Cardiac Arrhythmia Detection using Combination of Heart Rate Variability Analyses and PUCK Analysis

Faizal Mahananto, Tomohiko Igasaki, and Nobuki Murayama

Abstract— This paper presents cardiac arrhythmia detection using the combination of a heart rate variability (HRV) analysis and a "potential of unbalanced complex kinetics" (PUCK) analysis. Detection performance was improved by adding features extracted from the PUCK analysis. Initially, R-R interval data were extracted from the original electrocardiogram (ECG) recordings and were cut into small segments and marked as either normal or arrhythmia. HRV analyses then were conducted using the segmented R-R interval data, including a time-domain analysis, frequency-domain analysis, and nonlinear analysis. In addition to the HRV analysis, PUCK analysis, which has been implemented successfully in a foreign exchange market series to characterize change, was employed. A decision-tree algorithm was applied to all of the obtained features for classification. The proposed method was tested using the MIT–BIH arrhythmia database and had an overall classification accuracy of 91.73%. After combining features obtained from the PUCK analysis, the overall accuracy increased to 92.91%. Therefore, we suggest that the use of a PUCK analysis in conjunction with HRV analysis might improve performance accuracy for the detection of cardiac arrhythmia.

I. INTRODUCTION

Arrhythmia is a generalized term for a cardiac rhythm excited by abnormal electrical activity in the heart. Some arrhythmias produce a life-threatening condition that leads to cardiac arrest. Therefore, the detection of cardiac arrhythmia and its discrimination from the normal cardiac rhythm are important clinical tasks.

Various methods have been proposed to detect arrhythmia. These techniques can be classified as beat-based [1], [2] and heart rate variability (HRV) based [3], [4] methods. Beat-based methods provide a detailed beat per beat detection algorithm, and can distinguish more detail about the arrhythmia type. However, this method has a heavy computational load. Another proposed method analyzes distinguishing features extracted from the R-R interval data. Although that method is unable to distinguish details of the type of arrhythmia, it provides a fast detection alternative, especially when used under critical conditions such as in an emergency unit.

There are many kinds of HRV analysis. Each addresses the interpretation and quantification of the R-R interval data from a different point of view. In application, the individual analysis method does not provide good information about the patient's condition [5], [6]. The available methods are usually combined to produce better results. Many methods

Faizal Mahananto, Tomohiko Igasaki, and Nobuki Murayama are with the Department of Information Technology on Human and Environmental Science, Graduate School of Science and Technology, Kumamoto University, Kumamoto 860-8555, Japan (corresponding author phone: +81-96-342-3613; fax: +81-96-342-3630; email: faizal@st.cs.kumamoto-u.ac.jp)

are now being proposed to extract HRV to provide an alternative type of analysis result and interpretation. One promising method of time-series analysis was developed in the field of econophysics. This analysis considers the existence of potential forces in the time series that influence the fluctuation of the time series. Because its formulation is applicable for any time sequence and for apparently random data series, it may be applicable to R-R interval data.

In this study, we extract R-R interval data features using various HRV analyses and "potential of unbalanced complex kinetics" (PUCK) analysis to characterize normal cardiac rhythm and cardiac arrhythmia. A decision-tree algorithm was employed for classification purposes using training and testing data sets derived from R-R interval data features.

II. MATERIALS AND METHODS

A. Database

R-R interval data data used in this study were generated from the ECG signals obtained from the MIT-BIH Arrhythmia Database [7]. The MIT-BIH Arrhythmia Database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings obtained between 1975 and 1979. It contains various types of arrhythmia, such as premature ventricular contraction (PVC) and atrial premature contraction (APC). The database has been used as a reference standard for conducting research on the problem of cardiac arrhythmia detection and classification [1], [2], [3], [4]. Two or more cardiologists working independently manually annotated each of the approximately 109,000 beats. Then, to obtain the computer-readable reference annotations for each beat in the database, disagreements were resolved. R-R interval data and these annotations were extracted from the original ECG recording using RRLIST software provided by PhysioNet [8].

After obtaining all R-R interval data from all of the data sets, each data set was segmented into 64 R-R intervals, producing a total of 1693 R-R interval data segments. Each segment was classified as "normal" of "arrhythmia" if the segment contained no arrhythmia episode or contained the arrhythmia episode, respectively, resulting of total 706 normal segments and 987 arrhythmia segments. In addition, the artificial beat originating from a pacemaker device was not considered an arrhythmia [3].

B. HRV analysis

Generally, the cardiovascular system expresses linear and nonlinear behaviors. To clearly observe the system, different linear and nonlinear parameters should be used for R-R

interval data characterization. In this study, a combination of both linear and nonlinear features of the R-R interval data was considered.

Three types of HRV analyses were performed to extract features of R-R interval data segments including time-domain analysis, frequency-domain analysis, and nonlinear analysis. In addition, a new method was proposed for extracting the features of R-R interval data, termed the "potential of unbalanced complex kinetics" (PUCK) analysis [9]. Time-domain analysis provides a statistical value of the R-R time series. The mean of all R-R intervals (MeanRR), standard deviation of all R-R intervals (SDRR), coefficient of variance of R-R intervals (CVRR), and root-mean-square of the successive difference of R-R intervals (RMSSD) were calculated.

In the frequency domain, the power spectrum density was measured by discrete Fourier transformation from segmented R-R interval data. Prior to obtaining the power spectrum, data were interpolated using the linear interpolation method. Four features were considered: the low-frequency (LF) component (area under 0.04–0.15 Hz), high-frequency (HF) component (area under 0.15–0.4 Hz), total power (TP) component (area under 0.01–0.4 Hz), and LF/HF component.

Three nonlinear analyses were employed, including sample entropy (SampEn), Poincaré plot analysis (SD1 and SD2), and detrended fluctuation analysis (DFA). The SampEn analysis parameters, similarity tolerance r and epoch length m, were set at 2 and $0.2 \times$SDNN, respectively. SD1 and SD2 in the Poincaré plot analysis were calculated from the plot as the standard deviation of line $y = x$ and line $y = -x + 2$MeanRR, respectively. Two features from the DFA were obtained from the slope of the scaling exponent at fewer than 11 beats (DFA $\alpha1$) and more than 11 beats (DFA $\alpha2$).

PUCK analysis was introduced in the econophysics field as an analysis method to observe the time series of market data [9]. The main purpose of applying this method is to extract characteristics of the time-series data. It requires 500–2000 data points to obtain a feasible result. In our previous finding, a shorter R-R interval data (256–1024) was successfully used to extract the characteristics of some R-R interval data. In this study, we considered 64 data points from a segmented R-R interval data.

PUCK analysis was performed to extract features of the R-R interval data. Using the specific method, we estimated the best value of the order of the "optimal moving average" (k), and the order of the "super moving average" (M). We detected no significant change when combining different values of k and M. Thus, we assigned values of 5 and 8 to the parameter settings k and M, respectively. This combination was taken from the smallest possible parameter, considering that the small R-R interval data segment was used for PUCK input. PUCK features were calculated by observing scatter plot from those two moving averages in which one of them is assumed as the noise separated R-R interval data and another is smoother moving average representing long term trend of the R-R interval data. The first feature, scaling slope, was estimated from the best fit
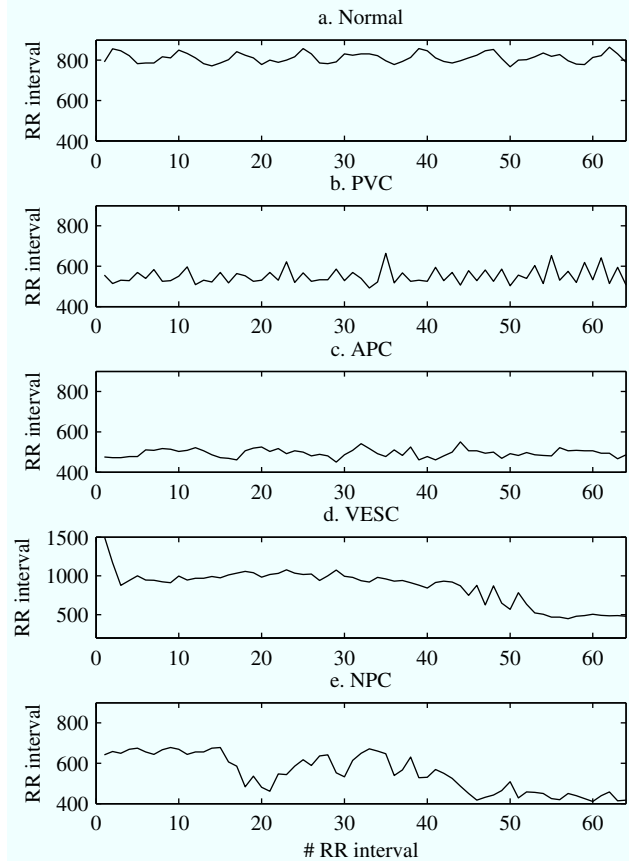

Fig. 1.   R-R interval data segment of a normal rhythm and arrhythmia.

line of the scatter plot. In addition, the standard deviations of the scatter plot toward perpendicular to the scaling slope and linear to the scaling slope were considered as SSD1 and SSD2.

C. Decision tree classification

All feature data obtained from the HRV analysis were later used for training and testing of the classification algorithm. For this purpose, a decision-tree classification algorithm, the C4.5 algorithm, was employed and calculated using WEKA classification software [10]. The training and testing data sets each were randomly extracted from 50% and 50%, respectively, of all R-R interval segment features data.

III. RESULTS

The R-R interval data segment of a normal rhythm and some types of arrhythmia such as premature ventricular contraction (PVC), atrial premature contraction (APC), ventricular escape beat (VESC), and junctional premature beat (NPC) are shown in Fig. 1. The R-R interval of a normal beat is comparatively higher than that of arrhythmic beats except for that of VESC. The R-R interval data containing PVC seemed to have a rapid fluctuation, and a fast R-R interval was observed in the signal containing APC beats. However, escaping beats noted as VESC produced a longer R-R interval.
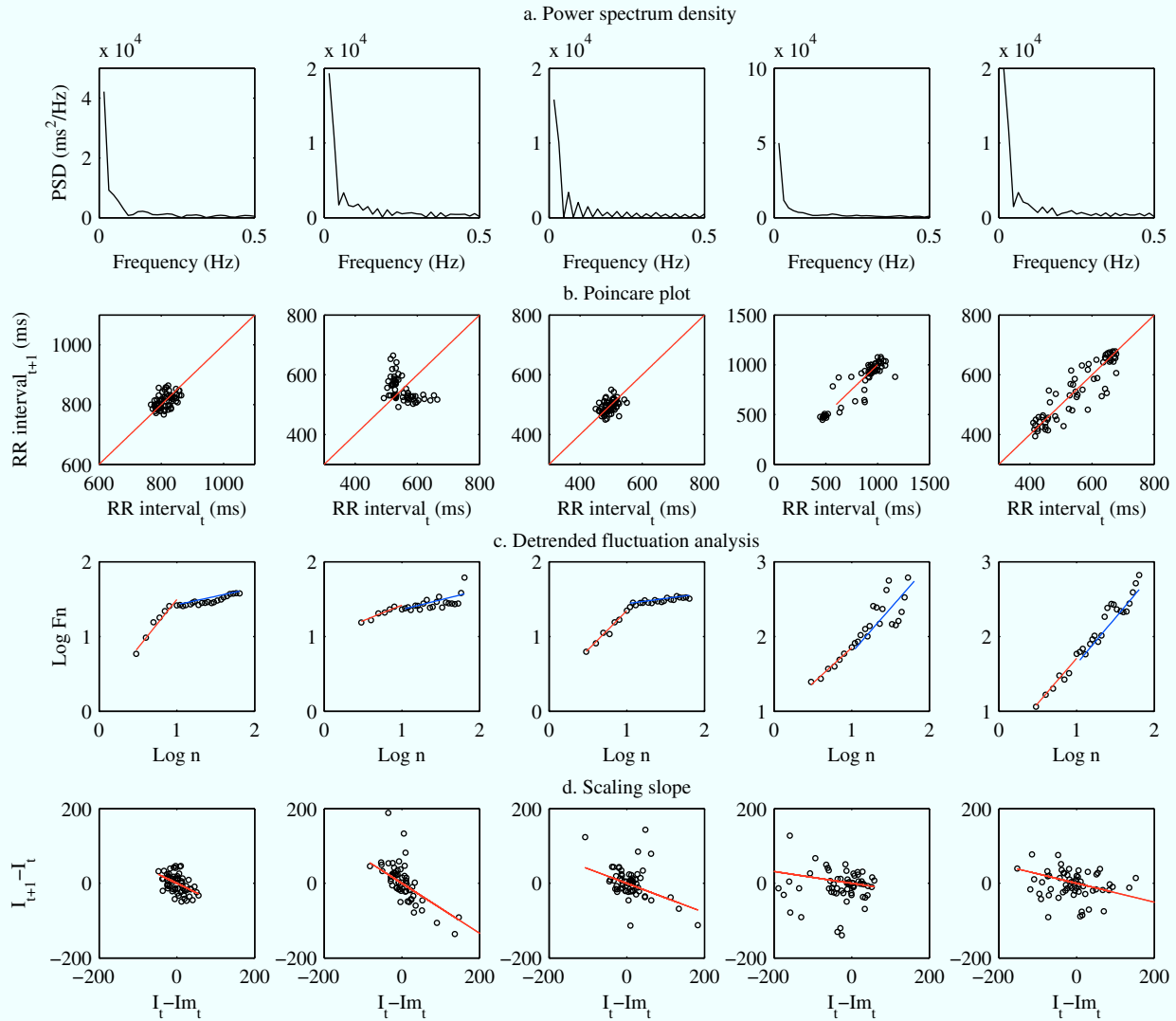
Fig. 2. HRV analysis results from left to right as normal, PVC, APC, VESC, and NPC respectively. The red line in the b. Poincaré plot is the line of identity. The red line in c., the detrended fluctuation analysis, shows the $\alpha 1$ slope whereas the blue line shows the $\alpha 2$ slope. Scaling slope was shown by the red line in d. while SSD1 and SSD2 represent standard deviation of the scatter plot, perpendicular and linear toward scaling slope, respectively

Fig. 2 shows a representation of the HRV and PUCK analyses in normal rhythms and arrhythmias. The power spectrum component of the normal rhythm is relatively lower than that of arrhythmia. The shape of the geometric plot in the Poincaré plot was easily distinguishable in the normal rhythm and arrhythmias. Although the normal rhythm and APC have similar Poincaré plot shapes, the SD1 and SD2 of the normal plot is larger than that of the APC. DFA analysis shows that DFA $\alpha 1$ and DFA $\alpha 2$ are strongly similar in arrhythmia, except for APC. This tendency was also found in the Poincaré plot. Clearer characteristics were found in the PUCK analysis. The SSD1 and SSD2 of normal rhythms are lower than those in arrhythmias. The scaling slope represented by the red line also distinguishes normal rhythms from arrhythmia.

After calculating all features from the HRV and PUCK analyses, a Wilcoxon rank-sum test was performed to deter-

mine the significance value. All features with a significance level $p < 0.05$ were used for classification training and testing. Table I lists mean $\pm$ standard deviation of each feature for normal rhythm and arrhythmia and its corresponding ROC area (AUC). All features have a significance level of $p < 0.001$ between the normal rhythm and arrhythmia, except the LF and TP component. AUC indicates individual parameter performance. RMSSD and SD1 have the highest AUC 90.96%.

Table II lists the classification accuracy using a decision-tree classifier. Classification accuracy was increased after combining one or more analysis-type features. Increasing accuracy was observed in almost all combinations of traditional HRV analysis features and additional PUCK features.

IV. Discussion

Arrhythmia detection using features extracted from the R-R interval data is presented in this study. Our motivation

Features	Normal	Arrhythmia	AUC (%)
MeanRR (ms)	846 ± 174	747 ± 158***	66.44
SDRR (ms)	45.5 ± 42.6	138 ± 118***	85.31
CVRR (ms)	5.31 ± 4.70	17.9 ± 12.7***	89.56
RMSSD (ms)	46.5 ± 42.5	206 ± 171***	90.96
LF (10^3ms^2)	12.3 ± 5.26	14.2 ± 8.72*	53.23
HF (10^3ms^2)	10.9 ± 4.07	15.1 ± 8.51***	66.27
LF/HF	1.11 ± 0.30	0.98 ± 0.27***	63.82
TP (10^3ms^2)	81.9 ± 13.2	81.9 ± 27.6**	54.26
SampEn	2.05 ± 0.41	1.36 ± 0.50***	85.90
SD1 (ms)	33.1 ± 30.3	147 ± 122***	90.96
SD2 (ms)	53.9 ± 53.2	121 ± 121***	77.66
DFA $\alpha1$	0.99 ± 0.32	0.60 ± 0.29***	81.84
DFA $\alpha2$	0.67 ± 0.42	0.41 ± 0.33***	68.20
Scaling slope	-0.54 ± 0.29	-0.80 ± 0.38***	69.87
SSD1 (ms)	31.1 ± 23.3	60.3 ± 50.3***	76.72
SSD2 (ms)	48.8 ± 34.6	123 ± 104***	83.40

Combination	Accuracy
Time domain	91.96
Freq. domain	80.26
Nonlinear	90.07
PUCK	79.08
Time domain, freq. domain	91.13
Time domain, nonlinear	92.91
Freq. domain, nonlinear	91.13
Time domain, PUCK	92.20
Freq. domain, PUCK	87.71
Nonlinear, PUCK	90.43
Time domain, freq. domain, PUCK	91.84
Time domain, nonlinear, PUCK	92.43
Freq. domain, nonlinear, PUCK	93.74
Time domain, freq. domain, nonlinear	91.73
Time domain, freq. domain, nonlinear, PUCK	92.91

for applying this method instead of a beat-based method is that the use of segmented R-R interval data have a lower computational load. Thus, the method can be applied in near-real-time situations. HRV itself has been proposed for clinically observing the autonomic nervous system [11] (ANS) in which ANS observation could be carried out at once.

Various HRV analyses are used. One challenge in this study is the use of a very short R-R interval data segment. The use of a specific length of R-R interval data has been proposed [11]. However, considering the problem, shorter signal segment is preferable. Especially PUCK analysis, the number of points used is much smaller than the suggested number [9].

Researchers have proposed various classification algorithms using HRV features for arrhythmia detection, including neural networks [3], support vector machines [4], and genetic programming [12]. We focus on the improvement of classification performance by empowering HRV analysis. This study attempted to provide alternative and/or additional HRV measures for enhancing the detection of cardiac arrhythmia. Therefore, we used an easily understood statistical classifier decision-tree algorithm to distinguish between HRV segment groups.

V. CONCLUSIONS

In this paper, cardiac arrhythmia detection using a combination of an HRV and PUCK analysis is presented. The available HRV analysis could extract the features of a short R-R interval data segment from the arrhythmia database. Although the use of a very short R-R interval data in PUCK analysis is still debated, this process characterized the R-R interval data containing arrhythmia with statistically significant accuracy. Our classification results showed that PUCK might improve the accuracy with which cardiac arrhythmia is detected. Compared to the use of just one type of HRV analysis, combining the features of these analyses with a PUCK analysis improved the accuracy of arrhythmia detection. Because a simple statistical classifier using a decision-tree algorithm was employed in this study, we believe that the accuracy and classification performance could be enhanced further using various classifier algorithms.

REFERENCES

[1] J.-S. Wang, W.-C. Chiang, Y.-L. Hsu, and Y.-T. C Yang, "ECG arrhythmia classification using a probabilistic neural network with a feature reduction method," *Neurocomputing*, 2012. doi:10.1016/j.neucom.2011.10.045

[2] M. R. Homaeinezhad, S. A. Atyabi, E. Tavakkoli, H. N. Toosi, A. Gaffari, and R. Ebrahimpour, "ECG arrhythmia recognition via a neuro-SVM-KNN hybrid classifier with virtual QRS image-based geometrical features," *Expert Syst. with Applicat.*, vol. 39, 2012, pp. 2047–2058.

[3] M. G. Tsipouras and D. I. Fotiadis, "Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability," *Comput. Methods and Programs in Biomed.*, vol. 74, 2004, pp. 95–108.

[4] M. A. Babak, K. S. Seyed, and M. Maryam, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artificial Intell. in Med.*, vol. 44, 2008, pp. 51–64.

[5] A. Jovic and N. Bogunovic, "Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features," *Artificial Intell. in Med.*, vol. 51, no. 3, 2011, pp. 175–186.

[6] J. F. Ramirez-Villegas, E. Lam-Espinosa, D. F. Ramirez-Moreno, P. C. Calvo-Echeverry, and W. Agredo-Rodriguez, "Heart rate variability dynamics for the prognosis of cardiovascular risk," *PLoS ONE*, vol. 6, no. 2, 2011. doi:10.1371/journal.pone.0017060

[7] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Eng. in Med. and Biol.*, vol. 20, no. 3, 2001, pp. 45–50.

[8] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. 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, 2000, pp. e215–e220.

[9] M. Takayasu, T. Mizuno, and H. Takayasu, "Potential force observed in market dynamics," *Physica A*, vol. 370, no. 1, 2006, pp. 91–97.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, 2009, pp. 10–18.

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

[12] M. Tavassoli, M. M. Ebadzadeh, and H. Malek, "Classification of cardiac arrhythmia with respect to ECG and HRV signal by genetic programming," *Canadian J. on Artificial Intell. Mach. Learning and Pattern Recognition*, vol. 3 no. 1, 2012, pp. 1–13.