A large section of the world’s population is affected by diabetes mellitus (DM), commonly referred to as “diabetes.” Every year, the number of cases of DM is increasing. Diabetes has a strong genetic basis, hence it is very difficult to cure, but can be controlled with medications to prevent subsequent organ damage. Therefore, early diagnosis of diabetes is very important. In this paper, we examine how diabetes affects cardiac health, which is reflected through heart rate variability (HRV), as observed in electrocardiography (ECG) signals. Such signals provide clues for both the presence and severity of diabetes as well as diabetes-induced cardiac impairments. Heart rate (HR) is a non-linear and non-stationary signal. Thus, extracting useful information from HRV signals is a difficult task. We review several sophisticated signal
 processing and information extraction methods in order to establish measurable relationships between the presence and the extent of diabetes as well as the changes in the HRV signals. Furthermore, we discuss a typical range of values for several statistical, geometric, time domain, frequency domain, time–frequency, and non-linear features for HR signals from 15 normal and 15 diabetic subjects. We found that non-linear analysis is the most suitable approach to capture and analyze the subtle changes in HRV signals caused by diabetes.

**Keywords:** Diabetes; heart rate variability; electrocardiography; higher order spectra; autonomic nervous system; Poincare plot; correlation dimension; sample entropy; approximate entropy; recurrence plot.

1. Introduction

Diabetes mellitus (DM), commonly referred to as “diabetes,” is a disorder which affects glucose processing of the metabolism. A consequent condition is called hyperglycemia, where the blood glucose level is abnormally high. The pathophysiology arises from the fact that (1) the pancreas is unable to produce enough insulin (known as Type 1 or insulin-dependent diabetes) and (2) the target cells are not able to respond to the produced insulin (known as Type 2 or non-insulin-dependent diabetes). Insulin is the key hormone that regulates blood sugar. Diabetes is difficult to cure because it has a strong genetic basis and therefore, the disease should be well managed to prevent consequent organ damage. Managing this disease mandates its early detection and medicinal treatment.

There were approximately 171 million diabetic people worldwide in 2000 and it is estimated that this number is likely to double by the end of 2030. The National Health and Nutrition Examination Survey found that in a population of people aged 65 years and above, 18–20% have diabetes and 40% have either diabetes or its early form of impaired glucose tolerance. If untreated, diabetes can lead to serious consequences. According to statistical data from the World Health Organization (WHO), the estimated mortality from diabetes worldwide for the year 2002 is 987,000 deaths, i.e., 1.7% of total world mortality. However, the actual mortality due to diabetes is much higher than the statistical reports, which usually retrieve data based on death certification. Often, diabetic patients die from diabetes-induced complications, mostly cardiovascular and renal diseases and not from hyperglycemia itself.

WHO statistics also underline the fact that a person with diabetes has a significantly increased risk of premature death due to cardiovascular disease. Lack of awareness and insufficient management of diabetes usually lead to such complications. The cause of the increased diabetes rate can be traced back to the changing lifestyles of the modern man, which has evolved to be much more sedentary and thus, has resulted in physical inactivity, weight increase, obesity, and related complications.

The global spread of diabetes has led the WHO to open a specific programme called WHO Diabetes Programme. The mission of this programme is to prevent diabetes whenever and wherever possible. If prevention cannot be achieved, the

complications due to diabetes have to be minimized by instituting medicinal treatment. The WHO Diabetes Programme sets standards to achieve the goal.

There are three main types of diabetes: Type 1 diabetes, Type 2 diabetes, and gestational diabetes. Type 2 diabetes accounts for about 90% of diabetes cases worldwide. In this paper, the term “diabetes” refers exclusively to Type 2 diabetes.

Symptoms of diabetes include excessive urination (polyuria), excessive thirst (polydipsia), increased hunger, weight loss, fatigue, and changes in vision. However, the key issue with its diagnosis is that the symptoms are not overt at the onset of the disease, but are likely to be visible once the disease worsens, leading to complications related to organ damage. Therefore, the disease may only be diagnosed several years after its actual onset and when some degree of damage has already been done.

The National Institute of Diabetes and Digestive and Kidney Diseases has summarized the consequences of diabetes as given below:

1. Diabetes causes an increased risk of heart disease and stroke. Almost half of the people affected by diabetes die of cardiovascular disease.
2. Diabetes can lead to neuropathy in the feet, impairing vascular supply leading to the severest extent, resulting in consequences such as limb amputation.
3. Nerve disorders due to diabetes, affect vital organs such as the eye. This specific complication is called diabetic retinopathy. It can cause blindness which results from long-term damage to retinal blood vessels. Statistics indicate the probability of occurrence of severe visual complications to a certain percentage of diabetes-affected people. Approximately 2% of diabetic people develop blindness and 10% develop severe visual impairment.
4. Ten percent to 20% of diabetic people die due to end-stage renal diseases leading to premature kidney failure.
5. Diabetes damages the peripheral nerves, which is known as diabetic neuropathy.

In this paper, as already mentioned, we focus on the effects of Type 2 diabetes on cardiac health, especially that of the electrophysiological properties of the heart. Diabetes causes complications in the autonomic nervous system (ANS). The symptoms displayed by the ANS are due to these complications and may serve as a good indicator of the extent of diabetes. For example, the activities of the ANS can be assessed effectively by analyzing heart rate variability (HRV), which displays the electrophysiological properties of heart. We discuss several methods and techniques where HRV measures are used to diagnose diabetes and its extent by detecting the deterioration in the cardiac electrophysiological property caused by diabetes.

1.1. Diabetes and HRV

High blood sugar levels, due to diabetes, lead to increased deposits of fatty material (dyslipidemia) on the inner blood vessel wall. It results in clogging (i.e., atherosclerosis) and hardening (i.e., arteriosclerosis) of blood vessels. This affects blood flow as the disease progresses. Diabetes also leads to in poor glycemic control,
causing precapillary damage. This will affect the normal functioning of endothelial cells and block the normal pathways of nitric oxide (NO) synthesis. Thus, the availability of NO, which is required for vasodilation, is grossly reduced. Diabetes causes the vessels to be in a constricted state, leading to poor circulation due to the impairments of microvascular endothelial cell metabolism. Thus, microvascular damage causes a retardation in local reflexes, NO production, and nerve blood flow, resulting in reduced HRV. Altogether, this is called diabetic vasculopathy.

Early works of Stamler et al. and Coutinho et al. clearly established the correlation between fasting plasma glucose and cardiovascular complications. Twenty percent to 25% of the patients with acute coronary disorder are diabetic patients as well. Viktor et al. explained how hyperglycemia causes cardiovascular dysfunction independent of other factors like arterial hypertension, dyslipidemia, and obesity. Hyperglycemia produces free oxygen radicals (known as superoxides) which inactivate NO (derived from the endothelium). Hyperglycemia also activates protein kinase C, which promotes vasoconstrictive prostanoid production. In addition, hyperglycemia causes endothelial dysfunction and increases platelet activity and their aggregability. The consequences are increased adhesion of platelets and monocytes, leukocytes in the endothelium, increased fibrinogen and coagulability of blood, decreased fibrinolitic activity, etc. Diabetic patients are found to have significantly lower values of left ventricular ejection fraction (LVEF) and higher values of left ventricular end-systolic diameter (LVESd) compared to normal people. It is important to note that the behavior of the diabetic heart depends on structural as well as functional changes, which influence its electrophysiological properties that can be captured by studying electrocardiography (ECG) signals.

ECG is a biosignal which represents the sum total of millions of cardiac depolarization potentials. It indicates the heart’s electrical activity during its structural and physiological states. ECG reflects the how the ANS regulates the rhythm of the heart. Abnormalities in specific segments of the ECG signal, such as the ST, QRS, and P-wave segments, indicate heart diseases. Hence, ECG signals can be used to diagnose a wide range of heart diseases. ECG is measured by placing sensors at the limb extremities and the chest of a person.

Diabetes causes nerve disorders, also called diabetic neuropathy. Diabetic neuropathy is mainly classified as peripheral, autonomic, proximal, and focal neuropathies. Autonomic neuropathy affects the heart and blood vessels, digestive system, urinary tract, sex organs, sweat glands, and lungs. The type of diabetic neuropathy that affects the nerves, which control the heart functions and regulate the blood pressure, is called cardiovascular autonomic neuropathy (CAN). An important risk factor for developing CAN is poor glycemic control due to diabetes. Diabetes-induced CAN leads to ECG alterations, such as sinus tachycardia, long QTc, QT dispersion, changes in HRV, ST–T changes, and left ventricular hypertrophy (LVH).

Stern et al. found indications of LVH in the ECG of a diabetic person showing no symptoms of CAN. The changes in the ECG indicated that the diabetic patient had a high risk for future cardiovascular disease. The patient was kept under...
measures to restrict his diet and control cardiac risk factors. Subsequently, a six-year follow-up study was conducted. The study showed that the ECG did not change further. Okin et al. observed appreciable QTc prolongation and ST segment depression in diabetic people. These symptoms are predictors of mortality. Sawicki et al. on the other hand, found that QT dispersion is the most important mortality predictor in diabetes.

ECG represents cardiac health, but there are a few inherent issues in predicting cardiac states manually/clinically. For example, subtle changes in the waveform are difficult to be perceived by the human eye. Another difficulty is that the morphologies of ECG waveforms show large variations, not only in different patients and groups, but also within the same patient. Moreover, the biosignals are non-stationary and probably even non-linear in nature. Due to these reasons, several biosignal analysis techniques do not perform well on ECG signals.

Another option is to analyze HRV signals, which are extracted from ECG signals, instead of directly analyzing ECG signals. The HRV signal is a discrete signal where waveform shape has no meaning. The complete information lies in the value of each discrete HRV data sample. The changes in the ANS, which are due to diabetes-induced hyperglycemia, are well reflected by HRV signals. In general, HRV is the variation of the heartbeat period measured in units of time. The time between two consecutive R-waves (RR interval) on the ECG is used to obtain the heart rate (HR) signal. The interval between adjacent QRS complexes is termed as NN or RR interval. The HR in terms of beats per minute is given by $HR = \frac{60}{t_{RR}}$. Thus, the HR signal is a time-series sequence of non-uniform RR intervals. HRV indicates variations of instantaneous HR. HR variation analysis (instantaneous HR against time) has become a very effective non-invasive tool, which helps to assess the condition of the ANS.

HRV signals are accurate, reliable, and reproducible, yet simple to measure and process. HRV analysis can thus be used as a reliable method for analyzing the cardiac health of a person and the state of the ANS. The parasympathetic nervous system and sympathetic nervous system are the two branches of the ANS. The sinoatrial (SA) node is the natural pacemaker of the heart. Spontaneous cardiac impulses generated by the SA node and the influence caused by the sympathetic and parasympathetic nervous systems on the conducting tissues of the heart together determine the normal HR. In the absence of any sympathetic or parasympathetic input, the SA node operates at its intrinsic rate. Normal resting HR is decided by the parasympathetic vagus nerve. Acceleration of the HR occurs by the inhibition of vagal influence and by stimulating the sympathetic nervous system. In other words, when vagal effects are increased, the HR is lower than the intrinsic HR and when sympathetic effects are dominant, the HR is greater than the intrinsic HR. Thus, it is possible to derive the ANS status from HRV signals. HRV gives information about the sympathetic–parasympathetic autonomic balance. The autonomic neural regulation of both the heart and circulatory system determines the normal HRV. The instantaneous HR is strongly dependent on a variety of neural, hormonal, and myocardial factors. Since ECG measurements are non-invasive, the extracted HR
is also a non-invasive method. Therefore, HRV analysis has numerous clinical applications, specifically in the fields of physiology, biochemistry, and pharmacology. The works of Kleiger et al. and Dingfei et al. indicate that non-invasive HRV measurements are easy to perform and have good reproducibility; if used, the measurements will be obtained under standardized conditions. 29,30

1.2. Diagnosis of diabetes by HRV analysis

Autonomic impairments, due to diabetes, are clinically detectable only after many years of the disease onset. However, subtle changes in HRV can occur much earlier and hence, this can be used as a marker to detect diabetes early. Traditional methods used to test autonomic dysfunction can document the presence of these abnormalities only when these impairments cause severe symptoms. In this scenario, the importance of HRV analysis lies in the fact that they can detect these impairments at a very early stage. HRV analysis helps in the early detection of subclinical autonomic impairments in diabetic people. There are tests, like autonomic reflex testing, 31 to assess the condition of the ANS and myocardial scintigraphy with I-meta idobenzyl guanidine (MIBG) to detect autonomic neuropathy in diabetic individuals. 32 Schroeder et al. have found that HRV-based analysis methods can detect cardiac autonomic impairments before traditional cardiovascular autonomic function tests like the Ewing battery. 33 Upon detection, the patient can opt to go for invasive and costly tests like angiography.

Wheeler et al. were the first to report a decreased beat-to-beat variability during deep breathing in the case of diabetic neuropathy. 34 Later, work from Pfeifer et al. confirmed the findings of Wheeler et al. by comparing cardiac autonomic function tests with HRV indices (extracted out of both short, i.e., 5-min and long, i.e., 24-h ECG recordings). 35 They found that diabetic patients who showed negative results for cardiac autonomic function tests had a lower HRV.

The work of Pfeifer et al., Singh et al., and Villareal et al. showed that parasympathetic autonomic activity declines in diabetic people even before clinical symptoms of neuropathy are evident. 35–37 Their results proved that HRV analysis is a strong indicator for diabetes.

2. Data Acquisition and Pre-Processing

2.1. Data acquisition

The ECG of 15 patients (10 males and five females) with diabetes and the ECG of 15 healthy volunteers (eight males and seven females) were recorded. The subjects were in a relaxed supine position for one hour per sitting. The diabetic subjects were in the age group of 50–70 years (58.5 ± 6.42 years) and the duration of diabetes for the patient group was 5–15 years. The normal subjects were in the age group of 40–60 years (50 ± 8.8 years). The ECG was recorded using BIOPACTM equipment and the inbuilt AcqKnowledge software was used to extract the HR. The measurements were done at Kasturba Medical Hospital, Manipal, India. The ethics
committee, comprising senior doctors, has approved the data used for this research. The ECG sampling rate was 500 Hz. The normal and diabetic data were split into 71 datasets, each 1,000 samples in length.

2.2. Pre-processing

The ECG signal noise was eliminated by passing the measurements through a low-pass filter with a 15-Hz cut-off frequency. Then, the baseline wander, present in the ECG signal, was removed by passing the signal through a high-pass filter with a cut-off frequency of 0.3 Hz. A median filter was used for extraction. Subsequently, the 50-Hz band-reject filter was used to remove the power-line interference noise. Finally, the R peaks of the ECG signal were detected. The RR interval is defined as the interval between two successive QRS complexes. HR is the time series

![Typical heart rate (HR) variation plots. (a) Normal and (b) diabetic subject.](image)
of unevenly spaced RR intervals. Figure 1 shows the typical HR signals of a normal and diabetic subject.

3. Methods Used

HRV signal analysis can be done by using several time domain, frequency domain, wavelet, and non-linear parameters. It was shown that variability in the time series changes with the ailments.\textsuperscript{40,41} The following sections briefly explain the techniques used for time domain, frequency domain, and non-linear analyses.

3.1. Time domain analysis

Short as well as long time variation indices can be calculated from the RR intervals. Hence, the mean value of the RR period (interval length) is an important HRV measure. This parameter measures the sum of parasympathetic and sympathetic influences.

A number of parameters can be found from RR intervals. Among them, there are measures of variance like SDNN (also known as SDRR), SENN, and the St. George’s index, where SDNN (unit: ms) is the standard deviation of all RR intervals about the mean RR for the entire recording. SDNN is an indicator of total variability. SENN is the standard error of the mean RR or NN interval. It is a measure of the standard deviation of the sampling distribution of means based on data. The St. George index (unit: ms) is based on the frequency histogram of 24-h ECG recordings. Long-term as well as short-term variabilities of mean RR interval can be assessed from 24-h SDRR and the St. George’s index. These parameters reflect vagal (i.e., parasympathetic) as well as sympathetic modulations within otherwise normal HR signals.

There are measures such as RMSSD and the Edinburgh index that analyze differences between adjacent intervals. RMSSD (unit: ms) is the square root of the mean of sum of the squares of the differences between adjacent normal RR intervals over the entire 24-h ECG recording. RMSSD indicates high-frequency variations in HR; it mainly reflects parasympathetic activity. The Edinburgh index (also called SNN50 counts) indicates the number of times when the differences between adjacent normal RR intervals are greater than 50 ms (NN50); it is calculated over the 24-h ECG recording. PNN50 (unit: %) is the percentage of the above stated differences. The statistical parameters SDNN, SENN, SDSD, RMSSD, NN50 (%), and pNN50\%\textsuperscript{42} can be used as time domain parameters. SDSD is the standard deviation of differences between adjacent NN intervals. RMSSD and NN50 are parameters which reflect short-term HRV, hence they show changes in vagal tones. Therefore, these parameters are known as vagal indices.

The time domain parameters SDNN, SENN, SDSD, RMSSD, and pNN50\% have higher values for abnormalities, like pre-ventricular contraction (PVC) and atrial fibrillation (AF), due to higher RR variation. For heart problems like complete heart block (CHB), ischemic/dilated cardiomyopathy, etc., these parameters show lower values since the RR signals are slowly varying for these abnormalities.
Besides time domain parameters, RR intervals can be represented in a geometrical form and HRV features can be extracted from these geometrical forms. Important geometrical measures are the triangular index and triangular interpolation of NN interval histogram (TINN). The RR triangular index is the integral of sample density distributions of RR intervals divided by the maximum of the density distribution. For this measure, a plot, where the x-axis is the length of the RR interval and the y-axis is the number of each RR interval length, is used. This is a RR interval frequency distribution diagram. The peak of this diagram is chosen as the length of the base of the triangle. TINN thus has the base of the triangle as the baseline width of the RR interval frequency distribution diagram. Here, the RR interval distribution is approximated by a linear function and the baseline width of this approximation triangle is taken as the HRV index measure. The RR triangular index has a high correlation with the standard deviation of all RR intervals. This measure is highly insensitive to artifacts and ectopic beats.

The time domain parameters, which are extracted from diabetic HRV samples, show considerable lower values when compared to normal samples, as shown in Table 1.

The principal disadvantage of time domain measures is that the mean value can be easily affected by artifacts and outliers; therefore, such measures require the elimination of artifacts from the data before analysis.

3.2. Frequency domain analysis

Time domain HRV analysis is relatively simple to conduct, but it has a drawback — sympathetic and parasympathetic influences cannot be clearly separated. Akselrod et al. discussed the typical power spectrum of the HR signal. They postulated that the HR signal has three main frequency regions:

1. the high-frequency (HF) power band (HF: 0.15–0.5 Hz),
2. the low-frequency (LF) power band (LF: 0.04–0.15 Hz), and
3. the very low-frequency (VLF) power band (VLF: 0.0033–0.04 Hz).

Respiratory sinus arrhythmia and cardiac vagal activity are reflected in the HF region, while LF shows baroceptor control mechanisms and the combined

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal (mean ± std)</th>
<th>Diabetes (mean ± std)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean RR (s)</td>
<td>884 ± 108</td>
<td>717 ± 97.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>STD (s)</td>
<td>50.4 ± 20.5</td>
<td>26.7 ± 27.6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>RMSSD (ms)</td>
<td>36.0 ± 16.7</td>
<td>25.8 ± 50.5</td>
<td>0.085</td>
</tr>
<tr>
<td>NN50 (count)</td>
<td>140 ± 166</td>
<td>31.5 ± 102</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>pNN50 (%)</td>
<td>14.5 ± 15.7</td>
<td>3.19 ± 10.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>RR triangular index</td>
<td>9.82 ± 2.89</td>
<td>4.40 ± 1.54</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TINN (ms)</td>
<td>209 ± 60.7</td>
<td>155 ± 123</td>
<td>0.0005</td>
</tr>
</tbody>
</table>
effect of the vagal and sympathetic systems. The VLF power region reflects the thermoregulatory and vascular mechanisms, and rennin—angiotensin systems. All these frequency bands are within the frequency range of respiration.

Spectral analysis of the short-term variability of HR allows a quantitative assessment of neurogenic oscillations that determine the instantaneous HR. HRV frequency domain studies use the fast Fourier transform (FFT) to estimate the power spectrum density (PSD). The negative effects of spectral leakage, due to windowing, can be avoided by using parametric or model-based PSD estimation methods. These methods have the added advantage in that they have a better frequency resolution when compared to classical or non-parametric methods.

Autoregressive (AR) modeling is another method for analyzing the frequency domain. In this method, data is modeled as the output of an all pole, causal discrete filter whose input is white noise. We have to estimate the AR parameters by solving linear equations. Another task is to select a suitable filter order. AR modeling was employed by Dingfei et al. to group cardiac arrhythmia into six classes. The disadvantage of FFT is its limitation in analyzing non-stationary signals. The reliability of PSD extraction methods decreases with a drop in both signal power and signal-to-noise ratio. Figure 2 shows the AR spectrum distributions for (a) normal and (b) diabetic HR signals (shown in Fig. 1).

Figure 2 indicates that the HF power is very low in diabetic HR signals due to a lower RR variability.

Table 2 shows the percentage of power distributions in various bands for normal and diabetic classes. The LF/HF value indicates the balance between the sympathetic and parasympathetic tones. This value is higher for the diabetic class than for the normal class.

### 3.3. Wavelet transform

Fourier transform (FT) techniques are not suitable for analyzing non-stationary signals because these techniques do not provide exact time localization. In short time FT (STFT), the signal is multiplied with a window function before evaluating

![AR Spectrum of HR signals. (a) Normal subject and (b) diabetic subject.](image-url)
the FT. The time–frequency plot (spectrogram) provides better time localization, but it has a poorer frequency resolution. The above-mentioned problem can be overcome by using wavelet analysis. Wavelet analysis involves comparing the signal with a chosen (finite duration) wavelet and recording the correlation coefficient. There are two types of wavelet analysis: continuous wavelet transform (CWT) and discrete wavelet transform (DWT).

A 3D diagram (scalogram) is obtained when a mother wavelet is dilated to different durations along the time scale and the signal to be analyzed is compared with each dilated version of the wavelet. The comparison operation results in the so-called “wavelet coefficients.” There are various types of mother wavelets available in the MATLAB toolbox; they are selected depending on the application. DWT is a sampled version of CWT in a dyadic grid. DWT coefficients are calculated for discrete values of scale factor and translation factors; the increments are in the dyadic scale.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal (mean ± std)</th>
<th>Diabetes (mean ± std)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power (%) LF</td>
<td>31.9 ± 12.8</td>
<td>17.6 ± 9.03</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Power (%) HF</td>
<td>16.8 ± 9.15</td>
<td>18.2 ± 21.9</td>
<td>0.59</td>
</tr>
<tr>
<td>Power (%) LF/HF</td>
<td>2.44 ± 1.48</td>
<td>2.02 ± 1.65</td>
<td>0.093</td>
</tr>
</tbody>
</table>

Figure 3 shows the scalogram, evaluated with the Morlet wavelet, for (a) normal and (b) diabetic HR signals. It can be seen from the figure that there are abrupt
changes in the normal scalogram (Fig. 3(a)) and these sudden changes are missing in the diabetic scalogram (Fig. 3(b)).

3.4. Non-linear analysis methods

Surrogate data analysis may be performed to test whether it is possible to rule out a non-linear origin of a particular signal. This test is conducted by extracting non-linear parameters from several surrogate data series. Then, their values are compared with those computed for the original HR signal\(^{56}\). A significant difference between the parameters, which are obtained from original and surrogate data, indicates that the presence of non-linear dynamics in the original HR signal cannot be ruled out. Surrogate data can be produced by evaluating the Fourier decomposition from the HR signal. The resulting amplitude samples are kept, but the phase components are randomized. The inverse FT will produce a surrogate signal which "looks" like a normal HR signal, but this signal does not contain phase information.

The theory of non-linear dynamics is based on the concept of chaos. By now, there is little disagreement that this method is very useful in analyzing bio-signals, because it is assumed that they are non-linear in nature\(^{57}\). Cohen et al. showed that chaos theory methods have been used for HRV signal analysis and prediction of events like ventricular tachycardia (VT).\(^{58}\) Allyson et al. used detrended fluctuation analysis (DFA) to identify changes in HRV.\(^{32}\) The numerical value \(\alpha\) was used to reflect changes in HRV. The aim of the study was to identify cardiac diseases (related or unrelated to diabetes) that were difficult to identify with current clinical methods.

Some other useful and tried non-linear parameters for the HRV-based detection of diabetes are Poincare geometry, correlation dimension (CD), approximate entropy (ApEn), sample entropy (SampEn), largest Lyapunov exponent (LLE), Hurst exponent \((H)\), fractal dimension (FD), DFA, recurrence quantification analysis (RQA) features, and higher order spectra (HOS) parameters.

3.4.1. State-space reconstruction

State-space (phase-space) reconstruction is the first step in non-linear analysis. It basically views one-dimensional data \(y(n)\) where \(n = 1, 2, 3, \ldots, N\), in an \(m\)-dimensional Euclidean space, \(\mathbb{R}^m\). The path that connects the HR signals in the state space form an attractor that preserves the topological properties of the original unknown attractor. The method of delays is a common procedure to reconstruct the state space.\(^{59}\) In this method, \(m\)-dimensional state-space vectors, \(x_n\), are formed from the time-delayed samples of the original signal, \(y(n)\), as follows:

\[
x_n = [y(n), y(n - d), y(n - 2d), \ldots, y(n - (m - 1)d)],
\]

where \(d\) is the embedding delay and \(m\) is the embedding dimension (number of coordinates).
3.4.2. *Estimation of embedding dimension (m)*

The false nearest neighbor (FNN) method can be used to determine the minimal sufficient embedding dimension \( m \).\(^{60}\)

We have obtained \( m \) as 10 for the normal HR signal shown in Fig. 1. Figure 4 shows the estimation of the embedding dimension for our data using the FNN method for the normal HR signal (Fig. 1).

3.4.3. *Estimation of delay time (\( \tau \))*

The time-delayed mutual information is used to determine a reasonable time delay. The method takes into account the non-linear correlations in the time series.\(^{61}\)

Mutual information function for the normal HR signal, shown in Fig. 1, is given in Fig. 5. It can be clearly seen that the mutual information reaches its first minimum

---

**Fig. 4.** Result of estimation of embedding dimension using the FNN method.

**Fig. 5.** Estimation of delay using average mutual information.
at $\tau = 6$. We have used a time delay of six as the embedding delay for the state-space reconstruction.

3.4.4. Poincare plot geometry (SD2)

The technique of Poincare plot is adopted from the field of non-linear dynamics. It gives an idea about the nature of RR interval fluctuations. Each RR interval is plotted as a function of the previous RR interval. Thus, it is a graphical representation of the correlation between consecutive intervals. Poincare plot analysis means to visually examine the shapes, shown in the Poincare plots, and then categorize them into functional classes. These classes can indicate the degree of heart failure in a subject. Woo et al. and Kamen et al. have shown that the plot provides detailed beat-to-beat information of the heart.62,63 Poincare plots are analyzed by determining the standard deviations of the distances of the RR$_n$ interval to the lines $y = x$ and $y = -x + 2 \times$ RR$_m$, where RR$_m$ is the mean of all RR$_n$ intervals.64 The short-term variability (or the fast beat-to-beat variability) of the HR signal is quantified by the standard deviation (SD1) of the points that are perpendicular to the line-of-identity. The long-term variability of RR data is quantified by the standard deviation (SD2) of the points along the line-of-identity. Figure 7 shows typical Poincare plots for both normal and diabetic subjects whose HR signals are shown in Fig. 1. It can be seen from Table 3 that the SD1 and SD2 values are lower for the diabetic class than the normal class; this is due to reduced variability.

The Poincare plot of normal HR (Fig. 6(a)) is elliptical-shaped and it is aligned at the center. It can be seen that SD2 is longer than the SD1, indicating more long-term variability than short-range variability.65 The plot is shifted upward and SD1 is reduced as compared to normal subjects for diabetic patients (Fig. 7(b)).

![Fig. 6. Poincare plots of the HR signals. (a) Normal subject and (b) diabetic subject (shown in Fig. 1).](image-url)
3.4.5. Correlation dimension (CD)

CD is a non-linear parameter which is a useful indicator of pathologies. CD is used to measure FD. The widely used algorithm for calculating CD was proposed by Grassberger et al. Here, an algorithm function \( C(r) \) is to be constructed. This is done by calculating the separation between every pair of \( N \) data points and sorting them into bins of width \( dr \) proportional to \( r \). The distance between each pair of points is calculated as: \( s(i,j) = |X_i - X_j| \).

The correlation function \( C(r) \) is computed as follows:

\[
C(r) = \frac{1}{N^2} \sum_{x=1}^{N} \sum_{y=1, x \neq y}^{N} \Theta(r - |X_x - X_y|) \ldots 
\]  

Table 3. Non-linear parameter results from HRV signals for normal and diabetic classes.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal (mean ± std)</th>
<th>Diabetes (mean ± std)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD1</td>
<td>25.4 ± 11.8</td>
<td>18.2 ± 35.7</td>
<td>0.085</td>
</tr>
<tr>
<td>SD2</td>
<td>66.4 ± 27.2</td>
<td>30.7 ± 19.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>CD</td>
<td>2.25 ± 1.35</td>
<td>0.249 ± 0.294</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>1.14 ± 0.228</td>
<td>0.98 ± 0.342</td>
<td>0.0004</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.904 ± 0.231</td>
<td>1.05 ± 0.287</td>
<td>0.0005</td>
</tr>
<tr>
<td>ApEn</td>
<td>2.04 ± 0.218</td>
<td>1.79 ± 0.391</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>SampEn</td>
<td>1.54 ± 0.369</td>
<td>1.38 ± 0.469</td>
<td>0.019</td>
</tr>
<tr>
<td>LLE</td>
<td>1.03 ± 1.19</td>
<td>0.541 ± 0.907</td>
<td>0.0035</td>
</tr>
<tr>
<td>( H )</td>
<td>0.222 ± 9.704E-02</td>
<td>0.223 ± 0.134</td>
<td>0.95</td>
</tr>
<tr>
<td>FD</td>
<td>-1.78 ± 0.103</td>
<td>-1.77 ± 0.135</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Fig. 7. DFA of HR signals. (a) Normal subject and (b) diabetic subject (shown in Fig. 1).
where $X_x$ and $X_y$ are the trajectory points in phase space, $N$ is the number of data points in the phase space, and $r$ is the radial distance around each point $X_i$.

CD is given by:

$$\text{CD} = \lim_{r \to 0} \frac{\log[C(r)]}{\log(r)}. \quad (3)$$

The CD value will be high for large or chaotic RR signal variations. The CD value is low for low or rhythmic RR signal variations. It can be seen from Table 3 that the CD value is higher for normal then diabetic signals due to a higher beat-to-beat variability.

3.4.6. Detrended fluctuation analysis (DFA)

DFA was proposed by Peng et al.\(^6^7\) The fractal scaling properties of signals with short-term RR intervals is assessed by the DFA parameter. The fluctuation is characterized by a factor $\alpha$ which indicates the time-series roughness. If the time series is smoother, then the value of $\alpha$ is larger. This value falls in different ranges for different types of cardiac abnormalities. Figure 7 shows both small-range correlation ($\alpha_1$) and long-range correlation ($\alpha_2$) for normal and diabetic HR signals (shown in Fig. 1).

It can be seen from Table 3 that a small-range correlation ($\alpha_1$) is higher for the normal class and a long-range correlation ($\alpha_2$) is higher for diabetic HR signals due to a lower beat-to-beat variability when compared to the normal class.

3.4.7. Approximate entropy (ApEn)

ApEn was proposed by Pincus.\(^6^8\) The technique results in a non-negative number which corresponds to the information in a time series. The value of ApEn is higher when the data is more complex or irregular.\(^6^9\) When applied to HR signals, ApEn measures disorder. The resulting values become lower as the HR variation reduces. Thus, the ApEn value is small for cardiac impairment cases.

Let $\mathbb{R}^m$ be the embedding space and $x(1), x(2), x(N)$ be the $N$ data points (RR intervals like RR(1), RR(2), RR(N)), then ApEn is calculated as:

$$\text{ApEn}(m, r, N) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log[C_i^m(r)] - \frac{1}{N - m} \sum_{i=1}^{N-m} \log[C_{i+1}^m(r)], \quad (4)$$

where

$$C_i^m(r) = \frac{2}{N_m(N_m - 1)} \sum_{i=1}^{N_m} \sum_{j=1, j \neq i}^{N_m} \Theta(r - \|x_i - x_j\|). \quad (5)$$

Table 3 shows that the ApEn value is higher for normal than for diabetic signals due to a higher beat-to-beat variability.
3.4.8. Sample entropy (SampEn)
Richman and Randall developed a parameter called SampEn. It measures both the complexity and regularity of time-series data.\textsuperscript{70} Compared to ApEn, it is an improved measure. Table 3 indicates that the SampEn value is higher for normal than for diabetic signals due to a higher beat-to-beat variability.

3.4.9. Largest Lyapunov exponent (LLE)
LLE defines the sensitivity of the system to initial conditions and it determines the predictability of the signal or system. A positive Lyapunov exponent indicates chaos as proposed by Rosenstien et al.\textsuperscript{71} The LLE extraction algorithm searches for the nearest neighbor of each point in the phase space and tracks their separation over a certain time evolution. LLE is evaluated by using a least-squares fit to “average” a line given by:

\begin{equation}
y(n) = \frac{1}{\Delta t} \langle \ln[b_i(n)] \rangle \ldots,
\end{equation}

where \( b_i(n) \) is the distance between the \( i \)th phase-space point and its nearest neighbor at the \( n \)th time step, and \( \langle \cdot \rangle \) denotes the average overall phase-space points. The values in Table 3 indicate that the LLE value is higher for normal than for diabetic signals due to a higher beat-to-beat variability.

3.4.10. Hurst exponent (\( H \))
\( H \) indicates both self-similarity and correlation properties of a signal. It quantifies the smoothness of a fractal time series based on the asymptotic behavior of the rescaled range of the process. \( H \)\textsuperscript{72} is defined as:

\begin{equation}
H = \frac{\log(R/S)}{\log(T)},
\end{equation}

where \( T \) is the duration of the sample of data, \( R/S \) is the corresponding value of the rescaled range, \( R \) is the difference between the maximum deviation and the minimum deviation from the mean, and \( S \) is the standard deviation. Table 3 shows that the \( H \) value is higher for diabetic as compared to the normal signal due to a higher beat-to-beat similarity.

3.4.11. Fractal dimension (\( FD \))
\( FD \) quantifies the time-series complexity. A fractal is a set of points which, when looked at smaller scales, resembles the whole set.\textsuperscript{73} \( FD \) is a powerful indicator of the so-called “transient detection.” This parameter has been used in ECG and electroencephalography (EEG) analysis to identify and distinguish specific states of physiologic functions.\textsuperscript{54}

The mathematical discussion of \( FD \) starts by defining \( S \) as a compact subset of a metric space. For each \( \varepsilon > 0 \), let \( N(\varepsilon) \) be the smallest number of circles, with a
radius \leq \varepsilon$, necessary to cover $S$:

$$\delta = \lim_{\varepsilon \to 0^+} \frac{\log[N(\varepsilon)]}{\log(\varepsilon)}.$$  \hfill (8)

$\delta$ is called the FD of $S$. In this work, we have used Higuchi’s algorithm to evaluate FD.$^{74}$ The entries in Table 3 indicate that the FD value is higher for diabetic than for normal signals due to a higher beat-to-beat similarity.

3.4.12. Recurrence quantification analysis (RQA)

The technique of recurrence plot was proposed by Eckmann et al.$^{75}$ Non-stationary time-series data can be analyzed with recurrence plots.$^{76}$ It is a graphical tool which helps us to find hidden periodicities in a time domain signal, which otherwise are not easily noticeable.

Let $x_i$ be the $i$th point on the orbit in an $m$-dimensional space. The recurrence plot is an array of dots in an $N \times N$ square. Whenever $x_j$ is sufficiently close to $x_i$, a dot is placed at $(i, j)$. The plots are symmetric along the diagonal $i = j$, because if $x_i$ is close to $x_j$, then $x_j$ is close to $x_i$.

Figure 8 shows typical recurrence plots of a normal subject and a diabetic subject. For normal cases, the plot indicates more variation in HR by displaying a diagonal line and fewer squares. In the case of abnormalities, like CHB and ischemic/dilated cardiomyopathy, the plot contains a higher number of squares; this indicates a lower HR variation.$^{65}$

(a) Fig. 8. Recurrence plots of HR signals. (a) Normal subject and (b) diabetic subject, for the signals shown in Fig. 1.
The RQA parameters are single-value measures which can be extracted from recurrence plots. They are briefly introduced below:

(1) Recurrence rate (REC) indicates the density of recurrence points in a recurrence plot. It is given by:

\[
REC = \frac{1}{N^2} \sum_{i,j=0}^{N} R_{i,j} \ldots ,
\]

where \( R_{i,j} \) is the representation of recurrence plot and \( N \) is the number of points.

(2) Determinism (DET) is the fraction of recurrence points that form diagonal lines. These lines represent epochs of similar time evolution of states in the system. Hence, DET indicates the determinism of the system. It is given by:

\[
DET = \frac{\sum_{l=l_{\text{min}}}^{N} l P(l)}{\sum_{i,j=1}^{N} R(i,j)} .
\]

(3) Mean diagonal line length \( L_{\text{mean}} \) or mean length of the diagonal lines. It is defined as:

\[
L_{\text{mean}} = \frac{\sum_{l=l_{\text{min}}}^{N} l P(l)}{\sum_{l=l_{\text{min}}}^{N} P(l)} \ldots .
\]

(4) Entropy (ENTR) measures the complexity of the recurrence structure. It is given by:

\[
\text{ENTR} = - \sum_{l=l_{\text{min}}}^{N} P(l) \ln[P(l)].
\]
5. Laminarity (LAM) is the fraction of recurrence points which form vertical lines and correspond to the amount of laminar states in the system. It is given by:

$$LAM = \frac{\sum_{v=v_{\text{min}}}^N vP(v)}{\sum_{v=1}^N P(v)} \ldots$$  

(13)

6. Trapping time (TT) corresponds to the mean length of the vertical lines. It measures the mean time that the system is either trapped in one state or changes very slowly. It is given by:

$$TT = \frac{\sum_{v=v_{\text{min}}}^N vP(v)}{\sum_{v=1}^N P(v)} \ldots$$  

(14)

7. Recurrence times (T1, T2): the first and second Poincare recurrence points are given by:

$$T1(i) = t_{i+1} - t_i, \quad t = 1, 2, \ldots, K \quad \text{and}$$

$$T2(i) = t'_{i+1} - t'_i, \quad t = 1, 2, \ldots, K.$$  

(15)

(16)

Table 4 lists the RQA features which were calculated for both normal and diabetic classes. The DET, LAM, TT, $L_{\text{mean}}$, LAM, and T1 are higher for the diabetic class than the normal class. This indicates slower HRV.

3.4.13. Higher order spectra (HOS)

HOS analysis is a powerful tool for the non-linear dynamical analysis of non-linear, non-stationary, and non-Gaussian physiological signals.$^{79,80}$ HOS analysis is capable of detecting non-linearity, deviations from Gaussianity, and phase relationships between harmonic components of the signal. HOS is used to analyze HRV signals and extract useful features which can be used to discriminate signal classes.

HOS (also called as polyspectra) is the spectral representation of higher order statistics, i.e., moments and cumulants of third and higher orders.$^{81}$ Since the HOS of Gaussian signals are statistically zero, HOS can measure non-Gaussianity and separate non-Gaussian signals from an additive mixture of independent

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal (mean ± std)</th>
<th>Diabetes (mean ± std)</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>REC</td>
<td>5.195E−02 ± 1.797E−02</td>
<td>0.106 ± 4.989E−02</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>DET</td>
<td>0.166 ± 7.096E−02</td>
<td>0.358 ± 1.41E−02</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$L_{\text{mean}}$</td>
<td>2.12 ± 7.721E−02</td>
<td>2.33 ± 0.239</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>ENTR</td>
<td>0.365 ± 0.143</td>
<td>0.687 ± 0.274</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>LAM</td>
<td>0.207 ± 9.383E−02</td>
<td>0.453 ± 0.166</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TT</td>
<td>2.13 ± 0.106</td>
<td>2.51 ± 0.492</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>T1</td>
<td>19.0 ± 6.0</td>
<td>10.1 ± 4.34</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>T2</td>
<td>21.5 ± 5.78</td>
<td>13.6 ± 4.31</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
non-Gaussian signals and Gaussian noise. Thus, HOS techniques provide high noise immunity. This property is particularly useful in cases where the signals are corrupted with additive white Gaussian noise. Another advantage is that HOS can preserve the true phase character of signals.

Most of the work done so far is on the third order; its statistics has the name bispectrum. Bispectrum $B(f_1, f_2)$ is the third order cumulant generating function. The bispectrum is also defined as FT of the third order correlation of a signal. It is given by:

$$B(f_1, f_2) = \mathbb{E}[X(f_1) X(f_2) X(f_1 + f_2)],$$

where $X(f)$ is the FT of the signal $X(nT)$, $n$ is an integer index, $T$ is the sampling interval, and $\mathbb{E}[\cdot]$ is the expectation operator. The bispectrum gives the cross-correlation between frequency components in a two-dimensional frequency plot. It is a triple product which is evaluated at two frequencies and their sum frequency. For deterministic signals, the expectation operator can be omitted. For deterministic sampled signals, $X(f)$ is the discrete-time FT and it is computed using the FFT algorithm. The frequency $f$ may be normalized by the Nyquist frequency (half of the sampling frequency) for values between 0 and 1. The bispectrum plot (Fig. 9) shows the non-linear interactions between the harmonic components of a signal. Bispectrum is a function of two frequencies unlike the power spectrum, which is a function of one frequency variable. The bispectrum can be normalized by the power spectra at the component frequencies to ensure that the magnitude falls between 0 and 1.

A normalized bispectrum $B_{\text{norm}}$ was defined by Haubrich as: \(^{82}\)

$$B_{\text{norm}}(f_1, f_2) = \frac{\mathbb{E}[X(f_1) X(f_2) X(f_1 + f_2)]}{\sqrt{P(f_1) P(f_2) P(f_1 + f_2)}},$$

where $P(f)$ is the power spectrum.

![Bispectrum plots of HRV signals. (a) Normal subject and (b) diabetic subject, for the signals shown in Fig. 1.](image)

Fig. 9. Bispectrum plots of HRV signals. (a) Normal subject and (b) diabetic subject, for the signals shown in Fig. 1.
Here, $P(f)$ is the power spectrum. The power spectrum gives the signal’s power within the frequency components considered. Bicoherence ($B_{co}$) is the squared-magnitude of the normalized bispectrum. Bicoherence is a chi-square distributed parameter.

Bicoherence takes the value of unity if the Fourier components at frequencies $f_1$, $f_2$ and $f_1 + f_2$ are perfectly phase-coupled in every data block. If they have completely random phases, then the bicoherence takes a value of zero. Thus, the bicoherence varies only between 0 and 1. The patterns of bispectrum and bicoherence plots are more important than finding out the values of bispectrum and bicoherence. Since both bispectrum and bicoherence are the products of three Fourier coefficients, they exhibit symmetry. The entire two-dimensional bi-frequency plane is covered by these plots, hence they exhibit six-fold symmetry.

Both bispectrum and bicoherence plots help us to differentiate diabetic and normal HR data visually by displaying clearly distinguishable patterns in the bi-frequency space. In general, the majority of bicoherence values from the bi-frequency space of each class will be zero, i.e., they indicate random phase and give a distinct visual pattern. Over this background pattern of zero bicoherence, there are peaks which are higher than this background. These peaks are highly likely to correspond to phase-coupled bicoherence values. The distribution and intensity of these peaks are distinctly different from class to class. Figures 9 and 10 show both bispectrum and bicoherence plots of normal and diabetic classes. It can be seen from these figures that these plots are unique for the two classes.

Mean bispectrum magnitude is defined as:

$$M_{ave} = \frac{1}{L} \sum_{(f_1,f_2)} |D(f_1,f_2)|,$$

where $B(f_1,f_2)$ is the bispectrum of the signal.

![Fig. 10. Bicoherence plots of HRV signals. (a) Normal subject and (b) diabetic subject (shown in Fig. 1).](image-url)
The parameter $M_{ave}$ clearly distinguishes classes which have similar power spectra, but different third order statistics. However, the effectiveness of this parameter is limited by its sensitivity to amplitude changes. For example, if the HRV range changes within a class, then the power spectrum and also the mean magnitude of the bispectrum will change. This limiting factor can be annulled by normalization.

The bispectrum phase entropy is:

$$P_e = \sum_n p(\Psi_n) \log[p(\Psi_n)],$$

where $p(\Psi_n)$ is defined as:

$$p(\Psi_n) = \frac{1}{L} \sum_{(\Omega)} 1\{\Phi[B(f_1, f_2)]\},$$

where

$$\Psi_n = \{\Phi/ -\pi + 2\pi n/N \leq \Phi < -\pi + 2\pi (n + 1)/N\}, \quad n = 0, 1, \ldots, N - 1.$$  

Here, $L$ is the number of points within the region shown in Fig. 9, $\Phi$ is the phase angle of the bispectrum, $\Omega$ is the space of the region defined in Fig. 11, and $1\{\cdot\}$ is the indicator function which has a value 1 when the phase angle $\Phi$ lies within the range of $\Psi_n$ in Eq. (21). The region $\Omega$ is known as the principal domain or the non-redundant region for the computation of the bispectrum for real signals.

The phase entropy parameter $P_e$ increases with an increase in the randomness of the underlying process. A time shift in the input signal will not affect the bispectral phase. In contrast, an input signal time shift will case a change in the Fourier phase. There are two additional bispectral entropies: normalized bispectral entropy ($P_1$) and normalized bispectral squared entropy ($P_2$).\cite{83,84}

Fig. 11. Non-redundant region $\Omega$ of the computation of the bispectrum for real signals (the frequencies are normalized by the Nyquist frequency).
Normalized bispectral entropy ($P_1$) is given by:

$$P_1 = - \sum_{\langle n \rangle} p_n \log(p_n),$$

(23)

where $p_n = \frac{|B(f_1, f_2)|}{\sum_{\langle n \rangle} |B(f_1, f_2)|}$.

The normalized bispectral squared entropy ($P_2$) is given by:

$$P_2 = \sum_{\langle i \rangle} q_i \log(q_i),$$

(24)

where $q_i = \frac{|B(f_1, f_2)|^2}{\sum_{\langle i \rangle} |B(f_1, f_2)|^2}$ and $\Omega = \text{region as shown in Fig. 11}$.

The normalized bispectral squared entropy ($P_3$) is given by:

$$P_3 = \sum_{\langle i \rangle} r_i \log(r_i),$$

(25)

where $r_i = \frac{|B(f_1, f_2)|^3}{\sum_{\langle i \rangle} |B(f_1, f_2)|^3}$ and $\Omega = \text{region as shown in Fig. 11}$.

The absolute magnitude and the square of the magnitude are known as the $L_1$ and $L_2$ norms of the bispectrum. Both entropies are normalized by the sum of the norm over $\Omega$, which is the complete non-redundant bi-frequency region. Thus, the entropies will be similar to a probability density function (PDF), which was estimated over the $\Omega$ region.

Table 5 shows the HOS features calculated for normal and diabetic classes. All HOS features are higher for the normal class than the diabetic class; this indicates a higher HR variability.

### 3.4.14. Surrogate data analysis

The aim of surrogate data is to examine whether the presence of non-linearity can be ruled out in the original HR signal. If and only if non-linearity cannot be ruled out, then we can apply non-linear methods on the HR data. The values, computed for the original HR data, are compared against the surrogate data and if significant differences in non-linear parameters values exist, this will then indicate that the presence of non-linearity in the original HR signal cannot be ruled out. The technique of using surrogate data in non-linear analysis is acquired by the phase randomization of the original data with spectral properties, which are almost the
same as the given data. Surrogate data possesses Fourier decomposition containing randomized phase components with identical amplitudes as the empirical data decomposition. To analyze the non-linear characteristics of HR signals, 15 sets of surrogate data were generated for two classes. CD was obtained for both original and surrogate data sets. The difference of the CD values between surrogate and original data was found to be more than 54% (shown in Table 6). Hence, this result rejects the null hypothesis, which indicates that we cannot rule out that the original data was the result of a non-linear process.

4. Discussion

Diabetes causes the early development of atherosclerosis and coronary disease. It also affects the neural control of the heart. The constraint of traditional measures for assessing cardiac health is that these can document the presence of neuropathy only at a worsened state, when neuropathy has started showing several symptoms. Pfeifer et al. have observed that diabetic patients, who showed negative results for cardiac autonomic function tests, had a diminished HRV. The work of Singh et al. and Villareal et al. also emphasized the fact that cardiac autonomic activity (i.e., parasympathetic activity) was reduced in diabetic patients a long time before clinical symptoms of neuropathy were evident.

All these work showed that diabetics causes detectable changes in HRV signals. This led to the development of non-invasive methods for the diagnosis of diabetes using HRV signals as the input. Malpas et al. showed that 24-h HRV recordings can be used to detect very minute changes in cardiac autonomic function and this technique is more sensitive than currently available tests. In fact, HRV can be used as an early indicator of diabetic neuropathy. Early detection is important to prevent the worsening of the disease, especially as it may lead to other serious complications. For HRV analysis, several processing methods such as time domain analysis, frequency domain analysis, geometric method, non-linear techniques, and HOS, were developed. These analysis methods yield specific parameters which indicate certain cardiac conditions. For example, HF power and RMSSD are HRV indicators of vagal HR modulation, while SDNN indicates total RR variability.

Through frequency domain methods, the RR cycle variation, within known frequency bands that correspond to sympathetic, vagal and non-neural modulation of intrinsic heartbeat duration, was analyzed. The difficulty with time domain

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Diabetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original data</td>
<td>2.25 ± 1.35</td>
<td>0.249 ± 0.294</td>
</tr>
<tr>
<td>Surrogate data</td>
<td>1.23 ± 0.79</td>
<td>0.137 ± 0.45</td>
</tr>
<tr>
<td>% Difference</td>
<td>54.75%</td>
<td>55.4%</td>
</tr>
</tbody>
</table>

Table 6. Surrogate data analysis results based on CD.
methods and parameters is that they are easily affected by artifacts and outliers. Before a time domain analysis, these artifacts should be eliminated from the data samples. The problem with frequency domain analysis is that the reliability of spectral power estimation diminishes as the power of the signal and the signal-to-noise ratio decreases. Furthermore, FT fails to represent the time a particular event occurs in the frequency domain.

Another method of HRV analysis is geometric analysis. It presents RR intervals in geometric patterns. There are different types of plots which represent different properties of HR signals. The disadvantage is their sensitivity to artifacts and ectopic beats.

Goldberger et al. have given evidence which suggested that the heart is not a periodic oscillator under normal physiologic conditions. This led to the development of several techniques based on non-linear dynamics and the chaos theory to quantify the dynamics of HR fluctuations. ECG and HRV are inherently non-linear and non-stationary. Recent studies have also pointed out that non-linear techniques are very much suited for HRV analysis because the second order moment statistics will not be able to detect subtle changes in a HR time series. Many non-linear techniques have been used for the analysis of ECG signals to develop cardiac arrhythmia detection.

HOS-based studies have been widely used in ECG signal analysis and recently, it has been used for the analysis of HRV. There are several reasons for the use of HOS in the area of HRV analysis. All higher order cumulant spectra (order three and higher) of Gaussian signals are zero. Hence, the higher order cumulant transform eliminates Gaussian noise. The true phase character of signals is preserved even after HOS (cumulant and moment) analysis. With their HOS-based analysis of respiratory and ECG signals, Jamsek et al. showed the coupling between cardiac and respiratory activities. Witte et al. have developed a technique for the time-variant analysis of quadratic phase-coupling of HR data. Chua et al. classified normal HRV signals and four other classes of arrhythmia with a very high accuracy using HOS.

Kirvela et al. analyzed HRV (taken from 24-h ECG recordings) in both time and frequency domains. They observed that in diabetic HRV samples, all time domain and frequency domain measures of HRV were markedly reduced ($p < 0.05$) when compared with normal samples. Mackay et al. measured HRV at different rates of regular and deep breathing in normal and diabetic patients with neuropathy. They observed that HRV was significantly smaller than normal in diabetic patients with neuropathy. Barry et al. started a work with the objective of finding the time it takes for diabetes to cause significant changes in HRV signals (either in time or frequency domain). They compared HRV signals taken from a diabetic and a normal (non-diabetic) person and found that HRV measures of patients who had a long history of diabetes (five years or more) are very much different from patients with a shorter history (three years or less) of diabetes. They also showed that the HRV of newly
diagnosed cases and those up to three-year diabetic cases is similar to non-diabetic cases. These works show the relation between HRV and diabetes.

Awdah et al. conducted a study with a group of diabetic and normal people. All were free of cardiovascular disease and with no previous history of ill health, except diabetes for the diabetic people. They measured time domain variables like SDRR, the St. George index, SNN50, pNN50%, and RMSSD of these groups and came to a conclusion that all these parameters were significantly decreased in all diabetic patients (with and without diabetic neuropathy) compared to the normal people.

Flynn et al. studied the effects of QT dispersion (QTd) and HRV on a group of people with and without Type 2 diabetes and with no recorded cardiac disease. They studied HRV in the form of a parameter called tone-entropy (TE), where tone represented sympatho-vagal balance and entropy indicated autonomic regularity. They found that there was a decrease in TE, which indicates a change in normal sympathetic-parasympathetic rhythm of diabetic people.

Chemla et al. used both AR spectral- and FFT-based methods for the study of HRV signals in diabetic people. Schroeder et al. showed that diabetic people have lower SDNN, RMSSD, and RR interval than non-diabetic people. Their conclusion was that these HRV changes in diabetic people suggest a decrease in ANS function in the early stage of diabetes itself. They found that diabetes led to a progressive decline in the ANS function.

Trunkvalterova et al. have shown that multiscale entropy (MSE) can detect even minute abnormalities in the cardiovascular system in young patients with Type 1 diabetes. They used SampEn and linear measures of HRV like RMSSD for their work. Ahamed Seyd et al. analyzed HRV signals from an equal number of diabetic and normal people with frequency and time domain methods. They found that the time domain parameters SDNN, RMSSD, NN50 count, pNN50 count, HRV triangular index, TINN, and mean RR interval decline for DM patients compared to normal ones. Frequency domain analysis showed significant difference in VLF power, LF power, and HF power between diabetics and normal people.

Ahsan et al. showed the importance of HRV analysis (HRV signals extracted from short-term ECG recordings) in detecting CAN in diabetic people. They found that Poincare plots and SampEn derived from HRV have great potential in discriminating diabetes-induced CAN. Nolan et al. reinforced the fact that the duration of Type 1 and Type 2 diabetes is inversely related to the HRV measures of total RR variability and vagal HR modulation. Faust et al. have analyzed normal and diabetic HRV data using time domain, frequency domain, and non-linear techniques and found that non-linear analysis methods of HRV are superior to time and frequency domain methods.

Acharya et al. extracted non-linear parameters like CD, SD2, REC, DET, and $L_{\text{mean}}$ out of HRV to distinguish between normal and diabetic subjects. They developed a novel diabetic integrated index by combining these parameters and they have used this index along with the Adaboost classifier to achieve an accuracy.
of 86%. Recently, HOS features, coupled with a Gaussian mixture model (GMM) classifier, were used differentiate normal and diabetic HR signals automatically, with an accuracy of 90.5%, a sensitivity of 85.7%, and a specificity of 95.2%.\textsuperscript{105}

Table 7 summarizes all the research work described above. The common observation in these works is that HRV is diminished and the complexity of the HR signals is reduced due to diabetes.

We have also conducted time domain, frequency domain, and non-linear analysis of normal and diabetic classes and we have obtained results which are shown in Tables 1–5. All the time domain parameters, displayed in Table 1, are lower for

<table>
<thead>
<tr>
<th>Author</th>
<th>Parameter and method used for HRV analysis</th>
<th>Results obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfeifer et al.\textsuperscript{35}</td>
<td>RR variations</td>
<td>Parasympathetic autonomic activity declined even before clinical symptoms of neuropathy are evident</td>
</tr>
<tr>
<td>Singh et al.\textsuperscript{36}</td>
<td>SDNN, high, and LF power</td>
<td></td>
</tr>
<tr>
<td>Villareal et al.\textsuperscript{37}</td>
<td>Linear and non-linear dynamics of HRV</td>
<td></td>
</tr>
<tr>
<td>Kirvela et al.\textsuperscript{34}</td>
<td>Time and frequency domain parameters</td>
<td>All time domain and frequency measures of HRV were markedly reduced</td>
</tr>
<tr>
<td>Al-Hazimi et al.\textsuperscript{97}</td>
<td>SDRR, NN50, RMSSD, and pNN50% (time domain variables)</td>
<td>All parameters significantly decreased</td>
</tr>
<tr>
<td>Flynn et al.\textsuperscript{32}</td>
<td>Tone-entropy (TE) parameter (method effect of QT dispersion)</td>
<td>TE parameter decreased</td>
</tr>
<tr>
<td>Chemla et al.\textsuperscript{98}</td>
<td>AR spectral method and FFT based method</td>
<td>Depressed HRV</td>
</tr>
<tr>
<td>Schroeder et al.\textsuperscript{33}</td>
<td>SDNN, RMSSD, and RR interval (time domain parameters)</td>
<td>All parameters showed lower value</td>
</tr>
<tr>
<td>Ahamed Seyd et al.\textsuperscript{100}</td>
<td>Several time and frequency domain parameters</td>
<td>All parameters reduced</td>
</tr>
<tr>
<td>Trunkvalterova et al.\textsuperscript{99}</td>
<td>SampEn, linear measures of HRV like RMSSD (method multiscale entropy)</td>
<td>Detected even minute abnormalities in cardiovascular system in young patients with Type 1 diabetes</td>
</tr>
<tr>
<td>Nolan et al.\textsuperscript{102}</td>
<td>Total RR variability and vagal–HR modulation</td>
<td>Duration of Type 1 and Type 2 diabetes is inversely related to the HRV measures</td>
</tr>
<tr>
<td>Faust et al.\textsuperscript{103}</td>
<td>Time domain, frequency domain, and non-linear analysis</td>
<td>Found that non-linear analysis of HRV is a superior method</td>
</tr>
<tr>
<td>Acharya et al.\textsuperscript{104}</td>
<td>Non-linear features like RQA parameters, CD, and Poincare geometry properties</td>
<td>Proposed an index to classify normal and diabetic HR signals and achieved an accuracy of 86%</td>
</tr>
<tr>
<td>Swapna et al.\textsuperscript{105}</td>
<td>HOS features and GMM</td>
<td>Accuracy 90.5%, sensitivity 85.7%, and specificity 95.2%</td>
</tr>
</tbody>
</table>
The analysis of HRV samples is more useful than extracting information directly from the HR samples. This shows that the analysis of HRV samples is more useful than extracting information directly from the HR samples.

Table 2 shows the power in both LF and HF frequency bands as well as the ratio parameter LF/HF. It can be observed that there is a significant difference between the normal and the diabetic frequency domain parameter values of LF and LF/HF. The non-linear features (Table 3) such as CD, ApEn, SampEn, LLE, SD1, SD2, $\alpha_1$, and $\alpha_2$ show higher values for the normal class as compared to the diabetic class due to a higher variability. However, $H$ and FD were lower for the diabetic class compared to the normal class due to a higher self-similarity in the time series.

Tables 4 and 5 show the variation of RQA and HOS for the two classes. Acharya et al. obtained the result of increased CD for normal individuals compared to diabetics. The CD parameter for normal HRV is about 10 times larger than that obtained from diabetics. The CD parameter is a measure of signal complexity. The parameter SD2 (indicates long-term variability), derived from Poincare plot, is also higher for the normal class. These results indicate that complexity of HRV signal reduces in diabetic people.

Faust et al. found that the values of SD2, CD, ApEn, and SampEn are higher for normal subjects compared to the diabetics. These non-linear parameters indicate the variability of the signal. The lower values for the diabetic samples implies that there is less variation in the HR of diabetic people. The obtained REC value is small for the normal class. The parameter REC is derived from recurrence plots, such as that shown in Fig. 8. It can be noted that the recurrence plots for the normal class contain bigger squares. This implies that the rate at which these squares occur is higher for diabetic samples. Thus, REC, which is the rate of recurrence, is small for normal and high for diabetic subjects.

Since it is widely accepted that HRV signals result from non-linear processes with the body, non-linear methods have the potential to capture the characteristics more accurately as compared to linear time and frequency domain methods. These are single, unique non-linear parameters like CD and SampEn, which are unique in their own fashion. According to the range over which the value falls for each of these parameters, a decision should be made about the disease. On the other hand, there are non-linear analysis techniques like HOS and RQA, which are composed of numerous of inter-related parameters. The values of these inter-related parameters combined will decide the nature of the disease. Thus, HOS and RQA can capture the subtle changes in the HRV signal much better than individual non-linear parameters of CD, ApEn, etc. Another improvement is the automation of the detection process using classifiers. The parameters are given as input to automated classifiers. The output of the classifier is one of the two class labels; these labels indicate the presence or absence of disease. Such automated detection of disease, which employs a classifier, is very useful for medical professionals. People with less medical expertise can operate the automated diagnosis system and they can still obtain accurate results. The only work which employed classifiers for the automated...
detection of diabetes using HRV analysis methods was done by Acharya et al.\textsuperscript{104} They used five parameters, SD2, REC, DET, $L_{\text{mean}}$, and CD, to extract information from HR signals and then combined these five non-linear parameters to form another parameter named the diabetic integrated index. The combination of these parameters was done in a manner such that the value of the index is significantly different for normal subjects and diabetic subjects. Ghista et al. have also proposed and used the concept of an integrated index for biomedical applications.\textsuperscript{106–108}

These HR signals and patient information can be interleaved within the images with the different error-correcting codes in a noisy environment without affecting the hidden information.\textsuperscript{109,110}

5. Conclusion

Diabetes is a metabolic disorder which affects the ability of the human body to process glucose effectively and hence results in hyperglycemia. Uncontrolled hyperglycemia can lead to serious long-term complications affecting vital organs and the normal functioning of the body. Furthermore, diabetes affects cardiac health. Early detection of diabetes is crucial to reduce the cardiac mortality in diabetic patients. HRV clearly reflects several physiological factors which affect the normal heart rhythm. HRV signals, extracted from ECG signals, can be used to diagnose diabetes. Clinically undetectable automatic impairments, associated with diabetes, can be detected with HRV analysis. Thus, HRV analysis can be considered as an effective, non-invasive test for detecting diabetes. We have discussed different HRV analysis methods and summarized the works related to diabetes/cardiac health/HRV analysis. Non-linear parameters and methods such as HOS can capture information present in the non-linear HR signal much better than traditional time and frequency domain methods.

**Acronyms**

- ApEn: Approximate Entropy
- ANS: Autonomic Nervous System
- AR: Autoregressive
- CAN: Cardiovascular Autonomic Neuropathy
- CD: Correlation Dimension
- CHB: Complete Heart Block
- CWT: Continuous Wavelet Transform
- DET: Determinism
- DFA: Detrended Fluctuation Analysis
- DM: Diabetes Mellitus
- DWT: Discrete Wavelet Transform
- ECG: Electrocardiography
- EEG: Electroencephalography
**Comprehensive Analysis of Normal and Diabetic Heart Rate Signals: A Review**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENTR</td>
<td>Entropy</td>
</tr>
<tr>
<td>FD</td>
<td>Fractal Dimension</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>FNN</td>
<td>False Nearest Neighbor</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier Transform</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian Mixture Model</td>
</tr>
<tr>
<td>H</td>
<td>Hurst Exponent</td>
</tr>
<tr>
<td>HF</td>
<td>High Frequency</td>
</tr>
<tr>
<td>HOS</td>
<td>Higher Order Spectra</td>
</tr>
<tr>
<td>HR</td>
<td>Heart Rate</td>
</tr>
<tr>
<td>HRV</td>
<td>Heart Rate Variability</td>
</tr>
<tr>
<td>LAM</td>
<td>Laminarity</td>
</tr>
<tr>
<td>LF</td>
<td>Low Frequency</td>
</tr>
<tr>
<td>LLE</td>
<td>Largest Lyapunov Exponent</td>
</tr>
<tr>
<td>LVH</td>
<td>Left Ventricular Hypertrophy</td>
</tr>
<tr>
<td>NO</td>
<td>Nitric Oxide</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PSD</td>
<td>Power Spectrum Density</td>
</tr>
<tr>
<td>REC</td>
<td>Recurrence Rate</td>
</tr>
<tr>
<td>RQA</td>
<td>Recurrence Quantification Analysis</td>
</tr>
<tr>
<td>SA</td>
<td>Sinoatrial</td>
</tr>
<tr>
<td>SampEn</td>
<td>Sample Entropy</td>
</tr>
<tr>
<td>STFT</td>
<td>Short Time Fourier Transform</td>
</tr>
<tr>
<td>TINN</td>
<td>Triangular Interpolation of NN Interval Histogram</td>
</tr>
<tr>
<td>TT</td>
<td>Trapping Time</td>
</tr>
<tr>
<td>VLF</td>
<td>Very Low Frequency</td>
</tr>
<tr>
<td>VT</td>
<td>Ventricular Tachycardia</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>

**References**

1. de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, Dupuis J, Florez JC, D’Agostino RB, Cupples LA, the MAGIC Investigators, the DIAGRAM Investigators, Meigs JB Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms, *Diabetes Care*, 2010 [Epub ahead of print].


82. Haubrich RA, Earth noise, 5 to 500 millicycles per second, 1, spectral stationarity normality, and non-linearity, J Geophys Res 70:1415–1427, March 1965.


