Construct a Discriminative System for Cardiopathy using Electrocardiogram Data

Chih-Hung Jen†
Department of Information Management
Lunghwa University of Science and Technology, Taoyuan City 33306, Taiwan (R.O.C.)
Tel: (+886) 2-8209-3211 ext. 6319, Email: f7815@mail.lhu.edu.tw

Yu-Ming Lu
Department of Information Management
Lunghwa University of Science and Technology, Taoyuan City 33306, Taiwan (R.O.C.)
Tel: (+886) 2-8209-3211 ext. 6313, Email: yuminglu@ms2.hinet.net

Yi-Min Huang
Department of Information Management
Lunghwa University of Science and Technology, Taoyuan City 33306, Taiwan (R.O.C.)
Email: cody41113@hotmail.com

Bo-Hua Lian
Department of Information Management
Lunghwa University of Science and Technology, Taoyuan City 33306, Taiwan (R.O.C.)
Email: jeff12345442@yahoo.com.tw

Abstract. Cardiovascular disease (CVD) in recent years has become one of the main causes of death in Taiwan, which shows that CVD is now a vital factor affecting the lives of many people. Therefore, the early warning system can quickly distinguish the related symptoms and prevent the sudden cardiopathy that will be an important topic for the mortality can be decreased. In this paper, the electrocardiogram (ECG) data of arrhythmia cases with normal persons would be collected for discriminating the extraordinary ECG signal. Firstly, the empirical mode decomposition (EMD) is used to decompose the ECG signal and selected the appropriate intrinsic mode functions (IMF) transformed to the MSE profile. Using the transformed curves, the research then used the polynomial function to fit these profiles and to execute the monitoring task. On the other hand, the classified result of cardiopathy symptom is also applied to construct the Hotelling T^2 control chart and monitor the estimated parameters in different MSE profiles. The experimental results show that the proposed system exhibited a better performance for detecting the exceptional situation. Moreover, it also provided a stable result for evaluating the cardiopathy.

Keywords: Cardiovascular disease (CVD), electrocardiogram (ECG), empirical mode decomposition (EMD), intrinsic mode functions (IMF)

1. INTRODUCTION

According to the relevant statistics of the Ministry of Health and Welfare in recent years, cardiovascular disease (CVD) still exerts a serious impact on patient health and consumes huge medical resources. Most current researches related to CVD explored the correct diagnosis of CVD via a single auxiliary instrument; while few probed how to prevent and control CVD through a simple and portable electrocardiograph and physiological indicators. Through waveform characteristics, traditional ECG signals can be classified into P, Q, R, S, and T waves. Doctors or professional medical staff can obtain important information from ECG clinical diagnosis. Studies of the previous decade proposed many algorithms to detect QRS waves, including neural networks (Hu et al. 1993; Dokur and

Ölmez 2001; Osowski and Linh 2001; Gűler and Űbeyli 2005; Özbaya et al. 2006), wavelet conversion (Mallat and Hwang 1991; Cuiwei et al. 1995; Dinh et al. 2001), genetic algorithms (Poli et al. 1995), filter groups (Afonso et al. 1999), etc.

Multiscale entropy (MSE) has been applied to analyze relevant signals. Kasper and Schuster (1987) developed the K-S algorithm, which simplified complexity. Richman et al. (2000) continued to improve the disadvantage of entropy, and put forward the concept of sample entropy (SampEn). Costa et al. (2003) developed multiscale entropy (MSE), and calculated its complexity from the multiscale perspective. In a study on ECG signals, Costa et al. (2005) pointed out that MSE analysis was imported after empirical mode decomposition (EMD). The complexity of entropy of the young and healthy elderly while awake was on the rise; and the complexity of the former was significantly higher than that of the latter. During sleep, the complexity showed a downward trend.

From the perspective of "disease surveillance", this study analyzed frequency domain through the ECG signals in a database. First, the original ECG signals were decomposed, via the EMD theory, and combined with appropriate intrinsic mode functions (IMF). Combined with the multiscale entropy (MSE) theory, linear classification, and a control chart, this combination of signals was used to monitor abnormal ECG signals and serve as a basis for disease classification.

2. RESEARCH METHOD

This study used EMD to decompose the original ECG signals, which were combined with screened IMF for MSE curve conversion. Then, the LSD theory was employed to distinguish between normal and abnormal ECG signals with the converted MSE, and abnormal signals were preliminarily classified and analyzed. Meanwhile, the control limit of the MSE curve was established in order to monitor the changes of diseases, and combined with profile monitoring theory (see Figure 1).

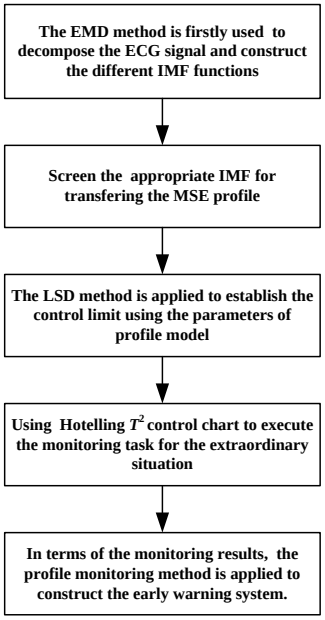

Figure 1: The flowchart for the proposed method

2.1 Empirical Mode Decomposition (EMD)

The process to decompose EMD to obtain each IMF is known as the shifting process. The original signals after the shifting process result in a set of IMF. Restricted conditions are checked; if restricted conditions are not met, decomposition and the shifting process shall be continued in order to obtain the next set of IMF. These steps shall be repeated till they are in line with the restricted conditions. The last set of IMF is the mean value trend. The steps of shifting process are described, as follows:

(1) Identify the local maximum and minimum values of the original signal $X(t)$. Cubic spline was employed to connect local maximum values into a maximum envelope line; and local minimum values into a minimum envelope line. Then, calculate the mean $m_1(t)$ of the two envelope lines. Subtract the original signal $X(t)$ by the mean envelope line $m_1(t)$ to obtain $h_1(t)$ weight, as shown in Eq. (1). Figure 2 refers to the steps from (1) to (2).

$$X(t) - m_1(t) = h_1(t) \quad (1)$$

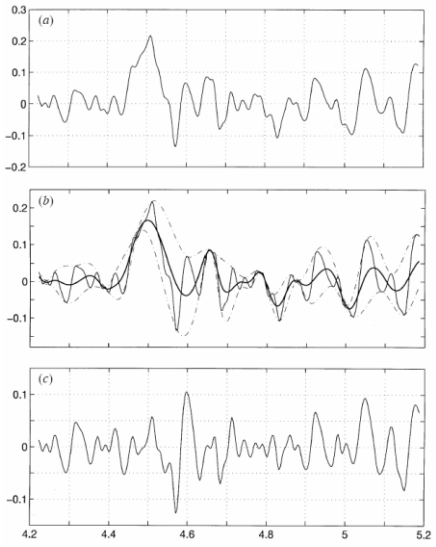

Figure 2: Shifting process diagram (see Huang et al. 1998)

In the above figure; (a) refers to the original signals; solid line (b) refers to the original signals, the dotted line to the maximum and minimum envelope lines and the bold solid line to the mean envelope line; (c) denotes the weight after the original signals were subtracted from the mean envelope line $h_1(t)$. (Huang et al., 1998)

(2) Check whether the $h_1(t)$ weight complies with the restricted constraints. If not, return to step (1). Regard $h_1(t)$ weight as the original signal to continue the second shift to obtain $h_{11}(t)$, as shown in Eq. (2).

$$h_1(t) - m_{11}(t) = h_{11}(t) \tag{2}$$

(3) If the original signal $X(t)$ after k times of repeated shifting reaches the restricted conditions of IMF, it can become a weight $h_{1,k}(t)$ of IMF, as shown in Eq. (3).

$$h_{1,k-1}(t) - m_{1,k}(t) = h_{1,k}(t) \tag{3}$$

(4) Setting convergence and stop conditions for the shifting process

(5) The original signal $X(t)$ was subtracted by the first IMF weight $c_1(t)$ to obtain co-function $r_1(t)$, as shown in Eq. (4).

$$X(t) - c_1(t) = r_1(t) \tag{4}$$

(6) If co-function $r_1(t)$ contains a long period weight, repeat steps (1) to (4) and continue shifting co-function $r_1(t)$ to decompose n IMF weight $c_n(t)$, as shown in Eq. (5).

$$\begin{aligned} r_1(t) - c_2(t) &= r_2(t) \\ &\vdots \\ r_{n-1}(t) - c_n(t) &= r_n(t) \end{aligned} \tag{5}$$

(7) When co-function $r_n(t)$ cannot decompose IMF weight, stop the shifting process. The final co-function of $r_n(t)$ is the mean trend. Finally, sum up all IMF weights $c_n(t)$ and the mean trend to obtain the original signal $X(t)$. In other words, Eqs. (4) and (5) were added to obtain Eq. (6).

$$X(t) = \sum_{k=1}^{n} c_k(t) + r_n(t) \tag{6}$$

2.2 Conversion of IMF function with multiscale entropy (MSE)

The MSE theory is imported to the screened IMF in order to analyze abnormal CVD. First, conduct the coarse-graining procedure on the screened IMF. Obtain one mean value of two value points in order to obtain another serial. Then, calculate the sample entropy, as shown in Eq. (7). The sample entropy of scale2 is obtained. Then, obtain one mean value of the three points of the values in order to obtain another serial. Then, calculate the sample entropy. The sample entropy of scale3 is obtained, the remaining entropy at different times can be completed in the same manner, and add them together to obtain the complexity index (CI). Figure 3 shows MSE curve, while Eq. (8) is the formula to calculate CI.

$$Sample\ En(m,r,N) = -\ln \frac{\sum C_{i(m+1)} / N-m-1}{\sum C_{i(m)} / N-m} \tag{7}$$

$$\begin{aligned} CI &= \sum_{i=1}^{n} Sample\ En(i), \\ i &= scale\ factor,\ n = total\ scale \end{aligned} \tag{8}$$

Then, import the sample entropy after data reduction with the multiscale method; this is the concept of multiscale entropy. Hence, this study assessed the changes of disease via this feature. Regarding operation, first convert the signal graphics of the screened IMF with the MSE theory to

the non-linear profile graphics in Figure 3. Then, select the eigenvalue and monitor the curve for the next stage.

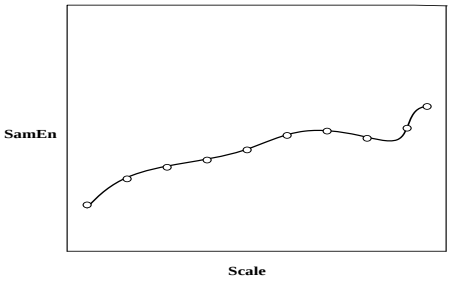

Figure 3: Converted MSE curve

2.3 Classification of abnormal MSE curves with LDA

Abnormal ECG signals were classified by linear classification according to the MSE curve parameters. First, the classification straight line of linear discriminate analysis (LDA) should be established. Hence, the following formula was used to calculate slope $\boldsymbol{w}_c$ and intercept the w_{c0} of the classification line:

$$g_c(\boldsymbol{x}) = \boldsymbol{w}_c'\boldsymbol{x} + w_{c0} \quad \begin{cases} > 0 & \text{if } \boldsymbol{x} \in \text{Group 1} \\ < 0 & \text{if } \boldsymbol{x} \in \text{Group2} \end{cases} \tag{9}$$

The Fisher theory was applied to establish the classification line, where the criteria were the intra-group and inter-group scale values of variation. The slope $\boldsymbol{w}_c$ of the classification line was obtained via the criteria (as shown in Eq. (10)).

$$J_F = \frac{\left|\boldsymbol{w}_c'(\boldsymbol{m}_1 - \boldsymbol{m}_2)\right|^2}{\boldsymbol{w}_c'\boldsymbol{S}_W\boldsymbol{w}_c} \tag{10}$$

Eq. (10) is mainly used to calculate $\boldsymbol{w}_c$ when J reached the maximum value, where $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ are the means of different types of groups; and $\boldsymbol{S}_W$ supposed the pooled within-class sample covariance matrix, which is represented, as follows:

$$\frac{1}{n-2}\left(n_1\hat{\Sigma}_1 + n_2\hat{\Sigma}_2\right) \tag{11}$$

In $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ of the above formula, some are Category 1 ($\omega_{\text{Group 1}}$) (e.g.: ventricular fibrillation) and Category 2 ($\omega_{\text{Group 2}}$), which are roughly similar to the estimated variation matrix $(n_1 + n_2 + 2 = n)$.

3. SIMULATION AND ANALYSIS

This study selected 24 lead I ECG original signals; wherein, two were arrhythmia signals. The original signals were decomposed to obtain several intrinsic mode functions (IMF) of different frequency bands and one residual. The original signals collected by this study could be decomposed into eight IMF and one residual. Through a combination of different IMF, this study found that IMF2, IMF3, and IMF4 have strong ability to distinguish abnormal signals. Hence, we used frequency bands from IMF2 to IMF4 to reconstruct the ECG signals for analysis. Then, we converted the reconstructed signals into MSE via Eqs. (7) and (8). The polynomial model of order 4 was used to conduct model fitting, which is described, as follows:

$$y_p = \beta_{0p} + \sum_{r=1}^{4}\beta_{rp}x^r + \varepsilon_p, \quad p = 1,\ldots,q \tag{12}$$

where β_{0p} and β_{rp} are the estimated parameters of the polynomial model of order 4; where r represents the number of scales. Through the operation of Eq. (12), Table 1 displays the fitting results of the 24 MSE profile models. Abnormal and normal ECG MSE curves were classified through the five parameters ($\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$) of the polynomial model of order 4 and the LSD theory. The results show that, the classification of the parameters can distinguish 100% of the differences between normal and abnormal ECG MSE curves. The results were used to establish the Hotelling T^2 control chart in order to detect on-line anomalies. The test results in Figure 6 clearly show that, profiles 6 and 11 exceed the control limits. The original information corresponding to the two profiles is determined as abnormal ECG. The results also show that the framework proposed by this study can be used to detect abnormal on-line ECG with excellent detection sensitivity.

Table 1: The fitting results of MSE curves

Sample	1	2	3	4	5	6
R^2	0.9902	0.9886	0.9903	0.9910	0.9874	0.9897
R^2_{adj}	0.9872	0.9813	0.9869	0.9896	0.9813	0.9815
Sample	**7**	**8**	**9**	**10**	**11**	**12**
R^2	0.9911	0.9901	0.9788	0.9888	0.9909	0.9910
R^2_{adj}	0.9897	0.9899	0.9704	0.9817	0.9889	0.9887

Sample	13	14	15	16	17	18
R^2	0.9913	0.9874	0.9901	0.9895	0.9899	0.9885
R^2_{adj}	0.9896	0.9827	0.9891	0.9808	0.9832	0.9803
Sample	**19**	**20**	**21**	**22**	**23**	**24**
R^2	0.9913	0.9905	0.9889	0.990	0.9876	0.9914
R^2_{adj}	0.9881	0.9843	0.9834	0.9811	0.9809	0.9897

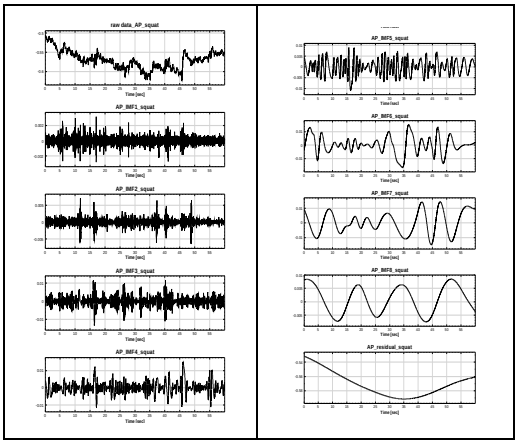

Figure 4: The schematic diagram for decomposition of the original signals

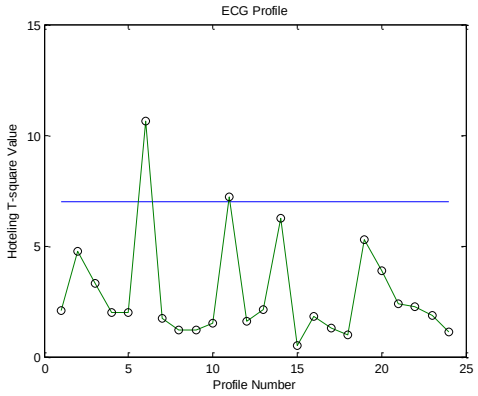

Figure 5: On-line monitoring results of ECG MSE profiles

4. CONCLUSION

In this study, as the ECG signals were non-steady and non-linear, Fourier analysis could not be applied. In view of this, this study adopted EMD combined with IMF to handle lead I original signals and analyze frequency domains. First, EMD was employed to filter the signals. The original COP signals were decomposed into several different frequency bands. The results show that, the weights from IMF2 to IMF4 after decomposition have good distinguishing ability. Then, the reconstructed signals were converted into MSE profiles, while abnormal signals were classified. According to the above simulation results, this study proposed relevant suggestions and notes, as follows:

(1) If the original signals are non-steady, EMD can be used to decompose them into steady signals for analysis with the noise filtering function.

(2) It is a key step to combine and reconstructed the signals of the screened IMF vectors. If IMF weights, which can clearly distinguish different features, cannot be selected, the ECG on-line monitoring cannot be successfully implemented.

(3) It is found that, the parameters of the polynomial model of order 4 converted from MSE profiles can correctly monitor abnormal signals.

(4) If the rebuilt signals have good distinguishing ability, then their multi-variable control chart shall be able to perform on-line monitoring. However, it is worth noting that different types of control charts have different sensitivities. How to choose control charts with fewer types I and II errors for monitoring will be topics worthy of future studies.

ACKNOWLEDGMENTS

This research was partially funded by the National Science Council of the Republic of China (Taiwan), under grant MOST-103-2221-E-262 -013-MY2.

REFERENCES

Hu, Y. H., Tompkins, W. J., Urrusti, J. L. and Afonso, V. X. (1993) Applications of artificial neural networks for ECG signal detection and classification, *Journal of Electrocardiology*, **26**, 66-73.

Dokur, Z. and Ö lmez, T. (2001) ECG beat classification by a novel hybird neural network, *Computer Methods and Programs in Biomedicine*, **66**, 167-181.

Osowski, S. and Linh, T. H. (2001) ECG beat recognition using fuzzy hybrid neural network, *IEEE Transaction on Biomedical Engineering*, **8**, 1265-1271.

Gűler, İ. and Űbeyli, E. D. (2005) ECG beat classifier designed by combined neural network model, *Pattern Recognition*, **38**, 199-208.

Ö zbaya, Y., Rahime, C. and Bekir, K. (2006) A fuzzy clustering neural network architecture for classification of ECG arrhythmias, *Computers in Biology and Medicine*, **36**, 376-388.

Mallat, S. and Hwang, W. L. (1991) Zero-crossing of a wavelets-transform, *IEEE Transaction Information Theory*, **37**, 1019-1033.

Cuiwei, L., Chongxun, Z. and Changfeng, T. (1995) Detection of ECG characteristic points using wavelet transforms, *IEEE Transactions on Biomedical Engineering*, **42**, 21-28.

Dinh, H. A. N., Kumar, D. K., Pah, N. D. and Buurton, P. (2001) Wavelets for QRS detection, *IEEE Transactions on Biomedical Engineering*, **2**, 1883-1887.

Poli, R., Cagnoni, S. and Valli, G. (1995) Genetic design of optimum linear and nonlinear QRS detectors, *IEEE Transactions on Biomedical Engineering*, **42**, 1137-1141.

Afonso, V. X., Tompkins, W. J., Nguyen, T. Q. and Luo, S. (1999) ECG beat detection using filter banks, *IEEE Transactions on Biomedical Engineering*, **46**, 192-202.

Wang, X. and Meng, H. (2008) A 2-D ECG compression algorithm based on wavelet transform and vector quantization, *Digital Signal Processing*, **18**, 179-188.

Kasper, F. and Schuster, H. G. (1987) Easily calculable measure for complexity of spatiotemporal patterns, *Physical Review A*, **36**, 842-848.

Richman, J. S. and Moorman, J. R. (2000) Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of Physiology Heart and Circulatory Physiology*, **278**, 2039-2049.

Costa, M., Hausdorff, J. M., Goldberger, A. L. and Peng, C. K. (2003) Multiscale entropy analysis of human gait dynamics, *Physica A*, **330**, 53-60.

Costa, M., Goldberger , A. L. and Peng, C. K. (2005) Multiscale entropy analysis of biological signals, *Physical Review E*, **71**, 021906-1 - 021906-18.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C. Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C. and Liu, H. H. (1998) The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society London*, **454**, 903-995.