



Rapid Detection of Heart Rate Fragmentation and Cardiac Arrhythmias: Cycle-by-Cycle rr Analysis, Supervised Machine Learning Model and Novel Insights

Ananya Rajagopalan¹(✉) and Marcus Vollmer²

¹ Redmond High School, Redmond, WA, USA
ananya.rajagopalan@gmail.com

² Institute of Bioinformatics, University Medicine Greifswald,
Greifswald, Germany
marcus.vollmer@uni-greifswald.de

Abstract. Heart rate dynamics are a macroscopic indicator of cardiac health. Sino-atrial degradation manifested as heart rate fragmentation (HRF) are analyzed using rr values (relative-RR intervals) derived from the inter-beat-intervals of ECGs. The rr-value is useful for the analysis of cycle-by-cycle variations such as HRF and arrhythmias. Three novel metrics developed in this work: CM20, Z3e20 and sPIP, along with two conventional metrics: SDNN and LFHF ratio are used for the detection of HRF and arrhythmias. The supervised machine learning technique of random forests is applied to develop the classification model. For this, we used a balanced dataset of 300 cases comprising of arrhythmic, non-arrhythmic coronary artery disease, and individuals without any medically significant cardiac conditions. The model was tested on 104 independent cases. The F1 score of the classifier is 91.1% without any adjustments for age, gender, prior medical conditions, etc. Insight into threshold values of heart rate dynamics for arrhythmic, heart rate fragmentation and normal cases are obtained from a single decision tree model.

Keywords: Heart rate fragmentation · Arrhythmia detection · Random forest · Machine learning

1 Introduction

Heart disease is responsible for 31% of deaths worldwide [1]. Thus, it is necessary to develop methods that accurately detect medically significant incipient cardiac conditions such as arrhythmias and heart rate fragmentation (HRF). HRF has been shown to be a dynamical biomarker of the neuroautonomic-electrophysiologic system breakdown: patients with greater HRF have been shown to be at an increased risk of coronary artery disease (CAD) [2]. HRF has been recently determined as an anomalous increase in heart rate due to short-term acceleration/deceleration [2, 10]; it has the potential to provide, falsely, values of heart rate variability (such as SDNN) that are deemed normal. Prior work [2] utilizes univariate predictors of HRF with limited discrimination; one

objective of this work was thus to improve the discrimination ability of HRF. In this work, heart rate fragmentation is synonymous with non-arrhythmic individuals who have been diagnosed as having CAD in the THEW database [4].

Work on the detection of arrhythmias outside of the clinical setting has shown tremendous progress in recent years, such as the 2017 Physionet/CinC Challenge [3] on single-lead atrial fibrillation detection. The work presented here seeks to improve upon existing research in the following ways: (1) accurate identification of heart rate fragmentation with high discrimination capability (2) separate classification of heart rate fragmentation, arrhythmias, and “normal” patterns (3) emphasis on the application of Random Forest (RF) for a computationally non-intensive methodology with straightforward implementation into battery-operated wearable devices.

2 Datasets and Methodology

2.1 Datasets

A total of 300 cases, representing a set of balanced class examples, was used for statistical analysis and ML model development: 90 cases of non-arrhythmic CAD from the University of Rochester THEW-project database [4], 115 examples of normal heart rate rhythm data from THEW-project database [4] and MIT-BIH nsr2db (normal sinus rhythm) database [5], and 95 examples of arrhythmic cases from MIT-BIH mitdb and afdb [5]. To test the model, 104 (30 non-arrhythmic CAD, 30 normal, 44 arrhythmic) new cases from the same databases were used [4, 5].

2.2 Data Visualization

In this work, rr (moving average window normalized RR intervals, [6] called relative RR intervals) is used, as it overcomes limitations of RR [7, 8]. The novelty of the usage of these rr -values for this work comes from the application to HRF and non-conduction type arrhythmia detection.

Data visualization of successive rr -values in a scatter plot [6] has yielded interesting insights: most cases of arrhythmias exhibited rr excursions $>20\%$. Thus, the “Zones of Cardiac Activity” concept was established: It was observed that individuals having an $|rr| < 5\%$ (Zone 1) most of the time is a warning of arrhythmia especially if the SDNN value is low (<75 ms); most healthy individuals usually have $5\% < |rr| < 20\%$ (Zone 2); rr -variations greater than 20% existing for $>1\%$ of the time (Zone 3) indicative of arrhythmias. Non-arrhythmic individuals with HRF from the THEW-Coronary Artery Disease datasets do not exhibit any discernible excursion signature on the rr scatter plot.

Table 1 shows the group statistics of the three patient classes and corresponding metrics summary represented by the median and interquartile range. To investigate the group specific distribution of the parameters, Random Forest was applied as the suitable classification technique (Fig. 1).

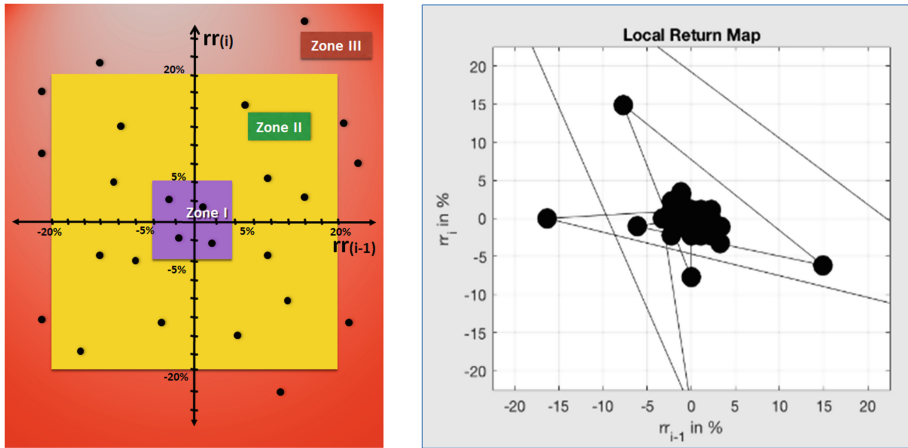


Fig. 1. Concept of zones of heart rate dynamics, rr-values in contour map (left); 60 s contour map of arrhythmic heart rate dynamics

Table 1. Group statistics (median and IQR) of training data

Class	Count	CM20	Z3e20	SDNN	sPIP	LFHF
Arrhythmic	95	1.91 [0.46,9.26]	5.90 [1.96,18.7]	133.17 [97,169]	60.16 [55,67]	0.50 [0.36,0.68]
CAD	90	0.06 [5e-3,0.16]	0.31 [0.09,0.58]	111.19 [86,129]	56.84 [52,62]	0.82 [0.61,1.03]
Normal	115	2.1e-3 [8e-4,0.11]	0.14 [0.04,0.38]	142.24 [115,168]	47.78 [45,50]	1.22 [0.93,1.33]

2.3 Methodology and Features Used for Machine Learning Model

- I. Analysis of datasets to determine critical rr-derived metrics that characterize arrhythmias; this step implies arrhythmia detection as detectable RR interval fluctuations (such as sinus arrhythmia, sinus tachycardia, sinus bradycardia, etc.).
- II. Extraction of features from the datasets (arrhythmia, THEW-cardiac and THEW healthy). In addition to conventional metrics such as SDNN and pNN50, the magnitude of cycle-to-cycle rr variation and consecutiveness of this magnitude was also analyzed.
- III. Elimination of statistically insignificant/redundant features using Spearman rank correlation through a pairwise analysis of the features.

Statistical analysis of data (Spearman rank correlation) revealed the following features as the five best representatives of cycle-by-cycle heart rate dynamics:

1. Z3e20: number of excursions into the 20% zone in the rr-contour map
2. CM20: Percentage of consecutive changes in opposite directions (acceleration ↔ deceleration), >20% in the rr scatter plot
3. sPIP: percentage inflection points of rr (positive ↔ negative value changes), accounting for quantization error in measurements
4. LFHF ratio and SDNN: as per conventional definitions [9]

IV Separation of the pre-classified datasets (arrhythmic, CAD, normal) into distinct non-duplicate training and testing data subsets. Analysis of the group distributions (i.e. median) for each of the features was performed to gain insight.

V. Development of random forest machine learning model: 300 datasets representing the three patient groups were used to train the models. 104 new datasets (see Sect. 2.1) were used to test the machine learning model (Fig. 2).

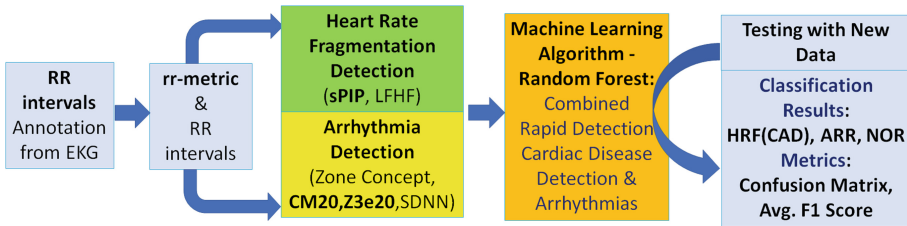


Fig. 2. Workflow methodology for HRF and arrhythmia detection

The combined incidence of HRF and arrhythmia in American adults is ~10% [12, 13]; training the detection model with ~90% “normal” patients for the sake of simulating real world scenarios is, thus, not desirable over the balanced datasets used in this work and hence the above balance of classes was chosen.

3 Results and Insights

A 30-tree ensemble of a random forest implementation [11] showed out-of-bag error leveling and was chosen. From Table 2, the F1_ARR score is 0.967, the F1_CAD score is 0.871 and F1_NOR score = 0.900; the overall F1 score is 0.911.

Table 2. Classification results of random forest classifier

Class	Actual ARR	Actual CAD	Actual NOR
Pred ARR	44	3 (FP)	0
Pred CAD	0	27	5 (FP)
Pred NOR	0	0	25 (TN)

Using a single decision tree algorithm, insights into thresholds for the various classes were identified: ARR cases were characterized by $Z_{3e20} > 1.1\%$; NOR and CAD cases were separated by a sPIP threshold of 54% and LFHF of ~ 1.00 .

4 Limitations and Future Work

The limitations of this work are analyzed by reviewing the mis-classifications among the results (Table 2). The classifier has a bias towards false positive results: 5 NOR cases are wrongly classified as CAD (sPIP $> 54\%$ and LFHF tending towards 1.00 in these cases); 3 CAD cases are wrongly classified as ARR – these cases have a $Z_{3e20} > 1.1\%$, sPIP $> 54\%$ and LFHF < 1.00 . The wrongly classified cases while justified by thresholds generated by the decision tree, are medically classified differently. Additional morphological features from the EKG might be required to decrease the false positive bias of the classifier.

References

1. Leading Causes of Deaths, Centers for Disease Control and Prevention (CDC). <https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>
2. Costa, M.D., Redline, S., Davis, R.B., Heckbert, S.R., Soliman, E.Z., Goldberger, A.L.: Heart rate fragmentation as a novel biomarker of adverse cardiovascular events: the multi-ethnic study of atherosclerosis. *Front. Physiol.* (2018). <https://doi.org/10.3389/fphys.2018.01117>
3. PhysioNet/CinC Challenge (2017). <https://physionet.org/challenge/2017/>
4. Thew-project, University of Rochester, Coronary Artery Patients database – <http://thew-project.org/Database/E-HOL-03-0271-002.html>, Healthy Individuals database – <http://thew-project.org/Database/E-HOL-03-0202-003.html>
5. Goldberger, A.L., et al.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23), e215–e220 (2000). <https://doi.org/10.1161/01.CIR.101.23.e215>
6. Vollmer, M.: Arrhythmia classification in long-term data using relative RR intervals, Computing in Cardiology (CinC), September 2017. <http://www.cinc.org/archives/2017/pdf/213-185.pdf>
7. Vollmer, M.: Ph.D. Dissertation, p. 63, Sect. 2.4.2. <https://d-nb.info/1124413723/34>
8. Lin, C.C., Yang, C.-M.: Heartbeat classification using normalized rr intervals and morphological features. *Math. Probl. Eng.* **2014**. <http://dx.doi.org/10.1155/2014/712474>
9. Schaffer, F., Ginsberg, J.P.: An overview of heart rate variability metrics and norms. *Front. Public Health* **5**, 258 (2017). <https://www.frontiersin.org/articles/10.3389/fpubh.2017.00258/full>
10. Costa, M.D., Davis, R.B., Goldberger, A.L.: Heart rate fragmentation: a new approach to the analysis of cardiac interbeat interval dynamics. *Front. Physiol.* **8**, 255 (2017). <https://doi.org/10.3389/fphys.2017.00255>
11. Mathworks: Matlab ver. 2018a, Random Forest Tree Bagger algorithm
12. Scripps Health. https://www.scripps.org/sparkle-assets/documents/heart_rhythm_facts.pdf
13. Sanchis-Gomar, F., Perez-Quillis, C., Leischik, R., Lucia, A.: Epidemiology of coronary heart disease and acute coronary syndrome. *Ann. Transl. Med.* **4**(13), 246 (2016). <https://doi.org/10.21037/atm.2016.06.33>