35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Application of Higher order Spectra for Accurate Delineation of Atrial Arrhythmia

Hari Prasad, *Graduate Student Member, IEEE*, Roshan Joy Martis, *Member, IEEE*, U. Rajendra Acharya, *Senior Member, IEEE*, Lim Choo Min, *Senior Member, IEEE, Jasjit S. Suri, Senior Member, IEEE*

***Abstract*— The electrocardiogram (ECG) is being commonly used as a diagnostic tool to distinguish different types of atrial tachyarrhythmias. The inherent complexity and mechanistic and clinical inter-relationships often brings about diagnostic difficulties to treating physicians and primary health care professionals creating frequent misdiagnoses and cross classifications using visual criteria. The current paper presents a methodology for ECG based pattern analysis for detection of atrial flutter, atrial fibrillation and normal sinus rhythm beats. ECG is an inherently non-linear and non-stationary signal; its variation may contain indicators of current disease, or warnings about impending cardiac diseases. Routinely used time domain and frequency domain methods will not be able to capture the hidden information present in the ECG beats. In the present study, we have used non-linear features of higher order spectra (HOS) to differentiate the normal, atrial fibrillation and atrial flutter ECG beats. The bispectrum features were subjected to independent component analysis (ICA) for data reduction. The ICA coefficients were subsequently subjected to K-nearest-neighbor (KNN), classification and regression tree (CART) and neural network (NN) classifiers to evaluate the best automated classifier. We have obtained an average accuracy of 97.65%, sensitivity and specificity of 98.75% and 99.53% respectively using ten-fold cross validation. Overall, the results show that application of higher order spectra statistics is useful for the classification of atrial tachyarrhythmias with reasonably high accuracies. Further validation of the proposed technique will yield acceptable results for clinical implementation.**

I. Introduction

Atrial fibrillation (AF) is the most common encountered arrhythmia in the clinical practice, yet it remains one of the greatest challenges in the field of heart rhythm disorders and for developing automated diagnostics. Approximately 1% of persons older than 60 years suffer from AF, increasing to more than 5% of persons aged 70 years and older, and to more than 8% of people over 80 years of age. AF accounts for 1/3rd of all hospital admission with arrhythmia as principal diagnosis. The rate of hospital admissions for AF has risen in recent years. The commonest causes of AF are rheumatic heart disease (66%), hypertension (10%), cardiomyopathy (9%), and ischemic heart disease (7%) [1]. The epidemiology of atrial flutter (AFL) is relatively less well known as compared to the literature available for AF. The incidence of "lone" AFL is considerably lower than that of AF and is roughly equal to 0.88 per 1,000 person-years. AFL, similar to AF, is more common in males and the incidence and prevalence of AFL rises markedly with age from 5 per 100,000 in those 50 years and younger to 587 per 100,000 in those aged 80 years and older [2].

The electrocardiogram (ECG) is being routinely used as a diagnostic tool to distinguish different types of atrial tachyarrhythmias. The ECG is a non-invasive transthoracic interpretation of the electrical activity of the heart over time. A typical ECG tracing of the cardiac cycle consists of a P wave, a QRS complex, and a T wave. Atrial fibrillation or AF is characterized by a lack of organized atrial activity without evidence of clear P waves before each QRS complex. The atrial rate is usually 300-600 bpm, while the ventricular rate may be 170 bpm or more. In addition, the rhythm is irregular. Atrial flutter or AFL, on the other hand, is relatively a more organized rhythm than atrial fibrillation and has characteristic "sawtooth" flutter waves. Typically, in AFL, the atrial rate is between 240 and 300 bpm without an isoelectric baseline. AV conduction is most commonly 2:1, which yields a ventricular rate of approximately 150 bpm. [3].

Atrial fibrillation is a growing public health problem, which has reached epidemic proportions. AF is predominantly associated with hypertension and valvular heart disease, and carries an increased risk of embolic stroke, heart failure, and cognitive dysfunction. The effect of AF on mortality, morbidity, and economic burden is substantial. Accurate differentiation of atrial fibrillation from atrial flutter is important. For patients with atrial fibrillation, anticoagulation should be established and maintained indefinitely, at least for high-risk patients. Studies have repeatedly shown that the guideline recommendation on anti-coagulation therapy is not being routinely followed in clinical practice. One major reason for the underutilization is believed to be the misdiagnosis of AF as AFL. Computer-aided cardiac arrhythmia detection and classification can play a significant role in the management of cardiovascular diseases. A recent prospective, blinded, multicenter study suggested that computerized algorithm to identify atrial tachyarrhythmias improved diagnostic accuracy and patient healing outcomes [4]. There is therefore a considerable necessity for the development of better diagnostic methods and classification techniques for differentiation of atrial tachyarrhythmias on a surface electrocardiogram.

Manuscript received February 04, 2013; accepted April 05, 2013.

Hari Prasad is with Cellular and Molecular Medicine Program and Department of Physiology, Johns Hopkins University School of Medicine, Baltimore 21205, MD, USA (e-mail: hprasad1@jhmi.edu).

Roshan Joy Martis is with Biomedical Center, Electronic and Computer Engineering Division, Ngee Ann Polytechnic, Singapore (corresponding author; phone: +65 93854120, e-mail: mrj2@np. edu.sg, roshanmartis@ieee.org).

U. Rajendra Acharya is with Biomedical Center, Electronic and Computer Engineering Division, Ngee Ann Polytechnic, Singapore and University of Malaya, Malaysia (e-mail: aru@np.edu.sg).

Lim Choo Min is with School of Engineering, Ngee Ann Polytechnic, Singapore (e-mail: lcm@np.edu.sg).

Jasjit S. Suri is with Global Biomedical Technologies, CA, USA (e-mail: jsuri@comcast.net).

Most proposed automated methodologies in the literature attempts classifying atrial fibrillation and atrial flutter separately. However, detection of a single type of atrial tachyarrhythmias alone has limited clinical value unless those methodologies help in discrimination of the atrial fibrillation and atrial flutter, since both these have different therapeutic options and cause frequent miss-classifications [5]. Automated discrimination of these two arrhythmias in surface electrocardiogram is still a challenge, particularly in atypical presentations. Recently, higher order spectra (HOS) statistics has been successfully utilized for detection of epilepsy [6], sleep stages [7], and beat classification in cardiac arrhythmia [8]. The HOS are spectral representations of higher order moments or cumulants of a signal that can be utilized to capture the hidden complexities present in nonlinear signals. In the current study, we present a methodology for ECG based pattern analysis for detection of atrial flutter, atrial fibrillation and normal sinus rhythm beats using HOS. The HOS bispectrum is used to obtain nonlinear features in ECG beats. The independent component analysis (ICA) is applied on the principal domain of the bispectrum. The independent components are subsequently used for the automated classification using classifiers. Section II provides methodology, section III gives results, section IV provides discussion and section V concludes the article.

II. METHODOLOGY

A. Dataset Used

In this study, two publicly available databases, MIT BIH arrhythmia database and the MIT BIH atrial fibrillation database were used [9]. The segments of ECG having episodes of atrial fibrillation and flutter were extracted from the database by a medical expert. The noise and baseline wander in these signal segments is removed by discrete wavelet transform based denoising. Then, the QRS complex middle point is extracted using Pan Tompkins algorithm [10]. After detection of QRS middle point, 74 samples to the left of QRS middle point and 75 samples to the right of QRS middle point are considered as a single cycle of ECG beat. These beats are used for subsequent bispectrum estimation. In total, 641 normal, 887 atrial fibrillation and 855 atrial flutter ECG beats are used for the study. *Figure 1* depicts the methodology, consisting of HOS bispectrum computation, independent components calculation using ICA and pattern identification using classifiers.

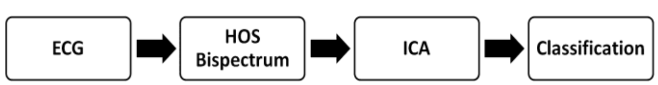

Figure 1: Proposed system for ECG classification.

B. HOS Bispectrum

Higher order spectra are the spectral representations of the higher order correlations of a signal [6-8]. In the current study, the third order spectra called as bispectrum is used for the analysis of normal, AF and AFL signals. The bispectrum $B(f_1, f_2)$ of an ECG beat is the Fourier transform of the third order correlation given by the averaged biperiodogram,

$$B(f_1, f_2) = E\left[X(f_1)X(f_2)X^*(f_1+f_2)\right] \quad (1)$$

where $X(f)$ is the Fourier transform of the signal, represents the complex conjugation operator, $E[\bullet]$ is the average over ensemble of realizations of the random signal known as expectation operator. The frequency f is normalized between 0 and 1. The bispectrum given by *Equation 1* is a complex valued function, and exhibits symmetry. The non-redundant region is called as principal domain of bispectrum. It is given by a triangle uniquely defined using the frequency $0 \le f_2 \le f_1 \le f_1 + f_2 \le 1$ [6-8].

C. Independent component analysis

Independent component analysis or ICA is a nonlinear dimensionality reduction technique. The electrocardiogram is an inherently non-linear and non-stationary signal. In order to capture the non-linear inter-relationships of ECG we have used ICA on bispectrum coefficients. The ICA model assumes that the observed signal x is linearly mixed with the source signal s [11]. The ICA model is given by,

$$x = As = \sum_{i=1}^{n} a_i s_i \quad (2)$$

where n is the number of mixtures. We need to solve for the elements of A to solve the ICA problem. In this study, computationally highly efficient fastICA algorithm [11] was used for computing the independent components.

D. Classification and regression tree (CART)

Classification and regression tree or CART is a non-parametric classifier where the training data is used to generate a set of rules based on the features [12]. The rules were selected such that the best split was obtained to discriminate the patterns belonging to different classes. The decision tree consists of a root node, connected by successive directional links called branches to other nodes. Each of the node splits into child nodes recursively until we reach terminal or leaf node. A rule splits the given node into two recursively until leaf node is reached. The splitting process is stopped when there is no further gain. Once the rules and the tree structure are learned, the testing set of the data is fed to the learned DT and as per the rules the test pattern is classified.

E. Neural network (NN)

Neural networks (NN) have been widely used for supervised pattern analysis and machine intelligence in a variety of fields including cardiac signal processing. In this study, a fully connected feed-forward neural network [12] is utilized. The input layer consists of 10 nodes corresponding to 10 independent component features, a hidden layer consisting of 6 neurons and an output layer of three neurons corresponding to three output classes was used. The NN weights are updated using error back-propagation method. A predefined threshold (0.0001 in our experiments) was set as termination condition for the iterations. Based on the trained NN (the node structure with the weights), the test data was fed and the testing patterns were classified.

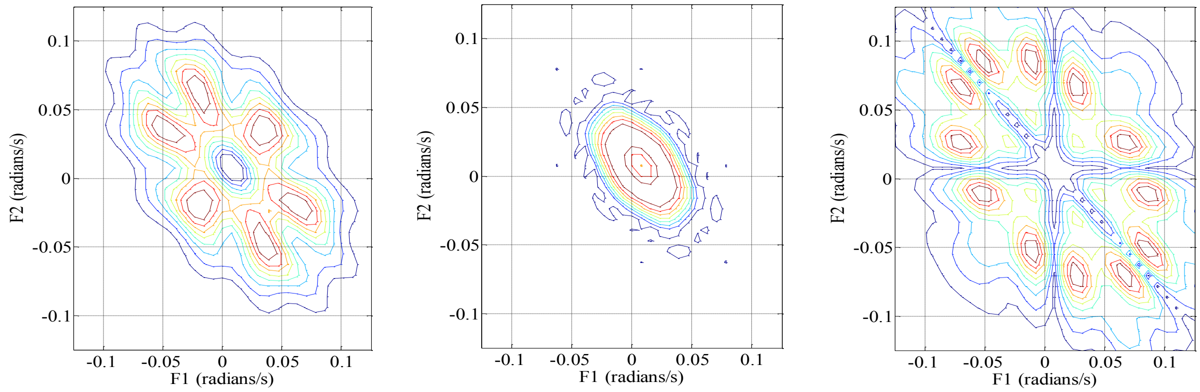

Figure 2: Typical bispectrum contour plot of ECG beat: (a) normal (b) atrial fibrillation and (c) atrial flutter.

F. K-NN classifier

K-NN classifier is a supervised classification method based on k nearest neighbor principle [12]. In the K-Nearest Neighbor classification algorithm, a set of training vectors with class labels are used as the starting point for assigning class labels to objects in the data to be classified. The classification is done based on a distance measure obtained by calculating a distance between the training object and the object of a class in consideration to be classified. This distance measure can be either Euclidean or a Mahalanobis distance or the user can define his own distance measure based upon the problem in question (K=2 in this case).

G. K-fold cross validation

In this study 10-fold cross validation was used for training and testing sets during classification. In this method the entire dataset was divided into ten non-overlapping sub sets such that almost equal number of data from each class belongs to each fold. In one fold of classification, one of the ten folds is used for testing and rest nine sub sets were together used for training the classifier. The process was repeated ten times so that each of the subset will be chosen for testing. The average accuracy, sensitivity, and specificity of the ten different folds were calculated to evaluate the performance of the proposed system.

III. Results

The proposed methodology on the three class ECG data is implemented in MATLAB. The bispectrum is computed for each pattern. The bispectrum of normal, AF and AFL ECG beat is represented as a contour plot in *Figure 2*. It can be seen from the figure that, these plots are unique for each class. The independent components of bispectrum are used as features for classification. *Table 1*, provides average sensitivity, specificity, PPV and accuracy for each classifier.

Table 1: Classification results using different classifiers.

Classifier	Sensitivity (%)	Specificity (%)	PPV (%)	Accuracy (%)
CART	98.68	92.35	97.24	93.45
NN	98.80	97.66	99.14	97.31
KNN	98.16	98.75	99.53	97.65

Abbreviations: CART= Classification and regression tree, NN= Neural network, KNN= k nearest neighbor, PPV=positive predictive value.

IV. Discussion

Summary of the studies on the identification of NSR, AF and AFL ECG beats obtained using MIT-BIH database is shown in *Table 2*. Logan *et al.* (2005) classified AF from the normal sinus rhythm (NSR) using the variance of RR intervals and reported 96% of sensitivity and 89% of specificity [13]. Tatento *et al.* (2001) applied Kolmogorov Smirnov test and detected AF with 91% of sensitivity and 96% specificity [14]. Cerutti *et al.* (1997) used auto-regressive modeling on RR intervals and detected AF with 96% sensitivity and 81% specificity [15]. Slocum *et al.* (1992) separated atrial activity to identify AF and obtained 62% sensitivity and 77% of specificity [16]. Recently, Martis *et al.* classified atrial fibrillation rhythm with atrial flutter using fractal dimension (FD) of continuous wavelet transform (CWT) with 100% accuracy using 100 ECG segments [5].

The proposed methodology is applied on ECG data belonging to three different classes. The three classes are normal, atrial fibrillation and atrial flutter. First, the Pan Tompkins algorithm is applied on each of these signals to obtain the QRS complex peak. Once QRS peak is accurately identified, 150 sample window is chosen around the R peak, such that 74 samples from the left and 75 samples from the right of R peak along with the R peak sample itself are chosen as one segment for further analysis. The data consists of 641 normal, 887 atrial fibrillation and 855 atrial flutter segments. The HOS analysis is used to obtain the nonlinear features. The ICA algorithms is applied on the principal domain of the bispectrum. The independent components were subsequently used for the automated classification using classifiers. Ten-fold cross validation is used to train and test the classifier. Overall, the classification accuracy is comparable to the studies reported in the literature.

In the present study, the normal, AF and AFL beats classified using independent components of bispectrum and KNN classifier resulted in maximum of 98.16% sensitivity, 98.75% specificity, 99.53% PPV and 97.65% accuracy. This prototype is developed in *Matlab*. We believe that the prototype model proposed in this study is of immense use for the primary care physicians and health practitioners, because it offers an automated method for classification of ECG into normal, atrial fibrillation and atrial flutter classes.

The system can also be used for screening for atrial tachyarrhythmias, particularly in elderly patients. Selection of patients with AF for prophylactic anticoagulation is evidence based and would lessen stroke risk in that age category [1].

Although the proposed methodology provides satisfactory results in our current research work, more study on a large population data set needs to be carried out on the generalization ability of the methodology. More exhaustive study over large population needs to be carried out to confirm the reproducibility of the results. Another key direction is to make use of better nonlinear features, diverse databases and robust classifiers and compare the accuracy.

Table 2: Summary of the studies on the identification of NSR, AF and AFL ECG beats obtained using MIT-BIH database.

Authors	Method	Classes	Performance
Logan *et al.*, 2005 [13]	RRI	NSR and AF	Sen= 96% Sp=89%
Tatento *et al.*, 2001[14]	RRI	NSR and AF	Sen= 91% Sp=96%
Cerutti *et al.*, 1997[15]	RRI	NSR and AF	Sen= 96% Sp=81%
Slocum *et al.*, 1992 [16]	AAA	NSR and AF	Sen= 62% Sp=71%
Current study	ICA of bispectrum	NSR, AF and AFL	Sen= 98.16% Sp=98.75% Ac=97.65%

Abbreviations: RRI= RR interval, AAA= Atrial activity analysis, ICA= Independent component analysis, NSR= Normal sinus rhythm, AF= Atrial fibrillation, AFL= Atrial flutter, Sen= Sensitivity, Sp= Specificity, Ac= Accuracy.

V. Conclusion

Globally, the number of patients with heart disease is growing every year and we see an increasing demand for development of computer aided diagnosis (CAD) system to obtain better outcomes for patients and lower health care costs. One of the major concerns with high dimensional biosignal datasets, including electrocardiogram, is the nonlinear, non-stationary, and non-Gaussian character of the signals. Non-linear features therefore need to be given important consideration during classification to enhance the diagnostic accuracy and clinical utility of CAD systems. Higher order spectra or HOS, also known as polyspectra, are spectral representations of higher order correlations. Many recent studies have successfully applied HOS analysis to extract the hidden complexities present in biosignals.

Discrimination of the atrial fibrillation and atrial flutter is important since both have different therapeutic options and cause frequent miss-classifications in day to day patient care. In the present scheme of analysis, nonlinear features of HOS were used to differentiate the normal, atrial fibrillation and atrial flutter ECG beats. The non-linear bispectrum features were subjected to independent component analysis (ICA) for data reduction. The ICA coefficients were subsequently subjected to K-nearest-neighbor (KNN), classification and regression tree (CART) and neural network (NN) classifiers to evaluate the best automated classifier. We have presented unique bispectrum contour plots for each cardiac class. We have obtained an average accuracy of 97.65%, sensitivity and specificity of 98.75% and 99.53% respectively using ten-fold cross-validation. The new paradigm of classification using non-linear features of HOS of ECG beats for atrial fibrillation, atrial flutter and normal sinus rhythm is the contribution of this paper. The accuracy can be further increased using better features and large population ECG for training and testing. We believe that the methodology based on application of higher order spectra analysis for discrimination of common atrial tachyarrhythmias provides valuable insights for future research and clinical applications.

References

[1] E. J. Benjamin, D. Levy, S. M. Vaziri, R. B. D'Agostino, A. J. Belanger, and P. A. Wolf, "Independent risk factors for atrial fibrillation in a population-based cohort. The Framingham Heart Study," *J. Am. Med. Assoc*., vol. 271, no. 11, pp. 840–844, Mar. 1994.

[2] J. Granada, W. Uribe, P. H. Chyou, K. Maassen, R. Vierkant, P. N. Smith, J. Hayes, E. Eaker, and H. Vidaillet, "Incidence and predictors of atrial flutter in the general population," *J. Am. Coll. Cardiol*., vol. 36, no. 7, pp. 2242–2246, Dec. 2000.

[3] J. T. Willerson, J. N. Cohn, H. J. J. Wellens, and D. R. H. Jr, *Cardiovascular Medicine*. 3rd ed. London: Springer, pp. 1955-1978, 2007.

[4] D. E. Krummen, M. Patel, H. Nguyen, G. Ho, D. S. Kazi, P. Clopton, M. C. Holland, S. L. Greenberg, G. K. Feld, M. N. Faddis, and S. M. Narayan, "Accurate ECG diagnosis of atrial tachyarrhythmias using quantitative analysis: a prospective diagnostic and cost-effectiveness study," *J. Cardiovasc. Electrophysiol*., vol. 21, no. 11, pp. 1251–1259, Nov. 2010.

[5] R. J. Martis, H. Prasad, C. Chakraborty, and A. K. Ray, "Automated detection of atrial flutter and fibrillation using ECG signals in wavelet framework," *J. Mech. Med. Biol*., vol. 12, no. 05, p. 1240023, Dec. 2012.

[6] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim, "Analysis of epileptic EEG signals using higher order spectra," *J. Med. Eng. Technol*., vol. 33, no. 1, pp. 42–50, 2009.

[7] U. R. Acharya, E. C.-P. Chua, K. C. Chua, L. C. Min, and T. Tamura, "Analysis and automatic identification of sleep stages using higher order spectra," *Int. J. Neural Syst*., vol. 20, no. 6, pp. 509–521, Dec. 2010.

[8] R. J. Martis, U. R. Acharya, K. M. Mandana, A. K. Ray, and C. Chakraborty, "Cardiac decision making using higher order spectra," *Biomed. Signal Process. Control*, vol. 8, no. 2, pp. 193–203, Mar. 2013.

[9] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–220, Jun. 2000.

[10] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng*., vol. BME-32, no. 3, pp. 230–236, 1985.

[11] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, Jun. 2000.

[12] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform," *Biomed. Signal Process. Control.* http://dx.doi.org/10.1016/j.bspc.2013.01.005, 2013.

[13] B. Logan and J. Healey, "Robust detection of atrial fibrillation for a long term telemonitoring system," *Computers in Cardiology*, 2005, pp. 619–622, 2005.

[14] K. Tateno and L. Glass, "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and deltaRR intervals," *Med. Biol. Eng. Comput*., vol. 39, no. 6, pp. 664–671, Nov. 2001.

[15] S. Cerutti, L. T. Mainardi, A. Porta, and A. M. Bianchi, "Analysis of the dynamics of RR interval series for the detection of atrial fibrillation episodes," *Computers in Cardiology* 1997, pp. 77–80, 1997.

[16] J. Slocum, A. Sahakian, and S. Swiryn, "Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity," *J. Electrocardiol.,* vol. 25, no. 1, pp. 1–8, Jan. 1992.