Study of Traditional and Enhanced Poincare Plot Descriptors for Atrial Fibrillation Detection

Sadaf Moharreri[1], Shahab Rezaei[2], Nader Jafarnia Dabanloo[3], Saman Parvaneh[4]

[1] Department of Biomedical Engineering, Khomeini Shahr Branch, Islamic Azad University, Isfahan, Iran
[2] Department of Biomedical Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran
[3] Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[4] Edwards Lifesciences, Irvine, CA, USA

Abstract

This article studies extracted from the Poincare Plot of RR intervals for atrial fibrillation (AF) detection. This study used AF, normal sinus rhythm (NSR), and other rhythms from the PhysioNet/Computing in Cardiology Challenge 2017 training dataset (NSR: 5,074, AF: 757, other rhythms: 2,415, and noise: 279). After QRS detection, 2D and 3D Poincare plots of RR intervals were constructed. Two traditional features from the 2D Poincare plot and seventeen geometric features extracted from the 3D Poincare plot were calculated. AutoML was used to find the best classifier, maximizing the f1-score. AutoML was trained on 80% of the data as a train set, and the f1-score was evaluated on 20% as a test set. Catboost was selected as a final model, which led to the f1-score of 0.87, 0.6, and 0.63 for NSR, AF, and other rhythms, respectively. Using AutoML with extracted geometric features facilitates finding the best model.

1. Introduction

Atrial Fibrillation (AF), the most common arrhythmia, is a heart condition characterized by an irregular and often rapid heart rate. Since AF is associated with an increased risk of stroke [1], detecting AF versus normal sinus rhythm (NSR) and other rhythms is an active research area [2], especially post-operation [3].

Heart Rate Variability (HRV) is the variation in time intervals between consecutive heartbeats [4] and was used successfully for AF detection [5-7].

In this paper, we evaluated the use of geometrics features extracted from the 3D Poincare Plots and traditional features from the 2D Poincare plot to distinguish AF versus NSR and other rhythms.

2. Data and Pre-processing

In this study, we have used single-lead ECG recordings (n= 8,525) with a sampling frequency of 300 Hz provided as a training set for Physionet/Computing in Cardiology Challenge 2017 [8, 9]. This database contains a single short ECG lead recording (between 30 s and 60 s in length) as normal sinus rhythm, atrial fibrillation, other rhythms, or noisy recording (NSR: 5,074, AF: 757, other rhythms: 2,415, and noise: 279).

Baseline wander removal was done using a moving average filter followed by QRS detection using the gqrs algorithm, publicly available in the WFDB toolbox [10]. Detected QRS peaks were used to create an RR time series.

3. Methods

This section introduces 2D and 3D Poincare plots, followed by extracting new geometric features based on the distribution of points in 3D space. These features are explained in detail in the following section. The extracted features are then utilized to distinguish AF versus NSR and other rhythms.

3.1. 2D and 3D Poincare Plot

Given a time series RR = (RR_1, RR_2, …, RR_n, RR_{n+1}), the standard 2D Poincare plot is constructed by locating points from the time series on the coordinate plane according to pairs (x_i, y_i) for i = 1, 2, 3, …, n:

$$x = (x_1, x_2, \ldots, x_n) = (RR_1, RR_2, \ldots, RR_n) \quad (1)$$

$$y = (y_1, y_2, \ldots, y_n) = (RR_2, RR_3, \ldots, RR_{n+1}) \quad (2)$$

Where *n* is the number of points in the Poincare plot, which is one less than the length of the RR time series [11]. Since some points in the 2D plot may overlap at specific coordinates, we considered the number of repeated points at specific coordinates as the third dimension in the 3D Poincare plot, as shown in Figure 1. So, the coordinates of points in the 3D Poincare plot are:

$$(RR_i, RR_{i+1}, the\ number\ of\ repetitions) \quad (3)$$

By analyzing the point's distribution in this new map, new geometric features have been introduced, explained in section 3.3.

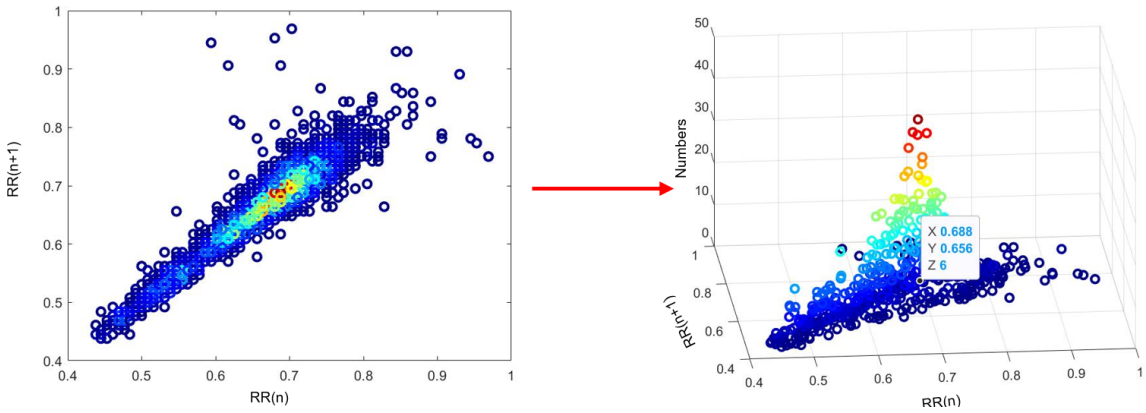

Figure 1. Construction of 3D Poincare Plot

3.2. Standard Descriptors of Poincare Plot

Two standard descriptors of the Poincare plot, *SD1* and *SD2,* are defined as the standard deviation of projection of the Poincare plot on the line perpendicular to the line of identity ($y = -x$) and ($y = x$), respectively [12]. So, we define them as:

$$SD1 = (Var((x-y) / (2)^{1/2}))^{1/2} \quad (4)$$

$$SD2 = (Var((x+y) / (2)^{1/2}))^{1/2} \quad (5)$$

3.3. Geometric Features in *3D Poincare Plot*

Analyzing the points distribution in a *3D Poincare plot*, a triangle is estimated (Figure 2), and its geometric properties are considered as extracted features in this mapping.

The first step in the geometrical analysis of a *3D Poincare Plot* is finding the coordination of three vertices of the triangle. There are multiple ways to measure the following geometric features of a triangle. But the methods we used in this paper are as follows:

The vertices *A* and *C* are the coordinates of points with the minimum and maximum distance to the line $y = -x$, respectively (xA, yA, xC, and yC). The vertice *B* is the point with a maximum number of repetitions in a *2D Poincare plot (xh and yh)*.

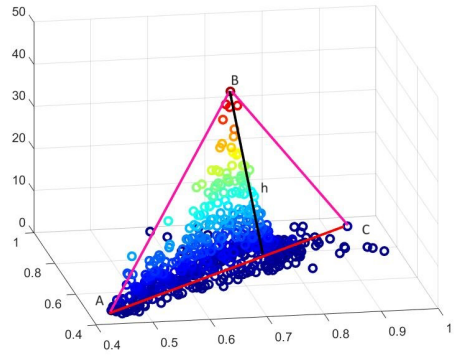

Figure 2. Estimation of a triangle based on the distribution of points in the 3D Poincare plot and its geometric properties.

After finding the coordinates of three vertices, the distance between two points was measured to find the length of the sides of the constructed triangle, as follows:

$$a = \sqrt{(x_B - x_C)^2 + (y_B - y_C)^2}$$
$$b = \sqrt{(x_A - x_C)^2 + (y_A - y_C)^2} \quad (6)$$
$$c = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

It is clear that the height of the triangle (h), as the next extracted feature, is equivalent to the maximum number of repeating points in the Poincare plot.

The perimeter of the triangle is defined by adding three sides of it:

$$Ptri = a + b + c \quad (7)$$

And the area of the triangle is considered as:

$$Area = \frac{1}{2}(h)(b) \quad (8)$$

In addition, as shown in Figure 3, by estimating a rectangle on the distribution of points in the *2D Poincare plot*, a pyramid-like shape with a rectangular base can be obtained in the *3D Poincare plot*, whose features are added

Table 1. Mean and standard deviation of extracted features across AF, NSR, and Other Rhythm

	Missing	Overall	AF	NSR	Other
n		8246	757	5074	2415
SD1, mean (SD)	0	0.1 (0.3)	0.2 (0.1)	0.1 (0.4)	0.2 (0.2)
SD2, mean (SD)	0	0.1 (0.3)	0.2 (0.1)	0.1 (0.3)	0.2 (0.3)
xh, mean (SD)	0	0.8 (0.3)	0.7 (0.3)	0.8 (0.3)	0.8 (0.4)
yh, mean (SD)	0	0.8 (0.3)	0.7 (0.3)	0.8 (0.2)	0.8 (0.3)
nh, mean (SD)	0	2 (1.9)	1.4 (1.2)	1.9(1.3)	2.4 (2.8)
b, mean (SD)	0	0.6 (0.9)	0.8 (0.4)	0.6 (1.0)	0.8 (0.9)
area, mean (SD)	0	0.5 (0.7)	0.5 (0.4)	0.5 (0.6)	0.7 (0.9)
xA, mean (SD)	0	0.5 (0.2)	0.4 (0.2)	0.6 (0.3)	0.5 (0.2)
yA, mean (SD)	0	0.8 (0.4)	0.7 (0.3)	0.8 (0.4)	0.8 (0.4)
xC, mean (SD)	0	1.1 (0.7)	1.2 (0.4)	1.1 (0.8)	1.2 (0.8)
yC, mean (SD)	0	0.9 (0.5)	0.7 (0.3)	0.9 (0.5)	0.8 (0.4)
a, mean (SD)	0	0.4 (0.8)	0.6 (0.4)	0.3 (0.9)	0.5 (0.8)
c, mean (SD)	0	0.3 (0.5)	0.3 (0.3)	0.3 (0.5)	0.4 (0.5)
Ptri, mean (SD)	0	1.4 (2.0)	1.6 (0.9)	1.2 (2.1)	1.7 (1.9)
xRect, mean (SD)	0	0.6 (1.0)	0.8 (0.4)	0.5 (1.1)	0.8 (1.0)
yRect, mean (SD)	0	0.6 (0.8)	0.7 (0.4)	0.5 (0.8)	0.7 (0.8)
Prect, mean (SD)	0	2.4 (3.5)	2.9 (1.6)	2.0 (3.7)	2.9 (3.4)
Srect, mean (SD)	0	1.1 (11.9)	0.7 (1.0)	1.0 (13.7)	1.2 (9.3)
V, mean (SD)	0	1.4 (12.1)	0.8 (1.3)	1.3 (13.8)	1.7 (9.7)

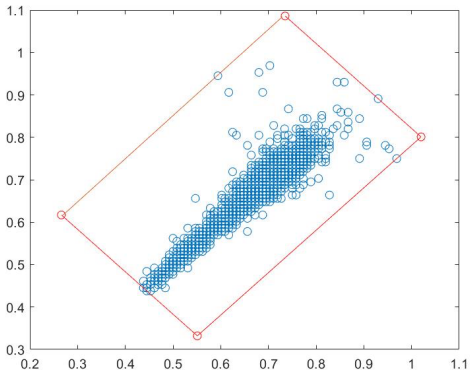

Figure 3. Estimation of a rectangle based on the distribution of points in *2D* Poincare plot

to the other extracted features. These features are the length of the rectangle (*xRect*), the width of the rectangle (*yRect*), the perimeter of the rectangle (*PRect*), and the area of the rectangle (*SRect*)

To find the length of the rectangle, we find the points that have the greatest distance from the top and bottom with the line $y = x$ and consider a line parallel to the same line with a slope of *1* that passes through each of these points. To find the width of the rectangle, we find the points that have the smallest and greatest distance from the line $y = -x$, and we consider a parallel line with a slope of *-1* that passes through each of these points. Then the perimeter and the area of the rectangle is defined as:

$$Prect = 2(xRect + yRect) \quad (9)$$

$$SRect = (xRect)(yRect) \quad (10)$$

The last obtained feature is the volume of estimated pyramid:

$$VRect = \frac{1}{3}(h)(SRect) \quad (11)$$

3.4. Model Development

The 80/20 stratified split was used to create train and test sets (training data=6,596 and test data=1,650). A 3-class classifier

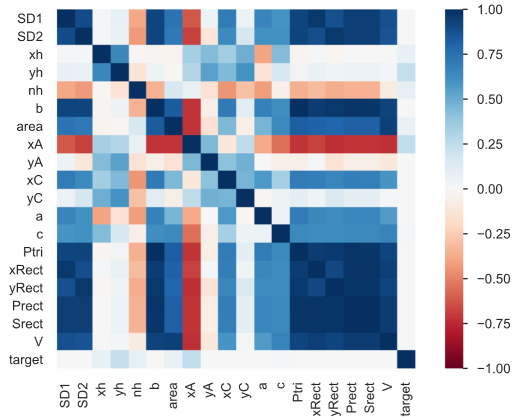

Figure 4. Pairwise correlation between extracted features

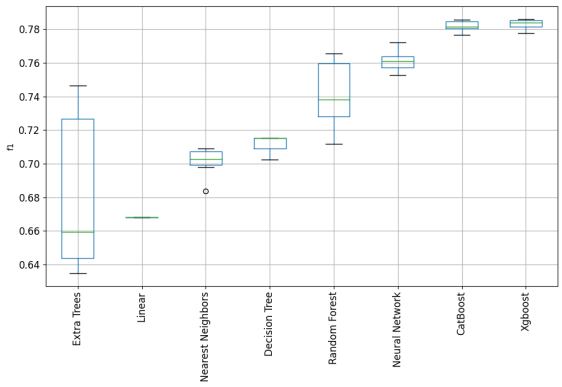

Figure 5. F1-score for different developed models using AutoML

(AF, NSR, and other rhythm) was developed using MLJAR [13], as noisy recordings were excluded from model development. MLJAR is AutoML python package find the best model using different algorithms including extra tree, linear, nearest neighbor, decision tree, random forest, neural network, Catboost, and Xgboost. Grid search was used by MLJAR to find the best hyper parameters leads to the best f1-score. Random shuffling of the data set was done before training.

Feature importance for each feature is computed using SHAP (Shapley Additive exPlanations) [14].

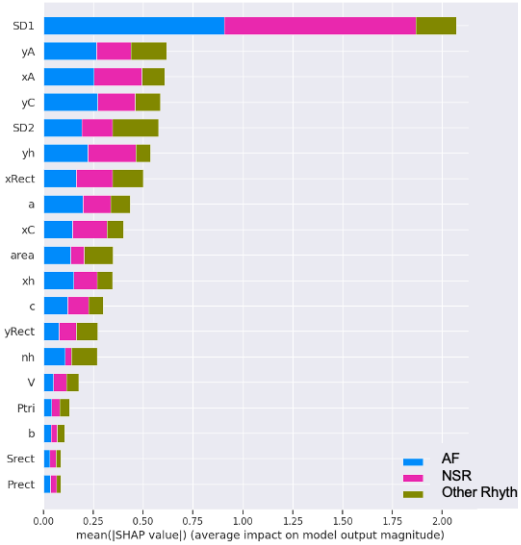

Figure 6. Computed feature importance using SHAP (Shapley Additive exPlanations)

3.5. Results

Nineteen traditional and geometric features were extracted from the 2D and 3D Poincare plots. The mean and standard deviation of these features across AF, NSR, and Other Rhythms are reported in Table 1. Furthermore, Figure 4 shows a pairwise correlation between features and a target variable. A positive high correlation was observed

between SD1, SD2, b, Ptri, xRect, yRect, Prect, Srect, and V. There was a high negative correlation of SD1 and xA as well as SD2 and xA.

Table 2. Performance metrics on the test set using the final Catboost model (AF: Atrial Fibrillation and NSR: Normal Sinus Rhythm)

	F1-score	Support
AF	0.60	152
NSR	0.87	1,015
Other Rhythm	0.63	483
Accuracy	0.78	1,650
Macro Avg	0.70	1,650
Weighted Avg	0.77	1,650

As shown in Figure 5, the Catboost and Xgboost got the best performance (highest f1-score) across different classifiers. Catboost was used as the final selected model. As shown in Figure 6, SD1, yA, xA, yC, and SD2 were the top five features for the selected classification.

A confusion matrix and performance metrics on the test set for Catboost are shown in Figure 7 and Table 2, respectively.

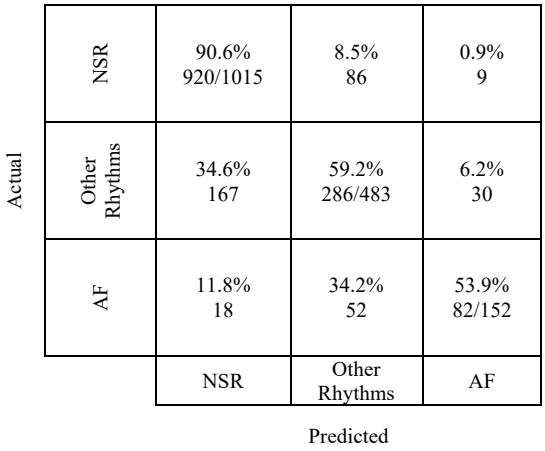

Figure 7. Confusion matrix for final Catboost model on the test set (AF: Atrial Fibrillation and NSR: Normal Sinus Rhythm)

3.6. Discussion

In this paper, traditional and enhanced Poincare plot descriptors are used for AF detection versus NSR and other rhythms. Results show the potential of extracted features for AF detection. However, enrichment of the feature set by adding other features, such as morphological features, is required to improve performance further.

For classification, AutoML is used to find the best model. AutoML facilitates finding the best model. However, as expected, a rich feature set is required to get the best AutoML results.

References

[1] R. G. Hart, and J. L. Halperin, "Atrial Fibrillation And Stroke: Concepts And Controversies," *Stroke,* vol. 32, no. 3, pp. 803-808, 2001.

[2] S. Parvaneh;, M. R. H. Golpaygani;, M. Firoozabadi;, and M. Haghjoo;, "Analysis Of ECG In Phase Space For The Prediction Of Spontaneous Atrial Fibrillation Termination," *Journal of Electrocardiology,* vol. 49 (6), pp. 936-937, 2016.

[3] R. Skaria, S. Parvaneh, S. Zhou, J. Kim, S. Wanjiru, G. Devers, J. Konhilas, and Z. Khalpey, "Path To Precision: Prevention Of Post-operative Atrial Fibrillation," *Journal of Thoracic Disease,* vol. 12, no. 5, pp. 2735, 2020.

[4] N. J. Dabanloo, S. Moharreri, S. Parvaneh, and A. M. Nasrabadi, "Application Of Novel Mapping For Heart Rate Phase Space And Its Role In Cardiac Arrhythmia Diagnosis," *Computing in Cardiology*, pp. 209-212, 2010.

[5] G. Hirsch, S. H. Jensen, E. S. Poulsen, and S. Puthusserypady, "Atrial Fibrillation Detection Using Heart Rate Variability And Atrial Activity: A Hybrid Approach," *Expert Systems with Applications,* vol. 169, pp. 114452, 2021.

[6] S. Rezaei, S. Moharreri, M. Abdollahpur, and S. Parvaneh, "Heart Arrhythmia Classification Using Extracted Features In Poincare Plot Of RR intervals." pp. 1-4.

[7] M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection Of Atrial Fibrillation In ECG Hand-held Devices Using A Random Forest Classifier." pp. 1-4.

[8] GD. Clifford, C. Liu, B. Moody, L-W Lehman, I. Silva, Q. Li, A. Johnson, and R.G. Mark, "AF Classification From A Short Single Lead ECG Recording: The PhysioNet/Computing In Cardiology Challenge 2017," *Computing in Cardiology*, 2017.

[9] A. L. Goldberger, L. A. Amaral, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, And PhysioNet: Components Of A New Research Resource For Complex Physiologic Signals," *Circulation,* vol. 101 (23), pp. 215-220, 2000.

[10] I. Silva, and G. B. Moody, "An Open-source Toolbox For Analysing And Processing Physionet Databases In Matlab And Octave," *Journal of open research software,* vol. 2, no. 1, 2014.

[11] 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, pp. 287, 2007.

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

[13] A. Płońska, and P. Płoński, "MLJAR: State-of-the-art Automated Machine Learning Framework For Tabular Data," *Version 0.10,* vol. 3, 2021.

[14] S. M. Lundberg, and S.-I. Lee, "A Unified Approach To Interpreting Model Predictions," *Advances in Neural Information Processing Systems,* vol. 30, 2017.

Address for correspondence:
Saman Parvaneh
1 Edwards Way, Irvine, CA, USA
parvaneh@ieee.org