16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

SELECTION OF HIGHER ORDER SUBBAND FEATURES FOR ECG BEAT CLASSIFICATION

Sung-Nien Yu and Ying-Hsiang Chen*

Department of Electrical Engineering, National Chung Cheng University
168 University Road, Ming-Hsiung, Chia-Yi County 621, Taiwan
phone: + (886) 5-2720411 ext 33205, fax: + (886) 5-2720862, email: yusn@ee.ccu.edu.tw

ABSTRACT

Five levels of discrete wavelet transform are applied to decompose the ECG beat signal into six subband components. Higher order statistics proceeds to calculate valuable features from the three midband components. These features together with three RR interval-related features construct the primary feature set. A feature selection algorithm based on correlation coefficient and Fisher discriminality is then exploited to eliminate redundant features from the primary feature set. The feedforward backpropagation neural network (FFBNN) is employed as the classifier to justify the capacity of the method.

The proposed method achieved an imposing ECG beat discrimination rate of more than 97.5%. By using the feature reduction method, the feature dimension can be readily reduced from 30 to 18 with negligible decrease in accuracy. Compared with other methods in the literature, the proposed method improves the sensitivities of most beat types, resulting in an elevated average accuracy. The results demonstrate the effectiveness and efficiency of the proposed method in ECG beat classification.

1. INTRODUCTION

Electrocardiogram (ECG) is a vital signal for clinical diagnosis of heart diseases. Many algorithms have been developed to improve the accuracy of ECG beat classification [1-7]. Among them, the application of higher order statistics (HOS) has been shown to be relatively insusceptible to the variation of ECG morphology among different patients [1, 2]. On the other hand, the wavelet transform (WT) opens another category of methods that represents the signal in different translations and scales [8, 9]. The discrete wavelet transform (DWT) decomposes a signal into a variety of subband components, which provides an opportunity to extract features from components with distinct frequency distributions. The advantages of the two techniques motivate the application of HOS to the DWT-decomposed subband components for the extraction of effective features.

However, considering the dramatic increase of computational complexity with feature dimension, we need to select the most representative features from the original feature set. Recent studies on feature selection can be virtually divided into two categories, namely filters and wrappers [10-12]. Both of them require recursive iterations in the selection process and are considered too complicated for this study.

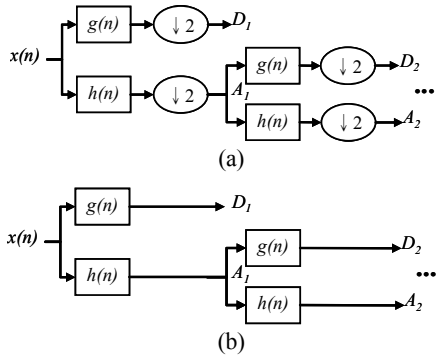

Figure 1 - Two schemes for DWT implementation, in which *g(n)* is a high-pass filter and *h(n)* is a low-pass filter. (a) Mallat's algorithm (b) *à trous* algorithms.

Consequently, we intend to design a one-pass filter which does not require iterative procedures. The classification results demonstrate that the proposed feature selection method is simple and effective in selecting a condensed feature set while preserving the classification accuracy.

2. THEORETICAL BACKGROUND

In this section, we review some background knowledge of discrete wavelet transform and higher order statistics. The interested reader should refer to [7-9] for more details.

2.1 Discrete Wavelet Transform

Wavelet transform (WT) is a powerful tool to represent a signal in different translations and scales. As for discrete-time signals, the dyadic discrete wavelet transform (DWT) can be implemented by low-pass, *h(n)*, and high-pass, *g(n)*, FIR filters. Figures 1 (a) and 1 (b) respectively illustrate the Mallat's and the *à trous* schemes for DWT implementation. The major difference between them is the downsamplers, denoted as ↓2, following the filters in Mallat's scheme, which removes the redundancy in the filtered signal, yet also reduces the temporal resolution [8]. On the contrary, the *à trous* algorithm, as depicted in Fig. 1 (b), reserves the temporal resolution at the expense of larger memories. Since the calculation of HOS requires signals with sufficient length, we employ the *à trous* algorithm in the study.

2.2 Higher Order Statistics

Higher order statistics plays an important role in statistical signal processing, especially for stationary random proc-

ess. Let *x(n)* be a real, discrete time random process. The moments of *x(n)* are defined as the coefficients in the Taylor expansion of the moment generating function [7]

$$\Phi_{\mathbf{x}}(w) = E\left[\exp(jw\mathbf{x})\right]. \quad (1)$$

In practice, the n^{th} order moment can be calculated by taking the expectation of the process multiplied by the *(n-1)* lagged version of the same process, such that

$$\begin{aligned} m_{1x} &= E\left[x(n)\right] \\ m_{2x}(\tau) &= E\left[x(n)x(n-\tau)\right] \\ m_{3x}(\tau_1,\tau_2) &= E\left[x(n)x(n-\tau_1)x(n-\tau_2)\right] \\ m_{4x}(\tau_1,\tau_2,\tau_3) &= E\left[x(n)x(n-\tau_1)x(n-\tau_2)x(n-\tau_3)\right] \\ &\vdots \end{aligned} \quad (2)$$

The second characteristic function of *x(n)*, defined as

$$\chi(w) = \ln\Phi_{\mathbf{x}}(w) = \ln E\left[\exp(jw\mathbf{x})\right] \quad (3)$$

is called the cumulant generating function, and the coefficients in its Taylor expansion are the n^{th} order cumulants of *x(n)*, represented as $c_{nx}(\tau_1,\tau_2,\cdots,\tau_{n-1})$. With zero-mean assumption, the first three cumulants are equivalent to their moment counterparts, i.e. $c_{1x} = m_{1x}$, $c_{2x} = m_{2x}$, and $c_{3x} = m_{3x}$. Furthermore, the 4th order cumulant can be calculated from the 2nd and 4th order moments [1], such that

$$\begin{aligned} c_{4x}(\tau_1,\tau_2,\tau_3) = {} & E\left[x(n)x(n-\tau_1)x(n-\tau_2)x(n-\tau_3)\right] \\ & -c_{2x}(\tau_1)c_{2x}(\tau_3-\tau_2) - c_{2x}(\tau_2)c_{2x}(\tau_3-\tau_1) \\ & -c_{2x}(\tau_3)c_{2x}(\tau_2-\tau_1) \end{aligned}. \quad (4)$$

3. PROPOSED METHOD

3.1 Feature Extraction

Features extracted directly from the cumulant functions, as in [1], are usually highly correlated, e.g. the 2nd cumulant is symmetric. It is, therefore, important to use cumulant-related features that are less correlated to one another. Additionally, the RR interval is important in characterizing many clinical ECG beat types. Therefore, we recruited four sets of cumulant-related features and three RR interval-related features in this study.

Denoting the j^{th} order cumulant of the D_i subband as c_{ij}, where $i \in \{3, 4, 5\}$ and $j \in \{2, 3, 4\}$, the features are defined as follows.

1) Standard Deviation of the Cumulant (CSD): The variance of a cumulant represents the energy within the cumulant. With time lag L, the variance of a cumulant is defined as

$$\sigma_{ij} = \sqrt{\frac{1}{2L}\sum_{l=-L}^{L}\left[c_{ij}(l) - \overline{c}_{ij}\right]^2} \quad (5)$$

where $\overline{c}_{ij}$ is the sample mean of the cumulant and l is the time shift ranging from $-L$ to $+L$.

2) Normalized Summation (NS): The normalized summation is defined as the summation of a cumulant divided by the area between the cumulant function and the zero line. For a cumulant c_{ij} the normalized summation NS_{ij} is defined as

$$NS_{ij} = \sum_{l=-L}^{L} c_{ij}(l) \Big/ \sum_{l=-L}^{L} \left|c_{ij}(l)\right| \quad (6)$$

which ranges between -1 and +1, depending on the relative allocation of the function in the negative and positive directions.

3) Number of Zero-Crossings (NZC): The number and position of zero-crossing are important in characterizing a signal. We considered the number of zero-crossing in cumulants c_{52}, c_{53}, and c_{54} in the feature vector.

4) Symmetry (SYM): The symmetry of a signal is defined as

$$SYM_{ij} = \sum_{l=1}^{L}\left|c_{ij}(l) - c_{ij}(-l)\right| \Big/ \sum_{l=-L}^{L}\left|c_{ij}(l)\right| \quad (7)$$

which equals zero with the 2nd order cummulent. Therefore, we consider only the *SYM* of the 3rd and 4th order cumulants.

5) RR Interval-related Features: The RR interval is defined as the time elapse between two adjacent R peaks. Certain ECG arrhythmias, such as PVC, APB, VEB, and VFW, show shorter or irregular RR intervals. In this study, we exploited three RR interval-related features, including the instantaneous RR interval, the ratio between the instantaneous and the previous ones, and the ratio between the pervious and the one before it.

In summary, the feature vector contains 30 features, including nine *CSDs*, nine *NSs*, three *NZCs*, six *SYMs*, and three RR interval-related features.

3.2 Normalization

A normalization process is necessary to standardize all the features to the same level. The hyperbolic tangent sigmoid function [4] is used to transform each feature to the same range of [-1, +1]. The mean and the standard deviation of each component in the feature vectors are calculated from the training dataset and used throughout the experiments.

3.3 Feature Selection

Feature selection explores the problem about how to select a subset of variables that can efficiently delineate the primary data. Features that are highly correlated with others are determined to be redundant. In this study, we apply the correlation coefficient to measure the dependency between two features and use the Fisher discriminality [13] as the measure of relevance between each pair of features and attributed classes.

If a feature is highly correlated with another, one of them must be redundant. The correlation coefficient between the k^{th} and the l^{th} features is defined as

$$\rho_{kl} = \frac{\mathrm{COV}(\mathbf{f}_k, \mathbf{f}_l)}{\sigma_k \sigma_l} \quad (8)$$

where $\mathrm{COV}(\mathbf{f}_k,\mathbf{f}_l)$ is the covariance, and σ_k and σ_l are the standard deviations of the k^{th} and l^{th} features, respectively. Since we concern only the predictability between features, we adopt the absolute correlation coefficient, $|\rho_{kl}|$, as the measure.

For the *c*-class problem, assume the feature vector under classification is denoted as $\mathbf{f}_k$ for $k=1, 2, \ldots, p$, where $\mathbf{f}_k=[f_{k1},$

$f_{k2}, \ldots, f_{kN}]^T$ and N is the number of samples in the feature set. The predictive power of the k^{th} feature based on the Fisher discriminality is defined as

$$S_k = \frac{S_B}{S_W} \tag{9}$$

where S_B and S_W are respectively the between-class and within-class scatters that are defined separately as

$$S_B = \sum_{i=1}^{c} n_i \left(\bar{\mathbf{f}}_{ki} - \bar{\mathbf{f}}_k \right) \tag{10}$$

$$S_W = \sum_{i=1}^{c} \sum_{f_{ki} \in D_i} n_i \left(f_{ki} - \bar{\mathbf{f}}_{ki} \right)^2 \tag{11}$$

where n_i is the number of features in class i, D_i is the feature set associated with class i, and $\bar{\mathbf{f}}_{ki}$ and $\bar{\mathbf{f}}_k$ are the means of the k^{th} features in class D_i and the entire feature set, respectively.

For a p-dimensional feature set, totally $p(p-1)/2$ correlation coefficients are calculated. These correlation coefficients are first sorted in descending order. Then, following the sorted order, each pair of features is examined carefully and the feature with lower predictive power is considered redundant and eliminated from the feature set. This process continues until the predetermined number of features is reached. This method eliminates the need for iterative feature selection and is considered much simpler than many other methods in the literature. Experiments are designated to justify the effectiveness and efficiency of the method.

3.4 Nonlinear Neural Classifier

Feed-forward backpropagation neural network (FFBNN) was employed as the classifier of the system. FFBNN is a well-known neural network which is powerful in generalizing training data into complex nonlinear discrimination functions [14]. The typical structure of a FFBNN classifier consists of three layers, namely the input, hidden, and output layers. Hyperbolic tangent sigmoid function is used as the activation function and the weights between layers are modified by propagating the error signals layer by layer backwardly. The number of neurons in the hidden layer was empirically chosen as sixty [4].

4. EXPERIMENTAL DESIGN

Experiments were designated to test the capacity of the proposed method. First of all, the capabilities of the proposed feature extractor and classifier were explored. Secondly, the effect of the feature selection procedure based on feature correlation was investigated.

4.1 Experimental Procedure

QRS segments with a length of 64 points centered at R peak were extracted from the MIT/BIH arrhythmia database [15] according to the sampling profiles listed in Table 1. The DC value of the non-zero mean QRS segment was subtracted. The DC-free signals were than decomposed into six subband signals by five levels of DWT. In view of its shape similarity to that of regular QRS complexes in ECG, the 'sym6' basis was employed as the mother wavelet [5]. To suppress the influences of different artifacts with white or colour spectra, only the three midband components, i.e. D3, D4 and D5, were considered. The higher order statistics were then applied to delineate them. Thirty features, including nine *CSDs*, nine *NSs*, three *NZCs*, six *SYMs* and three RR interval-related features described in section 3.1 were exploited as input features to the nonlinear classifier. These features were normalized as depicted in section 3.2. And finally, a non-linear FFBNN was employed to discriminate the ECG beats. Since the classification result is affected by initial weights, we repeated each experiment setup for five times and the results were averaged.

Table 1 - The profile for sample selection: multiple beat types are selected from each record

RECORD	ECG BEAT TYPES						
	NSR	LBBB	RBBB	PVC	APB	VEB	VFW
100	292	0	0	0	31	0	0
105	330	0	0	24	0	0	0
106	196	0	0	325	0	0	0
109	0	368	0	23	0	0	0
111	0	315	0	1	0	0	0
114	238	0	0	27	10	0	0
116	300	0	0	68	1	0	0
118	0	0	388	10	96	0	0
119	202	0	0	272	0	0	0
124	0	0	274	29	2	0	0
200	227	0	0	505	30	0	0
207	0	220	13	57	107	105	472
209	342	0	0	0	381	0	0
212	123	0	325	0	0	0	0
214	0	297	0	159	0	0	0
Total	2250	1200	1000	1500	658	105	472

The method of data sample selection can significantly influent the result in ECG beat classification. We adopt the profile that extracts multiple beat types from each of the records [16], as depicted in Table 1. Seven beat types were extracted from the 15 selected records with the profile. Half of the samples which are evenly distributed in the data sets, were used to train the neural classifier and the other half to justify the performance of the setups. It is worth noting that there is great conformity between the 2nd profile and that associated with FHyb-HOSA [1], such that the performance can be compared in a fairer manner.

4.2 Feature Dimension Reduction

The one-pass feature selection method described in section 3.3 was utilized to extract the most effective features for ECG beat discrimination. Due to the similarity between the profile and that used in FHyb-HOSA, we applied the method to reduce the feature dimensions to the same size and their results were compared. The redundancy of the feature space was measured by the pair-wised maximal correlation coefficient (MCC) of the remaining feature set. The capability of the feature selection method was justified by using both the averaged classification accuracies and pair-wised MCCs.

5. RESULTS AND DISCUSSION

5.1. Statistical Accuracy Measures

The performance of the classification system is determined by the following statistical measures:

$$specificity = \frac{\#\text{ of correct classified normal beats}}{\#\text{ of total normal beats}} \quad (12)$$

$$sensitivity = \frac{\#\text{ of correct classified abnormal beats}}{\#\text{ of total corresponding abnormal beats}} \quad (13)$$

$$accuracy = \frac{\#\text{ of correct classified beats}}{\#\text{ of total beats}} \quad (14)$$

where # represents "the number".

The classification results are summarized in Table 2. Also included is the method using higher order statistics and a more complicated fuzzy hybrid neural network, denoted as FHyb-HOSA [1]. The proposed method outperforms FHyb-HOSA in most beat types, resulting in 1.47% elevated average accuracy. The superiority is more significant especially in the recognition of RBBB, VEB, and VFW arrhythmias. Since both methods were applied to discriminate the same types of ECG beat types with similar sample selection profiles, it is evident that the application of higher order statistics in wavelet decomposed subband components can reliably improve the discrimination power of the classifier.

5.2. Reliability of the Feature Selection Method

The proposed feature selection method described in section 3.3 was applied to condense the feature space from a dimension of 30 to 24, 18, and finally 12, respectively. The comparative results are depicted in Fig. 2, in which the bars show the maximal correlation coefficients (MCCs) measured with the right vertical axis and the curve is the accuracies measured with the left vertical axis.

The MCC of pair-wised features reveals the redundant information of the remaining feature set. A large MCC implies that at least one pair of features is highly correlated and one of them can be removed from the feature set. The entire feature set produces a MCC larger than 0.95, which inferred the existence of redundancy in the feature set. After eliminating the redundant features, the MCC decreased dramatically while the classification accuracies only slightly affected. As depicted in Fig. 2, reducing the dimension of feature set to only 18, the MCC reduces to about 0.7, while high averaged accuracies of 97% still retained. Therefore, we can claim that the proposed feature selection method is effective in eliminating the redundant features, and preserving the effective features as the same time, for ECG beat classification.

Further examining the results in Fig. 2, both the accuracy and MCC decrease significantly when the feature dimension decreases from 18 to 12. This degradation implies that the preserved features are insufficient to discriminate different beat types. Thus, the size of 18 is determined an appropriate feature dimension to be used with the profile employed for ECG beat classification.

5.3. Comparison to Other Systems

Considering the similarity in the sample selection strategies of the proposed method and the well-know FHyb-HOSA method [1], it is possible to more objectively compare their performances in discriminating specific beat types. The results of using different feature dimensions with the method are listed in Table 2 for comparison. As the feature dimension have been reduced from 30 to 18, only minor deviation in the accuracies were observed in discriminating specific ECG beat types. The observation copes with the overall accuracies revealed in Fig. 2. It is worth noticing that when the considered feature dimension is restricted to 18, the proposed beat classification method outperforms FHyb-HOSA in most of beat types, resulting in 1.22% elevated average accuracy. The superiority is more significant when we highlight the higher than 4% sensitivity improvements in the recognition of RBBB, VEB, and VFW arrhythmias. Hence, we confidentially claim that the feature selection method is effective in removing the redundant features from the original feature set, which makes the proposed method even more efficient in discriminating ECG beat types.

It is also interesting to compare the proposed method with that of other studies. Three effective methods, including the FHyb-HOSA [1], MME [3], and Neuro-Fuzzy [2], were selected for comparison. Among them, only FHyb-HOSA described the detailed originality of the sampled beats. The feature dimension of the proposed method was restricted to

Table 2 - Comparative results of the proposed method with different feature dimension and the FHyb-HOSA method

		Proposed Method				FHyb-HOSA [1]
Dimension		30	24	**18**	12	18
Specificity (%)		97.83	97.69	**97.80**	96.30	98.10
Sensitivity (%)	LBBB	98.76	98.60	**98.60**	94.53	97.00
	RBBB	99.20	98.88	**98.68**	96.92	94.00
	APB	91.25	92.04	**91.92**	90.40	91.33
	PVC	97.65	97.12	**96.72**	93.79	96.57
	VEB	95.38	95.38	**95.00**	83.46	90.00
	VFW	98.56	98.05	**98.22**	95.93	94.50
Accuracy (%)		97.53	97.36	**97.28**	94.82	96.06

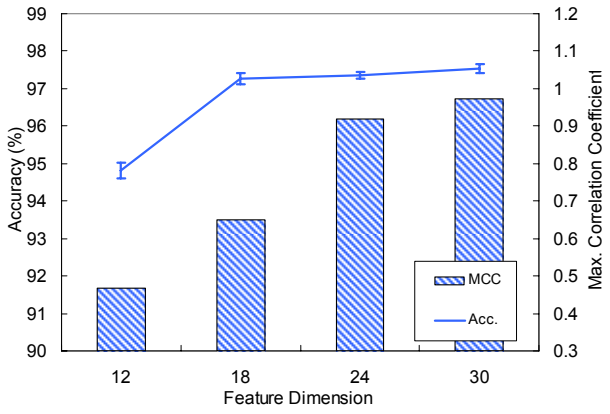

Figure 2 - Comparison of the classification accuracies (Acc., represented as bars referring to the left axis) and the maximal correlation coefficients (MCCs, represented as curves referring to the right axis), at different feature dimensions.

Table 3 - Comparison of the proposed method with the other ECG beat classification methods

Method	Number of beat types	Accuracy
Proposed	7	97.28 %
FHyb-HOSA [1]	7	96.06 %
MME [3]	5	97.78 %
Neuro-Fuzzy [2]	4	98.00 %

18 to conform to the FHyb-HOSA method. The comparative results are summarized in Table 3. It is obvious that the proposed method outperforms FHyb-HOSA. When compared with the other two methods, the proposed method is compatible with them in overall accuracy while possessing the capacity to discriminate more (seven) beat types than five in MME [3] and four in Neuro-Fuzzy [2]. The results support the effectiveness of using subband features based on higher order statistics in discriminating ECG arrhythmias.

With the success of the method, three related topics are to be included in our future studies. The first topic is about classifiers. We have used probabilistic neural network (PNN) in our other work [4]. PNN is characterized by its rapid learning in the training phase. However, it suffers from the requirement of huge memories for the training samples. In this study, we employed the popular FFBNN as classifier. The comparison of their effectiveness and efficiency in ECG beat classification is an interesting subject for further study. The second topic is about the noise-tolerant capacity of the method. Since the measurement of ECG signals usually suffers from environmental noises and experimental artefacts, it is important to develop ECG beat classifiers that are noise-tolerant. We will investigate the influence of noises on the ECG beat classification method in our future studies. Thirdly, we exploited a simple feature selection in this study. Since this method only applies simple linear correlation measures for feature selection, it is worthwhile to also compare its capacity to the other feature selection methods, which will be our further work.

6. CONCLUSION

In this study, we proposed a novel framework to characterize ECG betas by applying higher order statistics to subband components. A simple one-pass feature selection method was proposed to eliminate the redundant features in the primary feature set.

The results demonstrated that different pathological changes in ECG beats can be effectively represented by the features extracted from the higher order statistics of the wavelet transformed subband components. The feature selection method eliminate the redundant features from the original feature set while preserving high accuracies of more than 97% even at dimension of 18, which outperforms the well-know FHyb-HOSA method in a more objective manner.

ACKNOWLEDGMENT

This study was supported in part by the grant NSC94-2213-E-194-049 from the National Science Council, Taiwan.

REFERENCES

[1] S. Osowski and T. H. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng*. vol. 48, pp. 1265-1271, 2001.

[2] M. Engin, "ECG beat classification using neuro-fuzzy network," *Pattern Recognition Letters*, vol. 25, pp. 1715-1722, 2004.

[3] İ. Güler and E. D. Übeyli, "A modified mixture of experts network structure for ECG beats classification with diverse features," *Engineering Applications of Artificial Intelligence*, vol. 18, pp. 845-856, 2005.

[4] S.-N. Yu and Y.-H. Chen, "Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network," *Pattern Recognition Letters*, vol. 28, pp. 1142-1150, 2007.

[5] G. K. Prasad and J. S. Sahambi, "Classification of ECG arrhythmias using multi-resolution analysis and neural networks," in *Proc. IEEE Conf. on Convergent Technologies for the Asia-Pacific Region (TENCON 2003)*, vol. 1, 2003, pp. 227-231.

[6] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ECG complexes using Hermite function and self-organizing maps," *IEEE Trans. Biomed. Eng.* vol. 47, pp. 838-848, 2000.

[7] J. Jakubowski, K. Kwiatos, A. Chwaleba, and S. Osowski, "Higher order statistics and neural network for tremor recognition," *IEEE Trans. Biomed. Eng.* vol. 49, pp. 152-159, 2002.

[8] J.S. Sahambi, S.N. Tandon, and R.K.P. Bhatt, "Using wavelet transforms for ECG characterization," *IEEE Eng. Med. Biol. Mag.*, vol. 16, pp. 77-83, 1997.

[9] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha and P. Laguna, "A wavelet-based ECG delineator: evaluation on standard databases," *IEEE Trans. Biomed. Eng.* vol. 51, pp. 570-581, 2004.

[10] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Networks*, vol. 5, pp. 537-550, 1994.

[11] F. Fleuret, "Fast binary feature selection with conditional mutual information," *Journal of Machine Learning Research*, vol. 5, pp. 1531-1555, 2004.

[12] L. Yu and H. Liu, "Feature selection for high-dimensional data: a fast correlation-based filter solution," in *Proc. 20th Conf. Machine Learning*, 2003.

[13] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY: Wiley, 2001.

[14] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice Hall, 1994.

[15] Physiobank Archieve Index, MIT-BIH Arrhythmia Database: http://www.physionet.org/physiobank/database/mitdb/.

[16] Y.-H. Chen and S.-N. Yu, "Subband features based on higher order statistics for ECG beat classification," in *Proc. 29th Annual Int'l Conf. IEEE EMBS*, Lyon, France, August 23-26, 2007, pp. 1859-1862.