A non-invasive approach for the diagnosis of Type 2 diabetes using HRV parameters

P. Sankar*
Department of ECE,
Jerusalem College of Engineering,
Pallikaranai, Chennai 600100, India
Email: sankarp1@gmail.com
*Corresponding author

Marykutty Cyriac
Department of Biomedical Engineering,
Jerusalem College of Engineering,
Pallikaranai, Chennai 600100, India
Email: marycyriac123@gmail.com

Abstract: As per the records of World Health Organisation, diabetes is currently one of the major diseases faced by the world community. This necessitates the introduction of screening tools for diabetes. In this paper, a non-invasive approach is proposed to diagnose the presence of Type 2 diabetes by analysing the relationship between the Heart Rate Variability (HRV) parameters and the arterial blood glucose changes. The HRV analysis is performed using non-linear methods such as Detrended Fluctuation Analysis (DFA) and Poincare plot. A few parameters derived from these non-linear methods are used to introduce two metrics named as Standard Deviation Ratio (SDR) and alpha-ratio. These two metrics are given as input to a machine learning classifier to categorise the subjects as diabetic or non-diabetic. The accuracy analysis of the classification results shows that 94.7% of the subjects are categorised correctly. Therefore, the proposed metrics can be considered as non-invasive screening tools in predicting the presence of Type 2 diabetes.

Keywords: alpha-ratio; detrended fluctuation analysis; heart rate variability; Poincare plot; standard deviation ratio.


Biographical notes: P. Sankar received his BSc degree in Mathematics from University of Kerala in 1986, BE in Electronics and Telecommunication from the University of Poona, India, in 1994 and ME in Communication Systems from Madurai Kamaraj University in 1997. He received his PhD degree in Information and Communication from the Anna University, Chennai, India, in 2013. He is currently working as a Professor in the Department of Electronics and Communication Engineering in Jerusalem College of Engineering, Chennai, India. He is a member of the IEEE and life member of the ISTE. His areas of interests include multimedia information systems, image and video processing and computer networks.
Marykutty Cyriac completed her PhD from Anna University, Chennai, India, in 2014, ME in Applied Electronics from PSG College of Technology, Coimbatore, Tamil Nadu, in 2004 and BTech in ECE from TKM College of Engineering, Kerala, in 1985. She is currently working as a Professor in the Department of Biomedical Engineering, Jerusalem College of Engineering, Chennai. Previously, she worked with the computer production unit of Sanmar Group, Chennai, for 9 years. She has published papers in various international journals and presented papers in international conferences. Her research interests are image compression, pattern recognition and multimedia.

1 Introduction

The statistical data released by International Diabetes Federation (IDF) suggest that India has more number of people affected by diabetes than any other countries of the world. It is also predicted that, by the end of year 2030, more than 100 million Indians are anticipated to be suffering from this disease. Indians are more prone to have diabetic condition due to the genetic factors, diet habits and the ongoing migration of people from rural countries to urban areas. In this context, it is absolutely necessary to introduce diabetes screening tools that help in an early detection of diabetics.

Diabetic Autonomic Neuropathy (DAN) is one of the common and chronic complications of the Diabetes Mellitus (DM) condition connected with higher mortality and morbidity in symptomatic patients. However, symptoms of DAN generally do not appear immediately on the beginning of diabetes. Heart Rate Variability (HRV) allows early detection of autonomic dysfunction as it can detect changes before the clinical signs are evident and thereby enabling earlier treatment. The HRV provides information about the working of the nervous control on the heart rate. The clinical significance of HRV by observing changes in the RR intervals was reported very early in the literature (Hon and Lee, 1963). The normal HRV is due to the ANS regulation of the heart and the circulatory system of the body (Saul, 1990). Similarly, cardio-deceleration is due to a low sympathetic activity or a high parasympathetic activity (Acharya et al., 2006). The characteristic of HRV for various time intervals has been studied and it is observed that it differs significantly under different heart rate conditions (Lin et al., 2006). For instance, when HRV was characterised for different age groups using sample entropy, it was found that the sample entropy decreases in higher age groups, which is an indication of loss of HRV with the progression in age (Sunkaria et al., 2012).

Previous studies have demonstrated that it is possible to identify diabetic people with cardiac autonomic neuropathy through HRV analysis (Javoroka et al., 2005; Khandoker et al., 2009). This information is used by Roy and Ghatak (2013) to assess changes in the HRV in diabetic subjects. They concluded that Poincare plot and Detrended Fluctuation Analysis (DFA) are different in diabetic and non-diabetic subjects. Subjects from the same age group were considered in their study so that the variation in distribution would reveal changes in cardiovascular regulation occurring from the diabetic condition only. In a recent investigation, HRV parameters were suggested as a biomarker for evaluating the progress of DM. However, this experiment was conducted on mice only (Arroyo-Carmona et al., 2016).
HRV analysis methods are basically categorised into linear and non-linear methods. Many of the earlier works use linear analysis, which is further classified into time and frequency domain methods. HRV parameters of both domains are investigated previously for a group of diabetic patients with and without neuropathy and also for the healthy persons. The analysis did not show any considerable variations in the parameters between the healthy and diabetic patients without neuropathy. However, frequency domain parameters could distinguish diabetic patients with and without neuropathy (Guerrero et al., 1999). Another study was performed to examine the relationship between HRV of the ECG and arterial blood glucose changes in a female subject with Type 1 diabetes Mellitus disease during normoglycaemic and mild hyperglycaemic conditions. Power spectral analysis of HRV signals is performed in order to differentiate between healthy and diabetic subjects. The ratio of low frequency component to that of the high frequency component decreased when the blood glucose levels are changed from normoglycaemic to hypoglycaemic condition. This pilot study brought out only the brief relationship between the HRV spectral parameters and sugar levels during normoglycaemic and mildly hyperglycaemic conditions (Amanipour et al., 2012). In another study, the HRV in healthy individuals and diabetic patients is compared to find that a negative correlation exists between HbA1c values and the HRV parameters (Mirza and Laxmi, 2012). Schroeder, et al. (2005) conducted a study on diabetic and non-diabetic subjects to examine the relationship between diabetes and HRV. They concluded that decrease in autonomic function is present in the early stages of diabetes, and the progression of diabetes leads to declined autonomic function. Linear parameters of HRV were used in their analysis. Fast reduction of the standard deviation of the normal RR Intervals (SDNN) was observed in the diabetic subjects. In another study conducted on a large population after adjusting covariates such as age, heart rate, sex, systolic and diastolic blood pressures, body mass index, etc., LF/HF ratio and LF power were lesser in diabetic subjects than in non-diabetic subjects. They concluded that HRV and plasma glucose levels are inversely associated.

Compared to the linear methods, the interest in non-linear methods has increased in the recent years due to the observation that the HRV is non-linear. Therefore, it has been speculated that the non-linear analysis of the HRV might elicit useful information for the physiological explanation of HRV and for the assessment of the risk involved. The popular non-linear methods such as DFA, autocorrelation and Poincare plot have been used for the HRV analysis previously (Roy and Ghatak, 2013; Voss et al., 2009, Khandoker et al., 2008; Taouli and Bereksi-Reguig, 2012; Yeh et al., 2010; Kumar et al., 2013). A different approach used is based on the Principal Component Analysis (PCA) (Egorova et al., 2014). The results of HRV analysis based on these methods have been used to diagnose various types of medical conditions (Deepak et al., 2013; Dong et al., 2014; Ulanovsky et al., 2014; Tonello et al., 2014; Chu Duc et al., 2013). Many of the previous investigations have established the fact that there is a certain relation between diabetic conditions of a subject and the HRV parameters. Even for the non-diabetic young adults, there is an impact on the cardiac autonomic function if their parents have a history of Type 2 diabetes (Antino et al., 2014). A recent research compared HbA1c and FPG as a pre-diabetes screening tool (Goliska, 2012) for Indian population. However, this method is invasive. A recent non-invasive diabetes detection method used iris images with detection accuracy between 72% and 75% (Chaskar and Sutaone, 2012).
In this paper, a non-invasive screening tool is proposed for the early detection of diabetes by analysing HRV parameters and formulating two metrics. Non-linear HRV parameters are used to derive the metrics, and machine learning techniques are subsequently employed to calculate the accuracy of detection.

This paper is organised as follows: Section 2 explains the materials and methods. Section 3 is on the formulation of the proposed metrics and classification. Performance analysis is given in Section 4. Finally, Section 5 concludes the paper.

2 Materials and methods

The ECG signals of subjects are first obtained using the NI DAQ hardware. The subjects within the age group of 30–70 years are considered in the analysis. The ECG samples are taken from 75 subjects, out of which 25 are non-diabetic and the rest are diabetic without any other medical abnormalities such as neuropathy. Most of the subjects are institutional colleagues of the authors. LabVIEW software with Biomedical Toolkit is used for processing and analysing the ECG signals. Two popular non-linear methods such as Poincare plot and DFA are employed for HRV analysis. These two methods are briefly explained in the next subsections.

2.1 Poincare plot

Poincare plot is a type of non-linear analysis that is used to quantify the self-similarity. The Poincare plots are based on the idea that each RR interval is influenced by the previous one. A Poincare plot of RR intervals is a scatter plot of points RRj + 1 vs. RRj where RRj is the time between two adjacent R peaks and RRj + 1 is the time between next two adjacent peaks. The set of points generally appear in the form of an ellipse. Hence, an ellipse is fitted to the plotted data with the long axis of the ellipse along the line of identity. The plotted data generate two parameters: (i) the standard deviation of instantaneous beat-to-beat interval variation (SD1) or the short-term HRV, and (ii) the continuous long-term R to R interval variation (SD2) or the long-term HRV.

These points are usually characterised by the breadth across the line of identity (SD1) or the minor axis and the length along the line of identity (SD2) or the major axis of the ellipse. They are called the Poincare plot descriptors. SD1 is related to parasympathetic activity, while SD2 reflects total variability.

SD1 and SD2 are generally defined in terms of standard deviation of RR interval (SDRR), and standard deviation of the successive difference of RR interval (SDSD), which is expressed as given in equations (1) and (2) (Chandan et al., 2009).

\[
SD1^2 = \frac{1}{2}SDSD^2 \\
= \gamma_{xx}(0) - \gamma_{xx}(1) \\
\]

\[
SD2^2 = 2SDRR^2 - \frac{1}{2}SDSD^2 \\
= \gamma_{xx}(0) + \gamma_{xx}(1) - 2RR^2 \\
\]
where:
\[ \gamma_{RR}(0) \text{ and } \gamma_{RR}(1) \] – autocorrelation function for lag-0 and lag-1 RR interval.
\[ \bar{RR} \] – mean of RR intervals.

These equations are derived for Poincare plot with unit time delay.

2.2 Detrended fluctuation analysis (DFA)

DFA tries to quantify the self-similar properties of non-stationary time series (Goit et al., 2015). The root-mean-square fluctuation of an integrated and detrended time series is measured at different scales and plotted against the size of the scale. The scaling parameter alpha (\( \alpha \)) describes the autocorrelation properties of the signal. When \( \alpha < 0.5 \), the signal is anti-correlated, while \( \alpha > 0.5 \) shows positive autocorrelation. If \( \alpha = 0.5 \), then the signal is uncorrelated. When \( \alpha > 1 \), correlations exist but cease to follow the power law.

Often, two distinct linear regions on the log-log plot are used to describe the Scaling parameter. Depending upon the number of heartbeats, we have short range scaling exponent alpha \( \alpha_1 \), and long range scaling component alpha \( \alpha_2 \). Short-term scaling, \( \alpha_1 \), is measured over the range of 4–16 heartbeats and the long-term scaling, \( \alpha_2 \), is measured over the range of 16–64 heartbeats. \( \alpha_1 \) and \( \alpha_2 \) are used to enumerate the fractal property of the temporal series of the RRi. In healthy conditions, \( \alpha_1 \) value is higher than the \( \alpha_2 \) value (Wladimir et al., 2015).

2.2.1 DFA algorithm

Consider a 1-D signal \( S(i) \) where \( i = 1: N \). Let \( S_{avg} \) be the mean value of the signal and \( N \) is the number of samples. The integrated signal can be computed as:

\[
y(k) = \sum_{i=1}^{k} (S(i) - S_{avg})
\]

The signal is divided into segments of length \( n \) and each segment is linearly approximated using least squares fit to find the linear approximation \( y_n \). This will represent the trend for a given segment. The average fluctuation \( F(n) \) of the \( n \)th segment is given by:

\[
F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y(k) - y_n(k))^2}
\]

The average fluctuation is calculated for all the segments. In order to analyse the relation between \( n \) and \( F(n) \), a double logarithmic graph is created between log \( n \) and log \( F(n) \). This graph is referred as DFA. The slope of the line \( F(n) \) determines the scaling component and the linear dependence indicates the presence of self-fluctuations (Rajput, 2015). The scaling parameter alpha (\( \alpha \)) describes the autocorrelation properties of the signal. When \( \alpha < 0.5 \), the signal is anti-correlated, while \( \alpha > 0.5 \) shows positive autocorrelation. If \( \alpha = 0.5 \), then the signal is uncorrelated. When \( \alpha > 1 \), correlations exist but cease to follow the power law. Depending upon the value of \( n \), we have short range scaling exponent (fast parameter) \( \alpha_1 \) and long range scaling component (slow parameter) \( \alpha_2 \), which are used in this analysis.
3 Formulation of metrics and classification

The non-linear analysis has been carried out on the ECG signals acquired from diabetic and non-diabetic subjects of various age groups. Figure 1 (a) and (b) shows the DFA plots for the non-diabetic and diabetic subjects, respectively. It is evident that the DFA plot varies according to the diabetic condition. The non-linear parameters derived from this plot are used for formulating the alpha-ratio.

Figure 1  DFA plots for (a) non-diabetic subject and (b) diabetic subject (see online version for colours)

In the Poincare plot analysis, plot scatter increases with the lag number, yielding large values of width (SD1) and length (SD2). The incremental increase in the width of Poincare plot is smaller in the diabetes group than in the non-diabetic group. Marked differences in HRV pattern is observed between diabetic and non-diabetic subjects. Figure 2 (a) and (b) shows the Poincare plots of the non-diabetic and diabetic subjects, respectively.
A non-invasive approach for the diagnosis of Type 2 diabetes

Figure 2  Poincare plots of (a) non-diabetic subject and (b) diabetic subject

Using DFA and Poincare plots, non-linear parameters such as SD1, SD2, $\alpha_1$ and $\alpha_2$ are calculated for the ECG signals. These parameters obtained for non-diabetic group are shown in Table 1, while Table 2 shows the values for the non-diabetic group.

Table 1  Variation of parameters for non-diabetic subjects of different age groups

<table>
<thead>
<tr>
<th>Parameters</th>
<th>30–40</th>
<th>40–50</th>
<th>50–60</th>
<th>60–70</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD1 (ms)</td>
<td>200</td>
<td>150</td>
<td>120</td>
<td>280</td>
</tr>
<tr>
<td>SD2 (ms)</td>
<td>140</td>
<td>110</td>
<td>97</td>
<td>250</td>
</tr>
<tr>
<td>SDR</td>
<td>1.42</td>
<td>1.36</td>
<td>1.23</td>
<td>1.12</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0</td>
<td>0.6842</td>
<td>0.6106</td>
<td>0.6111</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-9.7E-18</td>
<td>0.28</td>
<td>0.49</td>
<td>0.42</td>
</tr>
<tr>
<td>Alpha-ratio</td>
<td>0</td>
<td>2.44</td>
<td>1.25</td>
<td>1.46</td>
</tr>
</tbody>
</table>
Table 2  Variation of parameters for diabetic subjects of different age groups

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Age in years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30–40</td>
</tr>
<tr>
<td>SD1 (ms)</td>
<td>53</td>
</tr>
<tr>
<td>SD2 (ms)</td>
<td>300</td>
</tr>
<tr>
<td>SDR</td>
<td>0.17</td>
</tr>
<tr>
<td>α1</td>
<td>-2.219E-17</td>
</tr>
<tr>
<td>α2</td>
<td>1.9E-17</td>
</tr>
<tr>
<td>Alpha-ratio</td>
<td>-1.17</td>
</tr>
</tbody>
</table>

The analysis based on SD1 and SD2 metrics between the diabetic group and the non-diabetic group shows significant variation in their values. The values of SD1 are higher for the non-diabetic group, whereas SD2 values are higher for the diabetic group. The lower SD1 for diabetic subjects indicates that parasympathetic regulation is weakened by the disease, presumably by peripheral neuropathy, whereas the large SD2 for diabetic subjects indicates increased long-term variability because of compensatory sympathetic input. The ratio of SD1 and SD2 is calculated and introduced as a metric named as Standard Deviation Ratio (SDR), which is given by:

\[
SDR = \frac{SD_1}{SD_2}
\]  

(5)

Figure 3 shows the SDR values for non-diabetic and diabetic subjects of various age groups. The SDR is above one for all non-diabetic subjects, while it is below one for the diabetic subjects. By setting a threshold value for SDR as one, it is now possible to differentiate between non-diabetic and diabetic subjects. Another observation on SDR is the gradual reduction in the SDR value with the increase in age. This is an expected result because the blood glucose level slowly rises with the age.

The α1 and α2 metrics also exhibit a pattern. In non-diabetic subjects, α1 is larger than α2, whereas α2 is larger than α1 for the diabetic group. This observation indicates that the long-term variations of HRV are more in the diabetic group than the normal subjects. The ratio of α1 and α2 is called as the metric alpha-ratio, which is given by:

\[
\text{Alpha-ratio} = \frac{\alpha_1}{\alpha_2}
\]  

(6)

Figure 4 shows the alpha-ratio for diabetic and non-diabetic subjects. It is observed that the alpha-ratio is always higher for a non-diabetic subject when compared to the diabetic subject irrespective of their age. Therefore, alpha-ratio can be used as a validation metric to confirm the diagnostic observations made using SDR.
A non-invasive approach for the diagnosis of Type 2 diabetes

Figure 3  Variation of SDR values with age for diabetic and non-diabetic subjects

Figure 4  Variation of alpha-ratio for diabetic and non-diabetic subjects
4 Performance analysis

The effectiveness of SDR and alpha-ratio metrics in the diagnosis of diabetic condition is validated by using an SVM classifier which uses a linear kernel function. SVM is very effective in cases where the number of dimensions is greater than the number of samples and is memory efficient during computations. It provides high true positive rate when used for diabetes data sets (Choubey and Paul, 2016). The key motivation for using SVM classifier is that heart rate signals are of low signal to noise ratio. As compared with neural network-based classifiers, SVM is robust to noisy ECG data. Previous research has already proved the effectiveness of using SVM in HRV analysis (Kampouraki et al., 2009).

The data set consists of features extracted from 25 diabetic (P) and 50 non-diabetic (N) subjects. When the subject is diabetic and classified as diabetic, it is taken as True Positive (TP) condition and when the diabetic subject is classified as non-diabetic, it is categorised as False Negative (FN). Similarly, when a non-diabetic subject is categorised as non-diabetic, it is True Negative (TN) condition and when the non-diabetic is categorised as diabetic it is False Positive (FP). The confusion matrix obtained for this data is shown in Table 3.

<table>
<thead>
<tr>
<th>Nature of subjects</th>
<th>Parameters</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Positives (TP)</td>
<td>False Positives (FP)</td>
<td>True Negative (TN)</td>
<td>False Negative (FN)</td>
</tr>
<tr>
<td>Diabetic</td>
<td>23</td>
<td>2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Non-diabetic</td>
<td>–</td>
<td>–</td>
<td>48</td>
<td>2</td>
</tr>
</tbody>
</table>

The values in the confusion matrix are used to calculate the accuracy, which is defined as:

\[
\text{Accuracy} = \frac{TP + TN}{P + N}
\]

where \( P + N \) is the total number of subjects considered for the analysis. An accuracy of 94.7% is obtained using this method. Therefore, based on the analysis results, it is concluded that the SDR and alpha-ratio are useful metrics for determining the presence of Type 2 diabetes in subjects without going for an invasive blood analysis.

One of the limitations of this study is the subject population. The diabetic subjects involved in this study are of at various stages of diabetes and are on medications. A small population is used in the study to confirm the results. A larger population who are at the beginning of the diabetes is required to validate the preliminary results reported through theoretic and analytic methods.

5 Conclusion

In this paper, an intelligent non-invasive approach based on HRV analysis is proposed to diagnose the presence of diabetes. The HRV analysis uses the non-linear methods such as Poincare and the DFA plots. The parameters extracted from these plots are used to
A non-invasive approach for the diagnosis of Type 2 diabetes

formulate two new metrics, namely the SDR and the alpha-ratio. These metrics are used to train an SVM classifier and an accuracy of 94.7% is obtained. Therefore, it is concluded that the SDR and the alpha-ratio are useful metrics in the preliminary screening and diagnosis of diabetes. By observing the SDR and alpha-ratio value alone, subjects prone to diabetics can be identified and remedial measures can be taken for preventing the onset of diabetes.

References


A non-invasive approach for the diagnosis of Type 2 diabetes
