

Heart Arrhythmia Classification using Extracted Features in Poincare Plot of RR Intervals

Shahab Rezaei¹, Sadaf Moharreri², Mostafa Abdollahpur³, Saman Parvaneh^{4,5}

¹Islamic Azad University, Central Tehran Branch, Tehran, Iran

²Islamic Azad University, Khomeini Shahr Branch, Isfahan, Iran

³K.N. Toosi University of Technology, Tehran, Iran

⁴Islamic Azad University, Science and Research Branch, Tehran, Iran

⁵Philips Research North America, Cambridge, MA, USA

Abstract

*Early diagnosis of cardiac arrhythmia is important for a better management of arrhythmia. The main goal of this article was to compare conventional and new features extracted from Poincare plot between normal sinus rhythm, atrial fibrillation, acute myocardial infarction, and congestive heart failure. Furthermore, extracted features from Poincare plot were used w *k*-nearest neighbor (KNN) classifier for classification of NSR and different arrhythmia. The classification sensitivity, specificity, and accuracy of 94.63%, 98.21%, and 97.31% were achieved using extracted features with KNN.*

1. Introduction

Cardiovascular disorders are one of the major causes of mortality [1]. Prevalence of heart arrhythmias, irregular electrical activities originate from heart, is increasing rapidly [1] and early diagnosis of cardiac arrhythmia makes it possible to choose appropriate anti-arrhythmic drugs and is important for improving arrhythmia therapy/management [2].

The Poincare plot is a tool developed by Henry Poincare for analyzing complex systems [3]. In the context of medical science, it is mainly used for HRV analysis and is proved to be an effective measure in different applications (e.g., arrhythmia detection and emotion assessment) [4]. Poincare plot is a geometrical representation of RR time series to represent patterns of heart rate dynamics [5]. Poincare plot analysis of RR time series allows a beat-to-beat analysis approach for HRV analysis [6]. Poincare plot is a known method for the analysis of two-dimensional nonlinear dynamic systems [7]. For quantification of Poincare plot, an ellipse fit to the distribution of points and *SD1* and *SD2* are used as

two conventional descriptors of Poincare plot [8].

In recent years, a lot of studies have been focused on Poincare plot and defined different features based on this plot that each of them can represent different aspects of RR intervals in this space.

In this article, the main aim is to compare features extracted from Poincare plot between normal sinus rhythm (NSR), atrial fibrillation (AF), acute myocardial infarction (MI), and congestive heart failure (CHF) and to create a classifier to discriminate these groups.

2. Data and Method

Block diagram of the proposed algorithm for arrhythmia classification is shown in Figure 1. In the first step, the RR time series of NSR, AF, MI, and CHF have been obtained from Physionet database [9]. In second step, different features in Poincare plot were extracted for HRV analysis, including new features introduced recently [10, 11]. In the last step, these features were used as input to *K*-nearest neighbor (*KNN*) classifier for classification of NSR and three arrhythmia (AF, MI, and CHF).

2.1. Data

The RR time series of NSR and three heart arrhythmia have been obtained from Physionet database [9]. Fourteen *RR* time series of subjects in NSR were extracted from Physionet Normal Sinus Rhythm database [12]. Furthermore, 14 *RR* time series of subjects with AF were obtained from MIT-BIH Atrial Fibrillation Database. Fourteen *RR* time series from NHLBI sponsored Cardiac Arrhythmia Suppression Trial (CAST) RR-interval sub-study database were used for the MI groups. Also, 14 *RR* time series of subjects with CHF were extracted from Physionet Congestive Heart Failure database [9].



Figure 1. Block diagram of proposed algorithm for classification of NSR and three arrhythmia. (NSR: normal sinus rhythm, AF: atrial fibrillation, MI: acute myocardial infarction, CHF: congestive heart failure)

2.2. Feature Extraction

In this study, RR intervals with length of 3500 were used for feature extraction.

2.2.1. Poincare Plot

Given RR time series as $(RR_1, RR_2, \dots, RR_n, RR_{n+1})$, the standard Poincare plot is constructed by plotting each RR interval against next RR interval. The number of points in the Poincare plot (n) will be one less than the length of the RR time series [13].

2.2.2. Standard Descriptors of Poincare Plot

Two standard descriptors of Poincare plot are $SD1$ and $SD2$ which are defined as the standard deviation of projection of the points in Poincare plot on the line perpendicular to the identity line ($y = -x$) and identity line ($y = x$), respectively [4].

2.2.3. Complex Correlation Measure (CCM)

Karmakar et al. for quantifying temporal variation in points' distribution in Poincare plot, developed a descriptor called Complex Correlation Measure (CCM) [11]. For measuring CCM, moving windows of three consecutive points from the Poincare plot is considered and the area of the triangle formed by these three points are computed [11]. By considering three consecutive points $a(x_1, y_1)$, $b(x_2, y_2)$, and $c(x_3, y_3)$, the area of the triangle (A) for i^{th} window can be computed using the following determinant [11]:

$$A(i) = \frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \quad (1)$$

where A is defined on the real line \mathfrak{R} and if points a , b and c are on a straight line, $A_i = 0$; if the points a , b and c have counter clock wise orientation, $A_i > 0$; and if the

points a , b and c are in clock wise orientation, $A_i < 0$.

By considering n points in Poincare plot, the parameter CCM is defined as [11]:

$$CCM(m) = \frac{1}{C_n} \sum_{i=1}^{n-2} A(i) \quad (2)$$

where m represents lag of Poincare plot and C_n is the normalizing constant which is defined as [11]:

$$C_n = \pi * SD1 * SD2 \quad (3)$$

2.2.4. Global Occurrence Matrix (GOM)

Moharreri et al. proposed GOM features to quantify the distribution of points in Poincare plot relative to the identity line ($y = x$) [10]. Number of points above, on, and below the identity line were extracted as features [10]:

$$GOM = [N_A \ N_O \ N_B] \quad (4)$$

Where N_A is the number of points above the identity line, N_O is the number of points on the identity line, and N_B is the number of points below the identity line.

2.2.5. Co-Occurrence Matrix (COM)

Moharreri et al. proposed nine COM features to quantify the behavior of two consecutive points in Poincare plot based on position of these two points relative to identity line and each other [10]:

$$COM = \begin{bmatrix} AA & AO & AB \\ OA & OO & OB \\ BA & BO & BB \end{bmatrix} \quad (5)$$

For example, OA is the number of consecutive points in Poincare plot that the first point is on the identity line and the second point is above identity line. More details about the COM can be found in [10, 14].

3. Statistical Analysis and Classification

After extracting above-mentioned features, statistical analysis was used to identify features that are significantly different between groups. Kruskal-Wallis test was used to compare extracted features between different groups; level of significance was set to 0.05.

Table 1. Mean and standard deviation of extracted features across groups with NSR and three different cardiac arrhythmia (NSR: Normal Sinus Rhythm, CHF: Congestive Heart Failure, AF: Atrial Fibrillation, MI: Acute Myocardial Infarction).

	NSR	CHF	AF	MI
<i>SD1</i>	0.02 ± 0.00006	0.53 ± 2.16	0.08 ± 0.003	1.98 ± 29.10
<i>SD2</i>	0.13 ± 0.002	0.53 ± 2.12	0.12 ± 0.003	1.98 ± 29.04
<i>CCM</i>	-0.07 ± 0.0004	-0.31 ± 0.0004	-0.23 ± 0.001	-0.28 ± 0.003
<i>C_n</i>	0.01 ± 5E-5	7.14 ± 0.006	0.03 ± 0.0001	97.18 ± 12.6
<i>N_A</i>	1470 ± 22	190 ± 1	1550 ± 19	1000 ± 88
<i>N_O</i>	486 ± 194	189 ± 4	187 ± 63	417 ± 3
<i>N_B</i>	1486 ± 133	19 ± 11	1419 ± 179	974 ± 74
<i>AA</i>	583 ± 98	0	657 ± 230	265 ± 32
<i>AO</i>	208 ± 35	17 ± 8	91 ± 12	158 ± 38
<i>AB</i>	683 ± 55	1 ± 0.02	802 ± 264	576 ± 259
<i>OA</i>	211 ± 249	2 ± 0.06	89 ± 120	166 ± 354
<i>OO</i>	90 ± 33	186 ± 10	13 ± 14	104 ± 42
<i>OB</i>	184 ± 149	2 ± 0.08	84 ± 17	146 ± 33
<i>BA</i>	679 ± 43	2 ± 0.07	804 ± 266	568 ± 245
<i>BO</i>	187 ± 117	3 ± 0.06	82 ± 166	154 ± 30
<i>BB</i>	618 ± 189	0	531 ± 387	250 ± 70

Table 2. *P*-value for GOM and COM features. Statistical significance is highlighted with *.

	<i>N_A</i>	<i>N_O</i>	<i>N_B</i>	<i>AA</i>	<i>AO</i>	<i>AB</i>	<i>OA</i>	<i>OO</i>	<i>OB</i>	<i>BA</i>	<i>BO</i>	<i>BB</i>
<i>NSR & CHF</i>	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
<i>NSR & AF</i>	0.059	< 0.001*	0.564	0.168	< 0.001*	0.005*	< 0.001*	< 0.001*	< 0.001*	0.003*	< 0.001*	0.154
<i>NSR & MI</i>	< 0.001*	0.462	< 0.001*	< 0.001*	0.080	0.093	0.080	0.550	0.089	0.066	0.107	< 0.001*
<i>CHF & AF</i>	< 0.001*	0.194	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
<i>CHF & MI</i>	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
<i>AF & MI</i>	< 0.001*	< 0.001*	< 0.001*	< 0.001*	0.015*	< 0.001*	< 0.001*	< 0.001*	0.015*	< 0.001*	0.001*	< 0.001*

For classification, K-nearest neighbor (KNN) classifier was used to classify NSR and different groups of arrhythmia. The classifier was trained on 70% of data as a train set, and the accuracy was evaluated on 30% of data as a test set.

4. Results

The means and standard deviations for all extracted features in four groups are reported in Table 1. The *p*-values from Kruskal-Wallis analysis for all 16 features are summarized in Table 2 and Table 3. Selected features were used with KNN classifier for heart arrhythmia classification. The classification performance on test set is reported in Table 4.

Table 3. *P*-value for *SD1*, *SD2*, *CCM* and *C_n*. Statistical significance is highlighted with *.

	<i>SD1</i>	<i>SD2</i>	<i>CCM</i>	<i>C_n</i>
<i>NSR & CHF</i>	1	0.151	< 0.001*	0.781
<i>NSR & AF</i>	< 0.001*	0.490	< 0.001*	0.013
<i>NSR & MI</i>	< 0.001*	0.080	< 0.001*	< 0.001*
<i>CHF & AF</i>	0.547	0.211	0.062	0.308
<i>CHF & MI</i>	0.016*	0.014	0.062	0.014
<i>AF & MI</i>	0.002*	0.014	0.098	0.006

Table 4. Classification performance on test data

	Accuracy	Sensitivity	Specificity
NSR	94.64%	92.85%	95.23%
CHF	100%	100%	100%
AF	96.42%	92.85%	97.61%
MI	98.21%	92.85%	100%

5. Conclusion

In this article, conventional and new features extracted from Poincare plot of RR intervals were compared between normal sinus rhythm, atrial fibrillation, acute myocardial infarction, and congestive heart failure. Furthermore, these features were used as an input to K-nearest neighbor classifier for heart arrhythmia classification. The promising results of this study warrant future study with a larger data set. Also, sensitivity of the proposed algorithm to signal length should be evaluated in future studies.

References

- [1] T. W. Kamarck and J. R. Jennings, "Biobehavioral factors in sudden cardiac death," *Psychological bulletin*, vol. 109, no. 1, p. 42, 1991.
- [2] S. Moharreri, S. Rezaei, N. Jafarnia Dabanloo, and S. Parvaneh, "Extended Parabolic Phase Space Mapping (EPPSM):

Novel Quadratic Function for Representation of Heart Rate Variability Signal," presented at the Computing in Cardiology (CinC 2014), 2014.

[3] P. Kamen, H. Krum, and A. Tonkin, "Poincaré plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans," *Clinical science*, vol. 91, pp. 201-208, 1996.

[4] J. Piskorski and P. Guzik, "Filtering Poincaré plots," *Computational methods in science and technology*, vol. 11, no. 1, pp. 39-48, 2005.

[5] V. Schechtman *et al.*, "Dynamic analysis of cardiac RR intervals in normal infants and in infants who subsequently succumbed to the sudden infant death syndrome," *Pediatric research*, vol. 31, no. 6, p. 606, 1992.

[6] M. Woo, W. Stevenson, D. Moser, R. Trelease, and R. Harper, "Patterns of beat-to-beat heart rate variability in advanced heart failure* 1," *American Heart Journal*, vol. 123, no. 3, pp. 704-710, 1992.

[7] M. Brennan, M. Palaniswami, and P. Kamen, "Do existing measures of Poincaré Plot geometry reflect nonlinear features of Heart Rate Variability?," *IEEE transactions on biomedical engineering*, vol. 48, no. 11, pp. 1342-1347, 2001.

[8] M. Tulppo, T. Makikallio, T. Takala, T. Seppanen, and H. Huikuri, "Quantitative beat-to-beat analysis of heart rate dynamics during exercise," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 271, no. 1, p. H244, 1996.

[9] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.

[10] S. Moharreri, S. Parvaneh, N. Jafarnia Dabanloo, and A. M. Nasrabadi, "Utilizing Occurrence Sequence of Heart Rate's Phase Space Points in order to Discriminate Heart Arrhythmia," in *the 17th Iranian Conference of Biomedical Engineering (ICBME2010)*, Isfahan, Iran, 2010, vol. 17: IEEE.

[11] C. K. Karmakar, A. H. Khandoker, J. Gubbi, and M. Palaniswami, "Complex Correlation Measure: A Novel Descriptor for Poincaré Plot," *Biomedical Engineering OnLine*, vol. 8, pp. 1-12, 2009.

[12] A. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, p. e215, 2000.

[13] J. Piskorski and P. Guzik, "Geometry of the Poincaré plot of RR intervals and its asymmetry in healthy adults," *Physiological measurement*, vol. 28, no. 3, p. 287, 2007.

[14] N. Jafarnia Dabanloo, G. Attarodi, S. Moharreri, S. Parvaneh, and A. M. Nasrabadi, "Emotion Recognition based on Utilizing Occurrence Sequence of Heart Rate's Phase Space Points," in *The IASTED International Conference of Biomedical Engineering (BioMed2012)*, Innsbruck, Austria, 2012, pp. 52-55: IEEE.

Address for correspondence.

Saman Parvaneh
2 Canal Park, 3rd floor, Cambridge, MA 02141
parvaneh@ieee.org