

Automatic Detection of Congestive Heart Failure and Atrial Fibrillation with Short RR Interval Time Series

Kwon-Ha Yoon*, Yunyoung Nam**, Tharoeun Thap***, Changwon Jeong***, Nam Ho Kim§, Joem Seok Ko§, Se-Eung Noh§§ and Jinseok Lee†

Abstract – Atrial fibrillation (AF) and Congestive heart failure (CHF) are increasingly widespread, costly, deadly diseases and are associated with significant morbidity and mortality. In this study, we analyzed three statistical methods for automatic detection of AF and CHF based on the randomness, variability and complexity of the heart beat interval, which is RRI time series. Specifically, we used short RRI time series with 16 beats and employed the normalized root mean square of successive RR differences (RMSSD), the sample entropy and the Shannon entropy. The detection performance was analyzed using four large well documented databases, namely the MIT-BIH Atrial fibrillation ($n=23$), the MIT-BIH Normal Sinus Rhythm ($n=18$), the BIDMC Congestive Heart Failure ($n=13$) and the Congestive Heart Failure RRI databases ($n=25$). Using thresholds by Receiver Operating Characteristic (ROC) curves, we found that the normalized RMSSD provided the highest accuracy. The overall sensitivity, specificity and accuracy for AF and CHF were 0.8649, 0.9331 and 0.9104, respectively. Regarding CHF detection, the detection rate of CHF (NYHA III-IV) was 0.9113 while CHF (NYHA I-II) was 0.7312, which shows that the detection rate of CHF with higher severity is higher than that of CHF with lower severity. For the clinical 24 hour data ($n=42$), the overall sensitivity, specificity and accuracy for AF and CHF were 0.8809, 0.9406 and 0.9108, respectively, using normalized RMSSD.

Keywords: Congestive heart failure, Atrial fibrillation, Normal sinus rhythm, Electrocardiogram, RR interval, RMSSD, Sample entropy, Shannon entropy

1. Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia and responsible for significant morbidity. Over 2.3 million individuals in the United States currently have AF, and approximately 5.6 to 12.1 million individuals are projected to be affected by 2050 [1, 2]. Through its association with increased risk for stroke and mortality, AF has a profound impact on the longevity and quality of life of a growing number of people [3, 4]. Although AF treatment strategies are available, a major challenge facing clinicians and researchers is the oftentimes paroxysmal and asymptomatic nature of AF, particularly in its early stages. Since even short episodes of AF are associated with adverse health outcomes, there is a pressing need to

develop methods for accurate detection of paroxysmal AF. Such a method would have important clinical applications for use of anticoagulants, antiarrhythmic agents, and ablative therapies to abrogate the effects of AF and its complications.

Similarly, congestive heart failure (CHF) is an increasingly frequent diagnosis as our population tends to age and is associated with significant morbidity [5, 6]. CHF affects nearly 5 million individuals and causes more than 200,000 deaths each year in the United States [7]. It is a rapidly growing clinical problem resulting in almost one million admissions yearly in the United States as a poor prognosis [8]. Hypertension, myocardial infarction, obesity, diabetes, renal insufficiency, anaemia, cardiac valvular abnormalities and depression cause the risk of CHF development. It is a chronic condition requiring careful monitoring over a person's lifetime. However, early detection of CHF significantly increases effectiveness in the treatment of CHF. For these reasons, the importance of developing AF and CHF detection technologies has been emphasized

With no doubt surface 12-leads ECG remains as one of the most useful tests in the diagnosis and prognosis of AF and CHF patients, providing data on the heart rhythm, heart rate, and morphological changes in subsequent ECG curve's components. However, contemporary

† Corresponding Author: Dept. of Biomedical Engineering, Wonkwang University College of Medicine, Korea. (gonasago@wku.ac.kr)

* Dept. of Radiology, Wonkwang University College of Medicine, Korea. (khy1646@wku.ac.kr)

** Dept. of Computer Science and Engineering, Soonchunhyang University, Korea. (ynam.sch@gmail.com)

*** Dept. of Biomedical Engineering, Wonkwang University College of Medicine, Korea. (bami1314@kwu.ac.kr(T.T.); mediblue@wku.ac.kr(C.J.))

§ Dept. of Cardiology, Wonkwang University College of Medicine, Korea. (cardionh@wku.ac.kr(N.H.K.); nemor@hanmail.net (J.S.K))

§§ Dept. of Rehabilitation Medicine, Wonkwang University College of Medicine, Korea. (wonrehab@wku.ac.kr)

Received: September 24, 2015; Accepted: May 9, 2016

methods of AF and CHF monitoring, although superior to the standard 12-lead ECG and symptom monitoring, are unable to sufficiently analyze beat-to-beat data to identify the presence of AF and CHF on a daily basis and are incapable of identifying and discriminating among AF, CHF and normal sinus rhythm (NSR). Thus, there is a need to develop accurate, automated and comprehensive detection algorithms using standard ECG recorders. In this work, we have developed a sensitive, real-time realizable beat-to-beat detection algorithm for AF, CHF and NSR using commercially available, clinically applicable ECG recordings.

Over the last decade, many algorithms have been proposed to detect AF using RR interval (RRI) variability [9-12]. The approach is based on the characteristic of AF as a random sequence of RRI series with markedly increased beat-to-beat variability and complexity, and most algorithms show high sensitivity, specificity and accuracy up to 97%. Furthermore, the approach has been recently applied to smartphone application [13, 14]. Indeed, a smartphone can be used to detect AF based on its ability to record a pulsatile photoplethysmogram signal from a fingertip using the built-in camera lens. The CHF detection using RRI time series also has been studied [15-20]. In previous work, cardiac chaos is prevalent in healthy heart, and a decrease in such chaos may be indicative of CHF [21]. Indeed, CHF is with very low variability and complexity in RRI time series while AF is with high variability and complexity. However, most previous studies used relative long RRI time series, which were unable to sufficiently analyze beat-to-beat data to identify the presence of CHF on a daily basis. Casolo *et al.* [15] evaluated HR variability by calculating the mean hourly HR standard deviation for discriminating CHF and NSR. Fei *et al.* [16] measured spectral analysis using fast Fourier transformation with 24 ambulatory ECG monitoring finding the CHF characteristics. Isler and Kuntalp [17] used wavelet entropy using each 5 minute segment. Thakre and Smith [18] observed a curvilinear association between lag and Pointcare plot indices using 50 RRI interval time series. Maestri *et al.* [19] used both empirical mode decomposition and Pointcare plots using each 5 minute RRI interval time series. Thuraisingham [20] also studied CHF by using standard deviation of 5 minute RR interval time series. To our best knowledge, the CHF detection algorithm with short RRI time series (16 beats) has not been studied and developed. Furthermore, the automatic and simultaneous AF, CHF and NSR discrimination has not been studied.

In this paper, we used simple statistical methods to discriminate AF, CHF and NSR with short RRI time series (16 beats). From the characteristics that AF exhibits high variability in RRI time series, and CHF exhibits very low variability in RRI time series, we used normalized root mean squared successive difference (RMSSD), sample entropy and Shannon entropy. Our expectation was that the statistical values of AF were the highest and those of CHF were the lowest among CHF, NSR and AF. Since

CHF is classified according to the severity, we analyzed two different classified databases of NYHA class I-III and NYHA class III-IV. For the performance analysis, the receiver operating characteristics (ROC) curves were used.

2. Materials and Methods

2.1 Databases

We used five databases to test the method on: the MIT-BIH AF, the MIT-BIH NSR, the BIDMC CHF, the CHF RRI and 24 hour clinical CHF database from Wonkwang University Hospital. For all databases, we used RR interval series. The MIT-BIH AF database contains 25 ECG recordings in which there are a total of 299 AF episodes. Each ECG recording is approximately 10 hours in duration. The data sets 4936 and 5091 were excluded from our study due to incorrect AF annotations. The MIT-BIH NSR database contains 18 ECG recordings, and the duration of each ECG record is approximately 24 hours. The NSR data do not contain AF and CHF episodes both. The BIDMC CHF database contains 15 ECG recordings with severe CHF (NYHA class III-IV). The individual recordings are each approximately 20 hours in duration. While treating a patient, physicians need to categorize how severely affected a patient is. The most widely used classification scheme is the NYHA (New York Heart Association) classification system. According to this system, a patient's heart failure is classified from I (mildest) to IV (severest). The CHF RRI database includes 29 ECG recordings with congestive heart failure (NYHA I-III). The individual recordings are approximately 20 to 24 hours each. The data sets 02 and 06 in BIDMC CHF, and 207, 213, 214 and 221 in CHF RRI were excluded from our study due to very high premature atrial contraction (PAC) and/or premature ventricular contraction (PVC) frequency (more than 50%), which may interrupt the correct detection. The clinical AF and CHF database consists of 24-hour Holter monitor data collected from 42 subjects using the GE Seer Light Holter monitor; 14 AF, 14 CHF (NYHA III-IV) and 14 NSR subjects. After obtaining informed consent, baseline clinical, demographic, laboratory, and electrophysiologic variables, as well as heart rate and blood pressure, were abstracted from participants' medical records by trained study staff. Data were acquired at 125 sampling rate with 10-bit resolution. We applied low pass filter with cutoff frequency of 50Hz. The annotation of CHF, AF and NSR was performed every beat by cardiologists. The extracted RR intervals were analyzed by the authors using MATLAB 2012b. These studies including data analysis were approved by the Wonkwang University Hospital Institutional Review Board.

2.2 Statistical analysis

ANOVA and Bonferroni t -test was used for analysis of difference between each pair of method. Statistical significant was identified at $p < 0.01$.

2.3 Beat-to-beat-analysis

We used a 16 beat RRI time series and computed statistical values of normalized RMSSD, sample entropy and Shannon entropy. Each statistical value from the RRI time series segment corresponds to the beat in the middle of the segment (i.e. 8th beat). The procedure is repeated by shifting the RRI time series segment every beat. In this way, the beat-to-beat classification of AF, CHF and NSR was performed. A CHF or NSR RRI time series including premature or ectopic beats may result in high statistical values. In addition, missing or false peak detection due to motion and noise artifact (MNA) in Holter monitor measurement also may results in incorrect classification of CHF, NSR and AF. In order to reduce the effects of the premature and ectopic beats as well as MNA, the two longest and two shortest RR values in the segment are considered outliers and are removed from consideration [9, 10]. The outlier removal can prevent from incorrect classification of AF and CHF from PAC and PVC as well.

2.4 Statistical Approach I - Normalized RMSSD

The RMSSD is used to quantify beat-to-beat variability. Since AF exhibits higher variability than CHF and NSR, and CHF exhibits lower variability than AF and NSR, the RMSSD is expected to classify CHF, AF and NSR. As subjects have different mean heart rates, we normalized by dividing the RMSSD by the mean value of the RR time series. Given RR interval time series segment $rr(i)$ with length l , the normalized RMSSD is calculated as

$$\text{normalized RMSSD} = \sqrt{\frac{1}{l-1} \sum_{i=1}^{l-1} (rr(i+1) - rr(i))^2} \quad (1)$$

Based on the databases of the MIT-BIH AF, the MIT-BIH NSR, the BIDMC CHF, the CHF RRI, we calculated the normalized RMSSD with 16 beat RRI time series. Each segment shifted by one beat throughout the whole databases. Fig. 1(a) shows the statistical value distribution of the normalized RMSSD distribution values from 18 NSR subjects from MIT-BIH NSR database, 25 CHF subjects from CHF RRI database, 13 CHF subjects from BIDMC-CHF database, and 23 AF subjects from MIT-BIH AF database. Note that the all statistical values were with 12 beat RRI time series after removing outliers. As shown in Fig. 1(a), the normalized RMSSD values for AF are higher and nearly distinct from those of CHF and NSR databases. In addition, the normalized RMSSD values for

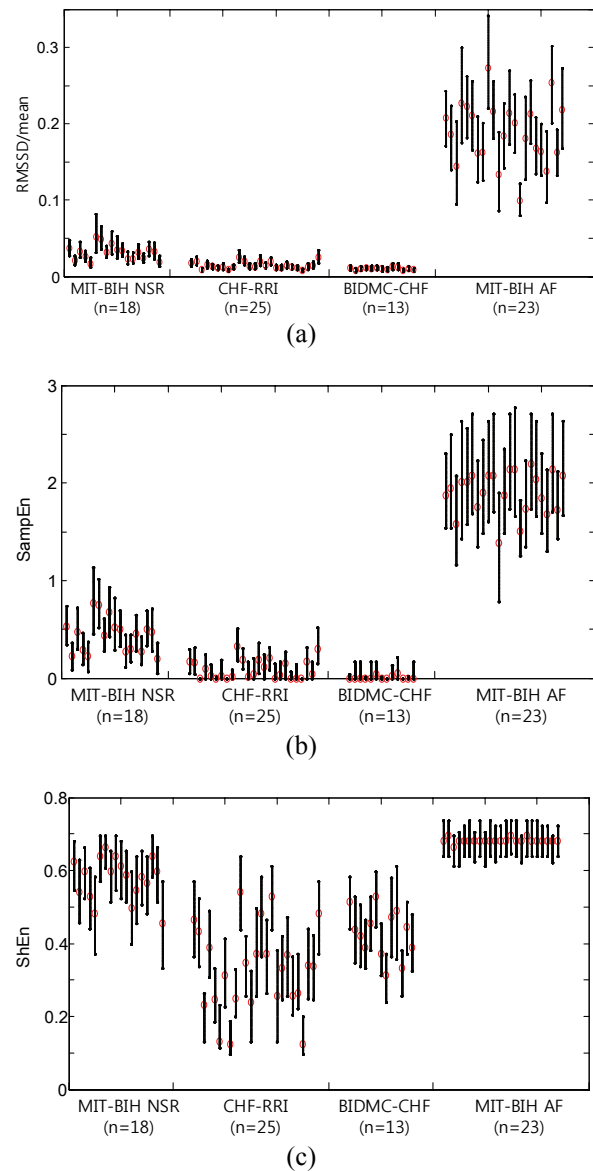


Fig. 1. Distribution of statistical value of : (a) Normalized RMSSD; (b) Sample entropy; (c) Shannon entropy on each data using 16 beats: MIT-BIH NSR, CHF-RRI(CHF NYHA I-III), BIDMC-CHF(CHF NYHA III-IV) and MIT-BIH AF. The bar above, middle, and below represent the 75th, the 50th, and the 25th percentiles, respectively.

CHF are lower and nearly distinct from those of AF and NSR databases. Table 1 shows mean, standard deviation, median and interquartile(IQR) values for NSR, CHF (NYHA I-III), CHF(NYHA III-IV) and AF. The mean value of CHF (NYHA I-III) was 9.08 and 1.69 times lower than those of AF and NSR, respectively. In the CHF (NYHA III-IV), the mean value was 14.13 and 2.62 times lower than those of AF and NSR, respectively. The median value of CHF(NYHA I-III) was 13.56 and 2.17 times lower than those of AF and NSR, respectively. In the CHF (NYHA III-IV), the median value of CHF(NYHA III-IV)

Table 1. Mean, standard deviation, median and interquartile (IQR) values for NSR, CHF (NYHA I-III), CHF (NYHA III-IV) and AF using normalized RMSSD

	Classification	Mean ± SD	Median	IQR
RMSSD/ Mean	AF	0.1907 ± 0.0783	0.1858	0.1090
	NSR	0.0354 ± 0.0274	0.0297	0.0221
	CHF (NYHA I-III)	0.0210 ± 0.0509	0.0137	0.0096
	CHF (NYHA III&IV)	0.0135 ± 0.0287	0.0102	0.0056

was 18.21 and 2.92 times lower than those of AF and NSR, respectively. The IQR value of CHF(NYHA I-III) was 11.35 and 2.30 times lower than those of AF and NSR, respectively. In the CHF(NYHA III-IV), the IQR was 19.46 and 3.96 times lower than those of AF and NSR, respectively. These results were statistically significant with $p < 0.01$ for all pairs: CHF(NYHA I-III) vs NSR, CHF (NYHA I-III) vs. AF, CHF(NYHA III-IV) vs NSR, CHF (NYHA III-IV) vs. AF, CHF(NYHA I-III) vs. CHF (NYHA III-IV) and NSR vs. AF. It also should be noted that the CHF with higher severity provided lower statistical values than the CHF with less severity in normalize RMSSD.

2.5 Statistical Approach II - Sample entropy

The sample entropy is the negative natural logarithm of an estimate of the conditional probability that a subseries that matches pointwise within a tolerance also match at the next point, where self-matches are not included in calculating the probability. Sample entropy is a measure based on comparing patterns within a time series to estimate its complexity. It has been fully characterized in the context of physiological time series analysis and has been successfully used as complexity measure in short-term physiological signal study. A high value of sample entropy indicates low similarity in the time series while a low value of SampEn indicates high similarity. Thus, sample entropy is a useful tool to assess randomness of RR time series.

For sample entropy calculation, the RR times series $rr(i)$ was first partitioned in patterns of its segment length l , and then computed as

$$SampEn(l,r) = -\log \frac{\sum_{i=1}^{N-l} N_i(l+1,r)}{\sum_{i=1}^{N-l} N_i(l,r)} \quad (2)$$

where $N_i(l,r)$ is the number of patterns of length l found at a distance smaller than the threshold r from RR times series with length l , and $N_i(l+1,r)$ is the number of patterns of length $l+1$ found at a distance smaller than r from RR times series with length $l+1$. The sums in (2) were extended to all the patterns of length $l+1$, and the distance was evaluated by the Euclidean norm [23].

Based on the same databases of the MIT-BIH AF, the

Table 2. Mean, standard deviation, median and interquartile (IQR) values for NSR, CHF (NYHA I-III), CHF (NYHA III-IV) and AF using sample entropy

	Classification	Mean ± SD	Median	IQR
SampEn	AF	2.0002 ± 0.6771	1.8971	0.9808
	NSR	0.4705 ± 0.3520	0.4055	0.4305
	CHF (NYHA I-III)	0.1467 ± 0.2123	0.0488	0.2144
	CHF (NYHA III&IV)	0.0623 ± 0.1316	0.0000	0.0783

MIT-BIH NSR, the BIDMC CHF, the CHF RRI, we calculated the sample entropy with 16 beat RRI time series, which was shifted by every beat. Fig. 1(b) shows the statistical value distribution of the SampEn distribution values from 79 same subjects from the same databases as shown for normalized RMSSD. As shown in Fig. 1(b), the sample entropy values for AF are higher and nearly distinct from those of CHF and NSR databases which are similar to the normalized RMSSD. In addition, the sample entropy values for CHF are also lower and nearly distinct from those of AF and NSR databases. Table 2 shows mean, standard deviation, median and interquartile(IQR) values for NSR, CHF(NYHA I-III), CHF(NYHA III-IV) and AF. The mean value of CHF(NYHA I-III) was 13.63 and 3.20 times lower than those of AF and NSR, respectively. In the CHF(NYHA III-IV), the mean value was 32.11 and 7.55 times lower than those of AF and NSR, respectively. The median value of CHF(NYHA I-III) was 38.88 and 8.31 times lower than those of AF and NSR, respectively. In the CHF(NYHA III-IV), the median value of CHF(NYHA III-IV) was zero while those of AF and NSR were 1.8971 and 0.4055, respectively. The IQR value of CHF(NYHA I-III) was 4.57 and 2.01 times lower than those of AF and NSR, respectively. In the CHF(NYHA III-IV), the IQR was 12.53 and 5.50 times lower than those of AF and NSR, respectively. These results were statistically significant with $p < 0.01$ for all pairs: CHF(NYHA I-III) vs NSR, CHF (NYHA I-III) vs. AF, CHF(NYHA III-IV) vs NSR, CHF (NYHA III-IV) vs. AF, CHF(NYHA I-III) vs. CHF(NYHA III-IV) and NSR vs. AF. In addition, similarly to normalized RMSSD, CHF with higher severity provided lower statistical values than the CHF with less severity in sample entropy.

2.6 Statistical Approach III - Shannon entropy

The Shannon entropy provides a quantitative measure of uncertainty for a random variable. Shannon entropy quantifies the probability that runs of patterns exhibiting regularity over some duration of data. For example, a random white noise signal is expected to have the value of Shannon entropy close to one while a simple sinusoidal signal is expected to have the value close to zero. From the characteristics, the Shannon entropy of CHF and NSR are expected to be lower than that of AF. In addition, the

Table 3. Mean, standard deviation, median and interquartile (IQR) values for NSR, CHF (NYHA I-III), CHF (NYHA III-IV) and AF using Shannon entropy

	Classification	Mean \pm SD	Median	IQR
ShEn	AF	0.6762 \pm 0.0634	0.6808	0.0991
	NSR	0.5605 \pm 0.1319	0.5818	0.1722
	CHF (NYHA I-III)	0.3296 \pm 0.1654	0.3247	0.2398
	CHF (NYHA III&IV)	0.4188 \pm 0.1340	0.4151	0.1824

Shannon entropy of CHF is expected to be lower than that of NSR. To calculate Shannon entropy, we constructed a histogram of the RR time series segment considered. The RR values are sorted into equally spaced bins whose limits are defined by the shortest and longest RR values. The probability distribution $p(n)$ is computed for each bin as the number of beats in that bin divided by the total number of beats in the segment, and the Shannon entropy is calculated as follows:

$$ShEn = -\sum_{n=1}^{N_{bin}} p(n) \frac{\log(p(n))}{\log(1/N_{bin})} \quad (3)$$

where N_{bin} is the number of bins.

Based on the same databases of the MIT-BIH AF, the MIT-BIH NSR, the BIDMC CHF, the CHF RRI, we calculated the Shannon entropy with 16 beat RRI time series, which was shifted by every beat. Fig. 1(c) shows the statistical value distribution of the Shannon entropy distribution values from 79 same subjects from the same databases as shown for normalized RMSSD and sample entropy. As shown in Fig. 1(c), the Shannon entropy values for AF are higher and nearly distinct from those of CHF and NSR databases which are similar to the normalized RMSSD and the sample entropy. In addition, sample entropy values for CHF are also lower and nearly distinct from those of AF and NSR databases. Table 3 shows mean, standard deviation, median and interquartile (IQR) values for NSR, CHF (NYHA I-III), CHF (NYHA III-IV) and AF. The mean value of CHF (NYHA I-III) was 2.05 and 1.70 times lower than those of AF and NSR, respectively. In the CHF (NYHA III-IV), the mean value was 1.61 and 1.34 times lower than those of AF and NSR, respectively. The median value of CHF (NYHA I-III) was 2.10 and 1.79 times lower than those of AF and NSR, respectively. In the CHF (NYHA III-IV), the median value of CHF (NYHA III-IV) was 1.64 and 1.40 times lower than those of AF and NSR, respectively. These results were statistically significant with $p < 0.01$ for all pairs except for CHF (NYHA I-III) vs. CHF (NYHA III-IV). It shows that Shannon entropy cannot efficiently discriminate CHF (NYHA I-III) and CHF (NYHA III-IV), but it is still a good tool for discriminating among CHF, NSR and AF.

2.7 Data analysis

The conditions for CHF and AF detection are based on threshold values of normalized RMSSD (RM), sample entropy (SE) and Shannon entropy (SA): TH_{RM}^C , TH_{SE}^C and TH_{SA}^C for CHF, and TH_{RM}^A , TH_{SE}^A and TH_{SA}^A for AF, where $TH_{RM}^C < TH_{RM}^A$, $TH_{SA}^C < TH_{SA}^A$ and $TH_{SE}^C < TH_{SE}^A$. The CHF detection is with normalized RMSSD $\leq TH_{RM}^C$, SampEn $\leq TH_{SA}^C$ and/or ShEn $\leq TH_{SE}^C$ while AF detection is with normalized RMSSD $\geq TH_{RM}^A$, SampEn $\geq TH_{SA}^A$ and/or ShEn $\geq TH_{SE}^A$. For the computation of sample entropy and Shannon entropy, we used $r=0.06$ and $N_{bin}=16$ [10, 11].

For the selection of each optimal threshold, we searched each possible threshold values with the following interval increments:

- 1) TH_{RM}^C varied from 0.001 to 0.050 at intervals of 0.001;
- 2) TH_{SA}^C varied from 0.100 to 0.200 at intervals of 0.001;
- 3) TH_{SE}^C varied from 0.200 to 0.800 at intervals of 0.001;
- 4) TH_{RM}^A varied from 0.010 to 0.150 at intervals of 0.001;
- 5) TH_{SA}^A varied from 0.800 to 1.500 at intervals of 0.001;
- 6) TH_{SE}^A varied from 0.500 to 1.000 at intervals of 0.001.

We first investigated the performance of each single method: normalized RMSSD, sample entropy and Shannon entropy by finding the number of true positives (TP), true negatives (TN), false positive (FP), and false negative (FN) from MIT-BIH AF, MIT-BIH NSR, BIDMC CHF and CHF RRI databases. Each set of TP, TN, FP and FN was calculated for AF and CHF, respectively. Subsequently, we calculated the sensitivity TP/(TP+FN), specificity TN/(TN+FP), and accuracy (TP+TN)/(TP+TN+FP+FN) for AF and CHF. For each statistical method, we found the threshold values using the receiver operating characteristics (ROC) curves for AF and CHF detection, respectively. Finally, the overall sensitivity, specificity and accuracy values were calculated by averaging from the results from AF and CHF.

For the combination method, CHF was detected when normalized RMSSD $\leq TH_{RM}^C$, SampEn $\leq TH_{SA}^C$ AND ShEn $\leq TH_{SE}^C$ while AF was detected when normalized RMSSD $\geq TH_{RM}^A$, SampEn $\geq TH_{SA}^A$ and ShEn $\geq TH_{SE}^A$. As a multivariable analysis, we first fixed as $TH_{SE}^C=0.200$ and $TH_{SA}^C=0.100$, and plotted TH_{RM}^C based ROC curve for CHF detection. Subsequently, we increased TH_{SA}^C from 0.100 to 0.200 at intervals of 0.001 and plotted TH_{RM}^C based ROC curve. The procedure was repeated by increasing TH_{SE}^C from 0.200 to 0.800 at intervals of 0.001. With the resultant ROC curves, we found the ROC curve providing the highest value of ROC area. Similarly, the

ROC analysis for the AF detection was performed. The values of sensitivity, specificity and accuracy from the combined statistical method and each single statistical method are summarized and compared in Table 4 and 5, which are for AF and CHF detection, respectively.

3. Results

3.1 Performance of Single Statistical Method and Its Analysis

Fig. 2(a) shows the normalized RMSSD based ROC curves for AF and CHF detection. The left panel is for AF and the right panel is for CHF. For AF detection, the sensitivity, specificity and accuracy were 0.9522, 0.9649

and 0.9607, respectively, with the threshold value of 0.075. For CHF detection, the sensitivity, specificity and accuracy were 0.7775, 0.9012 and 0.8600, respectively, with the threshold value of 0.0190. The overall sensitivity, specificity and accuracy for AF and CHF were 0.8649, 0.9331 and 0.9104, respectively. In addition, as we expected, the detection rate for CHF (NYHA I-III) was 0.7312 while the rate for CHF (NYHA III-IV) was 0.9113 with the selected threshold value of 0.0190. As a patient is more severe in CHF, the detection rate becomes higher.

Fig. 2(b) shows the sample entropy based ROC curves for AF and CHF detection. The left panel is for AF and the right panel is for CHF. For AF detection, the sensitivity, specificity and accuracy were 0.9380, 0.9684 and 0.9583, respectively, with the threshold value of 1.090. For CHF detection, the sensitivity, specificity and accuracy were 0.7718, 0.8957 and 0.8544, respectively, with the threshold value of 0.184. The overall sensitivity, specificity and accuracy for AF and CHF were 0.8549, 0.9321 and 0.9064, respectively. Similarly, in addition, the detection rate for CHF (NYHA I-III) was 0.7134 while the rate for CHF (NYHA III-IV) was 0.8902 with the selected threshold value of 0.184.

Fig. 2(c) shows the Shannon entropy based ROC curves for AF and CHF detection. The left panel is for AF and the right panel is for CHF. For AF detection, the sensitivity, specificity and accuracy were 0.8336, 0.7738 and 0.7771, respectively, with the threshold value of 0.620. For CHF detection, the sensitivity, specificity and accuracy were 0.7583, 0.8790 and 0.8388, respectively, with the threshold value of 0.474. The overall sensitivity, specificity and accuracy for AF and CHF were 0.7960, 0.8264 and 0.8587, respectively. In addition, the detection rate for CHF (NYHA I-III) was 0.7959 while the rate for CHF (NYHA III-IV) was 0.6776 with the selected threshold value of 0.474. It shows that Shannon entropy is not a good tool for CHF.

3.2 Performance of combined statistical methods and its analysis

Based on the performance of single statistical method as well as each statistical value distribution in Figs. 1(a-c), it was concluded that the normalized RMSSD and the sample entropy were better tools than Shannon entropy for CHF and AF detection, and the normalized RMSSD provided slightly better performance than the sample entropy. In order to investigate the effect of statistical method combination, we performed the ROC curve analysis with the combined normalized RMSSD and sample entropy. Regarding AF detection, the best accuracy was 0.9643, which was only 0.0036 and 0.0060 higher than the normalized RMSSD and the sample entropy, respectively. Regarding CHF detection, the best accuracy was 0.8587, which was even lower than the normalized RMSSD only. In addition, we performed the ROC curve analysis with the

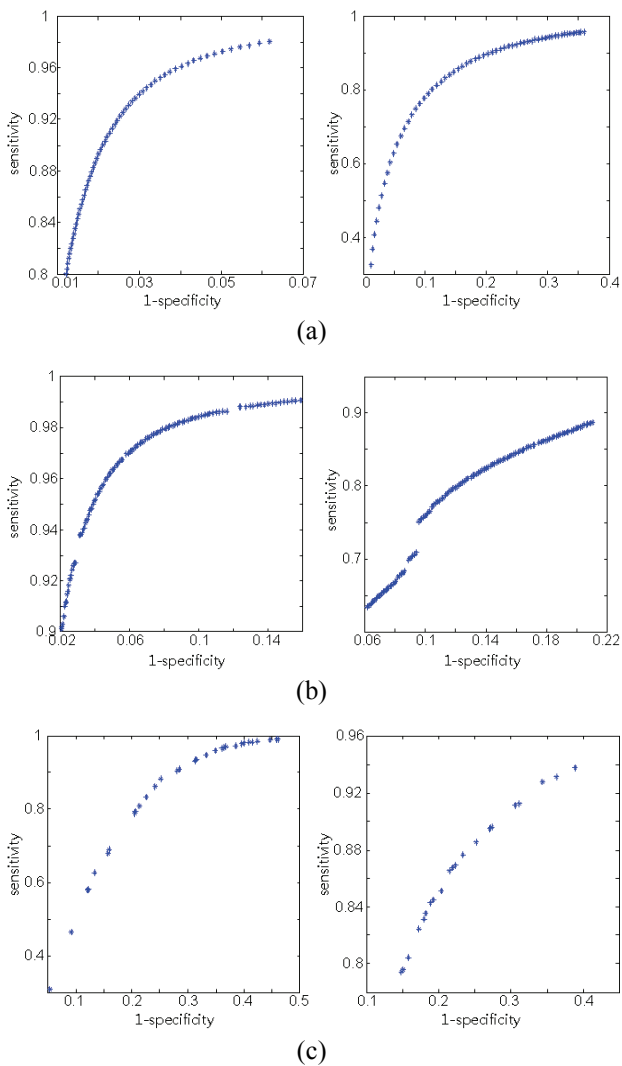


Fig. 2. (a) The normalized RMSSD; (b) The sample entropy; (c) The Shannon entropy based ROC curves providing the largest area with the best accuracy for AF and CHF detection. The left panel is for AF and the right panel is for CHF

Table 4. The values of sensitivity, specificity and accuracy with the combined statistical methods and single statistical methods for AF with four databases: the MIT-BIH AF, the MIT-BIH NSR, the BIDMC CHF, the CHF RRI

	Classification	Sensitivity	Specificity	Accuracy	Threshold
Single Method	Normalized RMSSD	0.9522	0.9649	0.9607	0.075
	SampEn	0.9380	0.9684	0.9583	1.090
	ShEn	0.8336	0.7738	0.7771	0.620
Combined Method	Normalized RMSSD& SampEn	0.9506	0.9712	0.9643	0.072 0.820
	All Three Combination	0.9495	0.9724	0.9648	0.054 0.740 0.515

Table 5. The values of sensitivity, specificity and accuracy with the combined statistical methods and single statistical methods for CHF with four databases: the MIT-BIH AF, the MIT-BIH NSR, the BIDMC CHF, the CHF RRI

	Classification	Sensitivity	Specificity	Accuracy	Threshold
Single Method	Normalized RMSSD	0.7775	0.9012	0.8600	0.019
	SampEn	0.7718	0.8957	0.8544	0.184
	ShEn	0.7583	0.8790	0.8388	0.474
Combined Method	Normalized RMSSD& SampEn	0.7535	0.9112	0.8587	0.020 0.220
	All Three Combination	0.7414	0.9124	0.8554	0.020 0.222 0.600

all three combined method and found the best accuracy with 0.9648 and 0.8554 for AF and CHF, respectively, which did not enhance the performance.

3.3 Clinical data result

With the same procedure, we performed the AF and CHF detection using each 16 beats for the entire 24 hour data, and the data segments were shifted by one beat for the entire Holter recording. As the result, using the normalized RMSSD with $TH_{RM}^C=0.019$, we obtained a sensitivity of 0.8104 and a specificity of 0.9014 for CHF. For AF detection, we obtained a sensitivity of 0.9514 and a specificity of 0.9798 with $TH_{RM}^A=0.019$. The results suggest that normalized RMSSD is applicable to clinical data as well. The other two statistical methods of sample entropy and Shannon entropy also provided similar high values of sensitivity and specificity compared to the databases of MIT-BIH AF, the MIT-BIH NSR, the BIDMC CHF, and the CHF RRI. However, the values were lower than the ones from normalized RMSSD. The clinical data results are summarized in Table 6 and 7.

Table 6. The values of sensitivity, specificity and accuracy with the combined statistical methods and single statistical methods for AF with clinical data set

		Sensitivity	Specificity	Accuracy	Threshold
Single Method	Normalized RMSSD	0.9514	0.9798	0.9656	0.075
	SampEn	0.9210	0.9690	0.9450	1.090
	ShEn	0.7976	0.7831	0.7904	0.620
Combined Method	All Three Combination	0.9510	0.9643	0.9577	0.054 0.740 0.515

Table 7. The values of sensitivity, specificity and accuracy with the combined statistical methods and single statistical methods for CHF with clinical data set

		Sensitivity	Specificity	Accuracy	Threshold
Single Method	Normalized RMSSD	0.8104	0.9014	0.8559	0.075
	SampEn	0.7923	0.8981	0.8452	1.090
	ShEn	0.7644	0.8892	0.8268	0.620
Combined Method	All Three Combination	0.7897	0.9001	0.8449	0.054 0.740 0.515

4. Discussions and Conclusions

The prevailing view considers AF to be high variability and CHF to be very low variability, and this assumption underlies the nonparametric statistical methods for both detections of AF and CHF even with a short RRI segment. From the characteristics of AF and CHF, many algorithms have been proposed to detect AF using RR interval (RRI) variability [9-12, 15, 17, 18]. However, most results were based on long term RR interval time series, especially for CHF detection. Furthermore, the automatic and simultaneous AF, CHF and NSR classification has not been studied. We tested three statistical methods using 16 beats based on four large well documented databases ($n=79$), and found that normalized RMSSD provided the best accuracy followed by sample entropy and Shannon entropy for both CHF and AF. Even any combination method did not outperform the normalized RMSSD method only. In addition, we found the similar values of sensitivity, specificity and accuracy from clinical data set ($n=42$). We are further collecting more extensive CHF and AF patient data, and validating our results. We hope to report about further investigation of the extensive database in a forthcoming publication.

Acknowledgements

Kwon-Ha Yoon and Yunyoung Nam contributed equally to this work. They are co-first authors. This study was partially supported by a grant from the Korean Health Technology R&D Project, Ministry of Health and Welfare, Republic of Korea (HI12C0110), Basic Science Research

Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1004033), the Bio & Medical Technology Development Program of the NRF funded by the Korean government, MSIP(NRF-2015M3A9D7067219), and the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8601-16-1009) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

- [1] A.S. Go, et al., "Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTi-coagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study," *JAMA : the journal of the American Medical Association*, vol. 285, no. 18, 2001, pp. 2370-2375.
- [2] Y. Miyasaka, et al., "Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence," *Circulation*, vol. 114, no. 2, 2006, pp. 119-125; DOI 10.1161/CIRCULATIONAHA.105.595140.
- [3] I. Hajjar and T.A. Kotchen, "Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988-2000," *JAMA : the journal of the American Medical Association*, vol. 290, no. 2, 2003, pp. 199-206; DOI 10.1001/jama.290.2.199.
- [4] E. Q. Wu, et al., "Economic burden and comorbidities of atrial fibrillation in a privately insured population," *Current medical research and opinion*, vol. 21, no. 10, 2005, pp. 1693-1699; DOI 10.1185/030079905X65475.
- [5] S.A. Lubitz, et al., "Atrial fibrillation in congestive heart failure," *Heart failure clinics*, vol. 6, no. 2, 2010, pp. 187-200; DOI 10.1016/j.hfc.2009.11.001.
- [6] M. M. Scheinman, "Atrial fibrillation and congestive heart failure: the intersection of two common diseases," *Circulation*, vol. 98, no. 10, 1998, pp. 941-942.
- [7] S.A. Hunt, et al., "ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to revise the 1995 Guidelines for the Evaluation and Management of Heart Failure)," *Journal of the American College of Cardiology*, vol. 38, no. 7, 2001, pp. 2101-2113.
- [8] M. H. Chin and L. Goldman, "Correlates of major complications or death in patients admitted to the hospital with congestive heart failure," *Archives of internal medicine*, vol. 156, no. 16, 1996, pp. 1814-1820.
- [9] J. Lee, et al., "Time-varying coherence function for atrial fibrillation detection," *IEEE transactions on bio-medical engineering*, vol. 60, no. 10, 2013, pp. 2783-2793; DOI 10.1109/TBME.2013.2264721.
- [10] S. Dash, et al., "Automatic real time detection of atrial fibrillation," *Annals of biomedical engineering*, vol. 37, no. 9, 2009, pp. 1701-1709; DOI 10.1007/s10439-009-9740-z.
- [11] D.E. Lake and J.R. Moorman, "Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices," *American journal of physiology. Heart and circulatory physiology*, vol. 300, no. 1, 2011, pp. H319-325; DOI 10.1152/ajpheart.00561.2010.
- [12] C. Huang, et al., "A novel method for detection of the transition between atrial fibrillation and sinus rhythm," *IEEE transactions on bio-medical engineering*, vol. 58, no. 4, 2011, pp. 1113-1119; DOI 10.1109/TBME.2010.2096506.
- [13] D.D. McManus, et al., "A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation," *Heart rhythm : the official journal of the Heart Rhythm Society*, vol. 10, no. 3, 2013, pp. 315-319; DOI 10.1016/j.hrthm.2012.12.001.
- [14] J. Lee, et al., "Atrial fibrillation detection using an iPhone 4S," *IEEE transactions on bio-medical engineering*, vol. 60, no. 1, 2013, pp. 203-206; DOI 10.1109/TBME.2012.2208112.
- [15] G. Casolo, et al., "Decreased spontaneous heart rate variability in congestive heart failure," *The American journal of cardiology*, vol. 64, no. 18, 1989, pp. 1162-1167.
- [16] L. Fei, et al., "Heart rate variability and its relation to ventricular arrhythmias in congestive heart failure," *British heart journal*, vol. 71, no. 4, 1994, pp. 322-328.
- [17] Y. Isler and M. Kuntalp, "Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure," *Computers in biology and medicine*, vol. 37, no. 10, 2007, pp. 1502-1510; DOI 10.1016/j.compbiomed.2007.01.012.
- [18] T.P. Thakre and M.L. Smith, "Loss of lag-response curvilinearity of indices of heart rate variability in congestive heart failure," *BMC cardiovascular disorders*, vol. 6, 2006, pp. 27; DOI 10.1186/1471-2261-6-27.
- [19] R. Maestri, et al., "Nonlinear indices of heart rate variability in chronic heart failure patients: redundancy and comparative clinical value," *Journal of cardiovascular electrophysiology*, vol. 18, no. 4, 2007, pp. 425-433; DOI 10.1111/j.1540-8167.2007.00728.x.
- [20] R. A. Thuraisingham, "Preprocessing RR interval time series for heart rate variability analysis and

estimates of standard deviation of RR intervals,” *Computer methods and programs in biomedicine*, vol. 83, no. 1, 2006, pp. 78-82; DOI 10.1016/j.cmpb.2006.05.002.

- [21] C. S. Poon and C. K. Merrill, “Decrease of cardiac chaos in congestive heart failure,” *Nature*, vol. 389, no. 6650, 1997, pp. 492-495; DOI 10.1038/39043.
- [22] H.M. Al-Angari and A.V. Sahakian, “Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome,” *IEEE transactions on bio-medical engineering*, vol. 54, no. 10, 2007, pp. 1900-1904; DOI 10.1109/TBME.2006.889772.
- [23] L. Faes, et al., “Time-varying surrogate data to assess nonlinearity in nonstationary time series: application to heart rate variability,” *IEEE transactions on bio-medical engineering*, vol. 56, no. 3, 2009, pp. 685-695; DOI 10.1109/TBME.2008.2009358



Kwan-Ha Yoon He received his BS degree at college of medicine from Wonkwang University in 1992, and Ph.D. degrees at college of medicine from Jeonbuk National University in 2002 in South Korea. Currently, he is a professor at dept. of Radiology in college of medicine, Wonkwang Uni-

versity. His main research interests include x-ray imaging, nano medicine, molecular imaging, and gastrointestinal imaging. He received awards for outstanding scientific paper published in Journal of Korean Radiological Society at April 1999 and Cum Laude from RSNA at December 2001. He is a member of KCR, RSNA, and ISMRM.



Yunyoung Nam He received the B.S., M.S., and Ph.D. degrees in computer engineering from Ajou University, Korea in 2001, 2003, and 2007 respectively. He was a Senior Researcher in the Center of Excellence in Ubiquitous System (CUS) 2007 to 2010. He was a Research Professor in Ajou

University from 2010 to 2011. He also spent time as a Visiting Scholar at Center of Excellence for Wireless & Information Technology (CEWIT), Stony Brook University, New York. He was a Postdoctoral Fellow at Worcester Polytechnic Institute, Massachusetts from 2013 to 2014. He is currently an assistant professor in the Department of Computer Science and Engineering at Soonchunhyang University. His research interests include multimedia database, ubiquitous computing, image processing, pattern recognition, context-awareness, wearable computing, intelligent video surveillance, cloud computing, and biomedical signal processing.



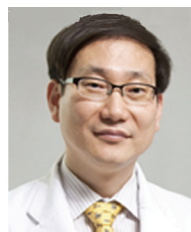
Tharocun Thap He received M.S. degree in Electrical Communication System Engineering from Soonchunhyang University, Korea, in 2013, and B.S. degree in Electronics Engineering from National Polytechnic Institute of Cambodia in 2009. He is currently doing his Ph.D. course in Biomedical

Engineering at Wonkwang University College of Medicine. His current research interests include biosignal processing and analysis, development of innovative algorithm for smartphone-based non-invasive physiological monitoring, underwater electrocardiogram (ECG) monitoring, and pulmonary function test using smartphone microphone.



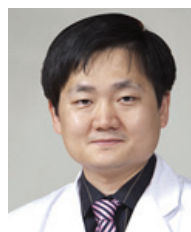
Changwon Jeong He received his BS degree in computer engineering from Wonkwang University in 1993, and M.S. and Ph.D. degrees at dept. of computer engineering from Wonkwang University in 1998 and 2003, respectively, in South Korea. Currently, he is a Research Fellow at Imaging Science

based Lung and Bone Disease Research Center in Wonkwang University. His main research interests include distributed object computing, middleware and u-healthcare.



Nam Ho Kim He received Ph.D. degree in medicine from Chonnam national University in 2003. He completed Cardiology fellowship training in Chonnam national university hospital in 2000 and research fellowship in Harvard medical school in 2007. Since 2012, he has been a professor of internal

medicine at Wonkwang University College of Medicine. His research interests are electrophysiology and arrhythmia.



Jeom Suk Ko He received Ph.D. degree in medicine from Chonnam national University in 2013. He completed Cardiology fellowship training in Chonnam national university hospital in 2011 and research fellowship in Indiana university medical school in 2016. Since 2012, he has been an

assistant professor of internal medicine at Wonkwang University College of Medicine. His research interests are electrophysiology and arrhythmia.



Se-Eung Noh He received the B.S. and M.S. degrees in medicine from Wonkwang University College of Medicine, Korea, in 2004 and 2008 respectively, and Ph.D. degree in rehabilitation medicine from Pusan National University in 2014. He was a clinical assistant professor in Wonkwang University

Hospital from 2010 to 2012. Since 2013, he has been an assistant professor in the department of rehabilitation medicine, Wonkwang University School of Medicine.



Jinseok Lee He received Ph.D. degree in Electrical Engineering from Stony Brook University in 2009, and dual B.S. degrees in Electrical Engineering from Stony Brook University and Ajou University in 2005. He completed postdoctoral training in Biomedical Engineering, Worcester Polytechnic

Institute in 2012. Since 2012, he has been an assistant professor of Biomedical Engineering at Wonkwang University College of Medicine. His research interests are medical instrumentation, biosignal processing, modeling, simulation and development of novel algorithms to understand dynamic processes and extract features of physiological systems.