IOP Publishing Physiol. Meas. 39 (2018) 074002 (18pp) https://doi.org/10.1088/1361-6579/aacc48

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

A comparison of entropy approaches for AF discrimination

Chengyu Liu[1,2], Julien Oster[3], Erik Reinertsen[5], Qiao Li[2], Lina Zhao[1,4], Shamim Nemati[2] and Gari D Clifford[2,5]

[1] The State Key Laboratory of Bioelectronics, Jiangsu Key Lab of Remote Measurement and Control, School of Instrument Science and Engineering, Southeast University, Nanjing, People's Republic of China
[2] Department of Biomedical Informatics, Emory University, Atlanta, GA, United States of America
[3] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom
[4] School of Control Science and Engineering, Shandong University, Jinan, People's Republic of China
[5] Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America

E-mail: chengyu@seu.edu.cn **and** gari@alum.mit.edu

RECEIVED
9 February 2018

REVISED
16 May 2018

ACCEPTED FOR PUBLICATION
13 June 2018

PUBLISHED
5 July 2018

Keywords: atrial fibrillation, ECG, entropy, normalized fuzzy entropy

Abstract

Objective: This study focuses on the comparison of single entropy measures for ventricular response analysis-based AF detection. *Approach*: To enhance the performance of entropy-based AF detectors, we developed a normalized fuzzy entropy, $\mathcal{H}_N^\theta$, a novel metric that (1) uses a fuzzy function to determine vector similarity, (2) replaces probability estimation with density estimation for entropy approximation, (3) utilizes a flexible distance threshold parameter, and (4) adjusts for heart rate by subtracting the natural log value of the mean RR interval. An AF detector based on $\mathcal{H}_N^\theta$ was trained using the MIT-BIH atrial fibrillation (AF) database, and tested on the MIT-BIH normal sinus rhythm (NSR) and MIT-BIH arrhythmia databases. The $\mathcal{H}_N^\theta$-based AF detector was compared to AF detectors based on three other entropy measures: sample entropy ($\mathcal{H}^\sigma$), fuzzy measure entropy ($\mathcal{H}^\theta$) and coefficient of sample entropy ($\mathcal{H}^c$), over three standard window sizes. *Main results*: To classify AF and non-AF rhythms, $\mathcal{H}_N^\theta$ achieved the highest area under receiver operating characteristic curve (*AUC*) values of 92.72%, 95.27% and 96.76% for 12-, 30- and 60-beat window lengths respectively. This was higher than the performance of the next best technique, $\mathcal{H}^c$, over all windows sizes, which provided respective *AUCs* of 91.12%, 91.86% and 90.55%. $\mathcal{H}^\sigma$ and $\mathcal{H}^\theta$ resulted in lower *AUCs* (below 90%) over all window sizes. $\mathcal{H}_N^\theta$ also provided superior performance for all other tested statistics, including the Youden index, sensitivity, specificity, accuracy, positive predictivity and negative predictivity. In conclusion, we show that $\mathcal{H}_N^\theta$ can be used to accurately identify AF from RR interval time series. Furthermore, longer window lengths (up to one minute) increase the performance of all entropy-based AF detectors under evaluation except the $\mathcal{H}^c$-based method. *Significance*: Our results demonstrate that the new developed normalized fuzzy entropy is an accurate measure for detecting AF.

1. Introduction

1.1. Atrial fibrillation and its prevalence

Atrial fibrillation (AF), defined as a 'tachyarrhythmia characterized by predominantly uncoordinated atrial activation with consequent deterioration of atrial mechanical function' by the American College of Cardiology, the American Heart Association and the European Society of Cardiology (Fuster *et al* 2001), is a disorder of the heart's electrical conduction system that leads to a fast and irregular heart rhythm, and acts as the most common sustained heart rhythm disorder, occurring in 1%–2% of the general population (Camm *et al* 2010, Lip *et al* 2016). AF is associated with significant mortality and morbidity because it increases the risk of death, stroke, hospitalization, heart failure and coronary artery disease, etc (Camm *et al* 2010, Colloca 2013). More than 12 million Europeans and North Americans are estimated to suffer from AF, and its prevalence will likely triple in the next 30–50 years (Savelieva and Camm 2008, Mozaffarian *et al* 2015). More importantly, the incidence of AF is destined to increase with the aging population (Wang *et al* 2003). The prevalence of AF increases with age, from less than 0.5% at 40–50 years of age, to 5%–15% for 80 year olds (Heeringa *et al* 2006, Naccarelli *et al* 2009).

1.2. Mechanism and clinical classification

The causes of AF are broadly cardiovascular and non-cardiovascular. The common cardiovascular risk factors for AF include hypertension, heart failure and ischaemic heart disease. The common non-cardiovascular risk factors for AF include sepsis, chest infection and obstructive sleep apnoea (Lip *et al* 2016). Although the precise pathobiological mechanisms of AF remain under investigation, they are thought to involve cardiac fibrosis and remodelling, which alter the way electrical impulses are propagated through the heart (Lip *et al* 2016). AF occurs because electrical signals are not systematically triggered via the sinoatrial node and follow an abnormal conduction pathway. The central feature of AF is very rapid and uncoordinated atrial activity at a rate up to 300–600 beats per minute. The ventricle's response depends on the atrial rate and on the filtering function of the atrioventricular (AV) node, which conducts only according to its refractory period. Irregular impulses to the ventricles cause a fast heart rate (up to 150 beats per minute) and less effective contractions, i.e. reduce cardiac output. AF is commonly classified into three types in clinical practice: paroxysmal if it self-terminates within 7 d, persistent if it lasts continuously for more than 7 d, and long-standing persistent if it is present continuously for more than 1 year, or as permanent (chronic) arrhythmia (Colloca 2013, Lip *et al* 2016).

1.3. AF is currently under-detected

The prevalence of AF is wide. However, its detection is insufficient in clinical practice. AF is usually diagnosed based on clinical grounds, but symptoms, including palpitations, fatigue, dizziness, light-headedness and dyspnea, are non-specific and are frequently absent (Savelieva and Camm 2000), especially in elderly patients. AF is commonly asymptomatic and the first presentation with AF is often in association with a devastating AF-related complication, including a four- to five-fold increased risk of stroke (Wolf *et al* 1991) and a two- to three-fold increased risk of heart failure (Wang *et al* 2003). The absence of symptoms is not associated with a lower consequential risk of myocardial infarction, heart failure or stroke; AF is thus under-detected and under-diagnosed (Lip *et al* 2016). AF diagnosis by symptoms leads to a large under-representation of the disturbance, since only one out of three patients may have been admitted to hospital (Hughes and Lip 2008).

In addition, according to data from the JAMA (Go *et al* 2001), 25%–62% of people with AF have paroxysmal AF. Meanwhile, more than 25% of people with paroxysmal AF will go on to develop persistent or permanent AF. However, paroxysmal AF is more difficult to recognize and detect than persistent or permanent AF (Gami *et al* 2007) since paroxysmal AF is usually accidental and could last only for several seconds, and paroxysmal AF is even less symptomatic, and it was estimated that only 1 in 12 paroxysmal AF patients are symptomatic (Go *et al* 2001). However, the type of AF—permanent or paroxysmal—does not statistically modify the risk of stroke occurrence (Go *et al* 2001, Gami *et al* 2007). Thus, detection for AF, especially for the short-term and less symptomatic paroxysmal AF, is still challenging in clinical applications. In the near future, increasing recognition of AF using ambulatory and mobile ECG might modify the global AF picture.

1.4. Atrial activity analysis-based and ventricular response analysis-based AF detector

AF is associated with rapid uncoordinated atrial activations. On an electrocardiogram (ECG), AF is characterized by the replacement of P waves with rapid oscillated f waves that vary in size, shape, and timing (Fuster *et al* 2001, Abusaada *et al* 2004). The rapid uncoordinated atrial activations also induce an irregular ventricular rhythm in ECG signals. Thus, AF detectors are commonly classified into two types: atrial activity analysis-based and ventricular response analysis-based methods.

Atrial activity analysis-based AF detectors focus on detecting rapidly oscillated f waves. It is similar to the diagnosis process in clinics and thus it is a direct AF detection method. Many AF detectors based on atrial activity analysis have been proposed. These methods are mainly based on the analysis of the absence of P waves or the fibrillatory f waves present in the TQ interval. Typical methods include an echo state neural network (Petrnas *et al* 2012), a P-wave absence (PWA) based detector (Ladavich and Ghoraani 2015), using the average number of f waves as an AF feature (Du *et al* 2014), P-wave-based insertable cardiac monitor application (Prerfellner *et al* 2014), wavelet sample entropy (Alcaraz *et al* 2006), wavelet entropy (Alcaraz and Rieta 2012, Rdenas *et al* 2015) and the relative wavelet energy method (Garca *et al* 2016).

Atrial activity analysis-based AF detectors can achieve high accuracy if the recorded ECG signals have a high signal quality, but it can be difficult when the ECG is distorted by noise, typical of ambulation, and can even be problematic in high fidelity monitoring (Colloca 2013). For atrial activity analysis, a stable, high quality ECG signal is required. However, a high quality ECG is difficult to obtain in real-time long-term recordings. In addition, even supposing high quality ECG is possible, sometimes a standard 12-lead ECG system may not be enough to study the atrial activity because the number of electrodes is too small and their location is not optimal to highlight f-wave peaks (Colloca 2013). Meanwhile, rapidly oscillated f waves vary from patient to patient (Fuster *et al* 2001, Abusaada *et al* 2004), resulting in the fact that no overall consensus exists when linking the amplitude of the f waves to the etiology of AF (Mainardi *et al* 2008).

In contrast, ventricular response analysis is based on the predictability of the inter-beat timing (RR intervals) of the ventricular contractions in the ECG. RR intervals are derived from the most obvious large amplitude feature in the ECG: the R-peak. This approach is robust to artifacts, and can be used in situations where the ECG is noisy or unavailable (by using pulsatile signals), and thus is most suitable for automatic, real-time screening AF detection applications (Park *et al* 2009, Carrara *et al* 2015).

Many ventricular response analysis-based AF detectors have also been proposed, including Poincare plot analysis (Park *et al* 2009, Ruan *et al* 2011), Lorenz plot analysis (Sarkar *et al* 2008), cumulative distribution functions (Tateno and Glass 2001), median absolute deviation of RR intervals (Linker 2009), density histogram of delta RR intervals (Huang *et al* 2011), minimum of the corrected conditional entropy of the RR interval sequence (Cerutti *et al* 2008), an 8-beat sliding window RR interval irregularity detector (Petrnas *et al* 2015a), symbolic dynamics and Shannon entropy (Zhou *et al* 2014), sample entropy, $\mathcal{H}^{\sigma}$ (usually written SampEn) (Alcaraz *et al* 2010), coefficient of sample entropy, $\mathcal{H}^{c}$, also referred to as COSEn (Lake and Moorman 2011, DeMazumder *et al* 2013).

It is worth noting that AF detectors with a combination of both atrial activity and ventricular response could enhance their performance. The combined AF detectors include RR interval Markov modeling combined with PR interval variability and a P-wave morphology similarity measure (Babaeizadeh *et al* 2009), achieving 92% sensitivity and 97% positive predictive value by adding atrial activity analysis, fuzzy logic classification using a combination of RR interval irregularity, PWA, f-wave presence, and noise level (Petrnas *et al* 2015b) resulting in a detection accuracy of 88% when using as few as five beats (compared with 82% only using the rhythm information). It is also worth noting that multivariable approaches based on machine learning and signal quality assessment that combine several of the above single features could also enhance the performance of AF detectors (more than 95% AF detection accuracy) (Colloca *et al* 2013, Oster and Clifford 2015).

1.5. Entropy methods and their applications in AF detection

Recently, entropy-based methods have been introduced as a quantification of regularity in time series, initially motivated by applications to the relatively short time series (Pincus and Goldberger 1994, Richman and Moorman 2000). Entropy refers to the degree of regularity or irregularity of a time series and is estimated by counting how many 'template' patterns repeat. Repeated patterns imply increased regularity in the time series and lead to low entropy values (Pincus and Goldberger 1994, Richman and Moorman 2000, Costa *et al* 2005). Pincus and Goldberger (1994) developed approximate entropy (ApEn), and Richman and Moorman (2000) developed SampEn ($\mathcal{H}^{\sigma}$) for short-term physiological signal analysis, defining entropy as the conditional probability that two short vectors of length m that match within a distance tolerance r will also match at the $m+1$st point. Because the calculation methods are relatively easy, these two methods have achieved wide applications in a variety of studies. Unlike ApEn, $\mathcal{H}^{\sigma}$ does not count self-matches (Richman and Moorman 2000), which significantly reduces bias but also lowers the counts of vector matching. This can result in infinite or indeterminate outputs and is especially problematic for a short-term time series (Lake and Moorman 2011). We previously studied the phenomenon of 'weak statistical stability' in $\mathcal{H}^{\sigma}$ and found that this weak statistical stability is due to the rigid determination rule (0–1 determination; vectors are either similar or dissimilar) (Liu and Zhao 2011, Liu *et al* 2013). We therefore replaced the 0–1 determination with the fuzzy rule, proposing a fuzzy measure entropy method, $\mathcal{H}^{\theta}$, also referred to as FuzzyMEn, for a univariate time series (Liu and Zhao 2011, Liu *et al* 2013) and a cross fuzzy measure entropy method ($\mathcal{H}^{\theta}_{X}$) for a bivariate time series (Liu *et al* 2015). The calculation process for $\mathcal{H}^{\theta}$ is similar to that of $\mathcal{H}^{\sigma}$ but employs a fuzzy function rather than a Heaviside function for vector similarity determination to reduce the sudden change of entropy values when threshold r changes slightly. Thus, the $\mathcal{H}^{\theta}$ approach is more statistically stable than previous entropy measures (Liu *et al* 2013, Zhao *et al* 2015). The SampEn-based developments also include its generalizations for multiscale analysis (Costa *et al* 2002, 2005) and multivariate multiscale analysis (Ahmed and Mandic 2011).

The entropy-based AF detectors also included two aspects, i.e. using the entropy method for atrial activity analysis and ventricular response analysis. For atrial activity analysis, entropy is usually used as the regularity measure of a P-wave episode in the TQ interval of an ECG signal to detect the appearance (or not) of the potential f waves. Analysis methods include using the wavelet sample entropy to directly measure the P-wave episode or using the wavelet entropy to measure the wavelet transform of the P-wave episode to detect AF (Alcaraz and Rieta 2012, Rdenas *et al* 2015).

As mentioned above, an atrial activity analysis-based AF detector relies significantly on the ECG signal quality which is difficult for real-time and high fidelity monitoring applications. For robust ventricular response analysis using the entropy method, Lake and colleagues reported that when using a coefficient of a sample entropy ($\mathcal{H}^{c}$) statistic, adapted from $\mathcal{H}^{\sigma}$, the performance of AF detection was significantly enhanced, even though only 12-beat RR segments were used (Lake and Moorman 2011, DeMazumder *et al* 2013). The authors reported an area under curve (*AUC*) of 92.8% when testing $\mathcal{H}^{c}$ on the MIT-BIH AF database. Improvements from $\mathcal{H}^{\sigma}$ to $\mathcal{H}^{c}$ were due to two key additions: the flexibility in choosing the distance threshold r, and adjusting for heart rate by

subtracting the negative natural log of the mean RR interval. Since AF events should usually last 30 s or longer in order to be considered clinically relevant (Heeringa *et al* 2006), the $\mathcal{H}^c$ was also applied to 30 s RR segments in Carrara *et al* (2015). This entropy metric could achieve positive predictive values higher than 90% when used on the University of Virginia Holter database.

1.6. Aim of this study

In this study, we propose a new entropy-based AF detector, normalized fuzzy entropy ($\mathcal{H}_N^\theta$), that combines the advantages of both $\mathcal{H}^\theta$ and $\mathcal{H}^c$ methods. We compare the robustness and accuracy of the AF detectors based on $\mathcal{H}_N^\theta$ versus those based on three existing entropy metrics: $\mathcal{H}^\sigma$, $\mathcal{H}^\theta$ and $\mathcal{H}^c$. In addition, recent studies report that an optimal window size of 41 beats was applied to the MIT-BIH AF database when using the multiple feature (ten)-based machine learning method for AF classification (Colloca *et al* 2013, Oster and Clifford 2015). Due to the wide variety of window lengths proposed as optimal time windows for entropy-based AF detectors, we also evaluate the performance of different entropies using three different window sizes: 12-beat, 30-beat and 60-beat RR segments.

We notice that the performance of the AF detector could benefit from the combination of atrial activity analysis or the machine learning-based multifeature fusion approach, especially after the PhysioNet/Computing in Cardiology Challenge 2017 entitled 'AF classification from a short single lead ECG recording' (Clifford *et al* 2017). However, in the current study, we mainly focus on the comparison of single entropy measures for ventricular response analysis-based AF detection, i.e. identifying an AF episode by performing the single entropy method on the short-term RR interval time series (within 1 min) to provide a simple and robust AF detection measure for automatic, real-time and ambulatory screening AF applications.

2. Entropy methods

2.1. Baseline algorithms

A sample entropy ($\mathcal{H}^\sigma$), fuzzy measure entropy ($\mathcal{H}^\theta$) and coefficient of sample entropy ($\mathcal{H}^c$) were taken as baseline algorithms in this study, and are described in the appendix. Herein, we pre-define two commonly used parameters in the calculation of entropy metrics: embedding dimension $m = 1$ and distance threshold $r = 0.1$ times the standard deviation of the RR interval time series. We set $m = 1$ because the appropriate embedding dimension m is suggested to deal with the time series with a length of 10^m to 10^{m+1} (Lake and Moorman 2011). Entropy is influenced with the setting of r, and $r = 0.1$ times the standard deviation is verified to provide more stable outputs for the short-term RR interval time series (Zhao *et al* 2015).

2.2. Normalized fuzzy entropy

The core calculation of $\mathcal{H}^c$, similar to $\mathcal{H}^\sigma$, uses a Heaviside function to classify vector similarity in a binary fashion. This rigid determination results in weak statistical stability. In contrast, $\mathcal{H}^\theta$ uses a fuzzy function to smooth the decision boundary and is far less sensitive to small changes in r (Liu and Zhao 2011, Liu *et al* 2013). In this study, we generate $\mathcal{H}_N^\theta$ from $\mathcal{H}^\theta$ using a similar method as was used to generate $\mathcal{H}^c$ from $\mathcal{H}^\sigma$. We now describe this process.

For an RR time series $x(i)(1 \leqslant i \leqslant N)$, firstly form the local vector sequences L_i^m and global vector sequences G_i^m respectively $(1 \leqslant i \leqslant N - m)$:

$$L_i^m = \{x(i), x(i+1), \cdots, x(i+m-1)\} - \bar{x}(i) \tag{1}$$

$$G_i^m = \{x(i), x(i+1), \cdots, x(i+m-1)\} - \bar{x}. \tag{2}$$

The vector L_i^m represents m consecutive $x(i)$ with the local mean, $\bar{x}(i) = \frac{1}{m}\sum_{k=0}^{m-1} x(i+k)$, subtracted. The vector G_i^m also represents m consecutive $x(i)$ values with the global mean, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x(i)$, removed.

The distance between the local vector sequences L_i^m and L_j^m, and the distance between the global vector sequences G_i^m and G_j^m are respectively defined as

$$dL_{i,j}^m = d[L_i^m, L_j^m] = \max_{0 \leqslant k \leqslant m-1} |(x(i+k) - \bar{x}(i)) - (x(j+k) - \bar{x}(j))| \tag{3}$$

$$dG_{i,j}^m = d[G_i^m, G_j^m] = \max_{0 \leqslant k \leqslant m-1} |(x(i+k) - \bar{x}) - (x(j+k) - \bar{x})|. \tag{4}$$

We calculate the similarity degree $DL_{i,j}^m(n_L, r_L)$ between local vectors L_i^m and L_j^m by the fuzzy function $\mu L(dL_{i,j}^m, n_L, r_L)$, and also calculate the similarity degree $DG_{i,j}^m(n_G, r_G)$ between global vectors G_i^m and G_j^m by the fuzzy function $\mu G(dG_{i,j}^m, n_G, r_G)$ as

$$DL_{i,j}^m(n_L, r_L) = \mu L(dL_{i,j}^m, n_L, r_L) = \exp\left(-\frac{(dL_{i,j}^m)^{n_L}}{r_L}\right) \tag{5}$$

$$DG_{i,j}^{m}(n_G, r_G) = \mu G(dG_{i,j}^{m}, n_G, r_G) = \exp(-\frac{(dG_{i,j}^{m})^{n_G}}{r_G}) \tag{6}$$

where n_L is the local similarity weight, r_L is the local tolerance threshold, n_G is the global similarity weight and r_G is the global tolerance threshold.

We define the functions $BL^m(n_L, r_L)$ and $BG^m(n_G, r_G)$ as

$$BL^m(n_L, r_L) = \frac{1}{N-m}\sum_{i=1}^{N-m}(\frac{1}{N-m}\sum_{j=1}^{N-m} DL_{i,j}^{m}(n_L, r_L)) \tag{7}$$

$$BG^m(n_G, r_G) = \frac{1}{N-m}\sum_{i=1}^{N-m}(\frac{1}{N-m}\sum_{j=1}^{N-m} DG_{i,j}^{m}(n_G, r_G)). \tag{8}$$

$BL^m(n_L, r_L)$ and $BG^m(n_G, r_G)$ measure the mean similarity degrees for the local and global vectors at dimension m respectively. Similarly, we define the functions of mean similarity degrees $AL^{m+1}(n_L, r_L)$ and $AG^{m+1}(n_G, r_G)$ for dimension $m+1$ respectively as

$$AL^{m+1}(n_L, r_L) = \frac{1}{N-m}\sum_{i=1}^{N-m}(\frac{1}{N-m}\sum_{j=1}^{N-m} DL_{i,j}^{m+1}(n_L, r_L)) \tag{9}$$

$$AG^{m+1}(n_G, r_G) = \frac{1}{N-m}\sum_{i=1}^{N-m}(\frac{1}{N-m}\sum_{j=1}^{N-m} DG_{i,j}^{m+1}(n_G, r_G)). \tag{10}$$

Then, unlike $\mathcal{H}^\theta$ which uses a probability-based method (equations A16 and A17 in the appendix), we use a density-based method to generate a quadratic fuzzy local measure entropy ($\mathcal{H}_L^\theta$) and a quadratic fuzzy global measure entropy ($\mathcal{H}_G^\theta$) using the volume of each matching region, i.e. $(2r)^m$ as[6]

$$\mathcal{H}_L^\theta = -\ln(\frac{AL^{m+1}(n_L, r_L)/(2r_L)^{m+1}}{BL^m(n_L, r_L)/(2r_L)^m}) = -\ln(\frac{AL^{m+1}(n_L, r_L)}{BL^m(n_L, r_L)}) + \ln(2r_L) \tag{11}$$

$$\mathcal{H}_G^\theta = -\ln(\frac{AG^{m+1}(n_G, r_G)/(2r_G)^{m+1}}{BG^m(n_G, r_G)/(2r_G)^m}) = -\ln(\frac{AG^{m+1}(n_G, r_G)}{BG^m(n_G, r_G)}) + \ln(2r_G). \tag{12}$$

We also subtract the natural log of the mean RR interval from $\mathcal{H}_L^\theta$ and $\mathcal{H}_G^\theta$ (as per Lake and Moorman (2011), and equation A21 in the appendix) as follows:

$$\mathcal{H}_L^\theta = -\ln(\frac{AL^{m+1}(n_L, r_L)}{BL^m(n_L, r_L)}) + \ln(2r_L) - \ln(RR_{mean}) \tag{13}$$

$$\mathcal{H}_G^\theta = -\ln(\frac{AG^{m+1}(n_G, r_G)}{BG^m(n_G, r_G)}) + \ln(2r_G) - \ln(RR_{mean}). \tag{14}$$

Finally, $\mathcal{H}_N^\theta$ is calculated as

$$\begin{aligned}\mathcal{H}_N^\theta =& \mathcal{H}_L^\theta + \mathcal{H}_G^\theta = -\ln(\frac{AL^{m+1}(n_L, r_L)}{BL^m(n_L, r_L)}) - \ln(\frac{AG^{m+1}(n_G, r_G)}{BG^m(n_G, r_G)}) \\ &+ \ln(2r_L) + \ln(2r_G) - 2\times\ln(RR_{mean})\end{aligned} \tag{15}$$

where RR_{mean} is the mean of RR intervals in the current window, RR_{mean}, r_L and r_G are expressed in units of s.

2.3. Determination of *r* values

Inaccurate probability estimates can often be avoided since both $\mathcal{H}^c$ and $\mathcal{H}_N^\theta$ allow the flexibility to vary r. Therefore determining a proper r value for $\mathcal{H}^c$ and $\mathcal{H}_N^\theta$ is crucial. The recommended method from the original work for $\mathcal{H}^c$ (called the minimum numerator count method) (Lake 2006, Lake and Moorman 2011) is to vary the r value from an initial value of 0.03 s until a specified number of matches for $A^{m+1}(r)$ is attained. In Lake and Moorman (2011), a minimum numerator count of five was recommended to maximize AF detection accuracy for a 12-beat RR time series.

[6] We note that the correct volume would require evaluating the sort of integral of the Gaussian kernel function, i.e. the fuzzy function with $n_L = 2$ and $n_G = 2$ (Lake 2009, Darmon 2016). However, to keep a constant expression with $\mathcal{H}^c$, herein we use the expression of $(2r)^m$. Although it will not affect the detection performance, it needs to be re-defined if the actual value of the quadratic entropy rate needs to be determined.

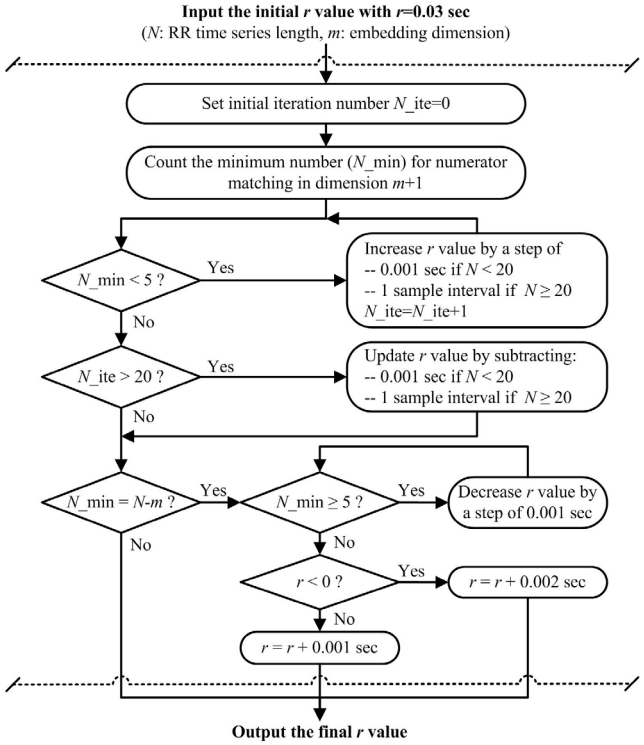

Figure 1. Flow chart of the detailed calculation process for determining the r values for $\mathcal{H}^c$ and $\mathcal{H}_N^\theta$ methods using the rule of 'minimum numerator count of five'.

However, the detailed description explaining how to vary the r value from the initial value of 0.03 s to achieve the minimum numerator count of five is unclear. Moreover, methods for dealing with the two potential risks due to infinite outputs were also not provided. Here, we describe a detailed calculation process for determining the proper r values for both $\mathcal{H}^c$ and $\mathcal{H}_N^\theta$ methods in figure 1.

In addition, for $\mathcal{H}_N^\theta$, we also use this 'minimum numerator count of five' method to determine a proper r value. The operation is the same with $\mathcal{H}^c$ by using the maximum absolute difference between two vectors. Once the appropriate r value is found, the distance matrix is constructed with the fuzzy function.

3. Experimental data and evaluation methods

3.1. Experimental data

Three standard datasets were obtained from PhysioNet (Goldberger *et al* 2000). We followed the method described previously in Colloca (2013), Colloca *et al* (2013) and Oster and Clifford (2015). First, the MIT-BIH AF database was used as the training dataset since it contains a sufficiently large number of AF and other rhythm episodes and thus can be used to determine stable and useful parameter values (Colloca 2013). This is also the most commonly used dataset and provides the widest comparability with other publications. Subsequently, the MIT-BIH normal sinus rhythm (MIT-BIH NSR) database and MIT-BIH arrhythmia database were used to test the performance of entropy measures. By using separate databases for training and testing, we significantly reduced the probability of over-fitting to recording methods (such as lead position and electrode preparation), patient selection, and recording parameters (such as background noise, bandwidth, analog filtering, sampling rate, electrode choice, etc).

(1) MIT-BIH AF database. The MIT-BIH AF database includes 25 long-term ECG recordings with rhythm and beat annotation files. Individual ECG recordings are 10 h in duration and were sampled at 250 Hz, resulting in a minimum temporal resolution of 0.004 s for the RR time series. Rhythm annotations were performed manually for four types: AF, AFL (atrial flutter), J (AV junctional rhythm) and N (used to indicate all other rhythms). Beat annotations were prepared using an automated detector with two recordings (no. 05091 and no. 07859) corrected manually.

(2) MIT-BIH normal sinus rhythm database. The MIT-BIH NSR database includes 54 long-term ECG recordings (roughly 24 h for each). In these recordings, no significant arrhythmias were diagnosed, except the presence of ectopic beats. Therefore, all rhythms in this database were labeled as non-AF rhythms. The sampling rate was 128 Hz, giving a minimum temporal resolution of approximately

Table 1. MIT-BIH database profile separated by the different rhythm types. For each rhythm type, the numbers and the corresponding percentages (%) are given. #: number of AF: atrial fibrillation, AFL: atrial flutter, J: AV junctional rhythm, N: rhythms except AF, AFL and J rhythms, NSR: normal sinus rhythm, AS100: arrhythmia series 100 database, AS200: arrhythmia series 200 database, unknown: beat classification was not available.

			Non-AF rhythm			
Database	Variable	AF rhythm	N	AFL	J	Total
AF	# rhythm episodes	299 (48.0%)	292 (46.9%)	14 (2.2%)	18 (2.9%)	324 (52.0%)
	Total time length (h)	93.5 (37.5%)	149.1 (59.8%)	1.4 (0.6%)	5.2 (2.1%)	155.7 (62.5%)
	# RR intervals	521 415 (42.6%)	663 202 (54.2%)	11 710 (1.0%)	26 818 (2.2%)	701.730 (57.4%)
	# RR intervals (⩽2 s)	521 359 (42.6%)	662 971 (54.2%)	11 710 (1.0%)	26 813 (2.2%)	701 494 (57.4%)
	# RR segments (12-beat)	43 307 (42.6%)	55 115 (54.2%)	969 (1.0%)	2227 (2.2%)	58 311 (57.4%)
	# RR segments (30-beat)	17 247 (42.6%)	21 968 (54.3%)	383 (0.9%)	886 (2.2%)	23 237 (57.4%)
	# RR segments (60-beat)	8552 (42.6%)	10 904 (54.3%)	190 (0.9%)	441 (2.2%)	11 535 (57.4%)
NSR	# rhythm episodes	0 (0%)	Unknown	Unknown	Unknown	54 (100%)
	Total time length (h)	0 (0%)	Unknown	Unknown	Unknown	1247 (100%)
	# RR intervals	0 (0%)	Unknown	Unknown	Unknown	5790 504 (100%)
	# RR intervals (⩽2 s)	0 (0%)	Unknown	Unknown	Unknown	5780 148 (100%)
	# RR segments (12-beat)	0 (0%)	Unknown	Unknown	Unknown	481 657 (100%)
	# RR segments (30-beat)	0 (0%)	Unknown	Unknown	Unknown	192 643 (100%)
	# RR segments (60-beat)	0 (0%)	Unknown	Unknown	Unknown	96 308 (100%)
AS100	# rhythm episodes	0 (0%)	Unknown	Unknown	Unknown	227 (100%)
	Total time length (h)	0 (0%)	Unknown	Unknown	Unknown	11.5 (100%)
	# RR intervals	0 (0%)	Unknown	Unknown	Unknown	48 200 (100%)
	# RR intervals (⩽2 s)	0 (0%)	Unknown	Unknown	Unknown	48 200 (100%)
	# RR segments (12-beat)	0 (0%)	Unknown	Unknown	Unknown	3904 (100%)
	# RR segments (30-beat)	0 (0%)	Unknown	Unknown	Unknown	1515 (100%)
	# RR segments (60-beat)	0 (0%)	Unknown	Unknown	Unknown	736 (100%)
AS200	# rhythm episodes	107 (7.1%)	Unknown	Unknown	Unknown	1394 (92.9%)
	Total time length (h)	2.2 (17.6%)	Unknown	Unknown	Unknown	10.3 (82.3%)
	# RR intervals	11 670 (18.1%)	Unknown	Unknown	Unknown	52 728 (81.9%)
	# RR intervals (⩽2 s)	11 670 (18.2%)	Unknown	Unknown	Unknown	52 605 (81.8%)
	# RR segments (12-beat)	922 (19.2%)	Unknown	Unknown	Unknown	3881 (80.8%)
	# RR segments (30-beat)	344 (19.7%)	Unknown	Unknown	Unknown	1403 (80.3%)
	# RR segments (60-beat)	157 (19.8%)	Unknown	Unknown	Unknown	636 (80.2%)

0.008 s for the RR time series. Beat annotations were obtained by automated analysis with manual review and correction.

(3) MIT-BIH arrhythmia database. The MIT-BIH arrhythmia database includes 48 short-term (30 min) ECG recordings chosen from a set of 4000 24 h ambulatory ECGs. This database can be divided into two sub-databases: the series 100 (MIT-BIH AS100) includes 23 subjects with non-AF rhythms; the series 200 (MIT-BIH AS200) includes eight AF subjects, containing both AF rhythm and a variety of non-AF rhythms such as atrial and ventricular bigeminy, ventricular trigeminy, AFL, ventricular flutter, and ventricular and supraventricular tachycardia. The sampling rate was 360 Hz, giving a minimum temporal resolution of about 0.003 s for the RR time series. Beats were annotated independently by at least two cardiologists.

3.2. Evaluation methods

For the MIT-BIH AF database, first, the RR time series corresponding to the four rhythm types (AF, AFL, J and N) were extracted to permit comparison between AF and N rhythm types while excluding the influences of AFL and J rhythm types. Then, the RR time series corresponding to the latter three rhythm types (AFL, J and N) were merged as non-AF rhythms to enable comparison against AF and non-AF rhythm types. For the MIT-BIH NSR and AS100 databases, all RR time series were regarded as non-AF rhythm types. For the MIT-BIH AS200 database, the RR time series were regarded as AF type during the AF rhythm episodes, and non-AF type during other rhythm episodes.

Data pre-processing was performed on the classified RR episodes. RR intervals greater than 2 s were removed to eliminate the influence of the missed QRS detection due to noise or ECG electrode drop out. Three types of

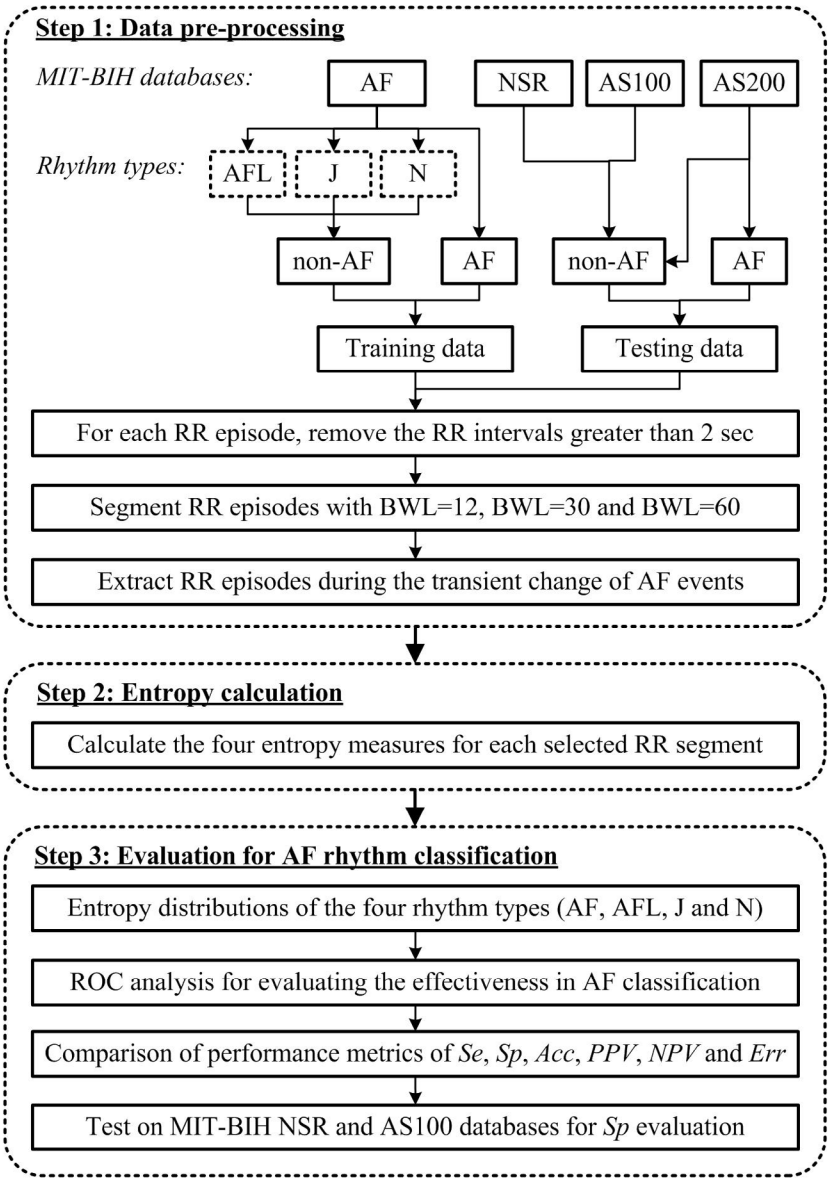

Figure 2. Block diagram of the evaluation procedure used in this study.

beat window length (BWL)—12, 30, and 60 beats—were used to segment RR episodes without overlap. Table 1 shows the detailed database profile.

$\mathcal{H}^{\sigma}$, $\mathcal{H}^{\theta}$, $\mathcal{H}^{c}$ and $\mathcal{H}^{\theta}_{N}$ values were calculated for each of the selected RR segments. These entropy measures were compared between the AF and N rhythm types for the MIT-BIH AF database, and between the AF and non-AF rhythm types for the MIT-BIH AF and AS200 databases. Entropy measures were also evaluated for how well non-AF rhythms were rejected using both the MIT-BIH NSR and AS100 databases.

Entropy values on one side of threshold c were labeled as AF rhythms and values on the other side of c were labeled as non-AF rhythms. Classifier accuracy was assessed via the following performance metrics:

Sensitivity: $Se = TP/(TP + FN)$

Specificity: $Sp = TN/(TN + FP)$

Accuracy: $Acc = (TP + TN)/(TP + FP + FN + TN)$

Positive predictive value: $PPV = TP/(TP + FP)$

Negative predictive value: $NPV = TN/(TN + FN)$

Total error: $Err = (FP + FN)/(TP + FP + FN + TN)$

where TP, TN, FP and FN are the numbers of true positives, true negatives, false positives and false negatives respectively.

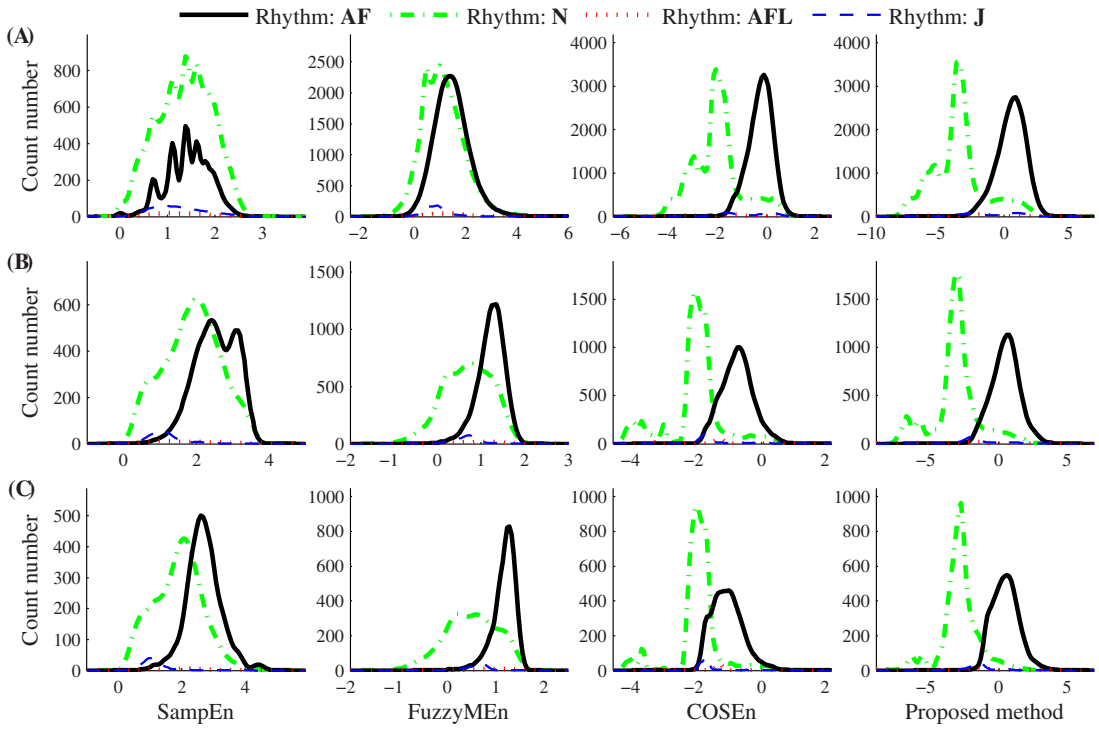

Figure 3. Distributions of the four entropy measures for the four rhythm types (AF, N, AFL and J) in the MIT-BIH AF database with the different time windows of RR segments: (A) 12-beat, (B) 30-beat and (C) 60-beat. Note that the proportions of AFL and J rhythm types are much smaller. The departure of the AF rhythm type from the N rhythm type is more obvious in the proposed method ($\mathcal{H}_N^\theta$) than the other three entropy measures: sample entropy (SampEn, $\mathcal{H}^\sigma$), fuzzy measure entropy (FuzzyMEn, $\mathcal{H}^\theta$), coefficient of sample entropy (COSEn, $\mathcal{H}^c$).

The receiver operating characteristic (*ROC*) curve was used to evaluate the effectiveness of each entropy measure in the AF classification. The *ROC* curve is a plot of (Se) versus ($1 - Sp$) for many possible values of c, which varied from the minimum to the maximum of the entropy outputs, with a step of 1% of the range. *AUC* was used to evaluate the performance of different entropy measures. The Youden index (J), another metric for assessing *ROC* curves, was also calculated as

$$J = \max_c\{Se(c) + Sp(c) - 1\}. \tag{16}$$

At the optimal cut-point c^*, J is maximized and the classifier equally weighs sensitivity and specificity. In this study, the opitic* values were determined (trained) from the MIT-BIH AF database, and then were used (tested) on other MIT-BIH databases. The aforementioned performance metrics of Se, Sp, Acc, PPV, NPV and Err were given at the point of c^*. Only Sp was evaluated for both the MIT-BIH NSR and AS100 databases, which lacked the AF rhythm. The evaluation procedure described above is illustrated in figure 2, which consisted of three major steps. Step 1: data pre-processing; step 2: entropy calculation; and step 3: evaluation for AF rhythm classification.

4. Results

Figure 3 shows the histograms of the four entropy measures for the four rhythm types (AF, N, AFL and J) in the MIT-BIH AF database when using 12-beat, 30-beat and 60-beat RR segments respectively. The departures of AF rhythm from the N rhythm are more obvious in $\mathcal{H}^c$ and $\mathcal{H}_N^\theta$ than those in $\mathcal{H}^\sigma$ and $\mathcal{H}^\theta$, and compared with $\mathcal{H}^c$, $\mathcal{H}_N^\theta$ exhibits a superior performance. In addition, it is worth noting that the AFL and J rhythm types only occupy small proportions.

Figure 4 illustrates the *ROC* curves with *AUC* values obtained using the four entropy measures and the MIT-BIH AF database for classifier testing. The upper plots (A1–C1) show the results for classifying AF and N rhythm types and the lower plots (A2–C2) for classifying AF and non-AF rhythm types using 12-beat, 30-beat and 60-beat windows (from left to right).

To classify AF and N rhythms and for each *BWL* type, $\mathcal{H}^\sigma$, $\mathcal{H}^\theta$, $\mathcal{H}^c$ and $\mathcal{H}_N^\theta$ result in the lowest to highest *AUCs*, in order. For the 12-beat RR segments, the *AUC* values are 54.09%, 63.89%, 92.16% and 93.59% respectively for the four entropy measures. For the 30-beat RR segments, they are 73.83%, 79.05%, 92.49% and 95.70% respectively, and for the 60-beat RR segments, they are 80.60%, 86.82%, 91.23% and 97.04% respectively. With an increase in *BWL*, the *AUC* values gradually increase for $\mathcal{H}^\sigma$, $\mathcal{H}^\theta$ and $\mathcal{H}_N^\theta$ but not for the $\mathcal{H}^c$. We also note that the *AUC* of $\mathcal{H}^c$ for the 60-beat RR segments is even lower than that for the 30-beat RR segments.

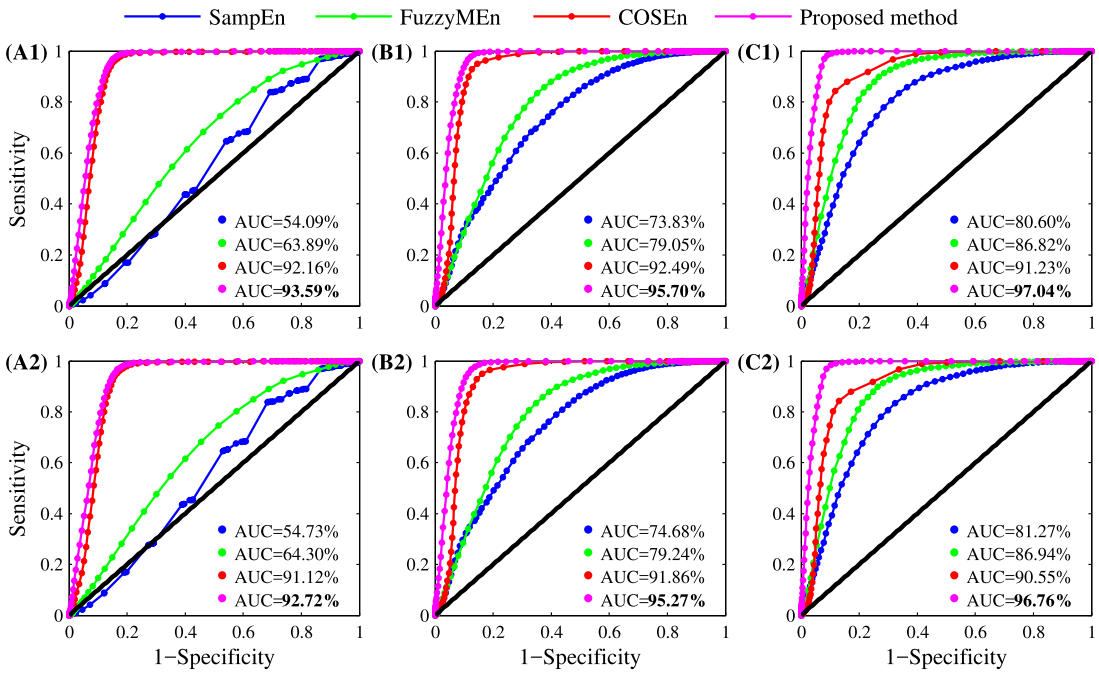

Figure 4. ROC curve plots with *AUC* values for the four entropy measures in the MIT-BIH AF database for classifying AF and N rhythm types (A1–C1) and for classifying AF and non-AF rhythm types (A2–C2). Three time window types of the RR segment were used: (A1) and (A2) for 12-beat, (B1) and (B2) for 30-beat, and (C1) and (C2) for 60-beat RR segments. The four methods tested were sample entropy (SampEn, $\mathcal{H}^{\sigma}$), fuzzy measure entropy (FuzzyMEn, $\mathcal{H}^{\theta}$), coefficient of sample entropy (COSEn, $\mathcal{H}^{c}$) and the proposed method ($\mathcal{H}_N^{\theta}$).

The *AUC* results for classifying AF and non-AF rhythms show a similar trend as seen in the classification of AF and N rhythms. The following *AUC* values are for classifiers using $\mathcal{H}^{\sigma}$, $\mathcal{H}^{\theta}$, $\mathcal{H}^{c}$ and $\mathcal{H}_N^{\theta}$, respectively. For the 12-beat RR segments, the *AUC* values are 54.73%, 64.30%, 91.12% and 92.72% respectively, for the 30-beat RR segments, the *AUC* values are 74.68%, 79.24%, 91.86% and 95.27% respectively, and for the 60-beat RR segments, the *AUC* values are 81.27%, 86.94%, 90.55% and 96.76% respectively. Because the AFL and J rhythm types are included in the non-AF rhythm, the AUC values for both $\mathcal{H}^{c}$ and $\mathcal{H}_N^{\theta}$ slightly decreased. The *AUC* values decreased by 1.04%, 0.53% and 0.68% respectively for the three *BWL* types when using $\mathcal{H}^{c}$. They decreased less for $\mathcal{H}_N^{\theta}$, with values of 0.87%, 0.23% and 0.28% respectively.

Table 2 summarizes the classifier performance metrics. Compared with the other entropy measures, $\mathcal{H}_N^{\theta}$ generally resulted in the largest values of *J*, *Se*, *Sp*, *Acc*, *PPV* and *NPV*, and the smallest values of *Err*, for each of the three *BWL* types. Specifically, $\mathcal{H}_N^{\theta}$ resulted in the highest *Se* values of 96.58%, 96.71% and 98.46% respectively for the three time window lengths for the MIT-BIH AF database. These values were even higher at 97.94%, 97.97% and 98.73% for the MIT-BIH AS200 database. In comparison, $\mathcal{H}^{c}$ resulted in *Se* values of 96.24%, 94.93% and 84.33% for the MIT-BIH AF database and 97.72%, 97.38% and 85.99% for the MIT-BIH AS200 database. Both $\mathcal{H}^{c}$ and $\mathcal{H}_N^{\theta}$ output larger *Se* values than $\mathcal{H}^{\sigma}$ and $\mathcal{H}^{\theta}$. We expect the increase in the time window length to increase the *Se* values of all entropy measures since more physiological signal information is used. This trend is observed for $\mathcal{H}^{\theta}$ and $\mathcal{H}_N^{\theta}$ but not for $\mathcal{H}^{\sigma}$ and $\mathcal{H}^{c}$.

$\mathcal{H}_N^{\theta}$ also resulted in the highest values for *Sp*: 83.31%, 87.52% and 89.85% respectively for the three time window types for the MIT-BIH AF database. *Sp* was higher for $\mathcal{H}_N^{\theta}$ when using the MIT-BIH NSR database, with values of 92.45%, 95.16% and 96.75%. Considering the large amount of the MIT-BIH NSR data (about 1247 h, see table 1), the high *Sp* for $\mathcal{H}_N^{\theta}$ demonstrates its ability to accurately exclude non-AF beats from AF. *Sp* of $\mathcal{H}_N^{\theta}$ was lower at 78.79%, 84.29% and 87.50% respectively for the MIT-BIH AS100 database, and even lower at 58.49%, 68.35% and 70.60% for the MIT-BIH AS200 database. Nevertheless, these values were superior to those obtained from the other entropy measures. The lower *Sp* values for $\mathcal{H}_N^{\theta}$ were mainly due to the presence of many other types of arrhythmia beats included in the non-AF rhythm type in addition to NSR beats. This complicated the task of classifying AF and non-AF rhythms. In comparison with $\mathcal{H}_N^{\theta}$, $\mathcal{H}^{c}$ produced similar *Sp* values for both the MIT-BIH AS100 and AS200 databases, but produced significantly lower *Sp* values for the MIT-BIH NSR database, indicating that $\mathcal{H}^{c}$ is not an optimal measure that should be used for excluding NSR beats from AF. Interestingly, for the MIT-BIH NSR database, *Sp* values using $\mathcal{H}^{c}$ were 87.97%, 81.69%, and 83.18% for $BWL = 12$, 30, and 60 beats respectively. Unlike the trend observed when using other entropy measures, *Sp* did not increase concurrently with window length for $\mathcal{H}^{c}$.

Table 2. Results of the performance metrics for the four entropy measures in all MIT-BIH databases. AF: atrial fibrillation, NSR: normal sinus rhythm, AS100: arrhythmia series 100 database, AS200: arrhythmia series 200 database, *BWL*: beat window length, c^*: the optimal cut-point, *J*: Youden index, *Se*: sensitivity, *Sp*: specificity, *Acc*: accuracy, *PPV*: positive predictive value, *NPV*: negative predictive value, *Err*: total error. The best performance figures are in bold.

Database	Metric / *BWL*	$\mathcal{H}^{\sigma}$			$\mathcal{H}^{\theta}$			$\mathcal{H}^{c}$			$\mathcal{H}^{\theta}_{N}$		
		12	30	60	12	30	60	12	30	60	12	30	60
AF	c^*	1.08	2.05	2.20	1.01	0.91	0.88	−1.32	−1.58	−1.60	−1.76	−1.42	-1.19
	J (%)	15.38	37.16	50.97	23.08	48.73	63.39	78.74	78.78	71.30	**79.89**	**84.23**	**88.30**
	Se (%)	83.80	78.88	83.41	74.54	85.82	89.80	96.24	94.93	84.33	**96.58**	**96.71**	**98.46**
	Sp (%)	31.58	58.28	67.56	48.63	62.91	73.58	82.49	83.85	86.97	**83.31**	**87.52**	**89.85**
	Acc (%)	44.25	66.37	74.31	59.62	72.67	80.49	88.35	88.57	85.85	**88.96**	**91.43**	**93.51**
	PPV (%)	28.19	55.03	65.59	51.82	63.20	71.60	80.33	81.35	82.75	**81.12**	**85.18**	**87.79**
	NPV (%)	85.89	81.00	84.60	71.97	85.67	90.63	96.73	95.70	88.22	**97.05**	**97.29**	**98.74**
	Err (%)	55.75	33.63	25.69	40.38	27.33	19.51	11.65	11.43	14.15	**11.03**	**8.57**	**6.49**
NSR	*Sp* (%)	17.86	61.44	71.77	51.26	72.80	84.27	87.97	81.69	83.18	**92.45**	**95.16**	**96.75**
AS100	*Sp* (%)	28.70	42.11	42.30	47.28	61.45	69.97	**79.25**	82.57	**87.64**	78.79	**84.29**	87.50
AS200	*J* (%)	16.76	31.30	46.79	27.34	54.02	64.54	54.13	61.10	56.11	**56.43**	**66.32**	**69.32**
	Se (%)	84.34	78.60	87.26	73.64	88.66	93.63	97.72	97.38	85.99	**97.94**	**97.97**	**98.73**
	Sp (%)	32.41	52.70	59.52	53.70	65.36	70.91	56.40	63.72	70.13	**58.49**	**68.35**	**70.60**
	Acc (%)	38.52	57.60	65.06	57.53	69.95	75.41	64.33	70.35	73.27	**66.06**	**74.18**	**76.17**
	PPV (%)	12.25	27.93	34.95	27.42	38.56	44.28	34.75	39.69	41.54	**35.92**	**43.15**	**45.32**
	NPV (%)	93.95	91.34	94.94	89.56	95.92	97.83	99.05	99.00	95.30	**99.17**	**99.28**	**99.56**
	Err (%)	61.48	42.40	34.94	42.47	30.05	24.59	35.67	29.65	26.73	**33.94**	**25.82**	**23.83**

Finally, $\mathcal{H}^{\sigma}$ exhibited the lowest *Se* and *Sp*. $\mathcal{H}^{\theta}$ was superior to $\mathcal{H}^{\sigma}$, although both were inferior to $\mathcal{H}^{c}$ and $\mathcal{H}^{\theta}_{N}$. Based on the observed *Se* and *Sp*, $\mathcal{H}^{\theta}_{N}$ provided the highest accuracy and largest Youden indices for both the MIT-BIH AF and AS200 databases. Moreover, increasing the *BWL* from 12-beat to 30-beat to 60-beat significantly increased the performance metrics, with a longer window resulting in higher classification accuracies when using $\mathcal{H}^{\theta}_{N}$. We note that accuracy only exceeds 90% when using $\mathcal{H}^{\theta}_{N}$ and a window size of 30-beat or 60-beat RR segments.

5. Discussion

5.1. Further explanation for entropy measures

Fuzzy function-based entropy measures ($\mathcal{H}^{\theta}$ and $\mathcal{H}^{\theta}_{N}$) were more consistent than Heaviside function-based entropy measures ($\mathcal{H}^{\sigma}$ and $\mathcal{H}^{c}$) (figure 3 and table 2). Specifically, classifier performance increased with window length for $\mathcal{H}^{\theta}$ and $\mathcal{H}^{\theta}_{N}$, but not for $\mathcal{H}^{\sigma}$ and $\mathcal{H}^{c}$ when *BWL* was increased from 12-beat to 60-beat. This may be due to the rigid membership degree determination (0–1 determination) in the Heaviside function—a limitation described earlier (Chen *et al* 2009, Liu *et al* 2013, Zhao *et al* 2015).

Entropy measures that measure similarity using the 0–1 determination run the risk of counting very few matching vectors, which can result in output values of infinity or undefined. In this study, $\mathcal{H}^{\sigma}$ produced many values of infinity or undefined when processing 12-beat RR segments. This issue occasionally arose even when processing 30-beat and 60-beat RR segments. For $\mathcal{H}^{c}$, we developed a process to determine suitable *r* values (see figure 1) and thus to successfully avoid the boring invalid output. The fuzzy function-based entropy measures ($\mathcal{H}^{\theta}$ and $\mathcal{H}^{\theta}_{N}$) consider actual distances between two vectors rather than the numbers of matched vectors, and avoid this limitation. We have summarized the numbers of valid and invalid values of $\mathcal{H}^{\sigma}$, as well as the percentages of the valid values of $\mathcal{H}^{\sigma}$ for each of the four MIT-BIH databases, as table D1 in the appendix.

5.2. Optimized threshold parameter *r* for SampEn

Although *m* and *r* are critical in determining the outcome of entropy estimation, no standard guidelines exist for optimizing their values (Lake *et al* 2002, Ramdani *et al* 2009). In Zhao *et al* (2015), $r = 0.1$ times the standard deviation of the RR time series was recommended. This study used this recommendation. Herein, we compared the SampEn results for different threshold *r* choices, which varied under two schemes. The first scheme is that *r* is equal to different ratios (i.e. 0.1–0.9 with a step of 0.1) of the standard deviation of the current RR segments (with 12 beats). The second scheme is that *r* is equal to different ratios (i.e. 0.1–0.9 with a step of 0.1) of the mean standard deviation of all RR segments (also with 12 beats), to stimulate the constant *r* value of the 30 ms suggested in Lake and Moorman (2011).

Table 3. Effect of varied threshold r on the SampEn results in the MIT-BIH AF database. The RR time series uses a window of 12 RR intervals. *SD*: standard deviation, c^*: the optimal cut-point, *J*: Youden index, *Se*: sensitivity, *Sp*: specificity, *Acc*: accuracy, *PPV*: positive predictive value, *NPV*: negative predictive value, *Err*: total error, *RIS*: ratio of RR segments with invalid SampEn calculation.

	$\mathcal{H}^{\sigma}$ with varied threshold r ($\times SD$ of the current RR segment)								
Metric	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
c^*	1.08	1.11	1.08	1.12	1.03	0.88	0.80	0.76	0.70
J (%)	15.38	22.50	20.85	17.90	16.69	17.27	17.19	16.67	14.78
Se (%)	83.80	80.20	83.77	72.68	72.32	75.10	72.98	67.80	56.19
Sp (%)	31.58	42.30	37.07	45.21	44.37	42.17	44.21	48.88	58.59
Acc (%)	44.25	58.01	58.28	57.65	56.87	56.74	56.89	57.20	57.54
PPV (%)	28.19	49.61	52.54	52.32	51.25	50.77	50.76	51.04	51.60
NPV (%)	85.89	75.09	73.32	66.68	66.46	68.08	67.50	65.87	62.99
Err (%)	55.75	41.99	41.72	42.35	43.13	43.26	43.11	42.80	42.46
	$\mathcal{H}^{\sigma}$ with fixed threshold r ($\times$mean *SD* of all RR segments)								
Metric	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
c^*	1.37	1.37	1.28	1.23	1.12	1.03	0.95	0.83	0.80
J (%)	41.72	60.42	63.58	64.17	64.40	66.03	65.61	67.87	65.00
Se (%)	62.97	67.94	67.85	66.96	66.86	68.64	68.18	71.10	68.01
Sp (%)	78.75	92.48	95.72	97.20	97.53	97.40	97.43	96.77	96.99
Acc (%)	76.50	86.30	86.68	86.06	85.40	85.53	85.04	85.73	84.41
PPV (%)	33.04	75.24	88.41	93.33	94.67	94.88	95.11	94.31	94.53
NPV (%)	92.74	89.55	86.11	83.44	81.80	81.54	80.65	81.62	79.82
Err (%)	23.50	13.69	13.12	13.94	14.60	14.47	14.96	13.56	15.59
RIS(%)	38.81	26.02	17.29	11.31	7.31	4.63	2.85	1.79	1.07

Table 3 shows the effect of varied threshold r on the SampEn results for 12-beat RR segments in the MIT-BIH AF database. When r increases from 0.1–0.9 times the standard deviation of the current RR segments, the classification accuracy firstly quickly increases to 58.01% at 0.2 times the standard deviation and then keeps at this constant level with a slight fluctuation. The optimal cut-point r decreases stably. When considering r as a constant ratio of the mean standard deviation of all RR segments, the classification accuracy reports high levels, which has an *Acc* of 76.50% for $r = 0.1$ times the mean standard deviation and increases to a peak of 86.68% for $r = 0.3$ times the mean standard deviation. The optimal cut-point r also decreases stably, from 1.37 to 0.80. However, although relatively high accuracy is achieved, the ratio of RR segments with invalid SampEn calculation is large, with reports as large as 38.81% for $r = 0.1$ times the mean standard deviation. Again, with an increase of r, the possibility of 'no matching' for vector similarity decreases for the 12-beat RR segments. Thus, the ratio of RR segments with invalid SampEn calculation decreases to 1.07% for $r = 0.9$ times the mean standard deviation.

5.3. Effect of 'minimum numerator count' on AUC Values

Figure 5 shows the effect of the 'minimum numerator count' on the AUC values for $\mathcal{H}^{c}$ and $\mathcal{H}_N^{\theta}$ measures for 30-beat and 60-beat RR segments respectively. The comparisons were performed between at least five vectors for vector matching and at least $Ratio \times N$ vectors for vector matching. We can determine which choice is better by comparing the corresponding AUC values. For the 30-beat RR segment, counting at least five vectors for vector matching is the best choice for $\mathcal{H}^{c}$. However, counting at least 50% N vectors for vector matching enhances the AUC value of 0.44% for $\mathcal{H}_N^{\theta}$. For the longer 60-beat RR segments, increasing the counting values could enhance the AUC value for both $\mathcal{H}^{c}$ and $\mathcal{H}_N^{\theta}$, and the increase in $\mathcal{H}^{c}$ is even larger than that in $\mathcal{H}_N^{\theta}$.

5.4. AF detection: from clinical to practical application

AF, the most common sustained arrhythmia, and particularly frequent in patients with heart disease, is often asymptomatic and carries a substantially increased stroke risk (Healey *et al* 2012). Simple and accurate methods to detect AF could increase the rate of early diagnosis, as well as reduce costs related to healthcare resource utilization. An AF screening device should be easily accessible, cost-effective and easy for clinicians to use. In this study, we describe a novel entropy measure $\mathcal{H}_N^{\theta}$ that achieves superior AF detection performance compared to previously described entropies. We validated $\mathcal{H}_N^{\theta}$ with different databases and time window lengths. AF detection based on $\mathcal{H}_N^{\theta}$ is both conceptually simple and computationally efficient, allowing it to be potentially used in real-time. This approach may be applied to patients with implantable devices and for screening patients at high risk for development of AF. In addition, the method for constructing $\mathcal{H}_N^{\theta}$ may also have other applications in

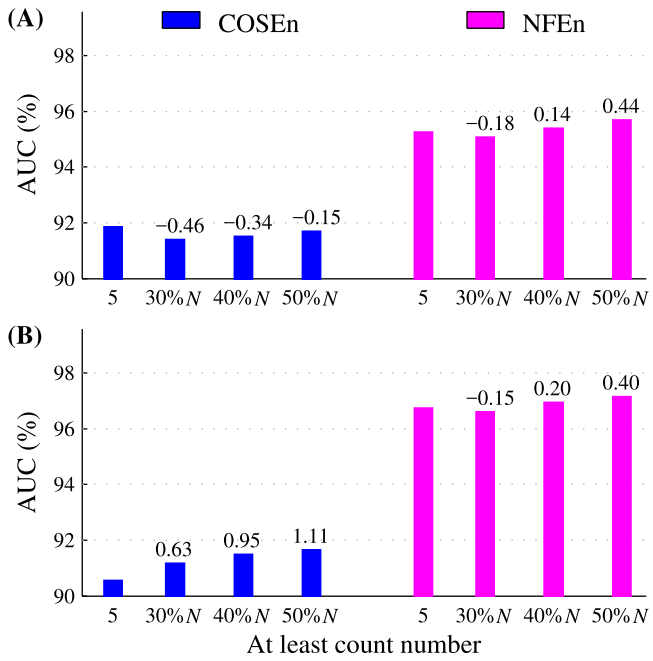

Figure 5. Effect of the minimum numerator count on the AUC values for $\mathcal{H}^c$ and $\mathcal{H}_N^\theta$ measures: (A) for 30-beat and (B) for 60-beat RR segments. N denotes the beat number of the RR segments.

biomedical signal processing, such as in developing specific entropy measures for detecting patients with heart failure, diabetes, etc.

5.5. Limitations

Atrial flutter has a similar clinical profile to AF but exhibits different RR interval dynamics. Some studies regard atrial flutter as an AF rhythm (Lake and Moorman 2011, DeMazumder *et al* 2013) whereas others regard it as a non-AF rhythm (Colloca 2013, Oster and Clifford 2015). In this study, we defined atrial flutter as a non-AF rhythm and found that there was little difference between classifying AF and non-AF (including atrial flutter) rhythms and classifying AF and normal sinus (not including atrial flutter) rhythms. This may be due to the small proportion of atrial flutter beats in the data. Second, sinus rhythm with frequent ectopic beats can resemble AF RR interval dynamics. We did not consider this in our study due to a lack of gold-standard labels for ectopic beats. Third, the R-peak annotations in this study were labeled by experts, so performance will likely drop when using real QRS detectors. Fourth, since the length of a 12-beat RR segment is short, when considering the large dimension m (such as $m = 2$, or 3), the ratio of RR segments with invalid SampEn calculation is very high, resulting in most of the 12-beat RR segments in the AF group output producing invalid SampEn values. Thus, only $m = 1$ is used in the current study. Finally, employing a new larger AF database can definitely be beneficial for systematically evaluating the proposed entropy method. In addition, we note that machine learning methods, which combine multiple metrics (see Colloca *et al* (2013) and Oster and Clifford (2015)), may improve AF classification performance compared to using $\mathcal{H}_N^\theta$ alone. We leave these issues to future studies.

6. Conclusion

Here, we propose normalized fuzzy entropy ($\mathcal{H}_N^\theta$), a novel entropy measure suitable for AF detection based on a short-term RR time series. $\mathcal{H}_N^\theta$ uses a fuzzy function to determine vector similarity, replaces a probability estimate with a density estimate for entropy approximation, utilizes a flexible distance threshold parameter, and adjusts for heart rate by subtracting the natural log of mean RR intervals. The performance of $\mathcal{H}_N^\theta$ was tested on the MIT-BIH AF, NSR and arrhythmia databases (containing more than 1500 h of RR interval time series) and compared against three existing entropy-based metrics: $\mathcal{H}^\sigma$, $\mathcal{H}^\theta$ and $\mathcal{H}^c$. Our results demonstrate that $\mathcal{H}_N^\theta$ is an accurate measure for detecting AF.

Acknowledgment

CL is grateful for the support from the Rhodes Trust, Emory University and the International Postdoctoral Exchange Programme of the National Postdoctoral Management Committee of China. This work was partly supported by the National Natural Science Foundation of China (61571113 and 61671275), the Key Research

and Development Programs of Jiangsu Province (BE2017735), and the Fundamental Research Funds for the Central Universities in Southeast University (2242018k1G010).

Appendix A. Sample entropy

The calculation process of SampEn, or $\mathcal{H}^{\sigma}$, can be summarized as follows (Richman and Moorman 2000).

For an RR time series $x(i)$ $(1 \leqslant i \leqslant N)$, given the parameters of embedding dimension m and distance threshold r, first form the vector sequences $X_i^m (1 \leqslant i \leqslant N - m)$:

$$X_i^m = \{x(i), x(i+1), \cdots, x(i+m-1)\}. \tag{A.1}$$

The vector X_i^m represents m consecutive $x(i)$ values. Then the distance between X_i^m and X_j^m based on the maximum absolute difference is defined as

$$d_{i,j}^m = d[X_i^m, X_j^m] = \max_{0 \leqslant k \leqslant m-1} |x(i+k) - x(j+k)|. \tag{A.2}$$

For each X_i^m, denote $B_i^m(r)$ as $(N-m)^{-1}$ times the number of X_j^m that meets $d_{i,j}^m \leqslant r$. Similarly, set $A_i^{m+1}(r)$ is $(N-m)^{-1}$ times the number of X_j^{m+1} that meets $d_{i,j}^{m+1} \leqslant r$ for all $1 \leqslant j \leqslant N-m$. The total number of matching vectors at dimension $m+1$ and m could be calculated as

$$A^{m+1}(r) = \sum_{i=1}^{N-m} A_i^{m+1}(r) \tag{A.3}$$

$$B^m(r) = \sum_{i=1}^{N-m} A_i^m(r). \tag{A.4}$$

Then $\mathcal{H}^{\sigma}$ is defined by

$$\mathcal{H}^{\sigma} = -\ln\left(\frac{A^{m+1}(r)}{B^m(r)}\right). \tag{A.5}$$

$\mathcal{H}^{\sigma}$ is therefore the negative natural log of the conditional probability of vector matching between dimension m and $m+1$.

Appendix B. Fuzzy measurement entropy

The calculation process of FuzzyMEn, or $\mathcal{H}^{\theta}$, can be summarized as follows (Liu and Zhao 2011, Liu *et al* 2013).

For an RR time series $x(i)(1 \leqslant i \leqslant N)$, first form the local vector sequences L_i^m and global vector sequences G_i^m respectively $(1 \leqslant i \leqslant N-m)$:

$$L_i^m = \{x(i), x(i+1), \cdots, x(i+m-1)\} - \bar{x}(i) \tag{B.1}$$

$$G_i^m = \{x(i), x(i+1), \cdots, x(i+m-1)\} - \bar{x}. \tag{B.2}$$

The vector L_i^m represents m consecutive $x(i)$ but removing the local baseline $\bar{x}(i) = \frac{1}{m}\sum_{k=0}^{m-1} x(i+k)$. The vector G_i^m also represents m consecutive $x(i)$ but removing the global mean value $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x(i)$.

The distance between local vector sequences L_i^m and L_j^m and the distance between global vector sequences G_i^m and G_j^m are respectively defined as

$$dL_{i,j}^m = d[L_i^m, L_j^m] = \max_{0 \leqslant k \leqslant m-1} |(x(i+k) - \bar{x}(i)) - (x(j+k) - \bar{x}(j))| \tag{B.3}$$

$$dG_{i,j}^m = d[G_i^m, G_j^m] = \max_{0 \leqslant k \leqslant m-1} |(x(i+k) - \bar{x}) - (x(j+k) - \bar{x})|. \tag{B.4}$$

Given the following parameters: local similarity weight n_L, local tolerance threshold r_L, global similarity weight n_G and global tolerance threshold r_G, we calculate the similarity degree $DL_{i,j}^m(n_L, r_L)$ between local vectors L_i^m and L_j^m by the fuzzy function $\mu L(dL_{i,j}^m, n_L, r_L)$, and also calculate the similarity degree $DG_{i,j}^m(n_G, r_G)$ between global vectors G_i^m and G_j^m by the fuzzy function $\mu G(dG_{i,j}^m, n_G, r_G)$ as

$$DL_{i,j}^m(n_L, r_L) = \mu L(dL_{i,j}^m, n_L, r_L) = \exp\left(-\frac{(dL_{i,j}^m)^{n_L}}{r_L}\right) \tag{B.5}$$

$$DG_{i,j}^m(n_G, r_G) = \mu L(dG_{i,j}^m, n_G, r_G) = \exp\left(-\frac{(dG_{i,j}^m)^{n_G}}{r_G}\right). \tag{B.6}$$

We define the functions $BL^m(n_L, r_L)$ and $BG^m(n_G, r_G)$ as

$$BL^m(n_L, r_L) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m}\sum_{j=1}^{N-m} DL_{i,j}^m(n_L, r_L)\right) \tag{B.7}$$

$$BG^m(n_G, r_G) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m}\sum_{j=1}^{N-m} DG_{i,j}^m(n_G, r_G)\right). \tag{B.8}$$

Similarly, we define the functions $AL^{(m+1)}(n_L, r_L)$ for $m+1$ dimension local vectors L_i^{m+1} and L_j^{m+1}, and the function $AG^{(m+1)}(n_G, r_G)$ for dimension $m+1$ global vectors G_i^{m+1} and G_j^{m+1} as

$$AL^{m+1}(n_L, r_L) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m}\sum_{j=1}^{N-m} DL_{i,j}^{m+1}(n_L, r_L)\right) \tag{B.9}$$

$$AG^{m+1}(n_G, r_G) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m}\sum_{j=1}^{N-m} DG_{i,j}^{m+1}(n_G, r_G)\right). \tag{B.10}$$

Then the fuzzy local entropy measure ($\mathcal{H}_L^\theta$) and fuzzy global entropy measure ($\mathcal{H}_G^\theta$) are respectively defined by

$$\mathcal{H}_L^\theta = -\ln\left(\frac{AL^{m+1}(n_L, r_L)}{BL^m(n_L, r_L)}\right) \tag{B.11}$$

$$\mathcal{H}_G^\theta = -\ln\left(\frac{AG^{m+1}(n_G, r_G)}{BG^m(n_G, r_G)}\right). \tag{B.12}$$

Finally, the $\mathcal{H}^\theta$ is calculated as the sum of these two latter measures:

$$\mathcal{H}^\theta = \mathcal{H}_L^\theta(m, n_L, r_L, N) + \mathcal{H}_G^\theta(m, n_G, r_G, N). \tag{B.13}$$

The parameters in equation (B.13) use the recommended settings in Zhao *et al* (2015): $n_L = 2, n_G = 2$, and r_L is equal to r_G, i.e. $r_L = r_G = r$. Equation (B.13) then becomes

$$\mathcal{H}^\theta = \mathcal{H}_L^\theta(m, r, N) + \mathcal{H}_G^\theta(m, r, N). \tag{B.14}$$

Appendix C. Coefficient of sample entropy

COSEn, or $\mathcal{H}^c$, was defined by Lake and Moorman (2011) and Carrara *et al* (2015) as an entropy measure derived from $\mathcal{H}^\sigma$, designed specifically to detect AF in very short RR time series (Lake and Moorman 2011, Carrara *et al* 2015). For $\mathcal{H}^\sigma$, if $A^{m+1}(r)$ and $B^m(r)$ in equation (A5) are equal, which means that the time series is very regular or predictable, the entropy value is zero, whereas if $A^{m+1}(r)$ is smaller than $B^m(r)$, this leads to a higher value of entropy. An important advantage of $\mathcal{H}^\sigma$ compared with approximate entropy is that the self-matches are not counted (Richman and Moorman 2000). This significantly reduces bias but could induce the problem of lowering the counts of vector matching to the point that $A^{m+1}(r)$ and even $B^m(r)$ can tend to zero, leading to infinite or indeterminate outputs (since $\mathcal{H}^\sigma$ is defined by their ratio). This becomes an increasing concern for the short time series (Lake and Moorman 2011).

To avoid this issue, a measure called quadratic sample entropy ($\mathcal{H}^q$), based on densities rather than probability estimates, was previously introduced (Lake 2006). $\mathcal{H}^q$ normalized $\mathcal{H}^\sigma$ by the volume of each matching region, i.e. $(2r)^m$, and equation (A5) is rewritten as

$$\mathcal{H}_L^q = -\ln\left(\frac{A^{m+1}(r)/(2r)^{m+1}}{B^m(r)/(2r)^m}\right) = -\ln\left(\frac{A^{m+1}(r)}{B^m(r)}\right) + \ln(2r) = \mathcal{H}^\sigma + \ln(2r). \tag{C.1}$$

In addition, regression analyses showed that heart rate was an important independent predictor of AF (Lake and Moorman 2011). Hence, the $\mathcal{H}^c$ measure uses the concept of density estimates of $\mathcal{H}^q$ but subtracts the natural log of mean RR interval from $\mathcal{H}^q$ as

$$\mathcal{H}^c = \mathcal{H}^q - \ln(RR_{mean}) = \mathcal{H}^\sigma + \ln(2r) - \ln(RR_{mean}) \tag{C.2}$$

where RR_{mean} is the mean of RR intervals in the current window, and both r and RR_{mean} use the unit s.

Appendix D. Analysis of valid and invalid values for entropy measures

Table D1. Numbers of the valid and invalid values of $\mathcal{H}^{\sigma}$, as well as the percentages of the valid values for all MIT-BIH databases. AF: atrial fibrillation, NSR: normal sinus rhythm, AS100: arrhythmia series 100 database, AS200: arrhythmia series 200 database, *BWL*: beat window length.

		$\mathcal{H}^{\sigma}$			$\mathcal{H}^{\theta}$, $\mathcal{H}^{c}$, $\mathcal{H}^{\theta}_{N}$
Database	*BWL* (beats)	Valid number	Invalid number	% of valid	% of valid
AF	12	39 692	61 926	39.1%	100%
	30	35 416	5068	87.5%	100%
	60	20 021	66	99.7%	100%
NSR	12	278 114	203 543	57.7%	100%
	30	181 763	10 880	94.4%	100%
	60	96 174	134	99.9%	100%
AS100	12	1129	2775	28.9%	100%
	30	1192	323	78.7%	100%
	60	721	15	98.0%	100%
AS200	12	1685	3118	35.1%	100%
	30	1507	240	86.3%	100%
	60	787	6	99.2%	100%

As shown in table D1, the MIT-BIH NSR database gave the largest percentages of valid $\mathcal{H}^{\sigma}$ values for each of the three time window types: 57.7% for $BWL = 12$, 94.4% for $BWL = 30$ and 99.9% for $BWL = 60$. This is because the irregular RR rhythms are much less than the other three databases. For the 12-beat RR segments, $\mathcal{H}^{\sigma}$ gave very low percentages of the valid numbers, only 39.1% for the MIT-BIH AF database, 28.9% for the MIT-BIH AS100 database and 35.1% for the MIT-BIH AS200 database.

ORCID iDs

Chengyu Liu https://orcid.org/0000-0003-1965-3020

References

Abusaada K, Sharma S B, Jaladi R and Ezekowitz M 2004 Epidemiology and management of new-onset atrial fibrillation *Am. J. Manag. Care* **10** 50–7

Ahmed M U and Mandic D P 2011 Multivariate multiscale entropy: a tool for complexity analysis of multichannel data *Phys. Rev.* E **84** 061918

Alcaraz R, Abásolo D, Hornero R and Rieta J J 2010 Optimal parameters study for sample entropy-based atrial fibrillation organization analysis *Comput. Methods Prog. Biomed.* **99** 124–32

Alcaraz R and Rieta J J 2012 Application of wavelet entropy to predict atrial fibrillation progression from the surface ECG *Comput. Math. Methods Med.* **2012** 245213

Alcaraz R, Vaya C, Cervigon R, Sanchez C and Rieta J J 2006 Wavelet sample entropy: a new approach to predict termination of atrial fibrillation *Comput. Cardiol.* **33** 597–600

Babaeizadeh S, Gregg R E, Helfenbein E D, Lindauer J M and Zhou S H 2009 Improvements in atrial fibrillation detection for real-time monitoring *J. Electrocardiol.* **42** 522–6

Camm A J *et al* 2010 Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC) *Eur. Heart J.* **31** 2369–429

Carrara M, Carozzi L, Moss T J, De Pasquale M, Cerutti S, Ferrario M, Lake D E and Moorman J R 2015 Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy *Physiol. Meas.* **36** 1873–88

Cerutti S, Mainardi L and Sornmo L 2008 *Understanding Atrial Fibrillation: the Signal Processing Contribution* (Williston, VT: Morgan and Claypool Publishers)

Chen W T, Zhuang J, Yu W X and Wang Z Z 2009 Measuring complexity using fuzzyen, apen, and sampen *Med. Eng. Phys.* **31** 61–8

Clifford G D, Liu C Y, Moody B, Lehman L, Silva I, Li Q, Johnson A E and Mark R G 2017 AF classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017 *Comput. Cardiol.* **44** 065–469

Colloca R 2013 Implementation and testing of atrial fibrillation detectors for a mobile phone application *Master's Thesis* Politecnico di Milano

Colloca R, Johnson A E W, Mainardi L and Clifford G D 2013 A support vector machine approach for reliable detection of atrial fibrillation events *Comput. Cardiol.* **40** 1047–50

Costa M, Goldberger A L and Peng C K 2002 Multiscale entropy analysis of complex physiologic time series *Phys. Rev. Lett.* **89** 068102

Costa M, Goldberger A L and Peng C K 2005 Multiscale entropy analysis of biological signals *Phys. Rev.* E **71** 021906

Darmon D 2016 Specific differential entropy rate estimation for continuous-valued time series *Entropy* **18** 190

DeMazumder D, Lake D E, Cheng A, Moss T J, Guallar E, Weiss R G, Jones S R, Tomaselli G F and Moorman J R 2013 Dynamic analysis of cardiac rhythms for discriminating atrial fibrillation from lethal ventricular arrhythmias *Circ. Arrhythm. Electrophysiol.* **6** 555–61
Du X, Rao N, Qian M, Liu D, Li J, Feng W, Yin L and Chen X 2014 A novel method for real-time atrial fibrillation detection in electrocardiograms using multiple parameters *Ann. Noninvasive Electrocardiol.* **19** 217–25
Fuster V *et al* 2001 ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: executive summary *J. Am. Coll. Cardiol.* **38** 1231–65
Gami A S, Hodge D O, Herges R M, Olson E J, Nykodym J, Kara T and Somers V K 2007 Obstructive sleep apnea, obesity, and the risk of incident atrial fibrillation *J. Am. Coll. Cardiol.* **49** 565–71
Garca M, Rdenas J, Alcaraz R and Rieta J J 2016 Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation *Comput. Methods Prog. Biomed.* **131** 157–68
Go A S, Hylek E M, Phillips K A, Chang Y, Henault L E, Selby J V and Singer D E 2001 Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the anticoagulation and risk factors in atrial fibrillation (atria) study *JAMA* **285** 2370–5
Goldberger A L, Amaral L A, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals *Circulation* **101** 215–20
Healey J S *et al* 2012 Subclinical atrial fibrillation and the risk of stroke *New Engl. J. Med.* **366** 120–9
Heeringa J, Van der Kuip D A, Hofman A, Kors J A, Van Herpen G, Stricker B H, Stijnen T, Lip G Y and Witteman J C 2006 Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study *Eur. Heart J.* **27** 949–53
Huang C, Ye S, Chen H, Li D, He F and Tu Y 2011 A novel method for detection of the transition between atrial fibrillation and sinus rhythm *IEEE Trans. Biomed. Eng.* **58** 1113–9
Hughes M and Lip G Y 2008 Stroke and thromboembolism in atrial fibrillation: a systematic review of stroke risk factors, risk stratification schema and cost effectiveness data *Thromb. Haemost.* **99** 295–304
Ladavich S and Ghoraani B 2015 Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity *Biomed. Signal Process. Control* **18** 274–81
Lake D E 2006 Renyi entropy measures of heart rate gaussianity *IEEE Trans. Biomed. Eng.* **53** 21–7
Lake D E 2009 Nonparametric entropy estimation using kernel densities *Methods Enzymol.* **467** 531–46
Lake D E and Moorman J R 2011 Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices *Am. J. Physiol. Heart Circ. Physiol.* **300** H319–25
Lake D E, Richman J S, Griffin M P and Moorman J R 2002 Sample entropy analysis of neonatal heart rate variability *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **283** R789–97
Linker D T 2009 Long-term monitoring for detection of atrial fibrillation *US Patent* 20060084883
Lip G Y H *et al* 2016 Atrial fibrillation *Nat. Rev. Dis. Primers* **2** 16016
Liu C Y and Zhao L N 2011 Using fuzzy measure entropy to improve the stability of traditional entropy measures *Comput. Cardiol.* **38** 681–4
Liu C Y, Li K, Zhao L N, Liu F, Zheng D C, Liu C C and Liu S T 2013 Analysis of heart rate variability using fuzzy measure entropy *Comput. Biol. Med.* **43** 100–8
Liu C Y, Zhang C Q, Zhang L, Zhao L N, Liu C C and Wang H J 2015 Measuring synchronization in coupled simulation and coupled cardiovascular time series: a comparison of different cross entropy measures *Biomed. Signal Process. Control.* **21** 49–57
Mainardi L, Sornmo L and Cerutti S 2008 *Understanding Atrial Fibrillation: the Signal Processing Contribution, Part I* (Williston, VT: Morgan and Claypool Publishers)
Mozaffarian D *et al* 2015 Heart disease and stroke statistics–2015 update: a report from the American Heart Association *Circulation* **131** e29–322
Naccarelli G V, Varker H, Lin J and Schulman K L 2009 Increasing prevalence of atrial fibrillation and flutter in the United States *Am. J. Cardiol.* **104** 1534–9
Oster J and Clifford G D 2015 Impact of the presence of noise on RR interval-based atrial fibrillation detection *J .Electrocardiol.* **48** 947–51
Park J, Lee S and Jeon M 2009 Atrial fibrillation detection by heart rate variability in poincare plot *Biomed. Eng. Online* **8** 38
Petrnas A, Marozas V and Soqrnmo L 2015a Low-complexity detection of atrial fibrillation in continuous long-term monitoring *Comput. Biol. Med.* **65** 184–91
Petrnas A, Marozas V, Srnmo L and Lukosevicius A 2012 An echo state neural network for QRST cancellation during atrial fibrillation *IEEE Trans. Biomed. Eng.* **59** 2950–7
Petrnas A, Sornmo L, Lukoeviius A and Marozas V 2015b Detection of occult paroxysmal atrial fibrillation *Med. Biol. Eng. Comput.* **53** 287–97
Pincus S M and Goldberger A L 1994 Physiological time-series analysis: what does regularity quantify? *Am. J. Physiol. Heart Circ. Physiol.* **266** H1643–56
Prerfellner H, Pokushalov E, Sarkar S, Koehler J, Zhou R, Urban L and Hindricks G 2014 P-wave evidence as a method for improving algorithm to detect atrial fibrillation in insertable cardiac monitors *Heart Rhythm* **11** 1575–158
Ramdani S, Seigle B, Lagarde J, Bouchara F and Bernard P L 2009 On the use of sample entropy to analyze human postural sway data *Med. Eng. Phys.* **31** 1023–31
Rdenas J, Garca M, Alcaraz R and Rieta J J 2015 Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms *Entropy* **17** 6179–99
Richman J S and Moorman J R 2000 Physiological time-series analysis using approximate entropy and sample entropy *Am. J. Physiol. Heart Circ. Physiol.* **278** H2039–49
Ruan X H, Liu C C, Liu C Y, Wang X P and Li P 2011 Automatic detection of atrial fibrillation using R–R interval signal *4th Int. Conf. on Biomedical Engineering and Informatics (Shanghai)* pp 644–7
Sarkar S, Ritscher D and Mehra R 2008 A detector for a chronic implantable atrial tachyarrhythmia monitor *IEEE Trans. Biomed. Eng.* **55** 1219–24
Savelieva I and Camm A J 2000 Clinical relevance of silent atrial fibrillation: prevalence, prognosis, quality of life, and management *J. Interventional Cardiac Electrophysiol.* **4** 369–82
Savelieva I and Camm J 2008 Update on atrial fibrillation: part I *Clin. Cardiol.* **31** 55–62
Tateno K and Glass L 2001 Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and Δrr intervals *Med. Biol. Eng. Comput.* **39** 664–71

Wang T J, Larson M G, Levy D, Vasan R S, Leip E P, Wolf P A, D'Agostino R B, Murabito J M, Kannel W B and Benjamin E J 2003 Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study *Circulation* **107** 2920–5

Wolf P A, Abbott R D and Kannel W B 1991 Atrial fibrillation as an independent risk factor for stroke: the Framingham Study *Stroke* **22** 983–8

Zhao L N, Wei S S, Zhang C Q, Zhang Y T, Jiang X E, Liu F and Liu C Y 2015 Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects *Entropy* **19** 6270–88

Zhou X, Ding H, Ung B, Pickwell-MacPherson E and Zhang Y T 2014 Automatic online detection of atrial fibrillation based on symbolic dynamics and shannon entropy *Biomed. Eng. Online* **13** 18