

Entropy-Based Data Mining on the Example of Cardiac Arrhythmia Suppression

Martin Bachler^{1,2}, Matthias Hörtenhuber², Christopher Mayer¹,
Andreas Holzinger³, and Siegfried Wassertheurer¹

¹ AIT Austrian Institute of Technology, Health & Environment Department,
Biomedical Systems, Donau-City-Str. 1, 1220 Vienna, Austria

{martin.bachler, christopher.mayer, siegfried.wassertheurer}@ait.ac.at

² Vienna University of Technology, Institute for Analysis and Scientific Computing,
Wiedner Hauptstr. 8-10, 1040 Vienna, Austria

{martin.bachler, e0927120}@student.tuwien.ac.at

³ Medical University Graz, Research Unit HCI, Institute of Medical Informatics,
Statistics and Documentation Auenbruggerplatz 2/V, 8036 Graz, Austria
a.holzinger@hci4all.at

Abstract. Heart rate variability (HRV) is the variation of the time interval between consecutive heartbeats and depends on the extrinsic regulation of the heart rate. It can be quantified using nonlinear methods such as entropy measures, which determine the irregularity of the time intervals.

In this work, approximate entropy (ApEn), sample entropy (SampEn), fuzzy entropy (FuzzyEn) and fuzzy measure entropy (FuzzyMEn) were used to assess the effects of three different cardiac arrhythmia suppressing drugs on the HRV after a myocardial infarction.

The results show that the ability of all four entropy measures to distinguish between pre- and post-treatment HRV data is highly significant ($p < 0.01$). Furthermore, approximate entropy and sample entropy are able to differentiate significantly ($p < 0.05$) between the tested arrhythmia suppressing agents.

Keywords: Data Mining, Entropy, Heart Rate Variability, Cardiac Arrhythmia Suppression.

1 Introduction

Heart rate variability (HRV) is the variation of the time interval between consecutive heartbeats. It highly depends on the extrinsic regulation of the heart rate (HR) and reflects the balance between the sympathetic and the parasympathetic nervous system [1]. In studies of HRV, both time- and frequency-domain measures are typically used by practitioners and researchers [1,2]. Additionally, there exist non-linear measures such as the Poincaré Plot [3] and entropy measures [4]. The later one were used in this study. We applied the entropy measures on recordings from the Cardiac Arrhythmia Suppression Trial (CAST), a large postinfarction trial, with data before and after cardiac arrhythmia suppression

treatments [5]. Our goal was to examine the effects of antiarrhythmic medication on various entropy measures.

2 Methods

2.1 Data and Study Population

All data used in this paper have been taken from Physionet.org [6], a free-access, on-line archive of physiological signals. Particularly, data are obtained from the CAST RR Interval Sub-Study Database [5], which consists of 1543 24-hour RR-interval records from 809 subjects. The database is divided into three sub-groups based on the cardiac arrhythmia suppression medication (Encainide, Flecainide and Moricizine) received by the subjects. For almost all subjects, there is a pair of records representing baseline and on-therapy data available. In total, 1464 records for 731 subjects (599 men and 132 women) have been used and 75 subjects have been excluded due to incompleteness of data (i.e., just baseline or just on-therapy data available), three subject were excluded additionally, because there were no recordings at the used time window. The age distribution of the subjects is represented in Figure 1. One-hour RR-intervals at 6pm have been extracted for all subjects to decrease computation time and to avoid daytime dependent variations. The Cardiac Arrhythmia Suppression Trial (CAST) was originally started to analyze the effect of suppressing ventricular arrhythmias by antiarrhythmic drugs after myocardial infarction (MI) on the survival rate [7]. The data are divided in three sub-groups depending on the treatment (Encainide, $N_E = 260$ (44 female, 216 male); Flecainide, $N_F = 207$ (43 female, 164 male); Moricizine, $N_M = 264$ (45 female, 219 male)).

2.2 Analysis of Heart Rate Variability

Heart rate variability is analyzed by the following entropy measures: approximate entropy (ApEn) [8], sample entropy (SampEn) [9], fuzzy entropy (FuzzyEn) [10] and fuzzy measure entropy (FuzzyMEn) [11].

2.3 Approximate Entropy (ApEn)

Approximate Entropy measures the logarithmic likelihood that runs of patterns that are close remain close on following incremental comparisons [12]. We state Pincus' definition [12,13], for the family of statistics $\text{ApEn}(m, r, N)$;

Definition 1. Fix m , a positive integer and r , a positive real number. Given a regularly sampled time series $u(t)$, a sequence of vectors $\mathbf{x}(1)^m, \mathbf{x}^m(2), \dots, \mathbf{x}^m(N - m + 1)$ in \mathbb{R}^m is formed, defined by

$$\mathbf{x}^m(i) := [u(t_i), u(t_{i+1}), \dots, u(t_{i+m-1})] \quad . \quad (1)$$

Define for each i , $1 \leq i \leq N - m + 1$,

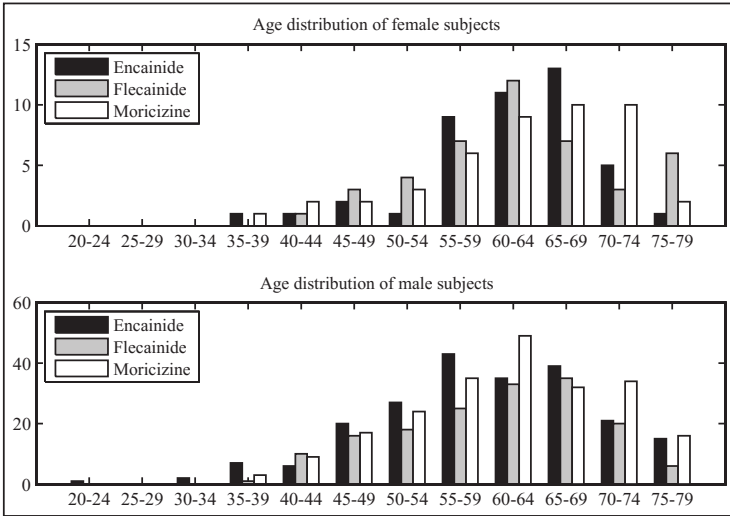


Fig. 1. Age distribution of subjects per gender

$$C_i^m(r) := \frac{\text{number of } j \text{ such that } d[\mathbf{x}^m(i), \mathbf{x}^m(j)] \leq r}{N - m + 1} , \tag{2}$$

where $d[\mathbf{x}(i), \mathbf{x}(j)]$ is the Chebyshev distance given by:

$$d[\mathbf{x}^m(i), \mathbf{x}^m(j)] := \max_{k=1,2,\dots,m} (|u(t_{i+k-1}) - u(t_{j+k-1})|) . \tag{3}$$

Furthermore, define

$$\phi^m(r) := (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log C_i^m(r) , \tag{4}$$

then the **Approximate Entropy** is defined as

$$\text{ApEn}(m, r, N) := \phi^m(r) - \phi^{m+1}(r) . \tag{5}$$

2.4 Sample Entropy (SampEn)

Richman and Moorman showed in [14] that approximate entropy is biased towards regularity. Thus, they modified it to Sample Entropy. The main difference between the two is that sample entropy does not count self-matches, and only the first $N - m$ subsequences instead of all $N - m + 1$ are compared, for both ϕ^m and ϕ^{m+1} [14]. Similar to ApEn above, SampEn is defined as follows:

Definition 2. Fix m , a positive integer and r , a positive real number. Given a regularly sampled time series $U(t)$, a sequence of vectors $\mathbf{x}^m(1), \mathbf{x}^m(2), \dots, \mathbf{x}^m(N - m + 1) \in \mathbb{R}^m$ is formed, defined by eq. (1). Define for each $i, 1 \leq i \leq N - m + 1$,

$$C_i^m = \frac{\text{number of } j \text{ such that } d[\mathbf{x}^m(i), \mathbf{x}^m(j)] \leq r \text{ and } i \neq j}{N - m + 1}, \tag{6}$$

where $d[(i), (j)]$ is the Chebyshev distance (see eq. (3)). Furthermore, define

$$\phi^m(r) := (N - m)^{-1} \sum_{i=1}^{N-m} C_i^m(r), \tag{7}$$

then the **Sample Entropy** is defined as

$$\text{SampEn}(m, r, N) := \log(\phi^m(r)) - \log(\phi^{m+1}(r)). \tag{8}$$

2.5 Fuzzy (Measure) Entropy (Fuzzy(M)En)

To soften the effects of the threshold value r , Chen et al. proposed in [15] Fuzzy Entropy, which uses a fuzzy membership function instead of the Heaviside function. FuzzyEn is defined the following way:

Definition 3. Fix m , a positive integer and r , a positive real number. Given a regularly sampled time series $U(t)$, a sequence of vectors $\mathbf{x}^m(1), \mathbf{x}^m(2), \dots, \mathbf{x}^m(N - m + 1) \in \mathbb{R}^m$ is formed, as defined by eq. (1). This sequence is transformed into $\bar{\mathbf{x}}^m(1), \bar{\mathbf{x}}^m(2), \dots, \bar{\mathbf{x}}^m(N - m + 1)$, with $\bar{\mathbf{x}}^m(i) := \{u(t_i) - u0_i, \dots, u(t_{i+m-1}) - u0_i\}$, where $u0_i$ is the mean value of $\mathbf{x}^m(i)$, i.e.

$$u0_i := \sum_{j=0}^{m-1} \frac{u_{i+j}}{m}. \tag{9}$$

Next the fuzzy membership matrix is defined as:

$$D_{i,j}^m := \mu(d(\bar{\mathbf{x}}_i^m, \bar{\mathbf{x}}_j^m), n, r), \tag{10}$$

with the Chebyshev distance d (see eq. (3)) and the fuzzy membership function

$$\mu(\mathbf{x}, n, r) := e^{-(\mathbf{x}/r)^n}. \tag{11}$$

Finally, with

$$\phi^m := \frac{1}{N - m} \sum_{i=1}^{N-m} \sum_{j=1, j \neq i}^{N-m} \frac{D_{i,j}^m}{N - m - 1}, \tag{12}$$

the **Fuzzy Entropy** is defined as:

$$\text{FuzzyEn}(m, r, n, N) := \ln \phi^m - \ln \phi^{m+1}. \tag{13}$$

Liu et al. proposed in [16] **Fuzzy Measure Entropy**, which introduces a distinction between local entropy and global entropy, based on FuzzyEn. It is defined as:

$$\text{FuzzyMEn}(m, r_L, r_F, n_L, n_F, N) := \ln \phi_L^m - \ln \phi_L^{m+1} + \ln \phi_F^m - \ln \phi_F^{m+1} \quad , \quad (14)$$

where the local terms ϕ_L^m and ϕ_L^{m+1} are calculated as in eq. (12) and the global terms ϕ_F^m and ϕ_F^{m+1} are calculated with eq. (10) and eq. (12), but with $\bar{x}^m(i) := \{u(t_i) - u_{\text{mean}}, \dots, u(t_{i+m-1}) - u_{\text{mean}}\}$, where u_{mean} is the mean value of the complete sequence $u(t)$.

Parameters of all entropy measures were selected according to [17].

2.6 Statistical Analysis

For better comprehension we will call from here on the Encainide group recordings before treatment EA and after treatment EB, the recordings of the Flecainide group pretreatment will be marked with FA and the posttreatment ones FB. The same scheme is used for MA and MB as abbreviations for the pre- and posttreatment recordings of the group receiving a medication with Moricizine.

First differences in the entropies of the groups' pretreatment recordings, i.e. between EA, FA, and MA (the baseline), were tested using the Kruskal Wallis test, since the three groups are assumed to be independent but not normally distributed. Afterwards we tested again with a Kurskal Wallis test for differences between the groups' entropies after the treatment, i.e. between EB, FB, and MA. Subsequently, the effect of the treatment was tested with a Wilcoxon signed-rank test for paired samples without normal distribution, by comparing the entropies of EA with EB, FA with FB, and MA with MB.

Further on we tested if there was any sex-based difference in the entropy values. Therefore we used the Wilcoxon rank sum test to test the female members of the group EA against the male members of the same group. The same was done for FA, MA and EB, FB, MB.

Finally, the connection of the various entropy values was investigated using scatter plots and Pearson's linear correlation coefficient on the results before and after the treatment.

For all tests their implementation in *The MathWorks MATLAB* was used. The test results were declared significant for $p < 0.05$ and highly significant for $p < 0.01$.

3 Results

Figure 2 shows the distribution of entropy values of the groups before and after treatment. All posttreatment recordings show lower entropy values than their respective pretreatment counterparts. No significant difference between the three medication groups could be detected in the entropies of the pretreatment recordings, as can be seen in Table 1. The same table shows, that ApEn and SampEn were significantly different for the recordings after the treatment between the

three groups ($p < 0.05$). This was not the case for FuzzyEn and FuzzyMEn. In the same table it can also be seen, that all four entropy measures show a highly significant difference between the subjects entropy of heart rate variability before and after treatment ($p < 0.01$ for all tests).

No significant difference was found between female and male subjects in the three groups before as well as after the treatment, see Table 2 for details.

Table 1. Results of the Kurskal-Wallis test and the Wilcoxon signed rank test of the recordings before and after treatment with different medications

	Entropy Measure	P-Value
EA vs. FA vs. MA:		
	ApEn	0.2974
	SampEn	0.1267
	FuzzyEn	0.2765
	FuzzyMEn	0.2635
EB vs. FB vs. MB:		
	ApEn	0.0307
	SampEn	0.0448
	FuzzyEn	0.1698
	FuzzyMEn	0.1055
EA vs. EB:		
	ApEn	< 0.01
	SampEn	< 0.01
	FuzzyEn	< 0.01
	FuzzyMEn	< 0.01
FA vs. FB:		
	ApEn	< 0.01
	SampEn	< 0.01
	FuzzyEn	< 0.01
	FuzzyMEn	< 0.01
MA vs. MB:		
	ApEn	< 0.01
	SampEn	< 0.01
	FuzzyEn	< 0.01
	FuzzyMEn	< 0.01

Figure 3 and Table 3 show the connection of the readings of the various entropy determining methods. Figure 3 contains scatter plots of the pairwise comparisons of the methods and the histogram for each method, respectively. It can be seen that the distribution of the results of all methods is highly asymmetrical. The scatter plots show that, in general, values of ApEn were higher than

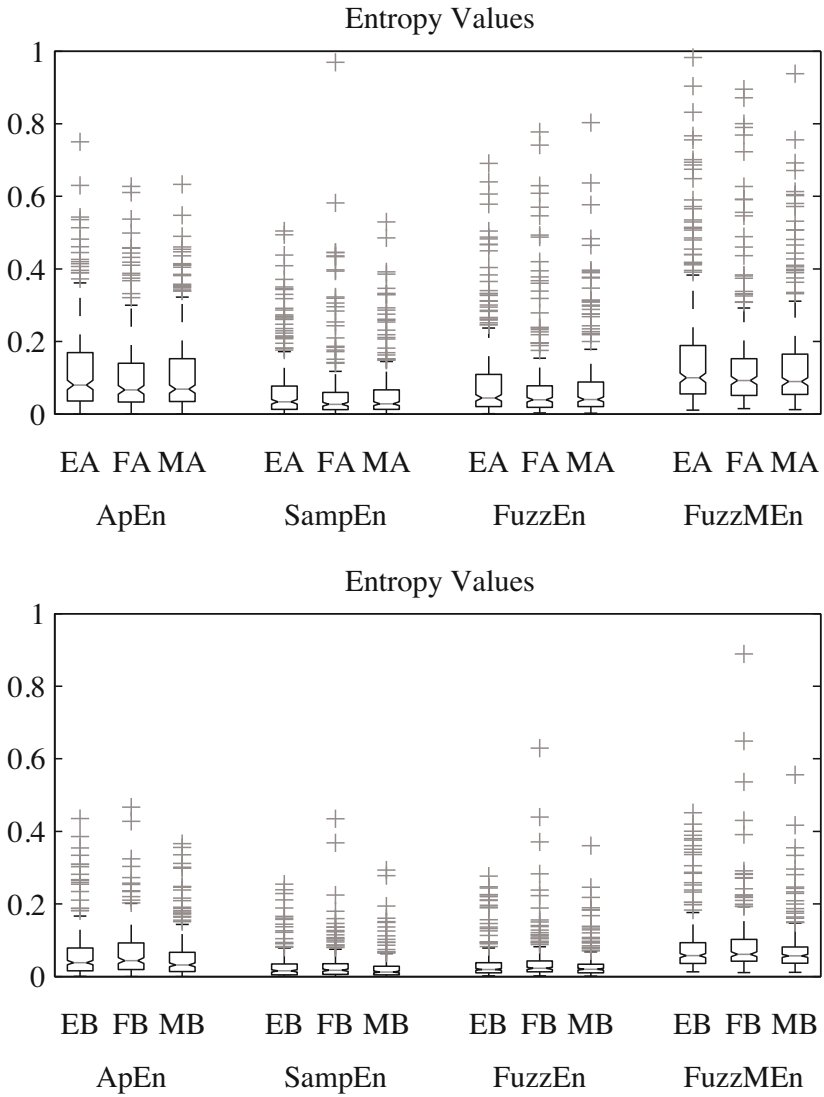


Fig. 2. Boxplot of the entropy values of the recordings before treatment (top) and afterwards (bottom)

Table 2. Results of the Wilcoxon rank sum test of the recordings before and after treatment with different medications separated by subject's sex

	Entropy Measure	P-Value
EA-female vs. EA-male:		
	ApEn	0.2080
	SampEn	0.1662
	FuzzyEn	0.4064
	FuzzyMEn	0.3250
FA-female vs. FA-male:		
	ApEn	0.3373
	SampEn	0.2893
	FuzzyEn	0.5245
	FuzzyMEn	0.3996
MA-female vs. MA-male:		
	ApEn	0.8994
	SampEn	0.6603
	FuzzyEn	0.7494
	FuzzyMEn	0.7204
EB-female vs. EB-male:		
	ApEn	0.8012
	SampEn	0.7465
	FuzzyEn	0.8820
	FuzzyMEn	0.7473
FB-female vs. FB-male:		
	ApEn	0.9715
	SampEn	0.9237
	FuzzyEn	0.4545
	FuzzyMEn	0.1055
MB-female vs. MB-male:		
	ApEn	0.7690
	SampEn	0.4302
	FuzzyEn	0.3772
	FuzzyMEn	0.3294

those of SampEn and FuzzyEn, whereas they are slightly lower than reading of FuzzyMEn. Results of SampEn tended to be lower than those of FuzzyEn and FuzzyMEn. The direct comparison of FuzzyEn and FuzzyMEn showed higher values for FuzzyMEn. Nevertheless, the scatter plots show a reasonably linear connection between all methods. Table 3 quantifies this observation, showing high correlation coefficients for the pairwise comparison of all methods.

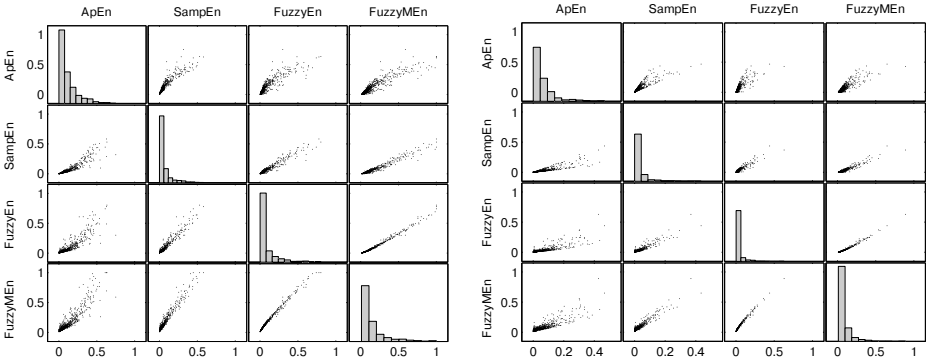


Fig. 3. Scatter plot matrix and histograms of the entropy values of the recordings before treatment (left) and afterwards (right) in order to visualize the connection between different entropy measures

Table 3. Pearson’s linear correlation coefficients for entropy values before and after treatment (all $p < 0.01$)

	Before Treatment (A)				After Treatment (B)			
	Ap	Samp	Fuzzy	FuzzyM	Ap	Samp	Fuzzy	FuzzyM
ApEn	1	0.9210	0.8822	0.9111	1	0.8736	0.8185	0.8601
SampEn	0.9210	1	0.9649	0.9750	0.8736	1	0.9341	0.9509
FuzzyEn	0.8822	0.9649	1	0.9936	0.8185	0.9341	1	0.9889
FuzzyMEn	0.9111	0.9750	0.9936	1	0.8601	0.9509	0.9889	1

4 Discussion

The results of the first test on the three pretreatment groups (EA vs. FA vs. MA) did not reveal any significant differences. Therefore the null-hypothesis of a consistent baseline population is accepted and further posttreatment comparisons are reasonable.

Testing the differences between the posttreatment groups (EB vs. FB vs. MB), though, yielded mixed results. While ApEn and SampEn showed significantly

different results for the diverse treatments, FuzzyEn and FuzzyMEn did not. Further investigation of these methods is necessary to determine the cause of these differences.

Cardiac arrhythmias are a disruption in the regularity of the heart rhythm. Therefore, they cause a distinct alteration in the HRV and in all measures quantifying it. Sethuraman et al. reported, that even one single ectopic beat (a certain type of cardiac arrhythmia) causes a striking alteration of the HRV [18]. As Encainide, Flecainide and Moricizine aim to suppress cardiac arrhythmia, the prominent reduction of the irregularity and therefore the highly significant difference between pre- and posttreatment as seen in Figure 2 and Table 1 was expected. It remains an open question, however, if these differences can be attributed solely to the reduction of cardiac arrhythmias, or whether the heart's sinus rhythm is changed as well. Methods of ectopic beat correction will be necessary for further data mining regarding this matter [19].

The fourth test, which looked for differences due to the sex of the subjects, showed no significant results. This is in accordance with Beckers et al., who showed in [20] a difference in the entropy based on sex, but reported that this effect vanishes for subjects older than 40 years. This is the case for most of the subjects in our used data.

The qualitative investigation of the connection between the four entropy measures using scatter plots (Figure 3) revealed obvious pairwise linear relationships. Therefore, the usage of Pearson's linear correlation coefficient is justified. As listed in Table 3, the highest correlations were found between FuzzyEn and FuzzyMEn. These readings suggest that the extension of FuzzyEn by a global term influences the results only by a constant value. This similarity between FuzzyEn and FuzzyMEn is also in accordance with the findings of the second test. Again, the striking reduction in all entropy measures between pre- and posttreatment is visible.

In general, Mäkikallio et al. found in [21] higher ApEn values for postinfarction patients compared to a healthy age matched control group. In our data this effect seems to be reduced after the treatment with any of the three medications. However, simply reducing the amount of cardiac arrhythmias and therefore reducing the entropy of the HRV for patients after a myocardial infarction does not reduce mortality. In fact, the postinfarction treatment with (Na^+) channel blocking antiarrhythmic agents (class I, e.g. Encainide, Flecainide and Moricizine) is associated with increased mortality [22]. This suggests the presence of more extensive alterations of the HRV than cardiac arrhythmias alone. As stated above, ectopic beat correction will be necessary for further investigation.

4.1 Limitations

Due to the lack of availability of the related survival outcome data, we could not evaluate the predictability of the tested entropy measures on mortality.

5 Conclusion

In our study, all four entropy measures (approximate, sample, fuzzy and fuzzy measure entropy) are significantly different before and after antiarrhythmic treatment. However, as also addressed by Holzinger et al. in [4], the problem of how to use entropy measures for the classification of pathological and non-pathological data still remains, as a simple reduction of entropy in HRV does not necessarily reduce mortality after a myocardial infarction. Further research using ectopic beat correction for entropy-based data mining in HRV will be necessary.

References

1. Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C., Suri, J.: Heart rate variability: a review. *Med. Biol. Eng. Comput.* 44, 1031–1051 (2006)
2. American Heart Association Inc.: European Society of Cardiology: Guidelines – heart rate variability. *Eur. Heart J.* 17, 354–381 (1996)
3. Smith, A.L., Reynolds, K.J., Owen, H.: Correlated Poincaré indices for measuring heart rate variability. *Australasian physical & engineering sciences in medicine / supported by the Australasian College of Physical Scientists in Medicine and the Australasian Association of Physical Sciences in Medicine* 30, 336–341 (2007)
4. Holzinger, A., Hörtenhuber, M., Mayer, C., Bachler, M., Wassertheurer, S., Pinho, A.J., Koslicki, D.: On entropy-based data mining. In: Holzinger, A., Jurisica, I. (eds.) *Knowledge Discovery and Data Mining. LNCS*, vol. 8401, pp. 209–226. Springer, Heidelberg (2014)
5. Stein, P.K., Kleiger, R.E., Domitrovich, P.P., Schechtman, K.B., Rottman, J.N.: Clinical and demographic determinants of heart rate variability in patients post myocardial infarction: insights from the cardiac arrhythmia suppression trial (cast). *Clinical Cardiology* 23, 187–194 (2000)
6. 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., Stanley, H.E.: *PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.* *Circulation* 101, E215–E220 (2000)
7. Epstein, A.E., Bigger, J.T., Wyse, D.G., Romhilt, D.W., Reynolds-Haertle, R.A., Hallstrom, A.P.: Events in the cardiac arrhythmia suppression trial (cast): mortality in the entire population enrolled. *Journal of the American College of Cardiology* 18, 14–19 (1991)
8. Pincus, S.M.: Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* 88, 2297–2301 (1991)
9. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049 (2000)
10. Chen, W., Zhuang, J., Yu, W., Wang, Z.: Measuring complexity using FuzzyEn, ApEn, and SampEn. *Med. Eng. Phys.* 31, 61–68 (2009)
11. Liu, C., Li, K., Zhao, L., Liu, F., Zheng, D., Liu, C., Liu, S.: Analysis of heart rate variability using fuzzy measure entropy. *Comput. Biol. Med.* 43, 100–108 (2013)
12. Pincus, S.M.: Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* 88, 2297–2301 (1991)
13. Pincus, S.: Approximate entropy (apen) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 5, 110–117 (1995)

14. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049 (2000)
15. Chen, W., Wang, Z., Xie, H., Yu, W.: Characterization of surface emg signal based on fuzzy entropy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 15, 266–272 (2007)
16. Liu, C., Li, K., Zhao, L., Liu, F., Zheng, D., Liu, C., Liu, S.: Analysis of heart rate variability using fuzzy measure entropy. *Comput. Biol. Med.* 43, 100–108 (2013)
17. Mayer, C., Bachler, M., Hörtenhuber, M., Stocker, C., Holzinger, A., Wassertheurer, S.: Selection of entropy-measure parameters for knowledge discovery in heart rate variability data (in press)
18. Sethuraman, G., Ryan, K.L., Rickards, C.A., Convertino, V.A.: Ectopy in trauma patients: cautions for use of heart period variability in medical monitoring. *Aviation, Space, and Environmental Medicine* 81, 125–129 (2010)
19. Mateo, J., Laguna, P.: Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Transactions on Biomedical Engineering* 50, 334–343 (2003)
20. Beckers, F., Verheyden, B., Aubert, A.E.: Aging and nonlinear heart rate control in a healthy population. *American Journal of Physiology-Heart and Circulatory Physiology* 290, H2560–H2570 (2006)
21. Mäkikallio, T.H., Seppänen, T., Niemelä, M., Airaksinen, K.J., Tulppo, M., Huikuri, H.V.: Abnormalities in beat to beat complexity of heart rate dynamics in patients with a previous myocardial infarction¹. *Journal of the American College of Cardiology* 28, 1005–1011 (1996)
22. Teo, K., Yusuf, S., Furberg, C.: Effects of prophylactic antiarrhythmic drug therapy in acute myocardial infarction: An overview of results from randomized controlled trials. *JAMA* 270, 1589–1595 (1993)