Automated Selection of Measures of Heart Rate Variability for Detection of Early Cardiac Autonomic Neuropathy

David Cornforth[1,2], Mika P Tarvainen[3,4], Herbert F Jelinek[5]

[1]University of Newcastle, Newcastle, Australia
[2]University of New South Wales, Canberra, Australia
[3]University of Eastern Finland, Kuopio, Finland
[4]Kuopio University Hospital, Kuopio, Finland
[5]Charles Sturt University, Albury, Australia

Abstract

Heart rate variability (HRV) analysis begins with the relatively non-invasive and easily obtained process of ECG recording, yet provides a wealth of information on cardiovascular health. Measures obtained from HRV use time-domain, frequency-domain and non-linear approaches. These measures can be used to detect disease, yet from the large number of possible measures, it is difficult to know which to select, in order to provide the best separation between disease and health.

This work reports on a case study using a variety of measures to detect the early stages of Cardiac Autonomic Neuropathy (CAN), a disease that affects the correct operation of the heart and in turn leads to associated co-morbidities. We examined time- and frequency-domain measures, and also non-linear measures. In all, 80 variables were extracted from the RR interval time series. We applied machine learning methods to separate participants with early CAN from healthy aged-matched controls, while using a Genetic Algorithm to search for the subset of measures that provided the maximum separation between these two classes. Using this subset the best performance was an accuracy of 70% achieved on unseen data.

1. Introduction

Cardiac Autonomic Neuropathy (CAN) is a disease that involves sympathetic and parasympathetic nerve damage leading to arrhythmias and heart attack. An open question is to what extent this condition is detectable by the measurement of Heart Rate Variability (HRV) in asymptomatic individuals or at the preclinical stage.

Heart rate, and its inversely related property heart rate variability, represents a non-stationary non-linear system [1]. In previous work [2] we have shown that entropy measures based on HRV allow people with CAN to be distinguished from controls with good sensitivity and specificity. In this work we report on detection of CAN at an earlier stage, which can greatly assist in management of this disease.

1.1. Heart Rate Variability (HRV)

HRV is commonly used in assessing the functioning of cardiac autonomic regulation [3]. The autonomic nervous system (ANS) regulates heart rate (HR) through sympathetic and parasympathetic branches. Sympathetic activity increases HR and decreases HRV, whereas parasympathetic activity decreases HR and increases HRV [4]. The heart rate (HR) is expressed as the number of beats per minute.

HRV provides information only on the changes in the interval length between heart beats (RR interval) using. using time and frequency based methods [5]. These methods have either focused on the magnitude of RR interval fluctuations around its mean, or on the magnitude of fluctuations in given frequency bands. However more recent analysis methods such as nonlinear methods have shown promise for identifying risk of future morbidity and mortality in diverse patient groups. For example, an estimate of HRV using the standard deviation of RR intervals found that this is higher in well-functioning hearts but can be decreased in coronary artery disease, congestive heart failure and diabetic neuropathy [6]. Although HRV is useful in disease detection, when only a simple derived measure is used, such as the standard deviation of the RR intervals (SDNN), it is no better than the average heart rate and in fact contains less information for risk prediction after acute myocardial infarction [7]. This indicates that more advanced measures of HRV should be explored for detection of .asymptomatic patients to reduce the incidence and prevalence of CAN [8]. Standard time and frequency-domain methods as well as different non-linear methods

have been proposed [9].

1.2. Time-domain measures

Time-domain measures include the SDNN, the number of pairs of successive intervals that differ by more than 50ms divided by the total number of intervals (pNN50%), the Root Mean Square of Successive Differences (RMSSD), the triangular index (Triang), and the triangular interpolation of the interval histogram (TINN).

The Poincaré plot is a visual representation of the time series and is constructed by plotting each consecutive RR interval as a point where $y = RR(t)$ and $x = RR(t-1)$. From this plot a fitted ellipse leads to estimating SD1 (short term variability) and SD2 (long term variability) [10]. An extension is the Recurrence Plot, which represents a sequence of length n as a point in n-dimensional space, then represents similar pairs as points on a two-dimensional space. The Recurrence Rate (REC) is the density of these similar points, Determinism (DET) is the percentage of recurring points identified by diagonal lines, and Lmean is the mean length of diagonal lines exceeding a threshold.

1.3. Frequency-domain measures

Frequency-domain methods divide the spectral distribution into very low, low and high frequency regions. Low frequency power (LF) is believed to be indicative of both parasympathetic and sympathetic activity. High frequency (HF) is indicative of parasympathetic activity. Very Low Frequency (VLF) is a sensitive indicator of management of metabolic processes and reflects deficit energy states [11]. The ratio of low to high frequency components, which is indicative of sympathovagal balance, may also be calculated as well as the total power [12]. A component may also be divided by the total power, to express it in normalized units (n.u.).

1.4. Non-linear measures

Non-linear measures include Detrended Fluctuation Analysis (DFA), which is an estimate of the fractal correlation of the RR interval series and provides an exponent expressing short-term correlations (alpha1) and another expressing long-term correlations (alpha2). The correlation dimension (D2) of fractal analysis was also used. The multi-scale Renyi entropy was introduced and applied to physiologic time series by [13]. Renyi entropy H is a generalization of the Shannon entropy:

$$H_\alpha(X) = \frac{1}{1-\alpha} log_2 \left(\sum_{i=1}^{n} p_i^\alpha \right)$$

where p_i is the probability that $X = x$ and α is the order of the entropy measure. This is the parameter that is varied to produce the multiscale entropy. In this work Renyi Entropy was calculated for using $-5 < \alpha < +5$ by estimating probabilities of sequences of RR intervals with length π (1, 2, 4, 8, 16), with a similarity parameter σ set to (0.01, 0.02, 0.04, 0.08, 0.16).

1.5. Moments

Moments are measures of distribution, in this work the distribution of R-R intervals. The familiar arithmetic mean and variance can be informally viewed as moments of order 1 and 2 respectively, where order refers to the exponent used in calculating these values. Higher order moments can be defined as:

$$m_k = E[(X-\mu)^k]$$

where $E[x]$ is the expectation of X, and μ is the arithmetic mean of the variable X.

1.6. Machine learning algorithms

Machine learning can be used to determine whether measurements pertaining to a study participant indicate Early CAN or Control. A number of algorithms for distinguishing between such classes are available using the Weka toolbox [14].

The Naïve Bayes (NB) algorithm [15] assumes that measures are independent. One would be skeptical of this, but the algorithm performs surprisingly well. It estimates prior probabilities by calculating simple frequencies of the occurrence of each value of each measure, given the class, then returning a probability of each class, given an unclassified set of measures.

Sequential Minimal Optimization (SMO) is a classifier based on the Support Vector Machine (SVM). The SVM builds a set of exemplars that define the boundaries of the different classes. SMO builds on this using polynomial kernels [16].

The Nearest Neighbor (NN) algorithm [17] simply stores samples. When an example is presented to the classifier, it looks for the nearest match from the examples in the training set, and labels the unknown example with the same class.

The Decision Table (DT) algorithm divides the dataset into cells, where each cell contains identical records. A record with unknown class is assigned the majority, or most frequent, class represented in the cell. The goal of training is to find a minimum set of measures that are optimal in predicting the class [18].

An implementation of the classic C4.5 decision tree algorithm [19], known as J48, was used. Numeric attributes are split using a measure of information gain, and this forms two or more branches of the tree. Subsequent splits are used until all branches contain only members of the same class. Then a pruning phase is used to reduce the complexity of the tree.

2. Methods

The aim of these experiments is to determine whether it is possible to build a model that can predict early CAN using measures derived from HRV. We used HRV data from the Charles Sturt University Diabetes Complication Study [21], which consists of 138 adult people, 67 with known CAN and 71 aged-matched controls. Both groups have been assessed for CAN using the Ewing battery. All participants provided a 20-minute recording using lead II ECG, from which the RR intervals were extracted and the measures discussed earlier were derived.

The number of possible interactions between the HRV measures makes it difficult to choose the correct subset by manual inspection alone. We therefore conducted a search both for the optimum classifier algorithm and for the optimum subset. We used an outside cross validation loop to eliminate bias. This was implemented as follows:

1. Perform Wrapper subset evaluation, for each classifier algorithm, using 90% of the dataset, selected at random. The outcome of this step is a subset of measures chosen by the Genetic Algorithm that maximises the classification accuracy on the dataset.
2. Using the 10% of the data not used during the Wrapper process, train and test each of five classifier algorithms using 10-fold cross validation.
3. Repeat 30 times.

Included in the Weka toolbox is a Wrapper Subset Evaluator. This searches through the 80 measures that are being evaluated, building a classifier on a different subset multiple times, and finding the set of measures that gives the best performance. The Genetic search method is based on a simple model of biological evolution [20].

3. Results

Each of the five classifier algorithms produced an optimum set of measures, and this subset was then used for each of the five classifiers to yield a success rate. This was repeated 30 times and the percentage correctly classified was expressed as an average over 30 repetitions. Table 1 shows a list of the most frequently selected measures. There was great variation in the subset chosen for different datasets, and for different classifiers. This suggests that some measures have higher discriminatory power than others. The 2nd moment, or variance, was selected in 85% of the trials, suggesting its important role. However, it was never selected as the only measure to use, indicating that alone it is unable to detect early CAN, and needs to be used in conjunction with other measures. Notice that variants of Renyi entropy were selected many times, indicating the importance of this measure.

Table 1. A list of attributes in descending frequency of use, including only the 15 attributes that were selected in more than 50% of trials. The numbers in brackets after Renyi indicate the values of parameters (π, α) where π is the sequence length and α is the exponent.

Attribute	Selected	Attribute	Selected
Moment 2	85%	Renyi (2, -5)	56%
Renyi (16, -1)	81%	pNN50 (%)	55%
Renyi (1, 1)	74%	Renyi (8, 4)	54%
Renyi (8, -2)	67%	LF power (n.u.)	53%
Renyi (4, 5)	65%	Moment 8	53%
Renyi (1, 4)	61%	Renyi (4, 3)	53%
Renyi (4, -5)	60%	Renyi (4, -2)	52%
Alpha 2	59%		

The results of classification using these subsets are shown in Table 2. Each row shows results for measures chosen using the classifier named in the row headings next to "Wrapper". Each column shows the mean accuracy for 30 trials made using the classifier named in the column headings. So for example, when the subset of measures was chosen using Naïve Bayes (NB) and then trialed using Sequential Minimal Optimization (SMO), the mean of 30 trials was 66.2% correct classification. The best mean result was 71.0% using SMO. This was unaffected by the Wrapper method used, except that Naïve Bayes seemed to consistently produce an inferior subset of measures. For most algorithms, the accuracy is encouraging, considering that a preclinical condition is being identified, and that a random choice would be expected to assign approximately half of all participants to the correct class. The single best result was 93% correct classification, and the corresponding feature set was (in no particular order): pNN50%, HFpower, HFpower(n.u.), SD1, Sample Entropy, Alpha1, DET%, Lmean, Moment2, Moment6, and Renyi entropy values H(1,1), H(1,2), H(1,3), H(2,-5), H(2,-4), H(2,-2), H(2,-1), H(2,1), H(2,5), H(4,-5), H(4,-2), H(4,-1), H(4,1), H(4,3), H(4,5), H(8,-4), H(8,-2), H(8,-1), H(8,1), H(8,3), H(8,4), H(16,-5), H(16,-2), H(16,-1), H(16,3).

4. Conclusions

Based on these results, it appears that it is possible to identify early CAN from analysis of HRV alone. Patients with early CAN could be correctly distinguished from normal participants 71% of the time

Diagnosis by this method is not 100% accurate, and depends upon careful selection both of the classifier algorithm used, and the subset of measures used.

Table 2. Performance of each classifier, expressed as the percentage correct predictions on unseen data records. Abbreviations are defined in section 1.5.

		Method used for testing				
		NB	SMO	NN	DT	J48
Wrapper	NB	67.4%	66.2%	66.4%	60.2%	61.2%
	SMO	65.0%	**71.0%**	70.2%	59.0%	66.7%
	NN	65.0%	**71.0%**	70.2%	59.0%	66.7%
	DT	65.0%	**71.0%**	70.2%	59.0%	66.7%
	J48	65.0%	**71.0%**	70.2%	59.0%	66.7%

The variation of success using different classifier algorithms when using a separate subset for each classifier supports the findings of Kohavi [18], by showing that the measures chosen should be regarded as part of the classifier algorithm. The best measure in the current experiments is the second moment, although this measure alone cannot provide any useful discrimination of diabetics from controls. Using a set of measures suited to the chosen classifier such as the Renyi entropy is very important in enabling this classification.

The success in discriminating early CAN from normal controls in HRV data suggests a methodology that can provide a very simple and quick test and if implemented, would bring great benefits in terms of early diagnosis and consequently a reduction in hospitalization and length of stay.

References

[1] Ivanov P, Rosenblum M, Peng CK, Mietus JE, Havlin S, Stanley HE, Goldberger AL. Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis. Nature 1996;383:323-327.

[2] Jelinek HF, Tarvainen MP, Cornforth DJ. Renyi entropy in identification of cardiac autonomic neuropathy in diabetes Comput Cardiol 2012;39:909–912.

[3] Flynn AC, Jelinek HF, Smith MC. Heart rate variability analysis: a useful assessment tool for diabetes associated cardiac dysfunction in rural and remote areas. Aust J Rural Health 2005;13:77-82.

[4] Berntson GG, Bigger JT Jr, Eckberg DL, et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiol 1997;34:623-48.

[5] Teich MC, Lowen SB, Jost BM, Vibe-Rheymer K, Heneghan C. Heart rate variability: measures and models. In Akay M, editor. Nonlinear Biomedical Signal Processing Vol. II Dynamic Analysis and Modelling. New York: IEEE Press, 2001.

[6] Kleiger RE, Miller JP, Bigger JT Jr, Moss AJ. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. Am J Cardiol 1987;59:256-62.

[7] Abildstrom SZ, Jensen BT, Agner E, et al. Heart rate versus heart rate variability in risk prediction after myocardial infarction. J Cardiovasc Electrophysiol 2003;14:168-73.

[8] TFESC/NASPE. Heart rate variability - Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur Heart J 1996;17:354-381.

[9] Khandoker AH, Jelinek H, Moritani T, Palaniswami M. Association of cardiac autonomic neuropathy with alteration of sympathovagal balance through heart rate variability analysis. Med Eng Phys 2010;32:161-7.

[10] Karmakar CK, Khandoker AH, Voss A, Palaniswami M. Sensitivity of temporal heart rate variability in Poincaré plot to changes in parasympathetic nervous system activity. BioMedical Engineering OnLine 2011;10(17). Available at http://www.biomedical-engineering-online.com/content/10/1/17, accessed Dec 2012.

[11] Fleishman AN. Slow hemodynamic oscillations. The theory, practical application in clinical medicine and prevention. Novosibirsk: Nauka, 1999. Siberian Enterprises RAS.

[12] Task Force of the European Society of Cardiology (TFESC), North American Society of Pacing Electrophysiology, heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circ 1996;93:1043-1065

[13] Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Phys Rev Lett 2002;89:068102.

[14] Witten IH, Frank E. Data mining: practical machine learning tools and techniques with Java implementations, Morgan Kaufmann, San Francisco, 2005.

[15] Bayes T. An essay towards solving a problem in the doctrine of chances. Philos Trans R Soc Lond 1763;53:370–418.

[16] Platt J. Fast training of support vector machines using sequential minimal optimization. In Schoelkopf B, Burges C, Smola A. editors. Advances in Kernel Methods - Support Vector Learning, MIT Press 1998.

[17] Fisher RA. The use of multiple measurements in taxonomic problems, Annu Eugen 7(II) 1936:179–188 (Reprinted in Contributions to Mathematical Statistics, Wiley, 1950).

[18] Kohavi R. The power of decision tables. In Proc Eur Conf Machine Learn, Lect Notes Art Intelligence 914, Springer Verlag, 1995: 174–189.

[19] Quinlan JR. Induction of decision trees. Mach Learn 1986;1:81–106.

[20] Goldberg DE. Genetic algorithms in search, optimization and machine learning. Addison-Wesley, 1989.

[21] Jelinek HF, Wilding C, Tinley P. An innovative multi-disciplinary diabetes complications screening programme in a rural community: A description and preliminary results of the screening. Aust J Primary Health 2006;12:14-20.

Address for correspondence.

David Cornforth.
School of Design, Communication and Information Technology, University of Newcastle, Callaghan 2308 NSW Australia
David.Cornforth@newcastle.edu.au