

Characterization of Cardiovascular Diseases Using Wavelet Packet Decomposition and Nonlinear Measures of Electrocardiogram Signal

Hamido Fujita¹(✉), Vidya K. Sudarshan², Muhammad Adam², Shu Lih Oh², Jen Hong Tan², Yuki Hagiwara², Kuang Chua Chua², Kok Poo Chua², and U. Rajendra Acharya^{2,3,4}

¹ Faculty of Software and Information Science, Iwate Prefectural University (IPU), Takizawa, Iwate 020-0693, Japan
HFujita-799@acm.org

² Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore, Singapore

³ Department of Biomedical Engineering, School of Science and Technology, SIM University, Singapore, Singapore

⁴ Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia

Abstract. Cardiovascular diseases (CVDs) remain as the primary causes of disability and mortality worldwide and are predicted to continue rise in the future due to inadequate preventive actions. Electrocardiogram (ECG) signal contains vital clinical information that assists significantly in the diagnosis of CVDs. Assessment of subtle ECG parameters that indicate the presence of CVDs are extremely difficult and requires long hours of manual examination for accurate diagnosis. Hence, automated computer-aided diagnosis systems might help in overcoming these limitations. In this study, a novel algorithm is proposed based on the combination of wavelet packet decomposition (WPD) and nonlinear features. The proposed method achieved classification results of 97.98% accuracy, 99.61% sensitivity and 94.84% specificity with 8 reliefF ranked features. The proposed methodology is highly efficient in helping clinical staff to detect cardiac abnormalities using a single algorithm.

Keywords: Coronary artery disease · Myocardial infarction · Congestive heart failure · Electrocardiogram · Entropies · Wavelet packet decomposition

1 Introduction

Cardiovascular diseases (CVDs) are the primary cause of cardiac deaths worldwide. In 2012, an approximated 17.5 million CVD-related deaths are reported and it represents 31% of the overall global deaths [34]. CVDs are the leading cause of disability, particularly among myocardial infarction (MI) and stroke survivors. Moreover, the burden of CVDs is commonly measured by the disability-adjusted life years (DALY), whereby one DALY is similar to losing one year of healthy life. Evidently, the CVD burden for

developing countries are comparably greater (DALY > 5100 per 100 000) than for the developed countries (DALY < 3000 per 100 000) [48].

Coronary artery disease (CAD) is caused due to coronary arteriosclerosis, an inflammatory of the arterial wall. Atherosclerotic plaques start to build up on the wall of the coronary arteries [12, 13, 47] and this progressive buildup of plaques eventually obstruct and reduce the blood flow to a region of the myocardium [14]. Thus, oxygen deprived myocardium due to CAD leads to a fatal myocardial infarction (MI) [46].

MI is an irreversible death of the heart tissue which further expands in sizes and damages the left ventricular (LV) function causing LV dysfunction. Improper management and treatment of LV dysfunction together with other cardiac abnormalities leads to a catastrophic stage called Congestive Heart Failure (CHF) [14, 19].

Thus, CHF is characterized by impaired ventricles and also, changes to the neuro-hormonal regulation. Moreover, CHF is a terminal stage of CVD whereby the heart fails to sufficiently circulate the blood throughout the body. This complex clinical syndrome causes hypoxia, congestion, and even death [11, 15].

The electrocardiogram (ECG) is a commonly preferred clinical diagnostic tool for most of the cardiac conditions such as CAD, MI, and CHF. ECG is comparably cheaper and noninvasive and it contains vital information relating to the functioning of the heart. The morphology of the ECG signals is diagnostically important during an episode of CAD, MI, and CHF. T-waves are abnormally tall, the QT intervals are longer and ST segment is elevated or depressed during CAD and MI [5]. However, visually examining these morphological changes of features in voluminous ECG signals is tedious and highly prone to errors. Hence, an automated computer-aided technique is necessary to overcome the drawbacks of manual analysis of ECG signals.

Several improved methodologies have been developed for automated detection of CAD, MI and CHF using ECG signals or heart rate (HR) signals by researchers. Analysis of ST segment variation using various methods such as Radial Basis Function (RBF) neural networks [25], Principal component analysis (PCA) [7], Binary Particle Swarm Optimization (BPSO), Genetic algorithm [8], Discrete Wavelet Transform [23] have shown good results for classification of CAD affected ECG signals from normal ones.

For classification of normal and MI affected ECG signals, researchers have employed various DWT [21], linear [6, 26, 42] and nonlinear [1, 20, 43, 45] methods to evaluate the signal characteristics such as QRS complex [7], ST segment [6, 44], and T wave amplitude [6]. Even the evaluation of ECG signal variation using detrended fluctuation analysis (DFA) [22], entropies [22], and autoregressive burg [26, 28] have shown to be useful for differentiation of normal and CHF classes.

In the study of normal and CAD ECG signals, Acharya et al. [4] proposed the application of Higher Order Spectra (HOS). In addition, they developed a Coronary Artery Disease Index (CADI) that could automatically characterize normal and CAD ECG signals using a single number.

Furthermore, Acharya et al. [2] studied the classification of normal, CAD, and MI. They compared three different techniques namely Discrete Wavelet Transform (DWT), Empirical mode decomposition (EMD), and Discrete cosine transform (DCT) to differentiate among the three classes (normal, CAD, and MI). It is reported that utilizing DCT technique yielded the highest accuracy of 98.50% is obtained with only seven features.

It is evident from the literature review that, the majority of the studies proposed are on automated classification of two classes (normal and either CAD, MI or CHF). To date, no study has been published an algorithm for characterization of four classes (normal, CAD, MI and CHF) using ECG beats. Thus, in comparison to the literature review, this work proposes a novel algorithm for automated classification of normal, CAD, MI and CHF using ECG beats. The flowchart of the proposed method is illustrated in Fig. 1.



Fig. 1. Block diagram of the proposed methodology.

2 Methodology

2.1 Materials

In this study, the ECG signals of normal, CAD, MI, and CHF are acquired from various Physionet databases, namely PTB Diagnostic ECG Database (for normal and MI), St.-Petersburg Institute of Cardiology Technics 12-lead Arrhythmia Database (for CAD) and BIDMC Congestive Heart Failure Database (for CHF) [18].

2.2 Pre-processing

The ECG signals from the BIDMC Congestive Heart Failure and St.-Petersburg Institute of Cardiology Technics 12-lead Arrhythmia databases are sampled at 250 Hz and 257 Hz respectively whereas PTB Diagnostic ECG database is sampled at 1000 Hz. Thus, to maintain the uniformity for all the databases, a standard sample frequency of 1000 Hz is selected. In addition, the baseline wander and noise from ECG signals are eliminated by using Daubechies 6 (db6) wavelet [33].

2.3 Beats Segmentation

The pre-processed ECG signals are segmented into ECG beats by first detecting the R-peak using Pan Tompkins algorithm [33, 37]. The visibly tall amplitude of the R peaks is chosen as the distinctive point for this study. Each ECG beat is segmented by taking 250 and 400 samples before and after the R peak respectively.

2.4 Wavelet Packet Decomposition (WPD)

The wavelet packet decomposition (WPD) is a wavelet transform technique that utilizes wavelets to transform the ECG signals. Wavelets are the result of the translated and

scaled shapes of the basic mother wavelet. Moreover, mother wavelets are localized in the time-frequency domain and have fluctuating amplitudes within a finite time [30]. As compared to DWT method, WPD decomposes the signal into both the low frequency components (approximations) and the high frequency components (details) at every level [30]. Hence, WPD provides more information as compared to DWT.

2.5 Features Extraction

In this study, twelve nonlinear features are obtained from thirty WPD coefficients. The twelve features are namely approximate entropy [38], sample entropy [40], fuzzy entropy [24], Kolmogorov-Sinai entropy [17], Renyi entropy [39], Tsallis entropy [10], fractal dimension [27], wavelet entropy [41], Signal energy [35], permutation entropy [9], recurrence quantification analysis [16], and bispectrum [36].

2.6 Features Selection – Sequential Forward Feature Selection (SFS)

Feature selection is a technique used to select a subset of significant features that yield minimal classification error [29]. Therefore, this step significantly enhances the performance of the classifier with the elimination of those redundant and insignificant features. In this study, sequential forward feature selection (SFS) is implemented for the feature selection process.

2.7 Features Ranking

In this study, ReliefF ranking technique is implemented. ReliefF estimates the significance of features by randomly sampling an instance and then consider the weighted value of features for the nearest instance of classes [32].

2.8 Classification – K-Nearest Neighbor (KNN)

KNN is an instance-based classification technique in which the unknown sample is classified according to either similarity or distance criteria [19]. In this work, the k value of the nearest neighbors is varied ranging from 5 to 10. We have experimented with different values of k and obtained maximum accuracy with $k = 5$.

3 Results

In total, 181,510 ECG beats of four classes are segmented from three separate ECG databases of 222 subjects. For each ECG beat, WPD of four level is implemented and resulted in 30 coefficients. Twelve different nonlinear features are obtained from the 30 coefficients, therefore, 1050 features are obtained from one beat. Further, out of the 1050 features extracted from each beat, 10 features are selected by SFS. The features have a p-value of $p < 0.0001$ which also means that these features are statistically significant.

The confusion matrix of the four classes is shown in Table 1.

Table 1. Confusion matrix of the 4 classes.

Original/ Predicted	N	MI	CAD	CHF	Acc (%)	PPV (%)	Sen (%)	Spec (%)
N	10002	440	103	1	99.34	93.81	94.84	99.61
MI	521	38826	835	0	98.15	95.12	96.63	98.59
CAD	139	1553	39845	8	98.51	97.56	95.91	99.29
CHF	0	0	59	89178	99.96	99.99	99.93	99.99

*Acc: Accuracy, Sen: Sensitivity, Spec: Specificity.

4 Discussion

In this study, a novel technique for automated characterization of various CVDs is proposed using WPD and nonlinear analysis of the ECG beats. A total of 181,510 ECG beats having normal, CAD, MI and CHF conditions are individually decomposed into 30 WPD coefficients. Twelve types of nonlinear features are then extracted from the coefficients. On the whole, the proposed methodology achieved maximum classification results of 97.98% accuracy, 99.61% sensitivity and 94.84% specificity with 8 ReliefF ranked features using KNN classifier.

The implementation of the various nonlinear analysis techniques is to measure the degree of complexity in the healthy and CVDs ECG signals. This relates to the presence of inherent patterns in the dynamics of nonlinear ECG signals [44]. Indeed, nonlinear techniques are highly sensitive towards the presence of subtle sudden changes in the ECG signals [3].

The main novelty of this work is the integration of WPD and nonlinear techniques into a computer-aided diagnosis (CAD) system which enhances the efficiency of the decision-making and diagnosis process. Thus, clinicians can expeditiously prescribe the relevant treatments to prevent the conditions from deteriorating further.

5 Conclusion

In this study, a novel technique is proposed for the identification and diagnosis of CAD, MI, and CHF by using nonlinear features that are obtained from the segmented ECG beats. The proposed automated diagnostic support system can reliably and efficiently assist clinical staff to detect and diagnose cardiac abnormalities. Thus, reducing the workload and the possibilities of manual errors in interpreting vital information during ECG data assessment. The integration of the proposed methodology with ECG system is cost effective and yield instantaneous results as compared to other conventional cardiac diagnostic modalities. Hence, the proposed cardiac diagnostic support system offers an alternative cheap cardiac screening especially for the developing countries whereby the majority of the CVDs deaths occur. For future work, authors aim to explore on different types of nonlinear feature extraction technique and a bigger database that can produce better accuracy with the lesser number of features. Subsequently, the work can be extended in the various stages of CAD, MI, and CHF. This helps in identifying

early indicators of the cardiac diseases from the ECG signals and thus, promptly suppressing the conditions from deteriorating further with the necessary clinical medications and treatments.

Acknowledgment. First author appreciates the support given by Japan Society for Promotion of Science (JSPS) KAKENHI Grant Number: 15K00439.

References

1. Acharya, U.R., Fujita, H., Sudarshan, V.K., Oh, S.L., Adam, M., Koh, J.E.W., Tan, J.H., Ghista, D.N., Martis, R.J., Chua, C.K., Poo, C.K., Tan, R.S.: Automated detection and localization of myocardial infarction using electrocardiogram: a comparative study of different leads. *Knowl.-Based Syst.* **99**, 146–156 (2016a)
2. Acharya, U.R., Fujita, H., Adam, M., Lih, O.S., Sudarshan, V.K., Hong, T.J., Koh, E.W., Hagiwara, Y., Chua, C.K., Poo, C.K., San, T.R.: Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: a comparative study. *Inf. Sci.* **377**, 17–29 (2016b)
3. Acharya, U.R., Kannathal, N., Krishnan, S.M.: Comprehensive analysis of cardiac health using heart rate signals. *Physiol. Meas. J.* **25**, 1130–1151 (2004)
4. Acharya, U.R., Sudarshan, V.K., Koh, E.W., Martis, R.J., Tan, J.H., Oh, S.L., Adam, M., Hagiwara, Y., Mookiah, M.R.K., Chua, K.P., Chua, K.C., Tan, R.S.: Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals. *Biomed. Sig. Process. Control* **31**, 31–43 (2017)
5. Arafat, S., Dohrmann, M., Skubic, M.: Classification of coronary artery disease stress ECGs using uncertainty modeling, 1-4244-0020-1. *IEEE* (2005)
6. Arif, M., Malagore, I.A., Afsar, F.A.: Detection and localization of myocardial infarction using k-nearest neighbor classifier. *J. Med. Syst.* **36**, 279–289 (2012)
7. Babaoglu, I., Findik, O., Bayrak, M.: Effects of principle component analysis on assessment of coronary artery diseases using support vector machine. *Expert Syst. Appl.* **37**, 2182–2185 (2010a). Elsevier
8. Babaoglu, I., Findik, O., Ulker, E.: A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine. *Expert Syst. Appl.* **37**, 3177–3183 (2010b). Elsevier
9. Bandt, C., Pompe, B.: Permutation entropy: a natural complexity measure for time series. *Rev. Lett.* **88**, 174102 (2002)
10. Bezerianos, A., Tong, S., Thankor, N.: Time dependent entropy of the EEG rhythm changes following brain ischemia. *Ann. Biomed. Eng.* **31**, 221–232 (2003)
11. Bui, A.L., Horwich, T.B., Fonarow, G.C.: Epidemiology and risk profile of heart failure. *Nat. Rev. Cardiol.* **8**, 3041 (2011)
12. Buja, L.M., Willerson, J.T.: The role of coronary artery lesions in ischemic heart disease: insights from recent clinicopathologic, coronary arteriographic, and experimental studies. *Hum. Pathol.* **18**, 451–461 (1987)
13. Buja, L.M., McAllister Jr., H.A.: Coronary artery disease: pathological anatomy and pathogenesis. In: Willerson, J.T., Cohn, J.N., Wellens, H.J.J., Holmes Jr., D.R. (eds.) *Cardiovascular medicine*, 3rd edn, pp. 593–610. Springer, London (2007)
14. Chee, J., Seow, S.C.: The electrocardiogram. In: Acharya, U.R., Suri, J.S., Spaan, J.A.E., Krishnan, S.M. (eds.) *Advances in Cardiac Signal Processing*, pp. 1–53. Springer, Heidelberg (2007)

15. Deedwania, P.C., Carbajal, E.V.: Congestive heart failure. In: Shanahan, J., Lebowitz, H. (eds.) *Current diagnosis & treatment, Cardiology*, pp. 203–232. McGraw-Hill Companies, USA (2009)
16. Eckmann, J.P., Kamphorst, S.O., Ruelle, D.: Recurrence plots of dynamical systems. *Europhys. Lett.* **5**, 973–977 (1987)
17. Farmer, J.D.: Information dimension and the probabilistic structure of chaos. *Naturforsch. Z.* **37**, 1304–1325 (1982)
18. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, PCh., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23), e215–e220 (2000)
19. Guyton, A.C., Hall, J.E.: *Text Book of Medical Physiology*, 11th edn. Elsevier, New York (2006)
20. Han, J., Kamber, M., Pei, J.: *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco (2005)
21. Jayachandran, E.S., Joseph, K.P., Acharya, U.R.: Analysis of myocardial infarction using discrete wavelet transform. *J. Med. Syst.* **34**, 985–992 (2010)
22. Kamath, C.: A new approach to detect congestive heart failure using detrended fluctuation analysis of electrocardiogram signals. *J. Eng. Sci. Technol.* **10**(2), 145–159 (2015)
23. Kaveh, A., Chung, W.: Automated classification of coronary atherosclerosis using single lead ECG. In: *IEEE Conference on Wireless Sensors*, Kuching, Sarawak (2013)
24. Kosko, B.: Fuzzy entropy and conditioning. *Inf. Sci.* **40**, 165–174 (1986)
25. Lewenstein, K.: Radial basis function neural network approach for the diagnosis of coronary artery disease based on the standard electrocardiogram exercise test. *Med. Biol. Eng. Comput.* **39**, 1–6 (2001)
26. Liu, B., Liu, J., Wang, G., Huang, K., Li, F., Zheng, Y., Luo, Y., Zhou, F.: A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection. *Comput. Biol. Med.* **61**, 178–184 (2015)
27. Mandelbrot, B.B.: *The Fractal Geometry of Nature*. W.H. Freeman and Company (1982)
28. Masetic, Z., Subasi, A.: Detection of congestive heart failures using C4.5 decision tree. *Southeast Eur. J. Soft Comput.* **2**, 74–77 (2013). ISSN 2233-1859
29. Masetic, Z., Subasi, A.: Congestive heart failure detection using random forest classifier. *Comput. Methods Programs Biomed.* **130**, 54–64 (2016)
30. Masiti, M., Masiti, Y., Oppenheim, G., Poggi, J.M.: *Wavelet toolbox for use with Matlab, User's Guide, Ver. 3*. The MathWorks, Inc. (2004)
31. Marcano-Cedeno, A., Quintanilla-Dominguez, J., Cortina-Januchs, M.G., Andina, D.: Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In: *IEEE, IECON 2010, 36th Annual Conference on IEEE Industrial Electronics Society* (2010)
32. Marko, R.S., Igor, K.: Theoretical and Empirical Analysis of ReliefF and RReliefF. *Mach. Learn. J.* **53**, 23–69 (2003). doi:[10.1023/A:1025667309714](https://doi.org/10.1023/A:1025667309714)
33. Martis, R.J., Acharya, U.R., Lim, C.M.: ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomed. Sig. Process. Control* **8**(5), 437–448 (2013a)
34. Mendis, S., et al.: *Global Status Report on Non-communicable Diseases 2014*. World Health Organization (2014)
35. Mookiah, M.R.K., Acharya, U.R., Lim, C.M., Petznick, A., Suri, J.S.: Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. *Knowl. Based Syst.* **33**, 73–82 (2012)

36. NIKIAS, C.I., RAGHUVEER, M.R.: Bispectrum estimation: a digital signal processing framework. *Proc. IEEE* **75**, 869–891 (1987)
37. Pan, J., Tompkins, W.J.: *A Real Time QRS Detection Algorithm*, 11th edn. WB Saunders Co, Philadelphia (2006)
38. Pincus, S.M.: Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci.* **88**, 2297–2301 (1991)
39. Renyi, A.: On measures of entropy and information. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561 (1961)
40. Richman, J.S., Mooran, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **278**, 2039–2049 (2000)
41. Rosso, O.A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schurmann, M., Basar, E.: Wavelet entropy: a new tool for analysis of short duration electrical signals. *J. Neurosci. Methods* **105**, 65–67 (2001)
42. Safdarian, N., Dabanloo, N.J., Attarodi, G.: A new pattern recognition method for detection and localization of myocardial infarction using T-wave integral and total integral as extracted features from one cycle of ECG signal. *J. Biomed. Sci. Eng.* **7**, 818–824 (2014)
43. Sharma, L.N., Tripathy, R.K., Dandapat, S.: Multiscale energy and eigenspace approach to detection and localization of myocardial infarction. *IEEE Trans. Biomed. Eng.* **62**(7), 1827–1837 (2015)
44. Sun, L., Lu, Y., Yang, K., Li, S.: ECG analysis using multiple instance learning for myocardial infarction detection. *IEEE Trans. Biomed. Eng.* **59**(12), 3348–3356 (2012)
45. Thuraisingham, R.A.: A classification to detect congestive heart failure using second-order difference plot of RR intervals. *SAGE-Hindawi access to research Cardiology Research and Practice*, article id 807379 (2009)
46. Townsend, N., Wickramasinghe, K., Bhatnagar, P., Smolina, K., Nichols, M., Leal, J., Luengo-Fernandez, R., Rayner, M.: *Coronary Heart Disease Statistics, a Compendium of Health Statistics*, 2012th edn. British Heart Foundation, London (2012)
47. Willerson, J.T., Hillis, L.D., Buja, L.M.: *Ischemic Heart Disease Clinical and Pathophysiological Aspects*. Raven, New York (1982)
48. World Health Organization (WHO). *Disease and injury country estimates*, Geneva, Switzerland (2009)