about 13 times higher [36]. Another issue related to the use of glycated haemoglobin is the fact that low values of *HbA1c* do not exclude diabetes [37], which means that the specificity of the method is increased by being tested in a population already at risk that has already had a two-hour OGTT test instead of a full OGTT [37].

In contrast, the ROC area of the method we propose in this paper is 70% (mean for all the CHCs), which, depending on the threshold (see Figures 6, 5 and 7), would give a sensitivity (true positive rate), of 80% and a specificity (1-false positive rate) of 48%. These results are slightly lower than the results obtained by using glycated haemoglobin such as the values reported in [38] [39], and [40]<sup>4</sup>.

Next, we compare the sensitivity and specificity obtained by means of the glycated haemoglobin tests reported in the before mentioned studies with the specificity that we would have with our method if the *decision threshold* were set at the same sensitivity point (note that in our results the percentages have been rounded down).

<sup>3</sup>Note that the ROC area is independent of the proportion of samples in each class. As the number of samples of non-diabetics class, was much higher than the diabetics class, we did the experiment of assigning the label 'non-diabetic' to all patients. The results were a sensitivity of a 100%, a specificity of 0% (100% rate of false positives), a ROC area of 50%, and a precision (positive predictive value of the class diabetic) of 30%.

<sup>4</sup>Note that these studies were done with different populations from different continents.

- [39], reported for a level of  $HbA1c \geq 7.0\%$  a sensitivity of 36% and a specificity of 100%. Our method for the same sensitivity of 36.0% gave a specificity of 84%.
- [38], reported for a level of  $HbA1c \geq 6.1\%$  a sensitivity of 65% and a specificity of 88%. Our method for the same sensitivity 65% gave a specificity of 64%.
- [40], reported for a level of  $HbA1c \geq 6.5\%$  a sensitivity of 72% and a specificity of 66.5%. Our method for the same sensitivity 72% gave a specificity of 57%.

The above comparison between performances is indicative because the performance depends on the prevalence of the disease in the underlying population. Note that the specificity and sensitivity of a screening method depends not only on the specific technique used but also on the prior probability that the tested individual has the disease. The sensitivity and specificity of a given test differ if the test is done on the general population or at a hospital or medical office. In one case, the prior probability is over the whole population, and in the other the probability is biased towards the case of disease.

In the [38] study, the size of the study sample was 1,972 subjects, which is nearly twice the size of our sample, of which 15.4% had diabetes (in our study, the prevalence was 35.72% for men and 23.81% for women). This lower prevalence of diabetes explains the fact that, with the same sensitivity, the specificity with  $HbA1c$  is better. However, the study is comparable to ours in the sense that the population was sampled in a CHC in a city in north India. [39] conducted a meta-analysis of 34 different studies, and the population consisted of 8,984 subjects. In this case, the studies are not directly comparable in relation to the underlying population. [40] was conducted with 315 patients; the population is comparable to the one presented in this study, though the number of patients is a third of the number of patients of the present study.

Although the results are comparable, the performance of our method is slightly lower. This could be explained by the fact that the target population in our study was patients who had routine blood extractions at a CHC and therefore were not expected to be tested for the presence of diabetes. The result is that the prior probability in our study of a patient being diabetic is lower than in studies where the target population has been selected based on the risk of diabetes.

There are two possible explanations for the low specificity.

- One is the fact that the prior probability of the reported diabetics is 28%, which increases that rate of false positives. In epidemiology, it is known that the performance of a screening technique depends greatly on the prior probability of the target class and that the sensitivity and specificity of a test degrades greatly, as the target class to be detected has a low frequency of occurrence.
- The other is the fact that there is a percentage of undetected diabetics in the general population, and our labeling of the diabetic cases were given by the clinical records of the CHCs, which only report previously diagnosed cases. The 2014 National Diabetes Statistics Report [41], stated that, in the USA 27% of the population with diabetes is undiagnosed. If the proportion is similar in Spain, this might explain in part the fact that there

is a higher percentage of false positives in our method. This also indicates that the performance of the screening method could be improved by using a database that is more accurately labelled.

The variability of the screening method can be inferred from Table V, with a mean ROC value of 70% and a standard deviation of 6%. As the statistics come from only five CHC, a better measure of the performance is the median value of 72.9% and the range [78.7% – 59.5%], which is 19.2%. We believe that this variability is representative in the sense that, as the five CHCs are similar with respect to age, BMI, weight, and smoking habits (see section V), the observed variability corresponds to other underlying factors, that are more difficult to measure in a standard setting.

## APPENDICES

### A. Heart rate variability

Heart rate variability (HRV) is considered a marker that reflects the sympathetic or vagal function of the autonomic nervous system. The following summary is based on [42] and [43]. HRV is defined as the variation in the instantaneous heart rate and  $RR$  intervals. The measure corresponds to the interval between consecutive beats and can be characterised in the time domain or in the frequency domain. Alterations in the HRV have been related to autonomic neuropathy in diabetic patients, foetal distress, and a higher risk of post-infarction mortality. The HRV also is affected by other factors such as age (diminishing with age), medication, circadian rhythm, etc. The most common options for measuring HRV variation are 'five-minute' recordings for the short term variability measurement and a '24 – hour' recording for long-term variability. The time domain measurements are the square root of the mean squared differences of successive  $RR$  intervals, the number of interval differences of successive  $RR$  intervals greater than 50ms, and the proportion of the intervals greater than 50ms to the total number of intervals, along with measurements related to the histogram of the  $RR$  interval. The frequency domain measurements are related to the relative energy of the signal at different bands: very low frequency (VLF), low frequency (LF), and high frequency (HF). The relationship between the LF and the HF components is related to changes in autonomic modulations of the heart period. The VLF component has not been used in this study because this component can only be estimated reliably from long recordings (24 hours), while the HF and LF components can be estimated from recordings of a few minutes. These frequency components are computed from the Fourier transform of the time series created from the  $RR$  intervals and from the absolute value of the Fourier transform at each frequency bin. A common way of characterising the HRV from the frequency domain measurements is to present the ratio of  $LF/HF$  power.

### B. Pulse oximeter

The Pulse oximeter that was used to record the signals was a CONTEC CMS50D+ with a USB connection. The recording of the signal from the USB connection was programmed for the recording software, which runs on a standard PC

with Windows 7. The technical characteristics of the pulse oximeter are shown in the Table B.

## REFERENCES

- [1] D. B. Sacks, D. E. Bruns, D. E. Goldstein, N. K. Maclaren, J. M. McDonald, and M. Parrott, "Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus," *Clinical Chemistry*, vol. 48, no. 3, pp. 436–472, 2002.
- [2] W. Consultation, *Definition, diagnosis and classification of diabetes mellitus and its complications*. Part, 1999, vol. 1.
- [3] M. I. Harris, K. M. Flegal, C. C. Cowie, M. S. Eberhardt, D. E. Goldstein, R. R. Little, H.-M. Wiedmeyer, and D. D. Byrd-Holt, "Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in u.s. adults: The third national health and nutrition examination survey, 1988/1994," *Diabetes Care*, vol. 21, no. 4, pp. 518–524, 1998.
- [4] J. P. Boyle, A. A. Honeycutt, K. V. Narayan, T. J. Hoerger, L. S. Geiss, H. Chen, and T. J. Thompson, "Projection of diabetes burden through 2050: Impact of changing demography and disease prevalence in the u.s.," *Diabetes Care*, vol. 24, no. 11, pp. 1936–1940, 2001.
- [5] A. G. Mainous, R. J. Tanner, R. Baker, C. E. Zayas, and C. A. Harle, "Prevalence of prediabetes in england from 2003 to 2011: population-based, cross-sectional study," *BMJ open*, vol. 4, no. 6, p. e005002, 2014.
- [6] F. Soriguer, A. Goday, A. Bosch-Comas, E. Bordiú, A. Calle-Pascual, R. Carmena, R. Casamitjana, L. Castaño, C. Castell, M. Catalá, E. Delgado, J. Franch, S. Gaztambide, J. Gírbés, R. Gomis, G. Gutiérrez, A. López-Alba, M. Martínez-Larrad, E. Menéndez, I. Mora-Peces, E. Ortega, G. Pascual-Manich, G. Rojo-Martínez, M. Serrano-Rios, S. Valdés, J. Vázquez, and J. Vendrell, "Prevalence of diabetes mellitus and impaired glucose regulation in spain: the diab.es study," *Diabetologia*, vol. 55, no. 1, pp. 88–93, 2012.
- [7] C. Castell, R. Tresserras, J. Serra, A. Goday, G. Lloveras, and L. Salleras, "Prevalence of diabetes in catalonia (spain): an oral glucose tolerance test-based population study," *Diabetes research and clinical practice*, vol. 43, no. 1, pp. 33–40, 1999.
- [8] T. Lauritzen, S. Griffin, K. Borch-Johnsen, N. Wareham, and B. Wolfenbutter, "The addition study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with type 2 diabetes detected by screening," *International Journal of Obesity & Related Metabolic Disorders*, vol. 24, 2000.
- [9] U. P. D. S. Group, "U.k. prospective diabetes study 16: Overview of 6 years' therapy of type ii diabetes: A progressive disease," *Diabetes*, vol. 44, no. 11, pp. 1249–1258, 1995. [Online]. Available: <http://diabetes.diabetesjournals.org/content/44/11/1249.abstract>
- [10] P. Zhang, M. M. Engelgau, R. Valdez, S. M. Benjamin, B. Cadwell, and K. V. Narayan, "Costs of screening for pre-diabetes among us adults a comparison of different screening strategies," *Diabetes Care*, vol. 26, no. 9, pp. 2536–2542, 2003.
- [11] B. Costa, F. Barrio, J.-J. Cabré, J.-L. Piñol, F.-X. Cos, C. Sole, B. Bolibar, C. Castell, J. Lindström, N. Barengo *et al.*, "Shifting from glucose diagnostic criteria to the new hba1c criteria would have a profound impact on prevalence of diabetes among a high-risk spanish population," *Diabetic Medicine*, vol. 28, no. 10, pp. 1234–1237, 2011.
- [12] S. Masaoka, A. Lev-Ran, L. R. Hill, G. Vakil, and E. H. Hon, "Heart rate variability in diabetes: relationship to age and duration of the disease," *Diabetes Care*, vol. 8, no. 1, pp. 64–68, 1985.
- [13] P. Shi, S. Hu, and Y. Zhu, "A preliminary attempt to understand compatibility of photoplethysmographic pulse rate variability with electrocardiographic heart rate variability," *Journal of Medical and Biological Engineering*, vol. 28, no. 4, pp. 173–180, 2008.
- [14] A. Schäfer and J. Vagedes, "How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram," *International journal of cardiology*, vol. 166, no. 1, pp. 15–29, 2013.
- [15] W. Shin, Y. D. Cha, and G. Yoon, "Ecg/ppg integer signal processing for a ubiquitous health monitoring system," *Journal of Medical systems*, vol. 34, no. 5, pp. 891–898, 2010.
- [16] N. Harris, R. Ireland, J. Marques, S. Hudson, C. Davies, S. Lee, R. Robinson, and S. Heller, "Can changes in qt interval be used to predict the onset of hypoglycemia in type 1 diabetes?" in *Computers in Cardiology 2000*. IEEE, 2000, pp. 375–378.
- [17] V. P. Sanon, S. Sanon, R. Kanakia, H. Yu, F. Araj, R. Oliveros, and R. Chilton, "Hypoglycemia from a cardiologist's perspective," *Clinical cardiology*, vol. 37, no. 8, pp. 499–504, 2014.
- [18] B. Suys, S. Heuten, D. De Wolf, M. Verherstraeten, L. O. de Beeck, D. Matthys, C. Vrints, and R. Rooman, "Glycemia and corrected qt interval prolongation in young type 1 diabetic patients what is the relation?" *Diabetes Care*, vol. 29, no. 2, pp. 427–429, 2006.
- [19] J. Taubel, U. Lorch, G. Ferber, J. Singh, V. N. Batchvarov, I. Savelieva, and A. J. Camm, "Insulin at normal physiological levels does not prolong qt interval in thorough qt studies performed in healthy volunteers," *British journal of clinical pharmacology*, vol. 75, no. 2, pp. 392–403, 2013.
- [20] J. P. Singh, M. G. Larson, C. J. O'Donnell, P. F. Wilson, H. Tsuji, D. M. Lloyd-Jones, and D. Levy, "Association of hyperglycemia with reduced heart rate variability (the framingham heart study)," *The American journal of cardiology*, vol. 86, no. 3, pp. 309–312, 2000.
- [21] M. Schönauer, A. Thomas, S. Morbach, J. Niebauer, U. Schönauer, and H. Thiele, "Cardiac autonomic diabetic neuropathy," *Diabetes and Vascular Disease Research*, vol. 5, no. 4, pp. 336–344, 2008.
- [22] E. Monte-Moreno, "Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques," *Artificial intelligence in medicine*, vol. 53, no. 2, pp. 127–138, 2011.
- [23] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning*. Springer, 2009, vol. 2.
- [24] C. M. Bishop, *Neural networks for pattern recognition*. Clarendon press Oxford, 1995.
- [25] D. J. Curcie and W. Craelius, "Recognition of individual heart rate patterns with cepstral vectors," *Biological cybernetics*, vol. 77, no. 2, pp. 103–109, 1997.
- [26] D. Curcie and W. Craelius, "Classification of heart rate variability patterns in diabetics using cepstral analysis," May 21 2002, uS Patent 6,390,986.
- [27] A. Oppenheim and R. Schaffer, *Discrete-Time Signal Processing*, ser. Pearson Custom Library. Pearson Education, Limited, 2013.
- [28] M. G. Rahim, B.-H. Juang, W. Chou, and E. Buhkrke, "Signal conditioning techniques for robust speech recognition," *Signal Processing Letters, IEEE*, vol. 3, no. 4, pp. 107–109, 1996.
- [29] R. W. Hamming, *Digital filters*. Courier Corporation, 1989.
- [30] V. J. Carey, E. E. Walters, G. A. Colditz, C. G. Solomon, W. C. Willet, B. A. Rosner, F. E. Speizer, and J. E. Manson, "Body fat distribution and risk of non-insulin-dependent diabetes mellitus in women the nurses' health study," *American journal of epidemiology*, vol. 145, no. 7, pp. 614–619, 1997.
- [31] P. Deurenberg, J. A. Weststrate, and J. C. Seidell, "Body mass index as a measure of body fatness: age-and sex-specific prediction formulas," *British journal of nutrition*, vol. 65, no. 02, pp. 105–114, 1991.
- [32] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [33] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [34] C. D.-S. *et al.*, "Vigilancia de la gripe en españa. temporada 2012-2013," <http://revista.isciii.es/bes/article/viewFile/840/974>, accessed: 2015-03-01.
- [35] D. C. Montgomery, *Design and analysis of experiments*. John Wiley & Sons, 2008.
- [36] F. J. Gomez-Perez, C. A. Aguilar-Salinas, P. Almeda-Valdes, D. Cuevas-Ramos, I. L. Garber, and J. A. Rull, "Hba1c for the diagnosis of diabetes mellitus in a developing country. a position article," *Archives of medical research*, vol. 41, no. 4, pp. 302–308, 2010.
- [37] E. Kilpatrick, "Haemoglobin a1c in the diagnosis and monitoring of diabetes mellitus," *Journal of clinical pathology*, vol. 61, no. 9, pp. 977–982, 2008.
- [38] P. R. Kumar, A. Bhansali, M. Ravikiran, S. Bhansali, P. Dutta, J. Thakur, N. Sachdeva, S. K. Bhadada, and R. Walia, "Utility of glycated hemoglobin in diagnosing type 2 diabetes mellitus: a community-based study," *The Journal of Clinical Endocrinology & Metabolism*, vol. 95, no. 6, pp. 2832–2835, 2010.
- [39] A. L. Peters, M. B. Davidson, D. L. Schriger, and V. Hasselblad, "A clinical approach for the diagnosis of diabetes mellitus: an analysis using glycosylated hemoglobin levels," *Jama*, vol. 276, no. 15, pp. 1246–1252, 1996.
- [40] W. Azim, M. Omaid, M. A. Khan, N. Shaheen, and S. Azim, "Correlation between glycated haemoglobin and random plasma glucose levels for the screening of diabetes mellitus," *International Journal of Pathology*, vol. 8, no. 2, pp. 59–62, 2010.
- [41] C. for Disease Control, Prevention *et al.*, "National diabetes statistics report: estimates of diabetes and its burden in the united states, 2014," *Atlanta, GA: US Department of Health and Human Services*, 2014.

- [42] T. F. of the European Society of Cardiology, T. F. of the European Society of Cardiology *et al.*, “the north american society of pacing and electrophysiology. heart rate variability: standards of measurement, physiological interpretation and clinical use,” *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.
- [43] C. M. van Ravenswaaij-Arts, L. A. Kollee, J. C. Hopman, G. B. Stoeltinga, and H. P. van Geijn, “Heart rate variability,” *Annals of internal medicine*, vol. 118, no. 6, pp. 436–447, 1993.

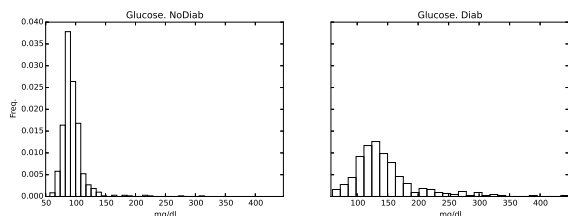


Fig. 1. Histogram of glucose. Left non-diabetics (mean 93.53 mg/dl, iqr 17.3), right diabetics (mean 144.88 mg/dl, iqr 6.31)

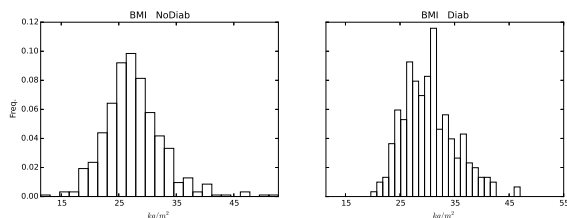


Fig. 2. Histogram of BMI. Left non-diabetics (mean 27.60 kg/m<sup>2</sup>, iqr 5.8), right diabetics (mean 30.36 kg/m<sup>2</sup>, iqr 6.3)

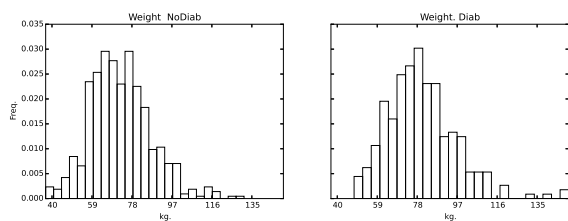


Fig. 3. Histogram of weight. Left non-diabetic (mean 72.7 kg., iqr 18.0), right diabetic (mean 80.1 kg., iqr 19.5)

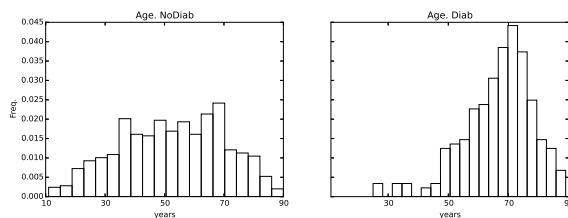


Fig. 4. Histogram of age. Left non-diabetic (mean 52.54 years, iqr 28), right diabetic (mean 66.48 years, iqr 14)

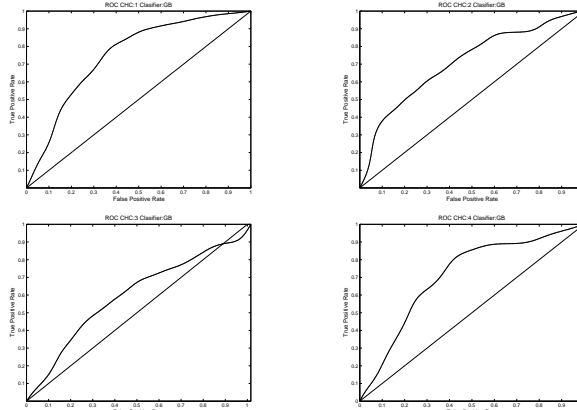


Fig. 5. ROC (false positive rate vs. true positive rate) for the community health centres number 1, 2 3 and 4. Results with gradient boosting.

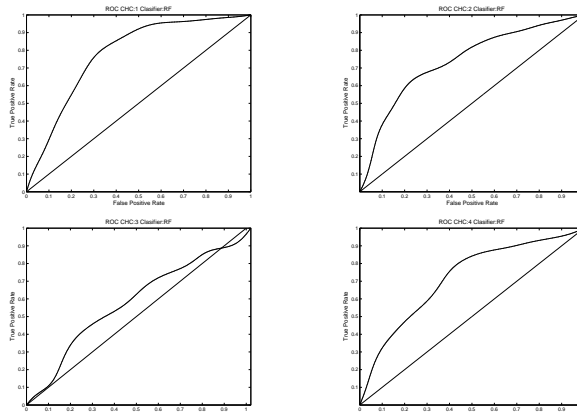


Fig. 6. ROC (false positive rate vs. true positive rate) for the community health centres number 1, 2 3 and 4. Results with random forest

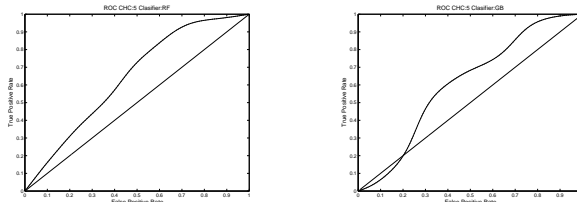


Fig. 7. ROC for the community health center number 5. Left: results with random forest, Right: results with gradient boosting.

Gender	Men	Women
Diabetic	184	156
Non-Diabetic	331	499
Diabetic( % by gender)	35.72%	23.81 %

TABLE I  
DISTRIBUTION BY GENDER

	CHC-1	CHC-2	CHC-3	CHC-4	CHC-5
Diabetic subjects	23.07 %	24.13 %	30.55 %	30.45%	34.45%

TABLE II  
PERCENTAGE OF DIABETICS AT EACH COMMUNITY HEALTH CENTER

Smokers	Total	Diabetic	Non-Diabetic
Percentage	16.18%	11.8%	17.71%

TABLE III  
DISTRIBUTION OF SMOKERS

	CHC-1	CHC-2	CHC-3	CHC-4	CHC-5
Diabetics	29.3 $kg/m^2$	29.6 $kg/m^2$	30.0 $kg/m^2$	30.3 $kg/m^2$	33.1 $kg/m^2$
Non-diabetics	26.5 $kg/m^2$	27.8 $kg/m^2$	28.1 $kg/m^2$	28.3 $kg/m^2$	27.6 $kg/m^2$

TABLE IV  
DISTRIBUTION OF THE BMI ( $kg/m^2$ ) AT EACH COMMUNITY HEALTH CENTER

	CHC-1	CHC-2	CHC-3	CHC-4	CHC-5
ROC RF	78.15 %	71.7 %	59.6 %	71.9%	61.8%
ROC GB	75.4 %	73.9 %	61.1 %	71.8%	64.6%
ROC LDA	56.3 %	57.2 %	53.2 %	55.2%	53.7 %

TABLE V  
ROC AREA AT EACH COMMUNITY HEALTH CENTER, **RF**: MEAN VALUE OF 68.6% AND STD OF 7.27% AND **GB**: MEAN VALUE OF 69.3% AND STD OF 6.2%

SpO2 Measuring Range	0% ~ 100%, (the resolution is 1%).
Accuracy	70% ~ 100%: $\pm 2\%$ ,Below 70% unspecified
PR Measuring Range	30bpm ~ 250bpm, (the resolution is 1bpm)
Accuracy	$\pm 2bpm$ or $\pm 2\%$
Measurement Performance in Weak Filling Condition:	SpO2 and pulse rate can be shown correctly when pulse-filling ratio is 0.4%. SpO2 error is $\pm 4\%$ , pulse rate error is $\pm 2$ bpm or $\pm 2\%$
Sampling rate	60 samples/second

TABLE VI  
FEATURES OF THE CONTEC CMS50D+ PULSE OXIMETER