HEART FAILURE DISCRIMINATION USING MATCHING PURSUIT DECOMPOSITION

Fausto Lucena[1]*, Takeuchi Yoshinori*[2]*, Allan Kardec Barros*[1]*, Noboru Ohnishi*[2]

(1) Laboratory for Biological Information Processing, Universidade Federal do Maranhão, Brazil
(2) Graduate School of Information Science, Nagoya University, Japan.

ABSTRACT

Congestive heart failure (CHF) is a cardiac disease associated with the decreases in cardiac output. As a measure to predict sudden death, we propose a framework for discriminating CHF subjects from normal sinus rhythm (NSR). This framework relies on matching pursuit decomposition to derive a set of features, which are tested in a hybrid genetic algorithm and k-nearest neighbor classifier to select the best feature subset. The performance of the proposed framework is analyzed using both Fantasia and CHF database from Physionet archives which are, respectively, composed of 40 NSR volunteers and 29 CHF subjects. The proposed methodology reaches an overall accuracy of 100% when the features are normalized and the feature subset selection strategy is applied. We believe that our method can be extremely useful to the clinician in primary health care as a support tool to discriminate healthy from CHF subjects.

1. INTRODUCTION

Every year, congestive heart failure (CHF) related diseases are responsible for the death of millions of people around the world. At the clinical level, conventional methods to diagnose heart failure are based on a combination of tests (i.e., Vasalvas manouver, electrocardiography, and chest radiograph) and clinical history to determine whether or not the patient is afflicted with heart failure. Among the tests used (i.e., Framingham, Duke, and Boston), the Boston criteria achieves sensitivity of 50% and specificity of 78% [1]. Electrocardiography methods, such as electrocardiogram (ECG), through the analysis of abnormal ECGs reach sensitivity of 81.14% and specificity of 51.01% [2]. As one can see, the current problem of the conventional diagnose methods is the considerable difference between the percentage of correct and incorrect initial diagnoses. A direct consequence is that false-positives will cause unnecessary tests, whereas the false-negatives will have late diagnostic.

The diagnoses reliability, however, might be increased if the screening test of heart failure could be assisted by signal processing techniques and biomedical analysis. In the past years, several works [3–6] have shown the possibility of classifying subjects with heart failure. For instance, Işler and Kuntalp (2009) using short-term heart rate variability (HRV) intervals have shown that normalizing classical HRV and entropy measures can lead to high levels of sensitivity (82.76%) and specificity (100%). Kampouraki *et al.* (2009) suggested that the classification accuracy of heartbeat time series can be highly improved and even reach maximum accuracy if support vector machines (SVM) are used. A joint wavelet and SVM, for example, yield one of the highest success rate (98.61%) during the task of classifying CHF from normal sinus rhythm (NSR) [5]. Thuraisingham (2009) using second-order difference plot of RR intervals reported the best success rate (100%), but at the cost of long-term RR intervals (24 hours). Despite the number of sample test and methodology used, the proposed techniques have different degrees of complexities. Specifically, they emphasize uncovering patterns that could be used to predict sudden death caused by heart failure. One interesting view of this problem is to find invariant representations that could be considered representative patterns. In [7], for example, the authors show that it is possible to segregate cardiopathies by scaling the behavior of heartbeat intervals using wavelets. Herein, we propose to derive invariant representations using a matching pursuit (MP) decomposition of short-term HRV intervals. The objective, therefore, is to use those features in a pattern classification to segregate pathological from non-pathological groups. Thus, the novelty of our work lies on decomposing HRV short-term intervals into invariant representations, which are used to extract few features to discriminate CHF from NSR according to a simple computational approach (MP algorithm).

2. METHODS

The heart rate variability is a straightforward data to access the neuroregulatory control of the heart by deriving discrete event series from the ECG. The advantages of analyzing the autonomous nervous system (ANS) using HRV are related to the computational simplicity and noninvasive aspects.

Several models of autonomic cardiac regulation are either based on the analysis of input-output relationship [8, 9] or the idea of selective frequency extraction [10]. Altogether, they often explore the standard frequency division suggested to analyze the HRV signals [11]. A simple way to accomplish this task is to use the Fourier transform or autoregressive methods (AR). A drawback, however, is that Fourier and AR meth-

ods are not robust to nonstationarity. An alternative way has been to use time and frequency transformations to overcome nonstationarity. Essentially, one can relaxe the nonstationarity problem by selecting a function that decomposes a signal into a sequence of bases using adaptive time-frequency transform (ATFT) algorithms. This approach is accomplished by scaling, translating, and modulating versions of the basis function, such that they represent the decomposed signal with a well-defined time and frequency distribution. For instance, ATFT algorithms have drawn a lot of attention in pattern classification and signal compression due to its capacity of reducing a higher dimension space to a few number of parameters. One of the most used ATFT algorithms exploits a matching pursuit (MP) decomposition [12]. The MP framework represents a signal $x(t)$ as a linear combination of N basis functions $\phi(t)$ drawn from an overcomplete dictionary $\mathbf{\Phi} = [\phi_1, \ldots, \phi_M]$ where $M \gg N$, or alternatively

$$x(t) \approx \sum_{n=1}^{N} c_n \phi_n(t), \phi(t) = Ae^{-\pi(\frac{t-u}{s})^2} cos(w(t-u)+\varphi), \quad (1)$$

where c_n means modulatory coefficient, s scale, w frequency modulation, u translation, φ phase, and A a normalization factor, such that $\|\phi(t)\| = 1$. Using Gabor functions have several advantages. One may recall that Gabor functions have a compact time-frequency localization and can yield a large variety of shapes. The MP decomposes $x(t)$ by finding the best orthogonal projections amongst a set of basis functions from a dictionary $\mathbf{\Phi}$ that matches the structure of $x(t)$. It results in a finite number of basis functions organized in decreasing order of energy. If the dictionary is a complete representation of the signal, then $x(t) = \sum_{n=1}^{N} \langle R^{n-1}x(t), \phi_n(t)\rangle \phi_n(t)$. One of the intrinsic properties of MP algorithm is regarded to how the signal is decomposed [13]. That is, because not all the signals are composed of well-defined (coherent) components, the MP tends to decompose first coherent underlying structures. And then, break random spike-like noise structures into a set of basis functions whose time and frequency distribution are less compact than coherent ones. Figure 1 illustrates an example of MP decomposition using CHF and NSR HRV waveforms followed by their time-frequency representation.

2.1. Dataset and Main Features

We applied the MP algorithm to decompose HRV intervals derived from CHF patients and NSR volunteers. The CHF dataset is composed of 29 ECG long-recording signals (24 hours) acquired from patients without any control protocol, whose age ranges from 34 to 79 years old. CHF is basically classified by the New York Heart Association [14] into four different classes, each one expressing how the CHF is evolved in terms of physical activity. The NSR dataset is used as a control group. It is composed of 40 ECG waveforms (two hours) recorded from healthy volunteers during supine resting while watching the movie Fantasia (Disney, 1940). This dataset was divided into two groups: young (21-34 years old) and elderly (68-85 years old). Each group contains the same amount of man and woman. Both CHF and NSR datasets were, respectively, digitalized at 128 Hz and 250 Hz. The beats from each ECG were carefully cataloged through unsupervised systems followed by visual inspection of experts.

Comparing CHF and NSR energy decay rate, it is possible to observe (figure not shown) that CHF has a faster decay when compared to NSR. Based on this observation, one can use the mean energy decay as a feature to differentiate between NSR and CHF. Thus, we define the mean energy decay rate as the average of the residual energy, which is derived from the difference between the signal being analyzed and its reconstructed version at each iteration. We express the residual energy rate in function of the iteration number m as $E_r^m(t,w) = E_x(t,w) - \sum_{n=1}^{m} |c_n|^2 \mathcal{W}_{\phi_n}(t,w)$, where $\mathcal{W}_{\phi_n}(t,w)$ is the Wigner-Ville of the $\phi_n(t)$ in the n-th MP iteration with t representing time, w frequency.

A standard measure to analyze the reciprocal relationship between the autonomic branches (SNS and PNS) is the ratio between the LF and HF bands[1] [11]. This ratio has been often used to show the degree of the modulatory mechanisms acting into the heart. It has been reported, however, that patients with CHF have a remarkable reduction of energy at HF bands following a high increase of energy at VLF bands. Therefore, one may expect that dividing the energy at HF by the VLF band causes an enhancement onto this ratio, such that the ratio value for CHF tends to be lower than NSR. The frequency ratio can be obtained by dividing the power spectrum density of HF by the VLF. Herein, we combine the $\phi(t)$ whose center frequencies are located at VLF, LF, and HF to construct sub-signals and thus obtain their PSD. Assuming that the $x(t)$ can be decomposed as a linear combination of N basis functions and weigth coefficients. We can express $x(t)$ as $x(t) \approx c_1\phi_1(t) + c_2\phi_2(t) + \ldots + c_N\phi_N(t)$, where the energy of each component $c_n\phi_n(t)$ is represented by $E_n = |c_n\phi_n(t)|^2$ with total energy $E_x = \sum_n E_n$. If the dictionary is complete, then the probability distribution of $x(t)$ can then be seen as the sum of individual probability contributions given by each component as $p_n(x) = E_n/E_x$. Using the definition of entropy given by Shannon [15], the entropy is obtained by $H_w(p) = -\sum_{n=1}^{N} p_n \log_2(p_n)$.

The frequencies obtained from the structures that decompose the HRV signal, can be used to reflect the frequency distribution of the basis functions according to the HRV frequency band division (i.e., VLF, LF, and HF). To capture these patterns, we use a feature based on the frequency distribution D represented by $\mathrm{D}(w) = M_{cf}(w)/M$, where M_{cf} accounts for the number of basis functions whose central frequency is

[1] High frequency, HF $= \{w | 0.15 < w \leq 0.40$ Hz$\}$, reflects both respiratory sinus arrhythmia (RSA) and vagal modulation (PNS). Low frequency, LF $= \{w | 0.03 < w \leq 0.15$ Hz$\}$, represents sympathovagal rhythms. Very low frequency (VLF $= \{w | 0.00 \leq w \leq 0.03$ Hz$\}$) band remains under ongoing study due to the absence of well-known physiological mechanisms.

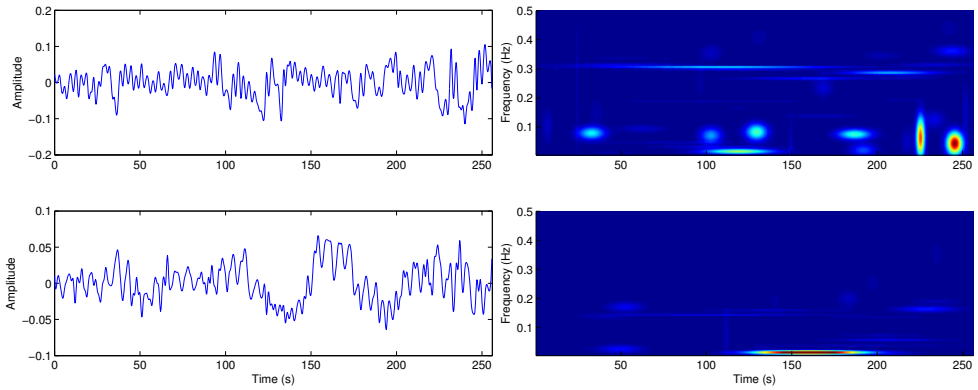

Figure 1: Example. Heart rate variability (zero mean) and its joint time and frequency domain. Each "circle" (in right side figures) represents a Gabor function chosen by the MP algorithm. (*Top*) Normal sinus rhythm. (*Bottom*) Congestive heart failure signal. Both time-frequency planes are normalized to have the same energy levels (scale omitted for better visualization).

either on VLF, LF, or HF bands. Moreover, M represents the total number of basis functions (constraint to 30) that are used to reconstruct the original signal. $\mathrm{D}(w)$ can be understood as how the dynamical behavior of the HRV is captured by the MP algorithm in relation to the frequency distribution of the basis functions.

3. FEATURE SUBSET SELECTION

Feature subset selection (FSS) is a process that deals with the problem of identifying quasi-optimal combination of patterns-representing features among a large set of features. In pattern classification problems, FSS has been used to improve the overall accuracy of the classifier. It can be considered a special case of feature selection where each feature is assigned to a weight value using binary strings. FSS is basically divided in filter or wrapper based-approaches. That is, if there is dependency between the classifier and the learning algorithm, FSS falls under the rubric of the *filter* approach, otherwise it is called *wrapper*. In this work, we use a *filter* approach based on genetic algorithms to select the most suitable subset of features to detect CHF from a control group composed of NSR volunteers.

In the proposed system, we use a genetic algorithm (GA) as learning algorithm. In brief, GAs use principles derived from natural selection and genetics to perform randomized search in complex landscapes. They have been largely used to provide quasi-optimal solutions in optimization problems, such as pattern recognition and machine learning [16].

A supervised classification system based on the k-nearest-neighbor (KNN) rule describes a method where a set of N labeled pattern vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$ (previously assigned to one of the M classes $C_1, C_2, \ldots C_M$) is used to determine which class C_i a new feature vector $\mathbf{x}$ belongs. That is, $\mathbf{x} \in C_i$ if $D(\mathbf{s}_i, \mathbf{x})$ where $D_i^2 = \|\mathbf{x} - \mathbf{s}_i\|^2$. This model uses a distance-weighted rule that does not depend on a true density model [17]. Therefore, the likehood ratio is not used. The classifier discrimination power can be increased by scaling feature values between 0 and 1 using a MinMax procedure.

3.1. Validation and Performance Assessment

A faithful way of estimating the system performance is to use a k-fold cross-validation. In this cross-validation version, the dataset is segregated into k subsets with (almost) of equal size, where k-1 subsets are used to train and the remaining to test. This process is repeated until all the folds are tested and their results averaged. Because the test set is disjoint of the training samples and used just once, the independence between training and test sets is maintained.

Herein the dataset is composed of 69 samples and they were divided into 23 folds, where 66 samples are used as training set and three samples as test per fold time. The averaged results of the test set are then used to evaluate the fitness value θ, which tries to minimize the error rate of the classifier according to, $\theta = 1 - \text{Correctly classified}/\text{Total}$.

Performance measures are results-based decisions traditionally organized into a confusion matrix. This matrix describes if the samples assigned by the classifier to the presence (true) or absence (false) of the disease are in fact correct (positive) or incorrect decisions (false). The three most common performance measures are: sensitivity [Se = TP/(TP+FN)], specificity [Sp = TN/(TN+FP)], and accuracy [Ac = (TP+TN)/(TP+TN+ FP+FN)], where TP, TN, FP, and FN correspond respectively to true positive, true negative, false positive, and false negative. Se, Sp, and Ac are, in this order, connected to the indicative presence or absence of illness, and general performance of the classifier.

The system is basically divided into two stages – preprocessing and processing – where the second stage is composed of three steps: 1) Feature extraction based on matching pursuit algorithm, 2) Feature subset selection using the KNN/GA algorithm, 3) Overall Classification.

In the *first step* of processing, the resulting HRV signal is decomposed using the MP algorithm and its reconstructed signal obtained using 30 basis functions. Using the decomposed basis functions 16 features were extracted, *viz.*, residual energy {E(VLF), E(LF), E(HF), E}, PSD based energy concentration {VLF, LF, HF, HF/VLF, HF+LF}, entropy {H_w(LF) and H_w(HF)}, and frequency distribution {D(LF),

D(HF), D(HF)/D(VLF), D(VLF)/D(LF), D(LF)/D(HF)}. In the *second step*, we used the combined KNN classifier and GA algorithm to simultaneous model optimization for feature subset selection based on the Bioinformatics and Genetic Algorithm Matlab Toolboxes (The Mathworks, 2007).The feature subset selection results are based on a 23-fold cross-validation method whose parameters settings for the binary population size is 300, number of generations is 100, crossover probability ($\mathrm{P_c}$) of 0.7 with a double string crossover, and mutation probability ($\mathrm{P_m}$) of 0.05.

Once the stop criteria is reached – either by succeeding the number of generations or when the fitness value does not decrease in the last 30 generations – the joint KNN/GA optimization algorithm yields the best selected feature subset; that is, the feature subset whose discriminative power has one of the lowest error rate to discriminate CHF from NSR. The *third step* consists of using the selected feature subset to validate the performance of the yielded features.

4. RESULTS

We have tested the discriminative power of the features derived from the MP decomposition with and without a strategy to select the best feature subset. We also investigated if scaling the features, which overcome exaggerated discrepancies among the numeric values, could improve the overall classification rate. The table 1 shows the results, namely accuracy, sensitivity, specificity, and number of features (used or selected). And, it is divided into different configurations where the used k-nearest neighbors in the classifier are 1, 3, 5, 7, 9, 11, and 13. The configurations are organized in (**a**) KNN classifier using all (16) features with feature scaling, (**b**) KNN classifier using all (16) features without feature scaling, (**c**) FSS based on KNN/GA algorithm with feature scaling, and (**d**) FSS based on KNN/GA algorithm without feature scaling. In configuration (**a**), the highest accuracy (95.65%) was obtained with $k = \{3\}$, followed closely by $k = \{1, 5, 7, 9, 11, 13\}$, whose accuracy is 92.75%. Configuration (**b**) yielded a lower accuracy rate (94.20%) than (**a**). Configurations (**c-d**) show a substantial improvement of system accuracy. Specifically, when compared to configuration (**a-b**), the system improvement ranges from 4.35% to 26.09%. For instance, the best accuracy is obtained in configuration (**c**), where the system reached its maximum performance (Ac = Se = Sp = 100%) using only five features. The selected features for $k = \{5\}$ are {D(HF)/D(VLF), D(LF), VLF, E, Hw(HF)}. We show the numeric values of the computed features to CHF and NSR after MinMax scaling in Fig. 2. Despite of their overlapping ranges, frequency distribution D(.) feature was selected as a being a "good" discriminant between NSR and CHF. Analysis of the individual features shows that D(HF)/D(VLF) was spanned over 0.32 ± 0.23 (mean $\pm$ SD) for CHF. The NSR, however, was spread in a much lower range (0.28 ± 0.15). At first sight, D(LF) seems to have a high discriminative power. In fact, their values are distribute over 0.38 ± 0.24 for CHF against 0.61 ± 0.17 for NSR. It has been also reported that energy-based measures derived from HRV signals are strong discriminant features between NSR and CHF. In our case, VLF and LF+HF were selected as subset features. In one hand, VLF has values at 0.10 ± 0.19 for NSR and 0.03 ± 0.07 for CHF. On the other hand, LF+HF has values at 0.18 ± 0.19 (NSR) and 0.02 ± 0.07 (CHF). Another selected feature was the residual energy decay rate (E), which is strongly dependent on the MP algorithm decomposition. Their values are 0.08 ± 0.09 (CHF) and 0.27 ± 0.19 (NSR). Nevertheless, the last feature selected by the joint KNN/GA algorithm is the entropy based on MP decomposition for HF bands with 0.15 ± 0.18 (NSR) and 0.02 ± 0.03 (CHF).

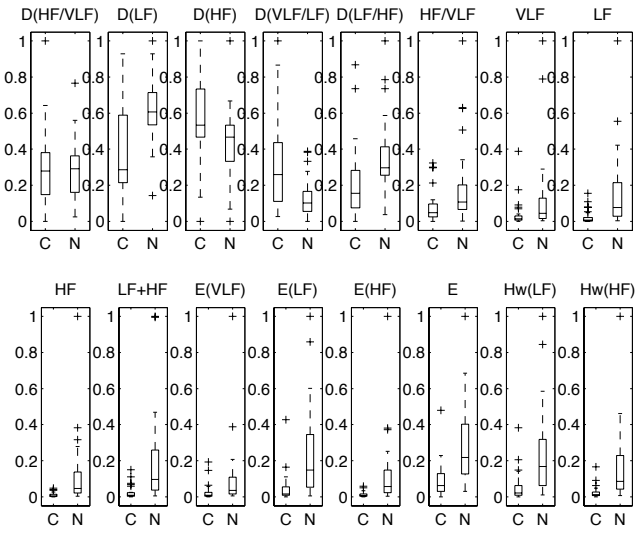

Fig. 2: (Box plot) of features computed for NSR and CHF heartbeat intervals. The central mark represents the median for 40 volunteers (NSR) and 29 subjects (CHF), the edges of the box are the 25th and 75th percentiles.

5. DISCUSSIONS AND CONCLUSIONS

One of the claimed challenges in discriminating CHF from NSR using short-term intervals is that five minutes (or less) may not be enough to fully characterize the day-life activity of the heart. We have shown that using an adaptive decomposition based on the MP algorithm, one can analyze the basis functions used to decompose the signal instead of the HRV signal itself. The novelty of this analysis lies in using the underlying structural complexities of NSR and CHF as discriminatory basis. That is, NSR requires a higher number of noncoherent structures than CHF to be decomposed, which causes a slower decay of energy (E). Moreover, each basis function corresponds to a specific position on the time and frequency plane (see Fig. 1). Their frequencies distribution (D) carries important information about the decomposed signal. We have

Algorithm		With MinMax Normalization				Without MinMax Normalization			
Method	k	Ac(%)	Se(%)	Sp(%)	Features	Ac(%)	Se(%)	Sp(%)	Features
KNN	01	92.75	78.26	100.0	16	73.91	78.26	71.73	16
KNN	05	92.75	78.26	100.0	16	89.85	82.60	93.47	16
KNN	09	92.75	78.26	100.0	16	91.30	82.60	95.65	16
KNN	13	92.75	78.26	100.0	16	94.20	86.95	97.82	16
KNN/GA	01	98.55	100.0	97.82	09	92.75	91.30	93.47	03
KNN/GA	05	**100.0**	**100.0**	**100.0**	**05**	94.20	91.30	95.65	08
KNN/GA	09	98.55	95.65	100.0	05	94.20	91.30	95.65	07
KNN/GA	13	98.55	95.65	100.0	06	94.20	86.95	97.82	07

Table 1: Classification Results. KNN (k-nearest-neighbor) with and without MinMax normalization. KNN/GA (genetic algorithm) optimization with and without MinMax normalization. The results use a 23-fold crossvalidation where Ac (accuracy), Se (sensitivity), and Sp (specificity) quantify the performance assessment of the classifier using N features.

also introduced a flexible way of measuring information from the HRV signals. Computing entropy (H_w) based on the MP algorithm allows one to estimate entropy directly from the decomposed basis functions. This method represents a flexible form to estimate entropy from the standard frequency division (VLF, LF, and HF) than using multiresolution decomposition.

Another relevant problem, which is related to feature selection, was circumvent by using a hybrid architecture (KNN/GA). In this regard, we have shown that configuration (**c**) with $k = 5$ has the lowest error rate, and one of the minor number of features among the other configurations. The selected features by the KNN/GA algorithm yields a subset selection containing five features with high discriminative power. According to Fig. 2 and mean $\pm$ SD of the features, one may organize selected features in decreasing order of discriminative power as Hw(HF), E, LF+HF, VLF,D(LF), and D(HF)/D(VLF). But, it should be noticed that the classification results may vary according to the number of k-nearest-neighbor used or different classifier methods.

REFERENCES

[1] F. Shamsha and J. Mitchel, "Essentials of the diagnosis of heart failure," *Am Fam Physician*, vol. 61, no. 5, pp. 1319–1328, 2000.

[2] C. Fonseca, T. Mota, H. Morais, F. Matias, C. Costa, A. G. Oliveira, and F. Ceia, "The value of the electrocardiogram and chest x-ray for confirming or refuting a suspected diagnosis of heart failure in the community," *Eur. J. Heart Failure*, vol. 4, pp. 807–812, 2004.

[3] Y. Işler and M. Kuntalp, "Heart rate normalization in the analysis of heart rate variability in congestive heart failure," *Proc. IMechE*, vol. 244, pp. 453–463, 2009.

[4] A. Kampouraki, G. Manis, and C. Nikon, "Heartbeat time series classification with support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 13, no. 4, pp. 512–518, 2009.

[5] E. D. Übeyli, "Ecg beats classification using multiclass support vector machines with error correction output codes," *Digital sig proc*, vol. 17, no. 3, pp. 675–684, 2007.

[6] R. A. Thuraisingham, "A classification system to detect congestive heart failure using second-order difference plot of RR intervals," *Cardiology Research and Practice*, 2009.

[7] P. Ch. Ivanov, M. G. Rosenblum, C.-K. Peng, J. Mietus, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Scaling behaviour of the heartbeat intervals obtained by wavelt-based time-series analysis," *Nature*, vol. 383, pp. 323–327, 1996.

[8] R. D. Berger, J. P. Saul, and R. J. Cohen, "Assessment of atonomic response by broad-band respiration," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 11, pp. 1061–1065, 1989.

[9] K. H. Chon, T. J. Mullen, and R. J. Cohen, "A dual-input nonlinear system analysis of autonomic modulation of the heart rate," *IEEE Trans. Biomed. Eng.*, vol. 43, pp. 530–544, 1996.

[10] R. Vetter, P. Celka, J. M. Vesin, G. Thonet, E. Pruvot, M. Fromer, U. Scherrer, and L. Bernardi, "Subband modeling of the human cardiovascular system: New insights into cardiovascular regulation," *Annals of Biomed. Eng.*, vol. 26, pp. 293–307, 1998.

[11] Task Force of the ESC and the NASPE, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, pp. 1043–1065, 1996.

[12] S. G. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. on Sig. Proc.*, vol. 41, pp. 3397–3415, Dec. 1993.

[13] K. Umapathy, S. Krishnan, V. Parsa, and D. G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Trans Biomed Eng*, vol. 52, pp. 421–30, 2005.

[14] The Criteria Committee of the NYHA, *Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels*, pp. 253–256, 9 edition, 1994.

[15] C. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.

[16] R. O. Duda, *Pattern Classification*, Wiley-Interscience, 2000.

[17] Rangaraj M. Rangayyan, *Biomedical Signal analysis: A Case Study Approach*, IEEE Press, 2001.