Discrimination Power of Long-Term Heart Rate Variability Measures

M. H. Asyali
Biostatistics, Epidemiology and Scientific Computing Department
King Fasial Special Hospital & Research Center, Riyadh, Saudi Arabia

Abstract—**Heart Rate Variability (HRV) can be assessed by time- or frequency-domain methods. The time-domain HRV measures are based on beat-to-beat intervals whereas frequency-domain analysis expresses HRV in terms of its constituent frequency components. HRV analysis has emerged as a diagnostic tool that quantifies the functioning of the autonomic regulation of the heart and heart's ability to respond. However, majority of studies on HRV report several different time and frequency domain HRV measures together, which may be redundant and confusing in many cases. The question of which HRV measures are the strongest overall indicators of the cardiac condition has not been addressed. In this study, using data from 52 normal subjects and 22 patients with congestive heart failure, and linear discriminant analysis, we investigated the class, i.e. normal versus abnormal, discrimination power of 9 commonly used long-term HRV measures and identified the one that indicates the cardiac condition with higher sensitivity and specificity. Our results revealed that the standard deviation of all normal-to-normal beat intervals (SDNN) has the highest class discrimination power and a Bayesian classifier based on this index achieves sensitivity and specificity rates of 81.8% and 98.1% respectively.**

Keywords—**Bayesian classification, heart rate variability, linear discriminant analysis**

I. INTRODUCTION

Heart Rate Variability (HRV) refers to the variations in the beat intervals or correspondingly in the instantaneous heart rate (HR). The normal variability in HR is due to autonomic neural regulation of the heart and the circulatory system [1-2]. The balancing action of the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) branches of the autonomic nervous system (ANS) controls the HR. Increased SNS or diminished PNS activity results in cardio-acceleration. Conversely, low SNS activity or high vagal, i.e. PNS, activity causes cardio-deceleration. The degree of variability in the HR provides information about the functioning of the nervous control on the HR and heart's ability to respond.

On electrocardiogram (ECG), the heartbeats are recognized through QRS complexes, i.e. the deflections corresponding to ventricular contractions. Since R peaks in QRS complexes indicate the zenith of ventricular contractions, the beat instants are taken at these points and consequently the beat-to-beat intervals are determined as the length in time from one R wave to the next one. Therefore, the term "beat-to-beat interval" refers to RR interval. The RR intervals are also referred as NN (normal-to-normal) intervals, as they are resulting from normal sinus rhythm. The subject of automatic detection of QRS complexes in digitized ECG recordings has been studied extensively and there exists algorithms that allow QRS detection with high precision [3].

The RR interval discrete time series (RRITS) is the basic material from which many different HRV *measures* or *indices* are derived. Some HRV measures quantify short-term variations in the HR, some correspond to the long-term changes, and yet some others represent both. The time-domain measures are calculated directly from the raw RRITS [4].

Frequency domain HRV measures are based on the power spectral density (PSD) analysis of the interpolated RRITS [5]. Interpolation is necessary to produce a uniformly sampled series out of the RRITS, which is an inherently nonuniformly sampled series. Frequency domain HRV measures have the advantage of relating power of variation in different frequency bands to different physiological modulating effects. The modulating effects of SNS, which increases the HR, are expected to be in the low frequency (0.04-0.15 Hz) range whereas the modulating effects of PNS, which decreases the HR, are in the high frequency (0.15-0.40 Hz) range. Recently some new HRV measures based on nonlinear analysis of the RRITS were also suggested but the use and interpretation of these new measures is not well established yet [6].

HRV has been a rapidly evolving diagnostic tool, as the research over the last 3 decades clearly demonstrated its clinical importance. Other than assessment of mortality risk after myocardial infarction [7], HRV is also used to judge to degree of neuropathy in diabetes [8] and for screening of obstructive sleep apnea [9]. For a detailed discussion of physiological origins and clinical significance of HRV, we refer the reader to [10] and the references therein.

The derivation of many popular HRV measures is quite simple. This also helped the HRV analysis become widespread among clinicians and researchers. Each of different time and frequency domain HRV measures developed so far is a potential quantitative marker representing ANS activity that reflects the individual's capacity to effectively develop appropriate reactions to changing external or internal conditions. However, this significance and popularity of HRV analysis has one objectionable consequence. Majority of the studies on HRV report many different HRV measures together, even though research has already revealed a significant correlation between many of these measures [10]. Therefore, simultaneous use of many highly correlated HRV measures unnecessarily complicates the analysis and the presentation of redundant information may be confusing in many cases. The issue of which HRV measures are the strongest overall indicators of the cardiac condition has not been addressed adequately in literature.

As the length of analyzed recording increases, the chances of observing more of the collective effects of the regulatory mechanisms increase. Correspondingly, the variance in the HR increases with the length of the recording. Therefore, the type of HRV analysis regarding the length of analyzed recording, i.e. short- or long-term, must be clearly distinguished and standardized to make reliable comparisons across different studies. The recommended recording lengths to compute short- and long-term HRV measures are 5 min and 24 h respectively [10].

Table I gives a list of commonly used HRV measures that are suitable for the analysis of long-term variations. In this study, we looked at these measures, aiming at identifying the ones that indicate the cardiac condition, i.e. normal versus abnormal, with high sensitivity and specificity. Because, using only those HRV measures will (i) further standardize the use of this emerging analysis tool in clinical studies (ii) avoid presentation of redundant and/or confusing information (iii) simplify instrumentation and analysis by reducing computational load. We used nominal 24 h long beat annotation data from normal subjects and patients with congestive heart failure, and linear discriminant analysis to investigate the class discrimination power of these measures.

II. Methodology

A. Experimental Data

The data used in this study are obtained from the "normal sinus rhythm" and "congestive heart failure" RR interval databases from the *PhysioBank* service of "The Research Resource for Complex Physiologic Signals" which is a joint project involving many research and healthcare institutions, under the support of the National Center for Research Resources of the National Institutes of Health.

TABLE I
LONG-TERM HRV MEASURES

Measure	Description	
SDNN	Standard deviation of all NN intervals	
SDANN	Standard deviation of the averages of NN intervals in all 5-minute segments of the entire recording	
SDNN index	Mean of the standard deviations of all NN intervals for all 5-minute segments of the entire recording	
SDSD	Standard deviation of differences between adjacent NN intervals	
Total power	Variance of all NN intervals	(≤ 0.4 Hz)
ULF	Power in the ULF range	(≤ 0.003 Hz)
VLF	Power in the VLF range	(0.003–0.04 Hz)
LF	Power in the LF range	(0.04–0.15 Hz)
HF	Power in the HF range	(0.15–0.4 Hz)

The *PhysioNet*, which hosts PhysioBank and many other services for the biomedical research community, can be reached at http://www.physionet.org. All data included in the PhysioBank are carefully reviewed by the experts.

The RR interval databases at PhysioBank include beat annotation files for long-term (~24 h) ECG recordings that were digitized at 128 samples per second. The beat annotations were obtained by automated analysis with manual review and correction. The normal sinus rhythm database has data from 54 normal subjects (30 men, 24 women) with mean (SD) age of 61.35 (11.65); range, 28 to 73 years. The congestive heart failure (CHF) database has data from 29 subjects (8 men, 2 women; gender is not known for the remaining 21 subjects) with mean (SD) age of 55.28 (11.60); range, 34 to 79 years.

B. Computation of HRV Measures

Essentially, RRITS defined only at the heart beat instants, which are of irregular nature in general. Defining t_i as the time in seconds at which the *i*th (i=0, 1, 2,…) beat occurs, the *i*th RR interval in seconds is $RR_i = t_i - t_{i-1}$. The instantaneous heart rate (IHR) in beats per minute (bpm) is simply $60 / RR_i$. It is clear from these definitions that RR interval and IHR time series contain the same information exactly. Therefore, the use of any of these time series in HRV analysis will produce similar results.

We computed the 9 different HRV measures listed in Table I, which are suitable for studying long-term variations in the HR. We carefully followed all the methodological guidelines listed in [10].

The detection of R waves or QRS complexes in ECG recordings may be hampered by some artifacts, e.g. recording noise, ectopic beats, arrhythmic events, etc. After automatic QRS detection, such artifacts, which may avoid proper detection of R waves, must be identified by experts during manual data revision. The beat annotation files we worked with included this "beat quality" information enabling us to remove the distorting effect of invalid beat times in RRITS computation. (A beat time marked as "non-normal" distorts the RR intervals values for the two neighboring intervals.) While improving data integrity, this so-called "RR interval rejection" operation may introduce a significant selection bias in HRV studies [10].

One way to deal with this problem is to use a suitable interpolation technique to fill in for the missing or rejected data. However, interpolation may affect the values of HRV measures, if used at too many places in the data set. We therefore kept a record of the percentage of RR intervals rejected for each case we studied. This percentage can be considered as an overall indicator of the quality of the data. We observed that in 2 out of 54 data files belonging normal subjects, and 7 out of 29 data files belonging CHF subjects, more than 10% of the RR intervals were eliminated during the RR interval rejection operation. We decided to exclude these cases in the subsequent analysis. In the remaining data sets, visual inspection of the RRITS plots revealed that the quality of the data were quite satisfactory, which led us to conclude that the use of any interpolation to compute the time domain HRV measures was not necessary.

For the computation of frequency domain HRV measures, the RRITS was first interpolated with cubic interpolation at rate of 4 samples per second. Then, the PSD of the interpolated series was estimated using Welch's averaged modified periodogram method [11]. The interpolated series was divided into overlapping segments of length 4000 points, corresponding to a frequency resolution of 0.001 Hz. Each segment was mean subtracted and Hanning windowed before the Fourier transform. The overlap was chosen to be 1200 points, i.e. 5 min.

All the computations were done using in-house software developed under Matlab (The MathWorks Inc., Natick, MA).

C. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a commonly used technique in data classification. LDA may be used for two objectives: either to assess the adequacy of classification, given the group memberships of the objects under study; or to assign objects to one of a number of (known) groups of objects. LDA may thus have a descriptive or a predictive objective.

When used for the predictive objective, LDA projects high-dimensional data onto a line and performs classification in this one-dimensional space. In a two-class problem with n_i (i=1,2) objects in each class, let **x** denote an observation, and **w** be vector in d dimensional space. Then the scalar product $y=\mathbf{w}^t\mathbf{x}$ gives the projected response y. We define the Fisher's criterion [12] that should be maximized over all linear projections **w**, as

$$LD(\mathbf{w}) = \frac{|m_1 - m_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \qquad (1)$$

where m and s represent projected class means and scatters, and the subscripts denote the two classes, i.e.

$$m_i = \frac{1}{n_i}\sum_{y\in \text{Class } i} y\,,\ \tilde{s}_i^2 = \sum_{y\in \text{Class } i}(y-m_i)^2\ .$$

This approach maximizes the distance between the means of the two classes while minimizing the scatter within each class, thereby guaranteeing maximal separability between projected classes. The optimal classification threshold t_{opt} is given as

$$t_{opt} = \frac{1}{2}(m_1+m_2) - \sigma^2\frac{\log(P_1/P_2)}{m_1-m_2} \qquad (2)$$

where P_1 and P_2 denote the class prior probabilities and σ is the pooled estimate of the variance.

When the class membership information is available, Fisher's linear discriminant given by (1) can be used for the descriptive objective, i.e. to assess the degree of separation between classes. The degree separation is simply the ratio of between class scatter to within class scatter.

III. Results

Table II shows the mean (SD) values for long-term HRV measures along with the p-values of the t-test comparing the averages of the two groups and the LDA results. (The HRV measures are sorted in descending order according to LDA results.) The mean (SD) RR interval length in milliseconds (ms) for the normal and the CHF groups are 784.08 (78.23) and 686.78 (86.63) respectively. The corresponding mean heart rates in bpm are 76.52 and 87.36, which indicates that the mean heart rate is elevated by 14.17 % for the CHF subjects. The LDA results indicate that SDNN and SDANN are the two long-term HRV measures with the highest class discrimination power. Fig. 1 shows box-plots of these HRV measures for the both groups. The box plots have horizontal lines at the 25, 50, and 75 percentiles. The vertical lines extending from each end of the boxes show the extent of the rest of the data. The data points marked with the "+" signs are outliers.

Since SDNN and SDANN are identified as the HRV measures with the highest class separation power, one would think that using two variables together would lead to a good classification of the groups. However, a quick examination reveals that there is already a very high correlation (ρ=0.9789) between the two variables and using both is not necessary.

As the next step, we designed a Bayesian classifier to check the performance of the classification based on SDNN only. Bayesian classification assigns objects into classes based on the likelihood, i.e. the class conditional densities, and the prior probabilities to minimize the probability of error. The class conditional densities were obtained by fitting normal densities to each group with a pooled estimate of the variance, which turned out to be 29.96.

TABLE II
CLASS DISCRIMINATION POWER OF LONG-TERM HRV MEASURES

HRV Measure (unit)	Normal G. Mean (SD)	CHF G. Mean (SD)	P-Value	Fisher's LD
SDNN (ms)	138.94 (27.35)	65.81 (35.51)	1.64E–14	0.0827
SDANN (ms)	127.92 (26.95)	57.47 (33.71)	2.24E–14	0.0815
SDNN index (ms)	49.83 (12.40)	28.25 (14.35)	8.23E–09	0.0382
SDSD (ms)	30.19 (25.08)	17.93 (7.85)	2.82E–02	0.0045
HF (ms^2)	312.02 (688.71)	82.91 (100.64)	1.26E–01	0.0022
LF (ms^2)	1137.67 (3194.94)	180.65 (267.65)	1.67E–01	0.0018
ULF (ms^2)	2703.11 (9012.90)	599.50 (560.07)	2.80E–01	0.0011
Total Power (ms^2)	11175.33 (53259.16)	1480.80 1393.27	3.98E–01	0.0006
VLF (ms^2)	7022.53 (40442.24)	617.74 (588.99)	4.62E–01	0.0005

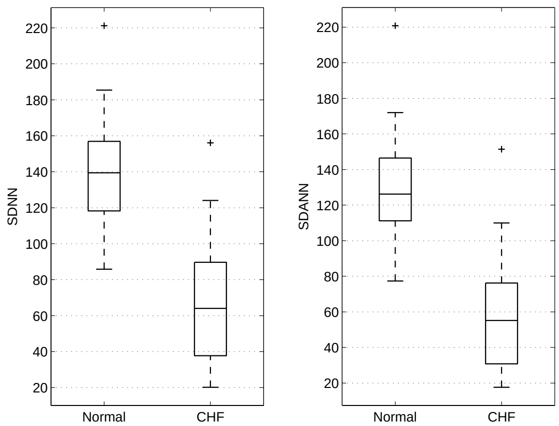

Fig. 1 Box-plots of the SDNN and SDANN HRV measures for the normal and CHF groups.

Fig. 2 shows the estimated class conditional densities scaled by the empirical prior probabilities estimated from the group relative frequencies. The SDNN data for the two groups are also shown. In order to improve the visibility, each class was given a small vertical offset.

The value of the optimal classification threshold can be either obtained as the abscissa of the intersection of the likelihood functions or computed analytically from (2). We computed the optimal threshold t_{opt} as 91.82, e.g. a subject with an SDNN value higher (lower) than t_{opt} is classified as normal (abnormal).

IV. Discussion and Conclusion

In this study, using data from an online resource and linear discriminant analysis, we investigated the class discrimination power of 9 different long-term HRV measures. We identified that the SDNN, standard deviation of all normal-to-normal beat intervals, has the highest class discrimination power. We consequently designed a Bayesian classifier that produced an optimal threshold for SDNN.

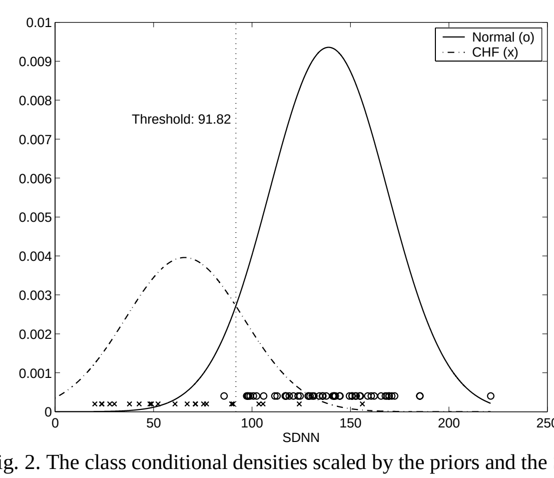

Fig. 2. The class conditional densities scaled by the priors and the SDNN values for the normal and CHF groups. To improve the visibility, each group/class is given a small vertical offset.

To judge the performance of the classifier with respect to the already known group memberships, we noted that 1 subject out of 52 in the normal and 4 subjects out of 22 in the CHF group were misclassified, whereas the remaining 69 subjects are classified correctly. These figures correspond to an observed agreement rate of 93.24%, and sensitivity (true positive) and specificity (true negative) rates of 81.82% and 98.08% respectively. We also computed the kappa statistics, which quantifies the level of agreement, as 0.832 (95% confidence interval: 0.689-0.974). The sensitivity and specificity rates, and the kappa value signify a high degree of agreement.

We only focused on the long-term HRV measures. As information about the sleep or physical activity status of the subjects was not available, we could not compare the class discrimination power of the short-term measures. In future work, we are planning to collect our own data, which will enable us to confirm the results of this study and make further assessments regarding the short-term HRV measures.

References

[1] J. P. Saul, "Beat to beat variation of heart rate reflect modulations of cardiac autonomic outflow," *News Physiol Sci.*, vol. 5, pp.32–37, 1990.

[2] P. B. Corr, K. A. Yamada, F. X. Witkowski, "Mechanisms controlling cardiac autonomic function and their relation to arrhythmogenesis," in *The heart and cardiovascular system*, H. A. Fozzard, E. Haber, R. B. Jennings, A. N. Katz, H. E. Morgan, Ed. New York: Raven Press, 1986, pp. 1343–1403.

[3] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no.3, pp. 230–236, 1985.

[4] V. L. Schechtman, K. A. Kluge, and R.N. Harper, "Time-domain system for assessing variation in heart rate," *Med Biol Eng Comput.*, vol. 26, pp. 367–373, 1988.

[5] M. V. Kamath, E. L. Fallen, "Power spectral analysis of heart rate variability: A noninvasive signature of cardiac autonomic function," *Crit Rev Biomed Eng.*, vol. 21, pp. 245–311,1993.

[6] I. S. Uzun, M. H. Asyali, G. Celebi, and M. Pehlivan, "Nonlinear analysis of heart rate variability," *Annals of Biomedical Engineering*, vol. 30, pp. 30–33, IEEE EMBS Society Annual Fall Meeting, Istanbul, 2001.

[7] R. E. Kleiger, J. P. Miller, J. T. Bigger, A. J. Moss, and the Multicenter Post-Infarction Research Group, "Decreased heart rate variability and its association with increased mortality after acute myocardial infarction," *Am J Cardiol.*, vol. 59, pp. 256–262, 1987.

[8] D. J. Ewing, C. N. Martin, R. J. Young, Clarke BF, "The value of cardiovascular autonomic function tests: 10 years' experience in diabetes," *Diabetes Care*, vol. 8, pp. 491–498, 1985.

[9] R. Frederic, J. M. Gaspoz, I.C Fortune *et al.*, "Screening of obstructive sleep apnea syndrome by heart rate variability analysis," *Circulation*, vol. 100, pp.1411–1415, 1999.

[10] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[11] S. M. Kay and S. T. Marple, "Spectrum estimation: A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification 2nd Edition*, New York: Wiley, 2000.