34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Multiscale Sample Entropy Based on Discrete Wavelet Transform for Clinical Heart Rate Variability Recognition*

Ming-Yuan Lee and Sung-Nien Yu, *Member, IEEE*

***Abstract*—Traditional multiscale method uses coarse grained average (CGA) to evaluate sample entropy (SE) parameters in different scales for signal characterization. In this study, we propose to use discrete wavelet transform (DWT) to decompose hear rate variability signals into multiscale sequences for the calculation of SE features for the recognition of congestive heart failure (CHF) and atrial fibrillation (AF) from normal sinus rhythm (NSR). The support vector machine (SVM) is used as the classifier and the capability of the features are justified with leave-one-out cross-validation method. The results demonstrate that the system using multiscale SE features calculated from both CGA and DWT with five dyadic scales outperforms that based on tradition multiscale method using CGA and 20 scales. Compared to the 5-scale CGA method, the proposed 5-scale DWT method achieved 6.7% and 0.77% increases in the recognition rates for CHF and AF, respectively, and resulted in an 8.35% raise in the overall recognition accuracy.**

I. Introduction

Heart rate variability (HRV) is a powerful tool for diagnosing disorders in the neural control of heart activities. Researchers regularly use linear methods, such as time-statistics analysis and Fourier transform, to characterize HRV signals for diagnosing heart diseases [1, 2]. However, HRV signals usually present more complex patterns which may not be completely described by linear methods. Consequently, researchers sought to apply nonlinear methods to uncover the nonlinear dynamics of HRV. Several approaches stemming from nonlinear dynamics have been employed for this purpose [3-6]. Amongst them, the concept of entropy is associated with the rate of information production and is usually used to quantify the complexity in physiological data [3, 4]. A practical measure of entropy for short data is sample entropy (SE) [7].

Recently, multiscale analysis method, namely coarse grained average (CGA), has been applied to calculate SE from different scales of the signals [8-9]. Complexity measures calculated in this manner was termed multiscale sample entropy (MSE). MSE based on CGA has been demonstrated to successfully differentiate HRV sequences of young from elderly subjects [8]. However, when applying MSE to differentiate the HRV of normal sinus rhythm (NSR) and two pathological states, congestive hearts failure (CHF) and atrial fibrillation (AF), the result was more complicated and may require considering also the specific values of the SE and their dependence on resolutions (scales) [8].

The CGA process shrinks a signal by first segmenting it into pieces of different lengths (scales) and calculates the average of the pieces as the values of the shortened signal. This process is similar to passing a signal through a low-pass filter and performs the down-sampling and mimics the use of discrete wavelet transform (DWT) for signal decomposition [10]. A signal can be decomposed into finer details by multi-level DWT using high-pass and low-pass filters. The DWT has the advantage that the process is reversible such that the signal can be decomposed into subband components and these components can then be reconstructed back to the original signal, which is not possible with CGA.

In the present study, we explore the discrimination power of multiscale SE features for CHF and AF recognition based on DWT. The discrimination power of the MSE features using DWT is compared to that of the original MSE features using CGA in the recognition of CHF and AF from NSR.

II. Multiscale Analysis of Signals

A. Coarse Grained Average (CGA)

The coarse grained average (CGA) method was proposed in [8]. Consider a time series of length N, $\{x_i\}=\{x_1, x_2, ..., x_N\}$, we can construct a coarse-grained averaged (CGA) time series $\{v_j(\tau)\}$, where τ is the scale factor, such that

$$v_j(\tau)=\frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau}x_i,\quad 1\le j\le N/\tau \tag{1}$$

For scale one, the time series $\{y_j(1)\}$ is simply the original time series. For scale factor τ larger than unity, the original time series shrinks into a coarse grained sequence of length N/τ with values calculated from the N/τ non-overlapping segments of length τ. In this manner, CGA performs smoothing and de-correlation of the sequence and is able to resolve complexity on different scales. We follow the CGA method proposed in [8] and generate the first twenty consecutive scales of signal for calculating the SE features.

B. Discrete Wavelet Transform (DWT)

Discrete wavelet transform (DWT) has been widely used in signal processing tasks [10]. The major advantage of the DWT is the providing of great time and frequency localizations. Moreover, DWT allows the decomposition of the signal into different scales, each of which represents particular coarseness of a signal. Amongst the various wavelet

*Resrach supported by National Science Council, Taiwan.

Ming-Yuan Lee is with the Department of Electrical Engineering, National Chung Cheng University, Chia-Yi County, Taiwan (e-mail: a39025309@hotmail.com).

Sung-Nien Yu is with the Department of Electrical Engineering, National Chung Cheng University, Chia-Yi County, Taiwan (corresponding author phone: +886-5-2720411 ext 33205; fax: +886-5-2720862; e-mail: ieesny@ccu.edu.tw).

bases, the db4 wavelet is most similar to the HRV signals and is chosen as the mother wavelet in this study.

The structure of a 4-level DWT for signal decomposition is illustrated in Fig.1 [10].With high-pass filter g(n), low-pass filter h(n), and down-sampling operator ↓ 2, the 4-level DWT was applied to decompose the input signal into subband components. The choice of 4-level DWT is based on the knowledge that subbands A_1 through A_4 correspond to the low-frequency parts of DWT decomposition with 2^1, 2^2, 2^3, and 2^4 points down-sampling, respectively, where $2^4 = 16$ is the largest dyadic integer which is smaller than the scale 20 used in CGA. By using DWT, 5-scale components, including subbands A_1 through A_4 and the original signal, are generated.

III. Calculation Of Sample Entropy

The concept of entropy has been widely used to quantify the complexity of a signal. Traditional entropy-based algorithm usually requires an infinite data series with infinitely accurate precision and resolution [7]. To deal with short and noisy time series, Pincus [11, 12] introduced the approximate entropy (ApEn). Then, Richmann and Moorman [7] modified ApEn and proposed the so-called sample entropy (SE). Costa et al. [8] further combined multiscale analysis with SE for quantifying the complexity of physiologic time series.

Given a time series of length N, $\boldsymbol{x}=\{x_1, x_2, \ldots, x_N\}$, define an m-dimension sequence vector $\boldsymbol{y^{(m)}(i)}=\{x_i, x_{i+1}, \ldots, x_{i+m-1}\}$. Two vector $y^{(m)}(i)$ and $y^{(m)}(j)$ are similar if their distance $d(i,j)=max\{|x_{i+k}-x_{j+k}|: 0 \leqq k \leqq m-1\}$ is smaller than a specified tolerance level δ. By repeating this process, the number of similar vectors can be determined for each of the $N-m+1$ vector $\boldsymbol{y^{(m)}(i)}$. Let $n_i^{(m)}$ represents the number of vectors similar to $\boldsymbol{y^{(m)}(i)}$. The relative frequency of finding a vector $\boldsymbol{y^{(m)}(j)}$ similar to $\boldsymbol{y^{(m)}(i)}$ within a tolerance level δ is calculated as

$$C_i^{(m)}(\delta) = \frac{n_i^{(m)}}{N-m+1} \tag{2}$$

Similar to the definition of entropy, we can taken logarithm of $C_i^{(m)}(\delta)$ and calculate the average value such that

$$H_N^{(m)}(\delta) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C_i^{(m)}(\delta) \tag{3}$$

The approximate entropy can be defined as

$$h_{approx}(\delta, m) = \lim_{N \to \infty} [H_N^{(m)}(\delta) - H_N^{(m+1)}(\delta)] \tag{4}$$

or equivalently using ergodicity

$$h_{approx}(\delta, m) = \lim_{N \to \infty} \frac{1}{m} H_N^m(\delta) \tag{5}$$

Since the evaluation of $h_{approx}(\delta,m)$ usually results in biased statistics [12]. Therefore, a related complex measure, the sample entropy, was usually used instead [7]. In the study, we adopt two sample entropies SE1 and SE2 calculated, respectively, from $h_{sample}(\delta,m)$ with m = 1, 2 and $\delta = 0.15\sigma$, where σ is the standard deviation of the original time series [8]. We then calculated the SE features from the multiscale signals generated by using both CGA and DWT methods.

Figure 1. Four filter-bank implementation of DWT.

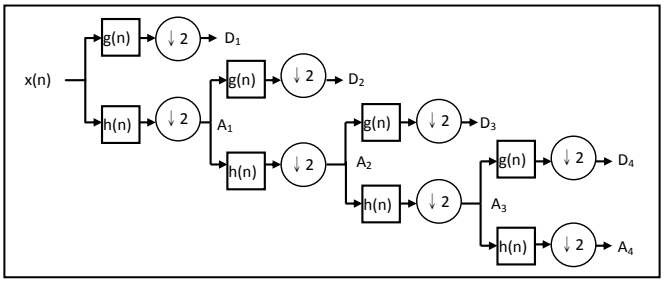
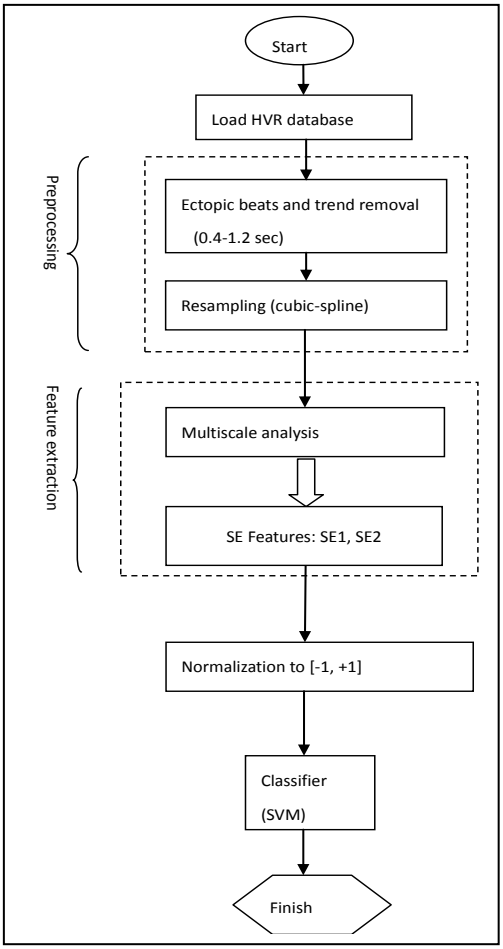


Figure 2. Block diagram of the proposed method.



IV. Experimental Design

A. Database

The block diagram of the proposed system is depicted in Fig. 2. Data for this study were selected from the congestive heart failure (CHF), artrial fibrillation (AF), and normal sinus rhythm (NSR) database; all of them were available on the PhysioNet [13]. Records acquired from 44 CHF, 80 AF, and 72 NSR subjects were obtained from the databases for analysis. Each record comprised also a beat annotation file which showed the occurring time of the specify R peaks confirmed by specialists.

The HRV sequences were generated by first extracting 15-minute segment data in the early morning from each of the records and then calculated the RR-interval (RRI) based on the annotation file. The 15-minute data length was motivated by a recent study [14] which explored the influence of segment length in differentiating pathological HRV. The authors have pointed out that using record segments of 15-minute in length was sufficient for recognition. We further confined the 15-minute data segments to be extracted in the same period of time as an attempt to minimize the influence of natural time cycle.

B. Preprocessing and Feature Extraction

We have developed preprocessors to remove the ectopic beats and trend in the original RRI sequences [15]. This procedure aimed to eliminate outliers, especially the extremely small-valued data, possibly induced by artifacts. The results in [15] demonstrated the effectiveness of the preprocessors in reducing the effect of artifacts while preserving the major properties of the RRI sequences for recognition. The filtered RRI sequence was first interpolated with cubic-spline method and resampled at a rate of 4 samples/sec. The interpolated and resampled RRI sequences were first analyzed with different multiscale method. The two SE features were then calculated from each of the multiscale signals.

C. Performance measures and validation

The performance of the classifier is measured by the recognition rates of CHF, AF, and NSR, respectively. The recognition rate of CHF is defined as the ratio of the number of correctly recognized CHF records to the total number of the CHF records, etc. Moreover, a measure of the classifier's performance, the accuracy (ACC), is the ratio of correctly classified test records to the total number of test records.

The leave-one-out cross-validation method was employed to evaluate the performance of a classifier. This method uses all, except one, samples to train the classifier and then uses the reserved sample to test the performance of the classifier. This procedure repeats until all the samples have been reserved once as testing sample and the percentage of true results is calculated as a measure of classifier's performance. This method tests over the entire database and allows each sample has the same opportunity to serve as the training and the testing sample.

D. Support Vector Machine Classifier

Support vector machine (SVM) maps the training samples from the input space into a higher-dimensional feature space via a mapping (kernel) function [16]. Any product between vectors in the optimization process can be implicitly computed to generate a hyperplan to categorize two classes.

For a training set of instance-label pairs (x_i, y_i), i=0,…l , where $x_i \in R^n$ and y_i=[-1,1] and a non-linear operator mapping with kernel function φ, the optimization problem becomes

$$\min_{w,b,\xi} \frac{1}{2} w^T w + S\sum_{i=1}^{l} \xi_i \text{ subject to } y_i\left(w^T \varphi(x_i) + b\right) + \xi_i - 1 \geq 0, \xi_i \geq 0 \quad (6)$$

where $S>0$ is the penalty parameter for the error term and ξ_i is the set of slack variables that is introduced when the training data is not completely separated by a hyperplane. To solve this problem, Vapink [16] has shown that the solution can be found by minimizing both the errors on the training set (empirical risk) and the complexity of the hypothesis space. Consequently, the decision found by SVM is a tradeoff between error and model complexity. Numerous studies have demonstrated the superiority of using SVM classifier over other classifiers in pattern classification tasks. Consequently, we employ the SVM classifier in the study. The radial basis function (RBF) was empirically selected as the kernel function of the SVM classifier.

E. Experimental protocol

The entire 196 records (44 CHF, 80 AF and 72 NSR) in the database were used in the study. All features were extracted from the records and normalized by first subtracting the mean and dividing with the standard deviation (s.d.) and then passing through a tangent sigmoid function, such that all the features were normalized to be bounded in the same range of [-1, +1]. The normalization process was performed prior to classification as an attempt to eliminate the influence of bias due to the use of different feature scales. The support vector machine (SVM) as classifier was then employed to discriminate the CHF, AF, and NSR HRV sequences with leave-one-out cross-validation method. The effects of different multiscale methods and the discriminating power of the SE features for recognition are investigated.

V. EXPERIMENTAL RESULTS

In recent studies [8, 9], multiscale sample entropy has been successfully used to distinguish CHF and AF from NSR. These studies applied CGA to characterize sample entropy (SE) in different scales and demonstrated an increase in the discrimination power of SE with the scale. In this study, we first quantitatively evaluate the discrimination power of multiscale SE features based on traditional CGA in recognizing pathological HRV and compare the results to that using other multiscale methods. The simulation results are demonstrated separately as follows.

We followed the results in [8] that used CGA to calculate the two sample entropies SE1 and SE2 from 20 scales. This process resulted in a total of 40 SE features. The other multiscale analysis method was the 4-level DWT which provide five dyadic scales, i.e. scales 1, 2, 4, 8, and 16, for the calculation of SE features. We also extracted SE features from the same five CGA scales for comparison. With two SE features calculated from each of the five scales, either of the 5-scale CGA or DWT method had ten SE features. The discrimination power of the SE features calculated using different multiscale analysis methods are summarized in Table 1. Also included for comparison are the recognition rates using SEs calculated only from the original signal.

It is notable that using only two SEs calculated from the original signal contributed to 0%, 67.50%, and 77.78% in recognizing CHF, AF, and NSR, respectively, resulting in 51.62% in ACC. Comparatively, applying SEs calculated using CGA method with the entire 20 scales achieved raised recognition rates of 47.73%, 58.75%, and 79.17% in recognizing CHF, AF, and NSR, respectively, with an ACC of 63.77%. Using the 5-scale CGA and DWT methods further improved the performance of the classifier to higher ACCs of 70.91% and 71.42%, respectively. The two 5-scale methods show miner difference in ACC, i.e. 70.91% with CGA and 71.42% with DWT. However, the 5-scale DWT outperforms the 5-scale CGA with 61.25% vs. 54.55% and 61.25% vs. 60.48% in recognizing CHF and AF, respectively. A 6.7% surplus in CHF recognition is achieved when replacing CGA with DWT in calculating the 5-scale SE features.

VI. Discussions

This study investigated the roles of different multiscale analysis methods in disclosing SE features from HRV signals for clinical purpose. The concept of multiscale analysis stemmed from coarse grained analysis (CGA) of sample entropy (SE) [8]. We expand this concept to subband decomposition based on discrete wavelet transform (DWT) for calculating multiscale SE features to characterize HRV signals for CHF and AF recognition. The results demonstrate that both multiscale analysis methods CGA and DWT are capable of promoting the discrimination power of SE features of the original RRI sequences. However, the original CGA methods with the entire 20 scales do not necessarily outperform that with only five dyadic scales in characterize SE features for classification. Instead, the CGA and DWT methods with five dyadic scales, even using smaller number of SE features, show superior discrimination power when compared to the traditional 20-scale CGA method. This observation may infer the existence of redundancy among the SE features calculated from 20 consecutive scales based on CGA. The use of five dyadic scales raise the effectiveness and efficiency of the SE feature set.

VII. Conclusion

In this study, we applied two multiscale analysis methods, coarse grained average (CGA) and discrete wavelet transform (DWT), to explore the capability of multiscale sample entropy (MSE) for congestive heart failure (CHF) and atrial fibrillation (AF) recognition. The results showed that multiscale SE features calculated from the 5-scale DWT were most promising in differentiating CHF and AF from normal sinus rhythm (NSR). Relatively high recognition rates of 61.25%, 61.25%, and 88.89% were achieved in recognizing CHF, AF, and NSR, respectively, resulting in an ACC of 71.42%. When compared to other well-known studies in the literature [8, 9] which only qualitatively described the discrimination power changes of multiscale SE using CGA, this study quantitatively assess the discrimination power of multiscale SE features using both CGA and DWT methods and highlights the efficiency of using SE calculated with 4-level DWT in characterizing RRI sequence for clinical diagnosis.

TABLE I. Performance of Sample Entropy (SE) Features Calculated from Different Multiscale Methods for CHF and AF Recognition

Feature group	Recognition Rate (%)				
	CHF	***AF***	***NSR***	***ACC***	***NF***
Origional signal	0	67.50	77.78	51.62	2
20-scale CGA	47.73	58.75	79.17	63.77	40
5-scale CGA	54.55	60.48	88.89	70.91	10
5-scale DWT	61.25	61.25	88.89	71.42	10

Acknowledgments

This study was supported in part by the grants NSC 99-2220-E-194-002 and NSC 100-2221-E-194-063 from the National Science Council, Taiwan.

References

[1] Electrophysiology, T.F.o.t.E.S.o.C.a.N.A.S.o.P.a., "Heart rate variability: standards of measurement, physiological interpretation and clinical use", Circulation 93, pp. 1043-1065, 1996.

[2] M. H. Asyali, "Discrimination power of long-term heart rate variability measures", Proceeding of the 25th Annual International Conference of the IEEE EMBS, Cancum, Mexico, 2003.

[3] N. Marwan, N. Wessel, U. Meyerfwld, A. Schirdewan, J. Kurths, "Recurrence-plot-based measures of complexity and their application to heart-rate-variability data", Physical Review E 66, 026702, (2002).

[4] M. Sakki, J. Kalda, M. Vainu, M. Laan, "The distribution of low-variability periods in human heartbeat dynamics", Physica A, vol. 338, pp. 255-260, 2004.

[5] J. E. Skinner, A. L. Goldberger, G. Mayer-Kress, R. E. Ideker, "Chaos in the heart: implications for clinical cardiology", Biotechnol. Vol. 8, pp. 1018-1024, 1990.

[6] C.-S. Poon, C. K. Merrill, "Decrease of cardiac chaos in congestive heart failure", Nature, Vol. 389, pp. 492-495, 1997.

[7] J. S. Richmann, J. R. Moorman, "Physiological time series analysis using approximate entropy and sample entropy", American Journal of Physiology - Heart and Circulatory Physiology, Vol. 278 pp.2039-2049, 2000.

[8] M. Costa, A. L. Goldberger. C.-K. Peng, "Multiscale entropy analysis of complex physiological time series", Physical Review Letters, Vol. 89, 068102, 2002.

[9] R. A. Thuraisingham, G. A. Gottwald, "On multiscale entropy analysis for physiological data", Physica A, Vol, 366, pp. 323-332, 2006.

[10] S. Mallat, "A Wavelet Tour of Signal Processing, 2nd ed"., San Diego: Academic Press, 2008.

[11] S. M. Pincus, "Approximate entropy as a measure of system complexity", Proceedings of the National Academy of Sciences USA Vol. 88, pp. 2297–2301, 1991.

[12] S. M. Pincus, "Approximate entropy (ApEn) as a complexity measure", Chaos Vol.5, pp. 110–117, 1995.

[13] http://www.physionet.org/physiobank/database.

[14] R.-G. Yeh, G.-Y. Chen, J.-S. Shieh, C.-D. Kuo, "Parameter investigation of detrended fluctuation analysis for short-term human heart rate variability", Journal of Medical and Biological Engineering, Vol. 30, pp.277-282, 2009.

[15] M.-Y. Lee, S.-N. Yu, "Improving discriminality in heart rate variability analysis using simple artifact and trend removal preprocessors", *IEEE EMBS* 32nd Annual International Conference, Buenos Aires, Argentina, 2010.

[16] V. N. Vapnik, "The nature of statistical learning theory", New York: Springer-Verlag, 1995.