Proceedings of the World Congress on Engineering 2016 Vol I
WCE 2016, June 29 - July 1, 2016, London, U.K.

Examination of Wavelet-Based Features for Congestive Heart Failure Classification Using SVM

Suparerk Janjarasjitt, *Member, IAENG,*

Abstract—**One of the most common heart disease is congestive heart failure. In this study, the detail coefficients of RR interval data obtained from the discrete wavelet transform are applied for congestive heart failure (CHF) classification. The wavelet-based feature examined is a difference between the logarithms of variances of detail coefficients of RR interval data corresponding to two consecutive levels, referred to as Δ_l. A feature vector used in CHF classification is formed by a pair of wavelet-based features, i.e., $[\Delta_m, \Delta_n]$. The classification is performed using SVM with a linear kernel function and its performance is validated using 10-fold cross-validation. From the computational results, the best performance on the CHF classification can be achieved using the feature vectors $[\Delta_2, \Delta_3]$ and $[\Delta_1, \Delta_2]$ where the best accuracy, the best sensitivity and the best specificity for the CHF classification are, respectively, 80.9324%, 81.4103% and 82.8938%.**

Index Terms—**congestive heart failure, heartbeat, wavelet analysis, classification.**

I. Introduction

Heart failure (HF), also called cardiac failure, and congestive heart failure (CHF), is a condition in which the heart is no longer able to pump enough blood to meet the body's needs [1]. Heart failure is a common but serious condition. There are about 5.1 million people who have heart failure in the United States [2]. The most common causes of heart failure are diseases that damage the heart including coronary heart disease, high blood pressure, and diabetes [1]. There is no single diagnostic test for heart failure [3]. An electrocardiogram (ECG) is one of the most common diagnostic tests that records the heart's electrical activity. One of crucial informations obtained from the ECG is the heart's rhythm. There have been a variety of computational methods and techniques applied to RR interval or heartbeat data for congestive heart failure classification.

Heart rate variability (HRV) analysis is a fundamental approach commonly used for assessing heart's health and also diagnosing heart diseases. HRV may be evaluated using a number of methods [4] such as time domain methods, statistical methods, frequency domain methods, and nonlinear methods. Diagnosis and classification of congestive heart failure are ones of applications to which HRV applied. There have been several studies applying measures obtained from various HRV analysis techniques, for example, [5]–[10]. Some computational techniques provide promising results on CHF classification with high sensitivity and specificity.

This work is supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission.

S. Janjarasjitt is with the Department of Electrical and Electronic Engineering, Ubon Ratchathani University, Ubon Ratchathani, 34190 Thailand e-mail: suparerk.j@ubu.ac.th

Recently, computational tools and techniques applied for complex systems analysis have been widely applied to applications in biology and medicine including cardiology. Some of measures derived from such complex systems analysis applied to characterize HRV include spectral exponent, scaling exponent, correlation dimension, and Lyapunov exponent [4]. Self-similarity or scale-invariance is one of intriguing properties that have been investigated in phyiological and biological signals. The $1/f$ processes is a class of statistical self-similar processes that exhibiting a power law behavior [11], [12]. The spectral exponent is an exponent characterizing such power law behavior of $1/f$ processes that specifies the distribution of power in $1/f$ processes from low to high frequencies.

In addition to frequency domain methods using various spectral analysis techniques, the spectral exponent can be estimated using wavelet coefficients, in particular detail coefficients, obtained from the discrete wavelet transform. The discrete wavelet transform is a natural tool for characterizing self-similar signals [11], [12]. In this study, the detail coefficients of RR interval data obtained from the discrete wavelet transform are applied as features for classifying between RR interval data associated with congestive heart failure and those associated with normal sinus rhythm. Vectors of wavelet-based features are classified using support vector machine (SVM) technique.

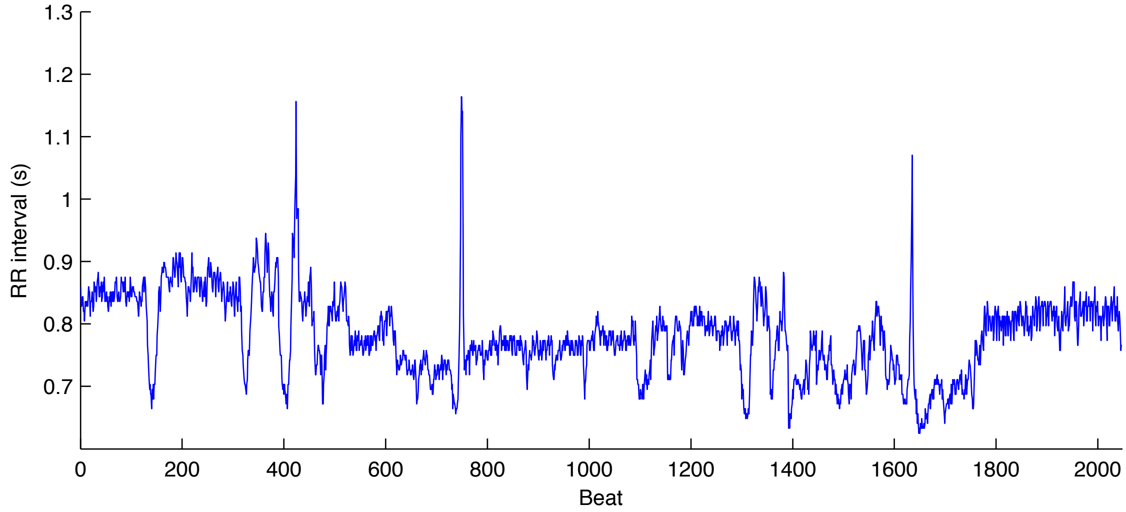
The rest of this paper is organized as follows. Section II presents data, feature extraction, and experimental setup. Section III details and discusses the CHF classification results. Finally, Section IV summarizes the paper and provides concluding remarks.

II. Methods

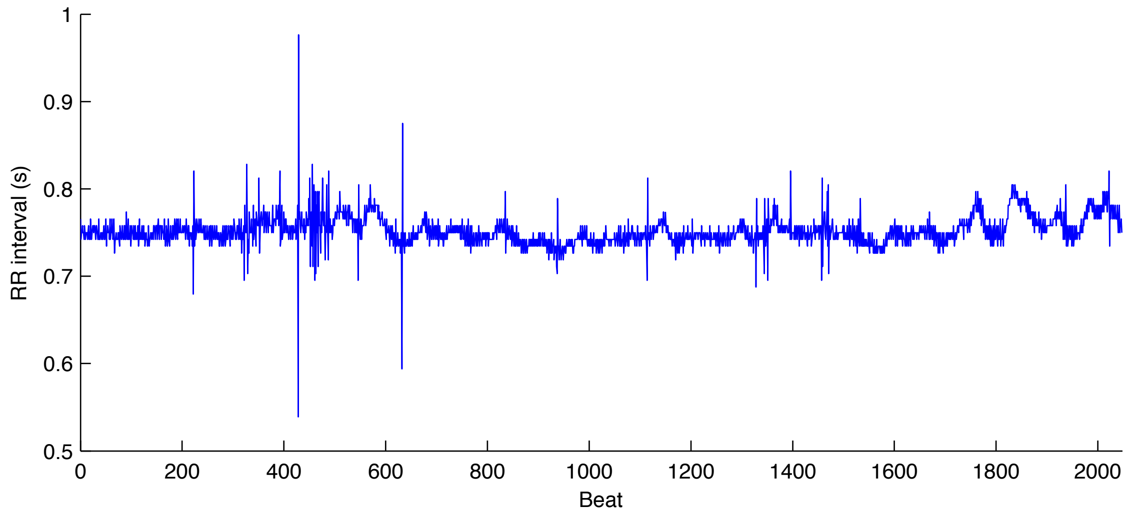
A. Data and Subjects

RR interval data examined in this study are composed of two data sets obtained from PhysioNet available online at http://www.physionet.org/physiobank/database/ [13]. The first set of RR interval data is obtained from the Normal Sinus Rhythm RR Interval Database (NSR2DB) available at http://physionet.org/physiobank/database/nsr2db/. Another set of RR interval data is obtained from the Congestive Heart Failure RR Interval Database (CHF2DB) available at http://physionet.org/physiobank/database/chf2db/. The Normal Sinus Rhythm RR Interval Database contains RR interval data of 54 subjects in normal sinus rhythm (NSR). The subjects included 30 men (aged from 28.5 to 76 years) and 24 women (aged from 58 to 73 years). The Congestive Heart Failure RR Interval Database contains RR interval data of 29 subjects with congestive heart failure (CHF).

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



(a) NSR subject



(b) CHF subject

Fig. 1. Epochs of RR interval data of NSR and CHF subjects.

The beat annotations in both databases were obtained by automated analysis with manual review and correction. Examples of 2048-beat epochs of RR interval data obtained from the NSR and CHF data sets are shown in Fig. 1(a) and Fig. 1(b), respectively.

B. Wavelet-Based Feature Extraction

The discrete wavelet transform (DWT) is a representation of a signal using a countably-infinite set of wavelets [14]. The discrete wavelet transform can be interpreted as a generalized octave-band filter bank [11], [15] as a wavelet is a bandpass filter [16]. A signal can be decomposed into approximations and details using the scaling function and the wavelet function that, respectively, correspond to halfband lowpass filter and halfband highpass filter [17]. For each single-level discrete wavelet decomposition, the approximation coefficients (a_l) and the detail coefficients (d_l) are obtained through the filtering processes. The detail coefficients obtained from the discrete wavelet decomposition are applied in this study.

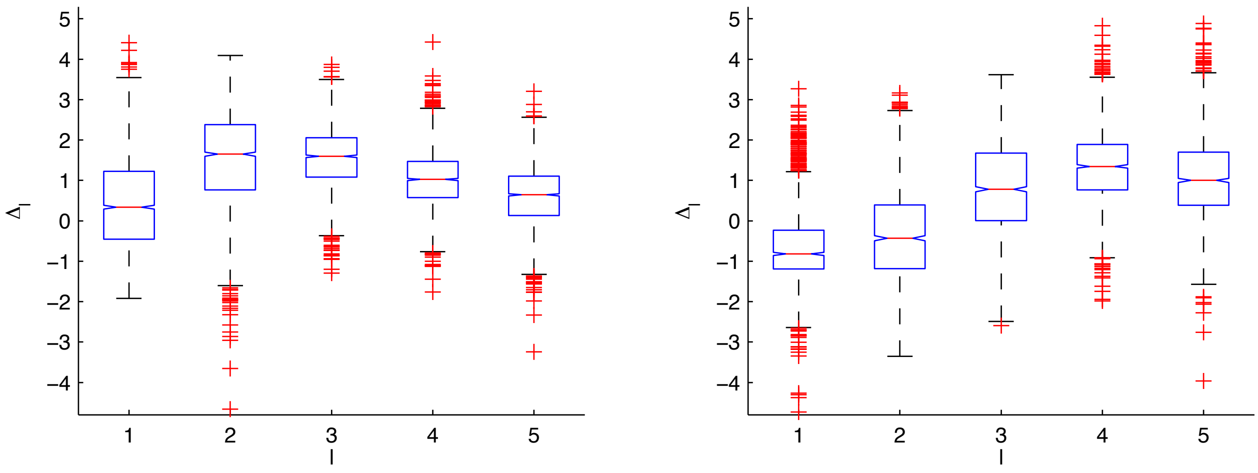
The Wavelet-based features of RR interval data proposed for congestive heart failure classification in this study can be obtained by the following steps.

1) Apply the wavelet transform to decompose RR interval data into L levels to obtain the detail coefficients d_l.
2) Compute the variance of detail coefficients d_l corresponding to each level l: $\mathrm{var}(d_l)$.
3) Take the logarith to base 2 of the corresponding variances of detail coefficients: $\log_2 \mathrm{var}(d_l)$.
4) Subtract the logarithm of variance of detail coefficients corresponding to the level l from that corresponding to the lower level: $\Delta_l = \log_2 \mathrm{var}(d_{l+1}) - \log_2 \mathrm{var}(d_l)$.

C. Experimental Setup

The RR interval data of all subjects obtained from both NSR and CHF data sets are partitioned into epochs with size of 2048 beats (or samples). The RR interval data are partitioned without overlap. This length of RR interval data is equivalent to approximately 30 seconds. There are 2730 epochs of RR interval data obtained from the NSR data set and there are 1560 epochs of RR interval data obtained from

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



(a) NSR data set (b) CHF data set

Fig. 2. Box plots of wavelet-based features Δ_l of epochs of RR interval data corresponding to various levels l.

the CHF data set. Note that the number of epochs of RR interval data of each subject is varied according to its original length. For the wavelet-based feature extraction, the 12th order Daubechies wavelets are used. The 2048-beat epochs of RR interval data are decomposed into 6 levels that are the maximum levels the 2048-beat epochs can be decomposed. Consequently, there are 5 wavelet-based features extracted from each epoch of RR interval data, i.e., Δ_1, Δ_2, Δ_3, Δ_4, and Δ_5.

A pair of wavelet-based features, i.e., $[\Delta_m, \Delta_n]$, is used as a feature vector applied for the CHF classification. Epochs of RR interval data obtained from the CHF data set are classified from epochs of RR interval data obtained from the NSR data set using SVM. The performance of the feature vector $[\Delta_m, \Delta_n]$ composed of all possible combinations of pairs of the wavelet-based features on the CHF classification are examined. A linear kernel function is used to train an SVM classifier. The performance of the CHF classification is validated using 10-fold cross-validation.

Furthermore, the performance of the CHF classification is evaluted using three conventional measures: accuracy (Ac), sensitivity (Se), and specificity (Sp) that are given by, respectively,

$$Ac = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Se = \frac{TP}{TP + FN}, \text{ and}$$
$$Sp = \frac{TN}{TN + FP}$$

where TP, TN, FP, and FN denote a number of true positives, a number of true negatives, a number of false positives, and a number of false negatives.

III. Results

A. Characteristics of the Wavelet-Based Features

The wavelet-based features Δ_l of all epochs of RR interval data obtained from the NSR and CHF data sets corresponding to various levels l are shown in Fig. 2(a) and Fig. 2(b), respectively. Table I summarizes the means and the standard deviations of the wavelet-based features Δ_l of all epochs of RR interval data obtained from the NSR and CHF data sets corresponding to various levels l. The wavelet-based feature Δ_l of epochs of RR interval data obtained from the CHF data set tends to increase as the level l increases, except for Δ_5. On the other hand, the wavelet-based feature Δ_l of epochs of RR interval data obtained from the NSR data set does not significantly change.

B. Performance of the CHF Classification

The accuracy (Ac), the sensitivity (Se) and the specificity (Sp) of the CHF classification by applying SVM to the feature vector $[\Delta_m, \Delta_n]$ composed of all possible combinations of pairs of the wavelet-based features are summarized in Table II. The performance of the CHF classification using 10-fold cross-validation corresponding to each feature vector $[\Delta_m, \Delta_n]$ is different. The performance of the CHF classification using the wavelet-based feature Δ_l extracted from the lower levels, i.e., $l = 1$, 2 and 3, corresponding to higher-frequency components tends to be better.

The best accuracy for the CHF classification is achieved using the feature vector $[\Delta_2, \Delta_3]$ with the accuracy of 80.9324%. The best sensitivity for the CHF classification is achieved using the feature vector $[\Delta_2, \Delta_4]$ with the sensitivity of 81.4103%. The best specificity for the congestive

TABLE I
Statistical Values (Mean±S.D.) of the Wavelet-Based Features Δ_l

Level l	Data Set	
	NSR	CHF
1	0.4161±1.08	-0.6220±1.00
2	1.5127±1.19	-0.3107±1.11
3	1.5583±0.72	0.7922±1.15
4	1.0185±0.69	1.3424±0.92
5	0.6039±0.74	1.0627±1.03

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

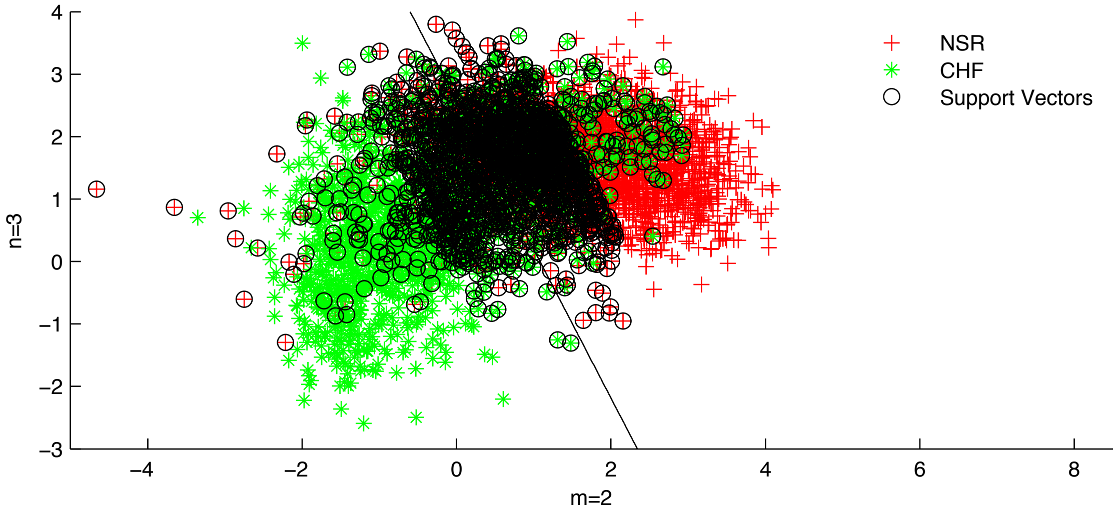


Fig. 3. The SVM classifier using the feature vector $[\Delta_2, \Delta_3]$.

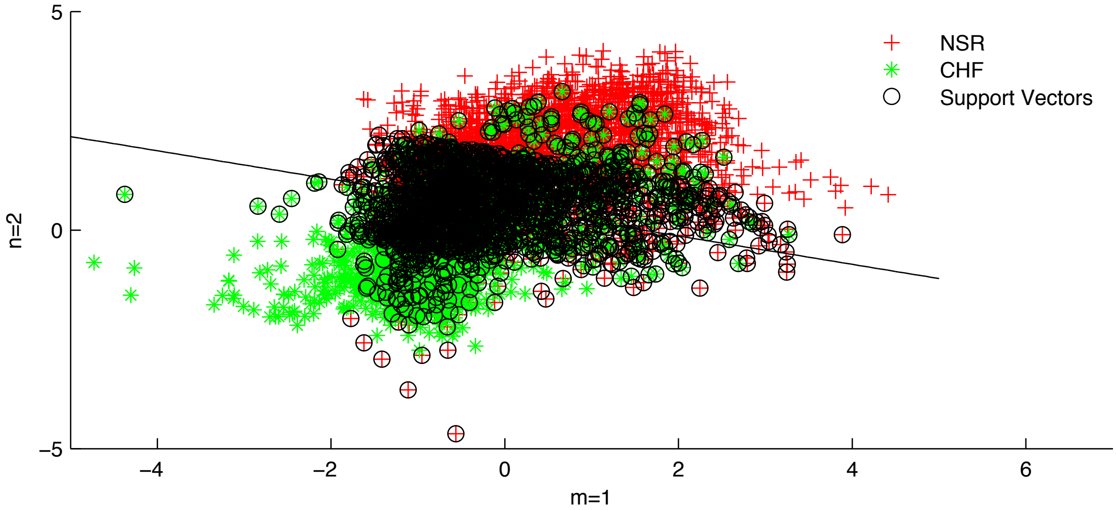


Fig. 4. The SVM classifier using the feature vector $[\Delta_1, \Delta_2]$.

heart failure is achieved using the feature vector $[\Delta_2, \Delta_3]$ with the specificity of 82.8938%. The best overall performance, i.e., the average of all three performance measures, is achieved using the feature vector $[\Delta_2, \Delta_3]$. However, the feature vector $[\Delta_1, \Delta_2]$ provides the CHF classification with all three performance measures greater than 80.0% and the second best overall performance. The support vectors and the corresponding separating line obtained using the feature vectors $[\Delta_2, \Delta_3]$ and $[\Delta_1, \Delta_2]$ are shown in Fig. 3 and Fig. 4, respectively.

TABLE II
PERFORMANCE OF THE CHF CLASSIFICATIONS

Feature Vector	Accuracy (Ac)	Sensitivity (Se)	Specificity (Sp)
$[\Delta_1, \Delta_2]$	80.3030	80.0641	80.4396
$[\Delta_1, \Delta_3]$	77.5058	68.4615	82.6740
$[\Delta_1, \Delta_4]$	69.3473	77.8205	64.5055
$[\Delta_1, \Delta_5]$	70.3730	76.8590	66.6667
$[\Delta_2, \Delta_3]$	80.9324	77.5000	82.8938
$[\Delta_2, \Delta_4]$	79.0210	81.4103	77.6557
$[\Delta_2, \Delta_5]$	79.3473	78.7179	79.7070
$[\Delta_3, \Delta_4]$	71.6317	64.7436	75.5678
$[\Delta_3, \Delta_5]$	72.3776	57.8846	80.6593
$[\Delta_4, \Delta_5]$	65.3147	59.6154	68.5714

IV. CONCLUSIONS

The feature of RR interval data obtained from subjects with normal sinus rhythm and those obtained from subjects with congestive heart failure is extracted using the discrete wavelet transform. The difference between the logarithms of variances of detail coefficients of RR interval data corresponding to two consecutive levels, Δ_l, is examined. The classification of RR interval data associated with the congestive heart failure is performed using SVM with a linear kernel function applied to the feature vectors formed by a pair of wavelet-based features $[\Delta_m, \Delta_n]$. All possible combinations of pairs of the wavelet-based features are applied for the

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

CHF classification. From the computational results, it is shown that the wavelet-based feature is potentially a good feature for the CHF classification. The best accuracy, the best sensitivity, and the best specificity for the CHF classification are 80.9324%, 81.4103%, and 82.8938%, respectively. The wavelet-based feature Δ_l extracted from lower levels tends to provide a better performance on the CHF classification. The feature vectors providing the best performance on the CHF classification are $[\Delta_2, \Delta_3]$ and $[\Delta_1, \Delta_2]$.

REFERENCES

[1] Heart Failure. National Heart, Lung, and Blood Institute, Bethesda, MD, 2014 [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0063056/

[2] A.S. Go, *et al.*, "Heart disease and stroke statistics — 2013 update: a report from the American Heart Association," *Circulation*, vol. 127, pp. e6–e245, 2013.

[3] C.W. Yancy, *et al.*, "2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines," *Circulation*, vol. 128, pp. e240–e327, 2013.

[4] Electrophysiology, Task Force of the European Society of Cardiology the North American Society of Pacing, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, pp. 1043–1065, 1996.

[5] G. Liu, *et al.*, "A new approach to detect congestive heart failure using short-term heart rate variability measures," *PLoS ONE*, vol. 9, pp. e93399, 2014.

[6] R.K.G. Moore, *et al.*, "HRV spectral power and mortality in chronic heart failure (CHF): 5 year results of the UK heart study," *Heart*, vol. 90, pp. A6, 2004.

[7] S. Guzzetti, E. Magatelli, E. Borroni, and S. Mezzetti, "Heart rate variability in chronic heart failure," *Auton. Neurosci.*, vol. 90, pp. 102–105, 2001.

[8] J. Nolan, *et al.*, "Prospective study of heart rate variability and mortality in chronic heart failure: results of the United Kingdom heart failure evaluation and assessment of risk trial (UK-Heart)," *Circulation*, vol. 98, pp. 1510–1516, 1998.

[9] M.H. Asyali, "Discrimination power of long-term heart rate variability measures," in *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2003, pp. 200–203.

[10] Y. Isler and M. Kuntalp, "Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure," *Comput. Biol. Med.*, vol. 37, pp. 1502–1510, 2007.

[11] G.W. Wornell, "Wavelet-based representations for the $1/f$ family of fractal processes," *Proceedings of the IEEE*, vol. 81, pp. 1428–1450, 1993.

[12] G.W. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*. Prentice Hall: New Jersey, 1995.

[13] A. L. Goldberger, *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215–e220, 2000.

[14] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press: San Diego, 1998.

[15] G.W. Wornell and A.V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Signal Process.*, vol. 40, pp. 611–623, 1992.

[16] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Trans. Signal Process.*, vol. 40, pp. 2207–2232, 1992.

[17] S. Janjarasjitt, "Temporal characteristics of wavelet subbands of epileptic scalp EEG data based on the number of local min-max," *Lecture Notes in Electrical Engineering*, G.-C. Yang, S.-I. Ao, X. Huang, O. Castillo, Eds. Netherlands: Springer, 2014, pp. 55–69.

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)