Rapid Detection of Heart Rate Fragmentation and Cardiac Arrhythmias:

Cycle-by-Cycle rr Analysis,
Supervised Machine Learning Model,
and Novel Insights

Ananya Rajagopalan

Student, Redmond High School, Redmond, WA 98052, USA

Dr. Marcus Vollmer

Institute of Bioinformatics, University Medicine Greifswald,
University of Greifswald,
Greifswald, Germany

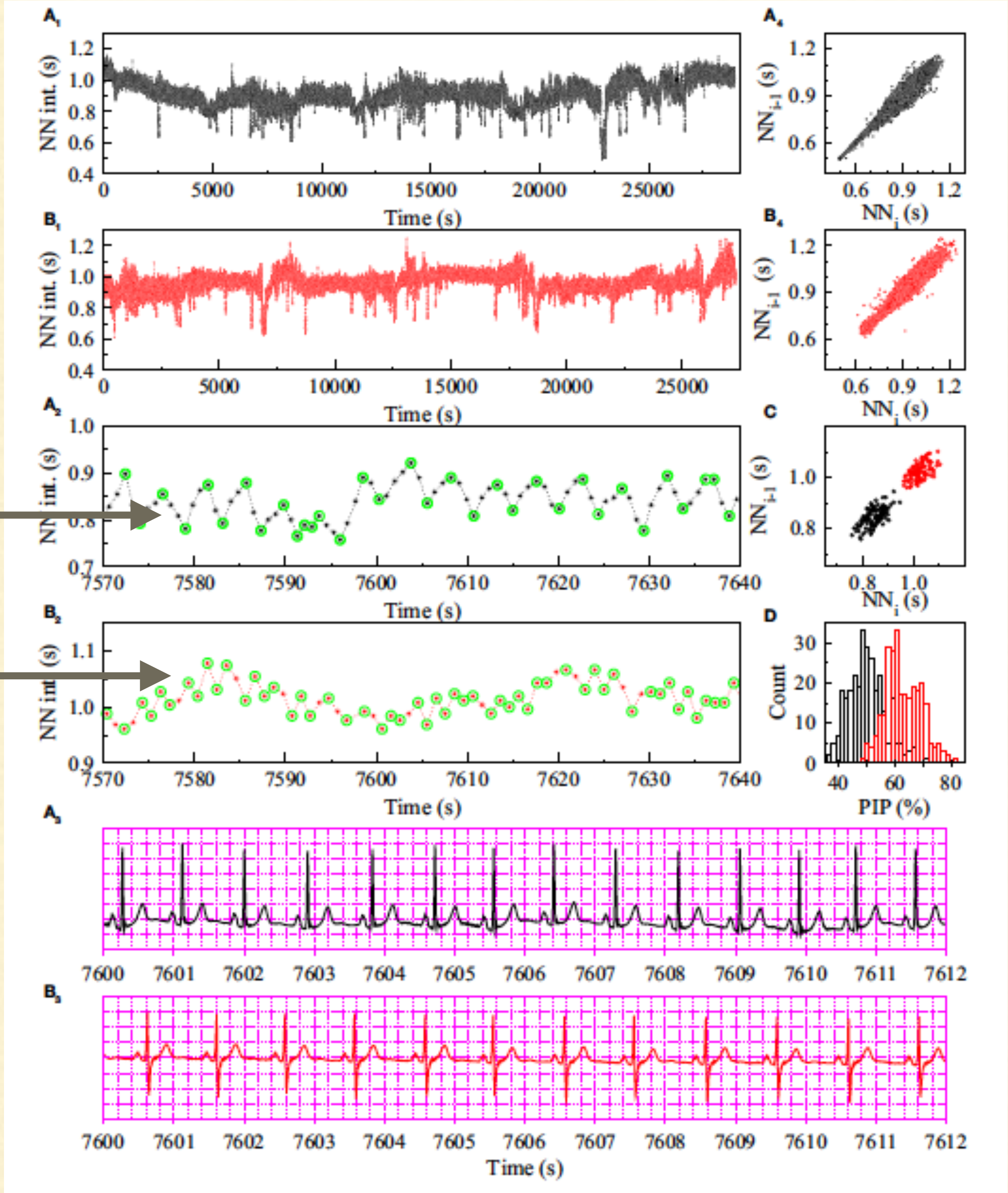
Presentation at
AIME 2019 Poster Session
Poznan, Poland

Outline

- Background & Motivation: Heart Rate Fragmentation
- Datasets
- Analysis Method: Feature Selection & Random Forest
- Results & Analysis
- Summary & Outlook

What is Heart Rate Fragmentation (HRF)?

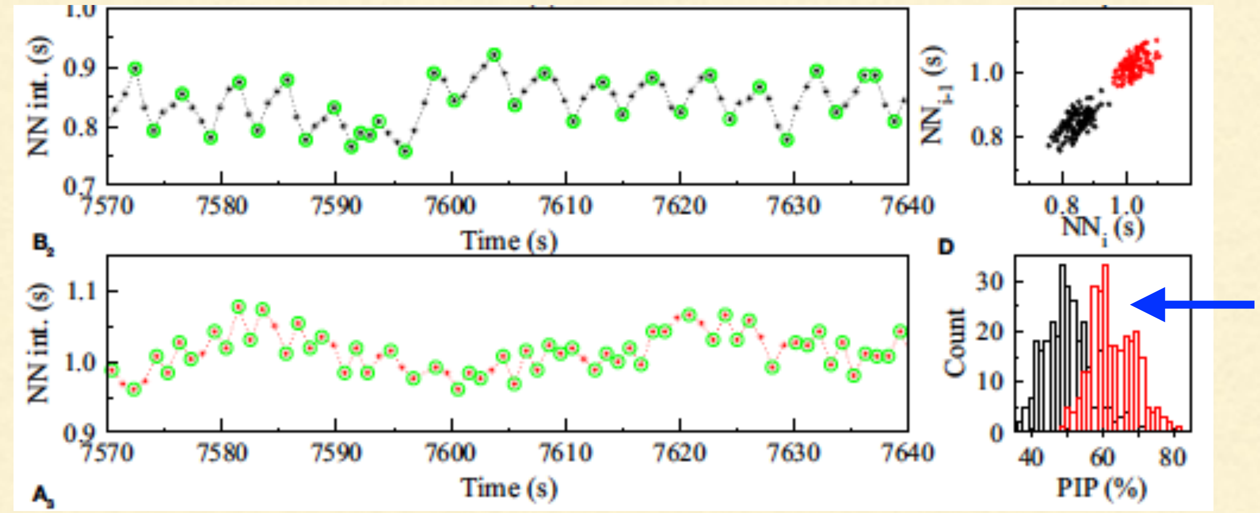
- Increased short-term HRV with frequent changes in heart-rate acceleration [Ref. 1]
- Proposed in [Ref. 1,2,3] for:
  - discrimination of ostensibly health subjects from patients with coronary artery disease
- HRF indices added significant value to Framingham and MESA CV indices [Ref.3]
  - HRF increased with patients age in both healthy and population with CAD
  - Fragmentation was higher in patients with CAD than healthy subjects
  - Traditional HRV metrics did not discriminate between the 2 groups



[Ref. 3]

Using HRF Metrics to discrminate between Healthy and Cardiac patients

- In work by Costa et. al. individual metrics were applied
  - PIP, IALS, PSS, PAS (new)
  - AVNN, SDNNIDX, rMSSD, pNN50, HF, LF/HF (traditional)
  - DFA (traditional non-linear)
- How can we improve discrimination between healthy vs. CVD patients?
  - Cycle-by-Cycle analysis
  - Normalized metrics applied
  - Classifier system based on suitable combination-of-metrics



[Ref. 3]

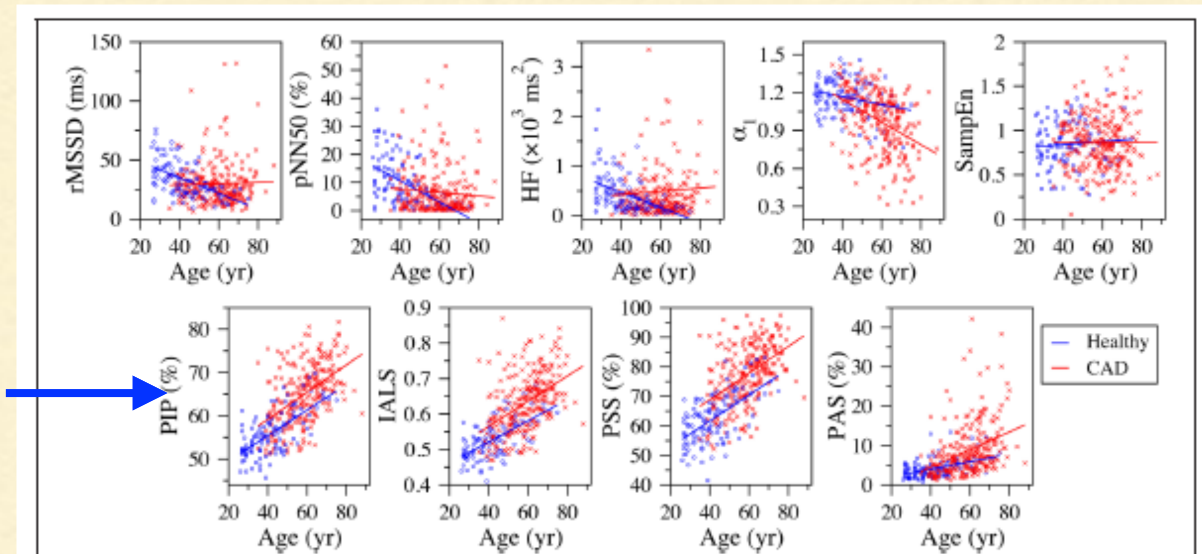


FIGURE 2 | Scatter plots of the traditional heart rate variability (rMSSD, pNN50, and HF), nonlinear (α_1 and SampEn) and fragmentation (PIP, IALS, PPS, and PAS) indices vs. the participants' age for the group of healthy subjects and patients with coronary artery disease (CAD), derived from the analysis of the full (~ 24-h) period. The solid lines are the linear regression lines. rMSSD, root mean square of the successive differences; pNN50, percentage of differences between successive NN intervals above 50 ms; HF, high frequency spectral power; α_1, detrended fluctuation analysis short-term exponent; SampEn, sample entropy; PIP, percentage of inflection points; IALS, inverse of the average length of the acceleration/deceleration segments; PPS, percentage of NN intervals in short segments; PAS, percentage of NN intervals in alternation segments.

[Ref. 1]

Datasets used in this work

Databases from University of Rochester, Telemetric and Holter ECG Warehouse ([4], used in Ref. 2)

1. **Healthy Cardiac Non-Arrhythmic datasets**:
    - E-HOL-03-0202-003 database
    - Healthy individuals, no overt cardiovascular disease, no history of BP>150/90, no medication, no chronic illness.
2. **Non-Arrhythmic Coronary Artery Patients**:
    - E-HOL-03-0271-002 database
    - **All patients are in normal sinus rhythm, without any evidence of congestive heart failure**
    - **Early stages of Atherosclerosis, Coronary Artery Diesease, myocardial infarction, etc.**

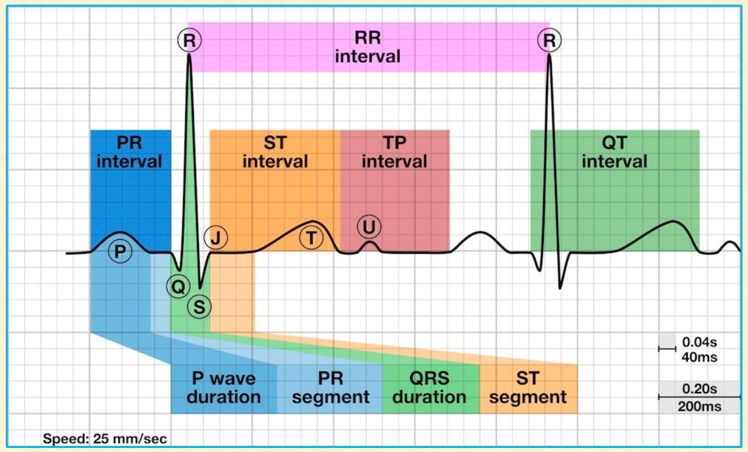
Devices in real-world would need to classifies Arrhythmic cases suitably [5]

3. **MIT-BIH Arrhythmia Database** mitdb
4. **MIT-BIH Atrial Fibrillation Database** afib, afib2
5. **MIT-BIH Normal Sinus Rhythm Database** nsr2db

For all the ECG readings
- Entire length of ECG recordings used (no segmentation into short recordings)
- ECG recordings with noise or artifacts (as per annotations) were discarded

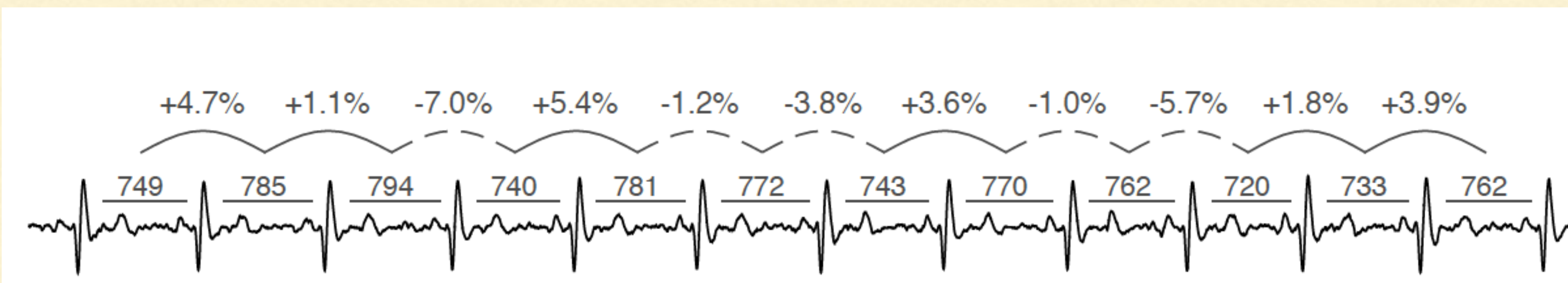
Basic Calculation Unit: rr-Metric [6]

RR intervals → rr-metric



Relative RR intervals rr_i are defined as a **normalized** difference of RR

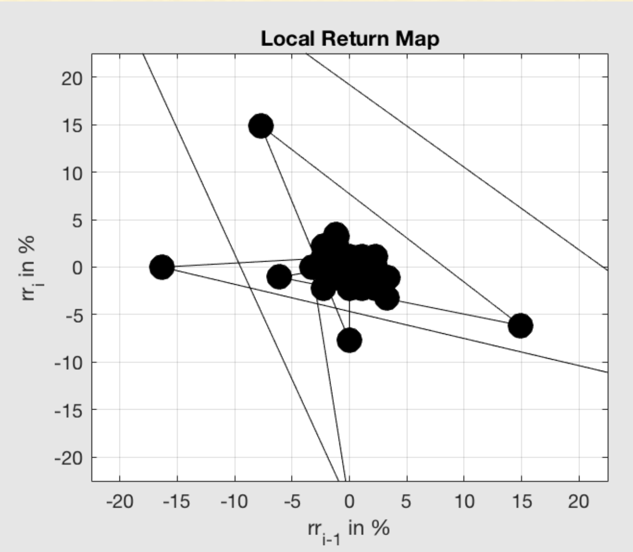
$$rr_i = \frac{RR_i - RR_{i-1}}{\frac{1}{2}(RR_i + RR_{i-1})}, i \in \{2, \ldots, n\}$$



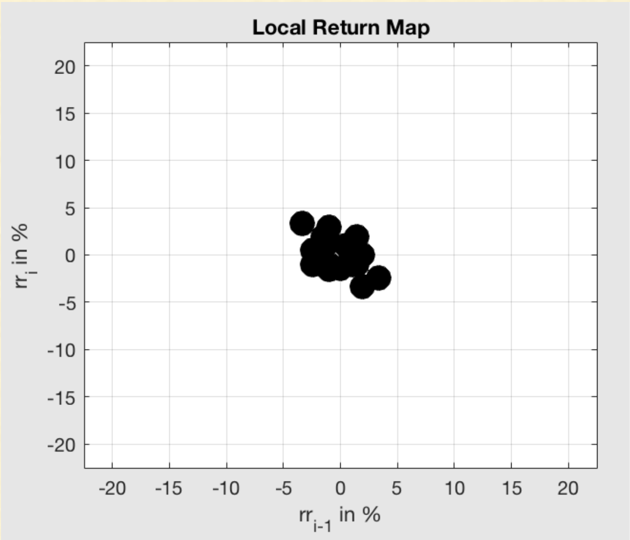
Why normalized RR intervals?

[Ref.6] "…The study results demonstrate that the use of the normalized RR interval features **greatly improves the positive predictive accuracy** of identifying the normal heartbeats and the **sensitivity for identifying the supraventricular ectopic heartbeats in comparison with the use of the nonnormalized RR interval feature**"

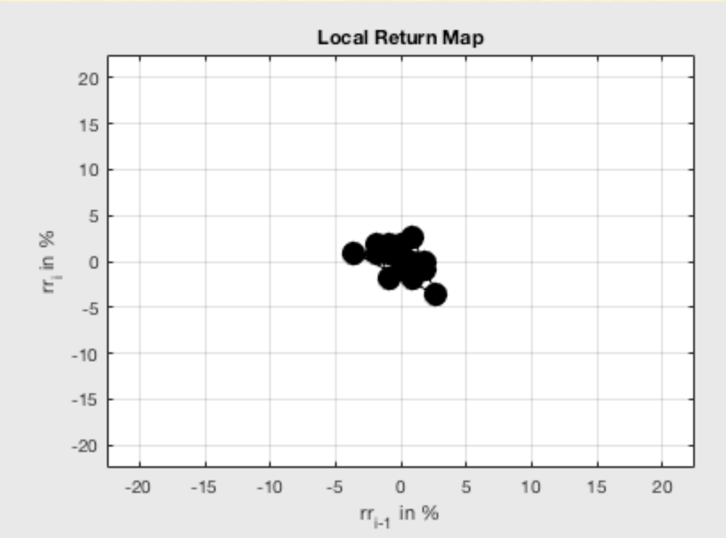
Zone Concept using scatter map of Relative-RR rr_i



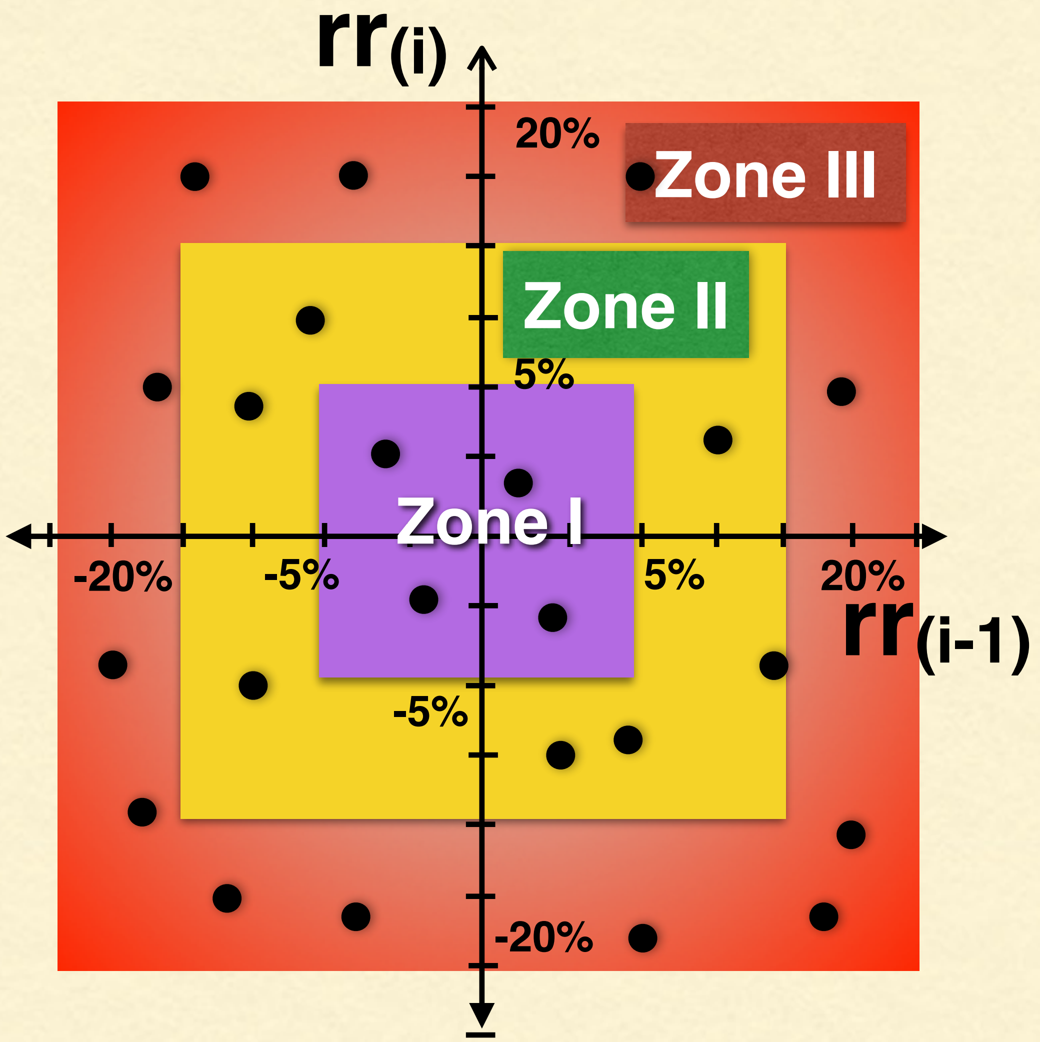
ARR (Amplitude variation large)



HRF: No discernable amplitude signature



NSR (Amplitude variation small)



Feature Selection for HRF and Arrhythmia Detection

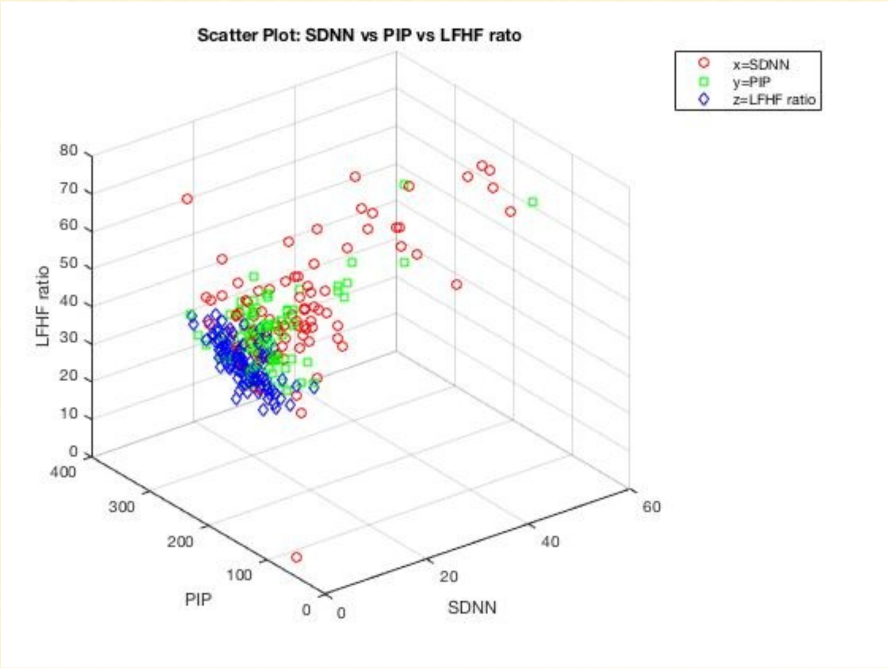
SDRR	pNN50	Triangular Index (TINN)	rr Histogram	2%, 5%, 20% rr-excursions	SD1, SD2	CM5, CM10, CM20	Z3e20	sPIP	LFHF ratio

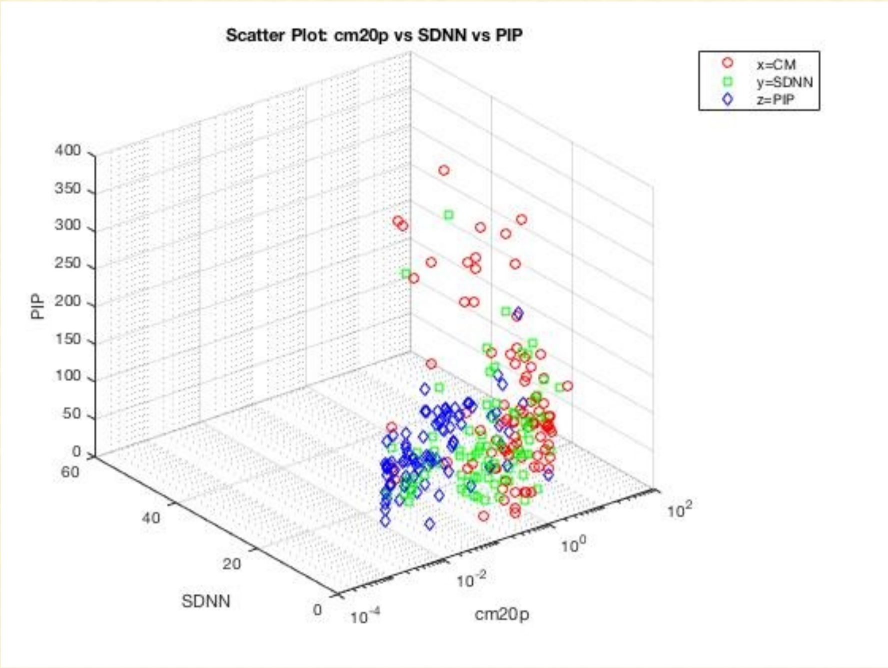
Spearman Ranked Correlation Analysis

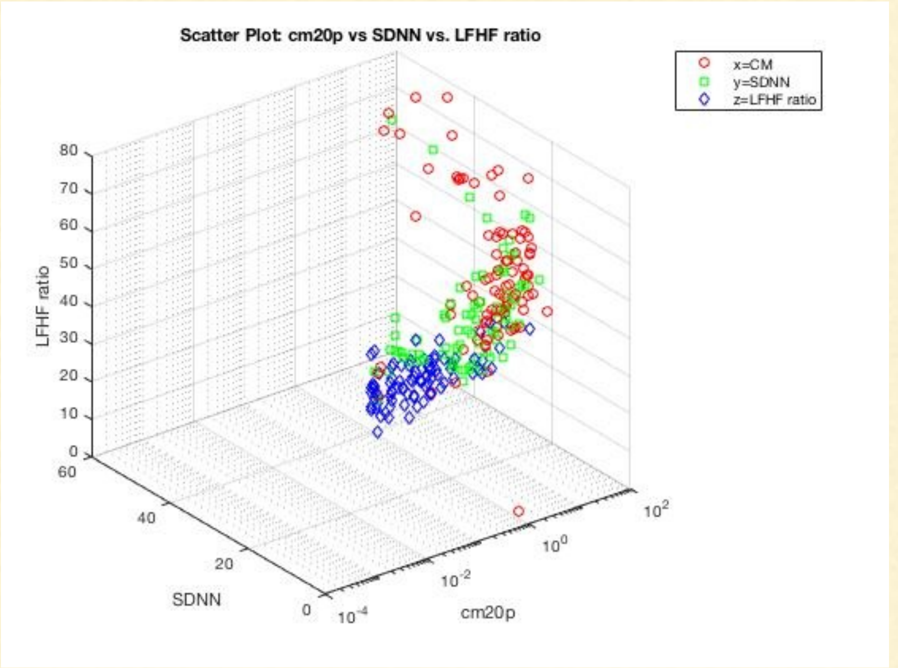
SDRR	20% rr-excursions	CM20	Z3e20	sPIP	LFHF ratio

Definition of Features used:
CM20, Z3e20, sPIP are newly defined and use the mathematically robust rr metric
CM20: % of Consecutive changes in opposite direction (acceleration ↔ deceleration), >20%, rr-contour map
Z3e20: number of excursions into the 20% zone in the rr contour map
sPIP: Smoothed Percentage Inflexion Points of rr (positive ↔ negative value changes)
SDNN (of RR intervals) and **LFHF** ratio are used as per the standard definitions

Data Distributions: Normal - HRF - Arrhythmias



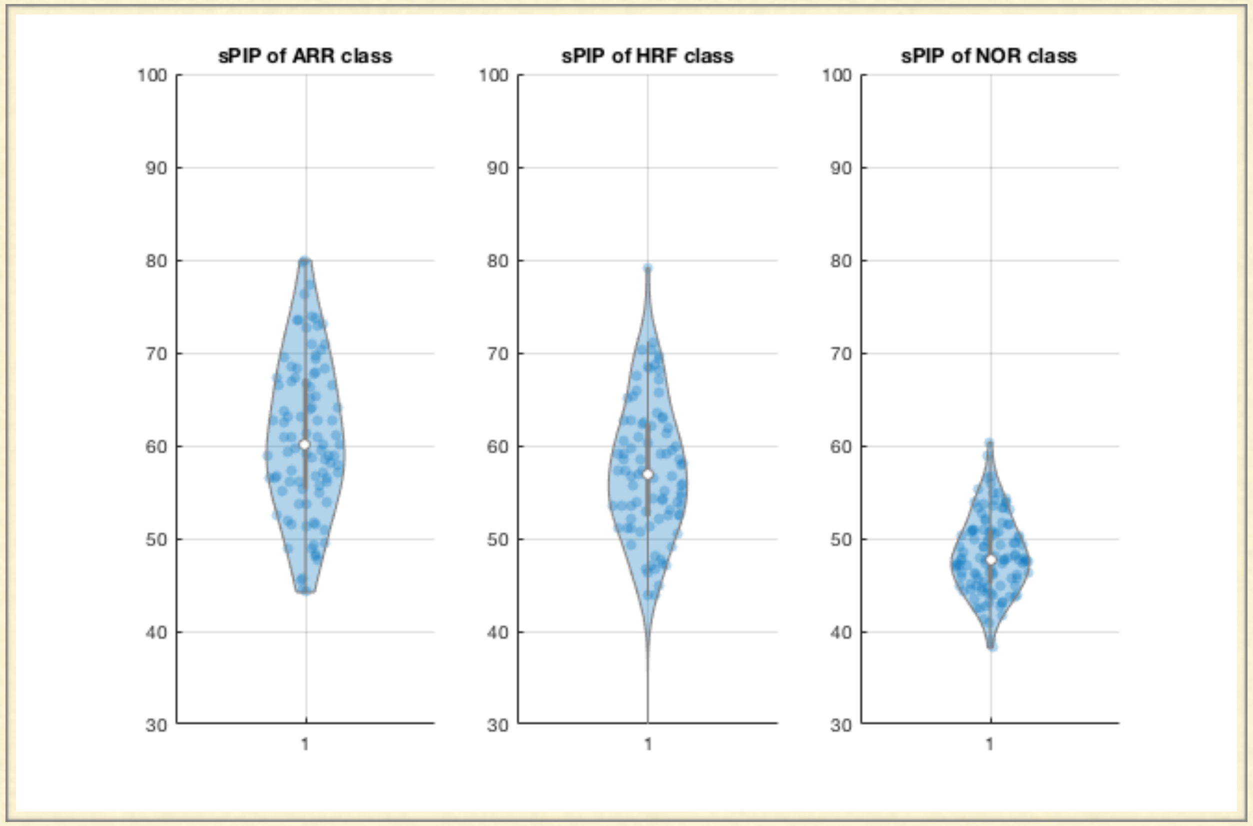


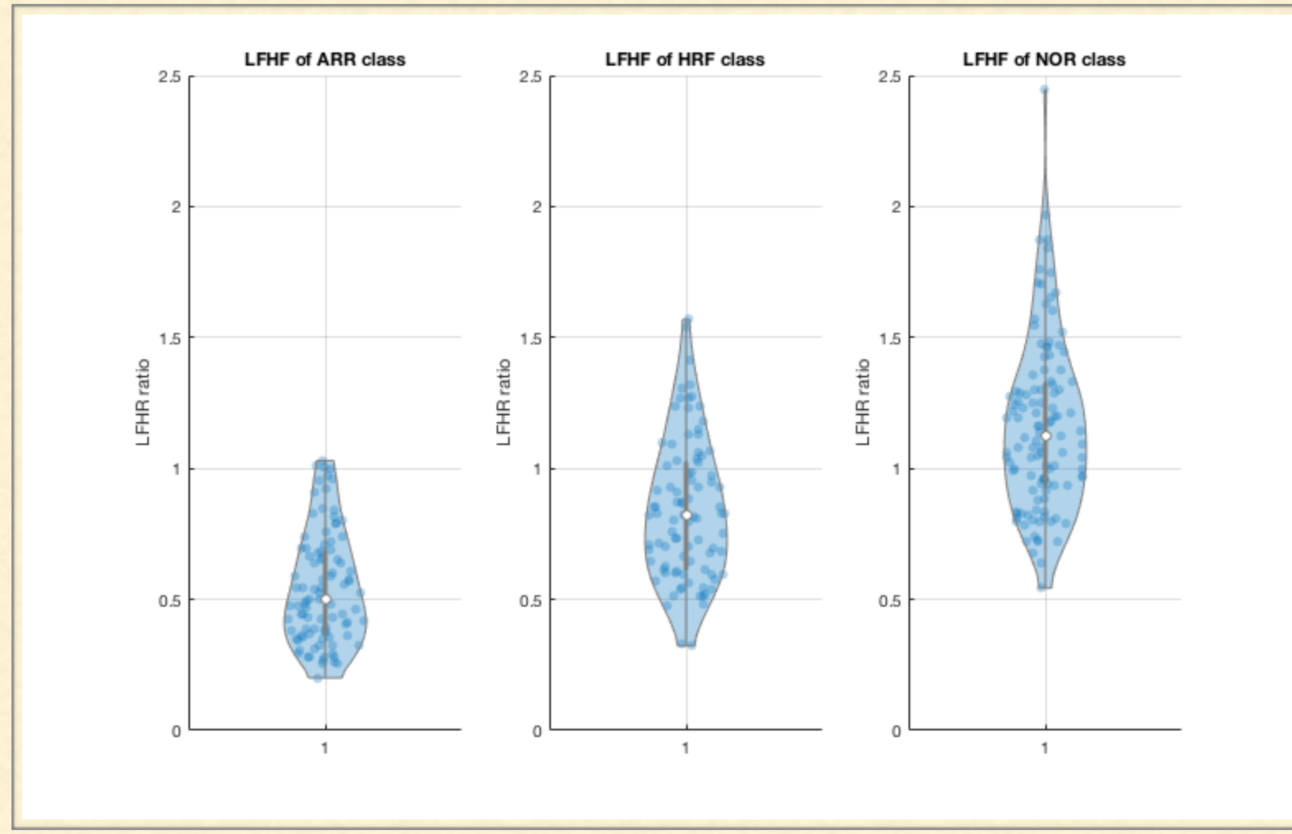


Group Statistics: Median & IQR of Data

Class	Count	cm20p	z3e20	SDRR	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 [0.46, 0.68]
Heart Rate Fragmented	90	0.06 [5.1e-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]

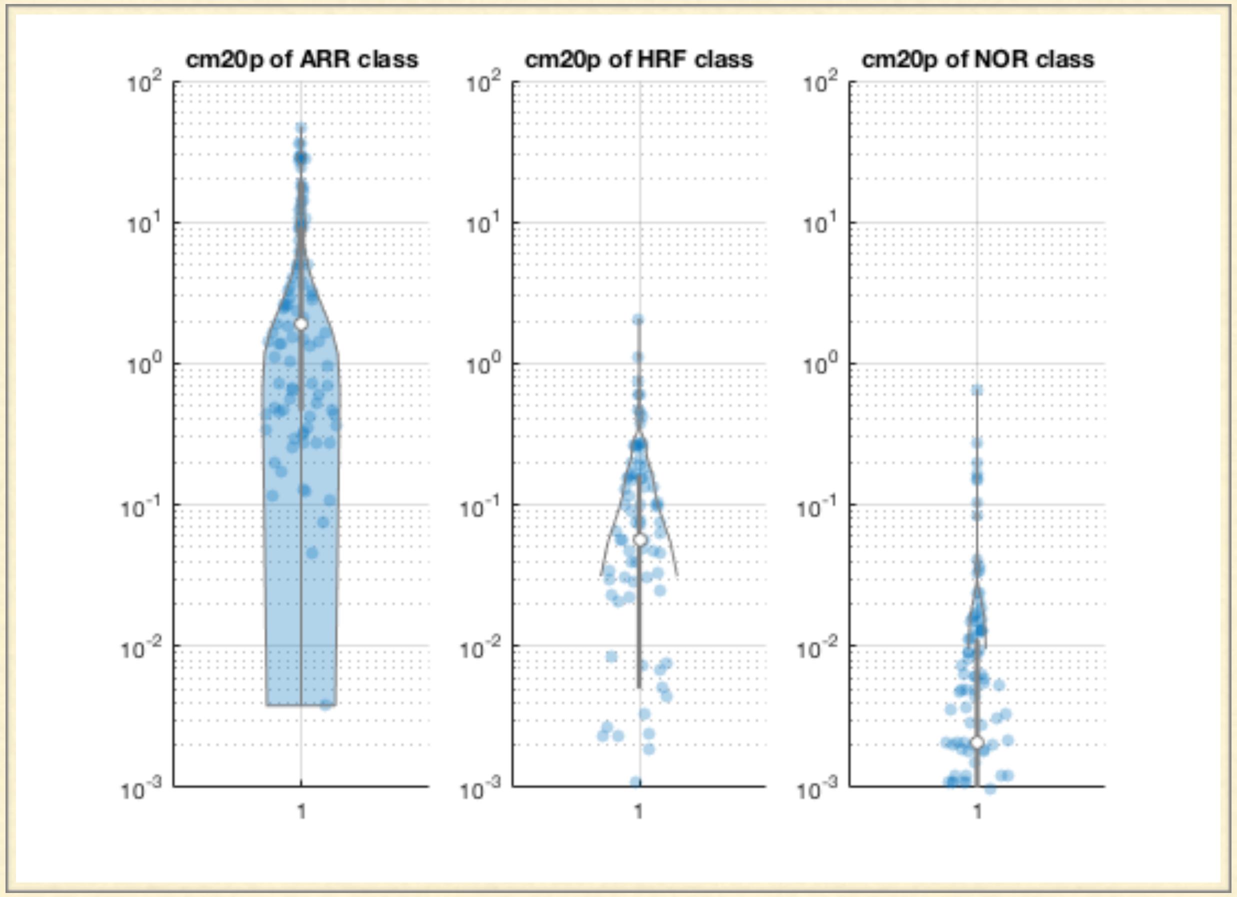
Violin Plots: LFHF and sPIP

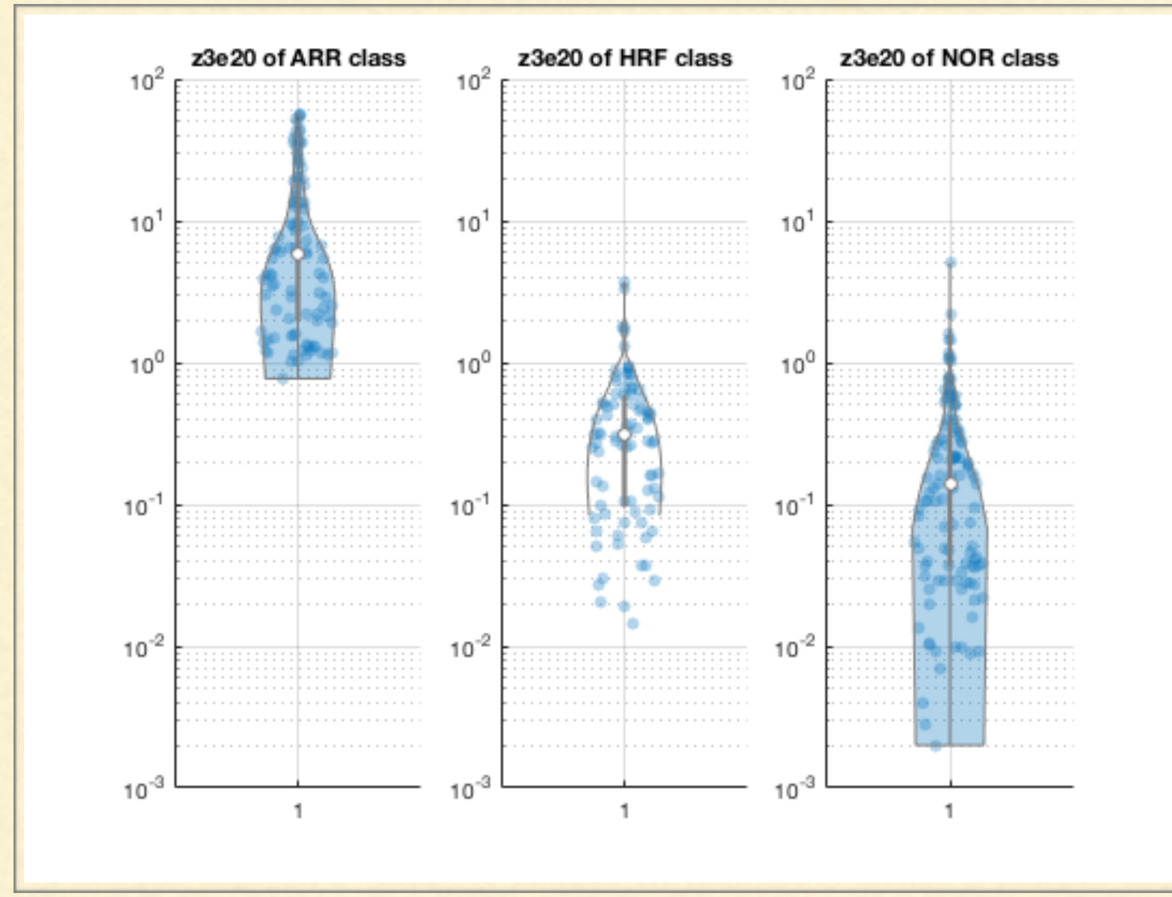




Class	Count	cm20p	z3e20	SDRR	sPIP	LFHF
Arrhythmic	95	1.91 [0.46, 9.26]	5.90 [1.96, 18.7]	133.17 [97, 169]	60.16 [55, 67]	0.50 [0.46, 0.68]
Heart Rate Fragmented	90	0.06 [5.1e-3, 0.16]	0.31 [0.09, 0.58]	111.19 [86, 129]	56.84 [52, 62]	0.82 [0.61, 1.03]
Normal	115	2.1e-3 [8e-4, 0.11]	0.14 [0.04, 0.38]	142.24 [115, 168]	47.78 [45, 50]	1.22 [0.93, 1.33]

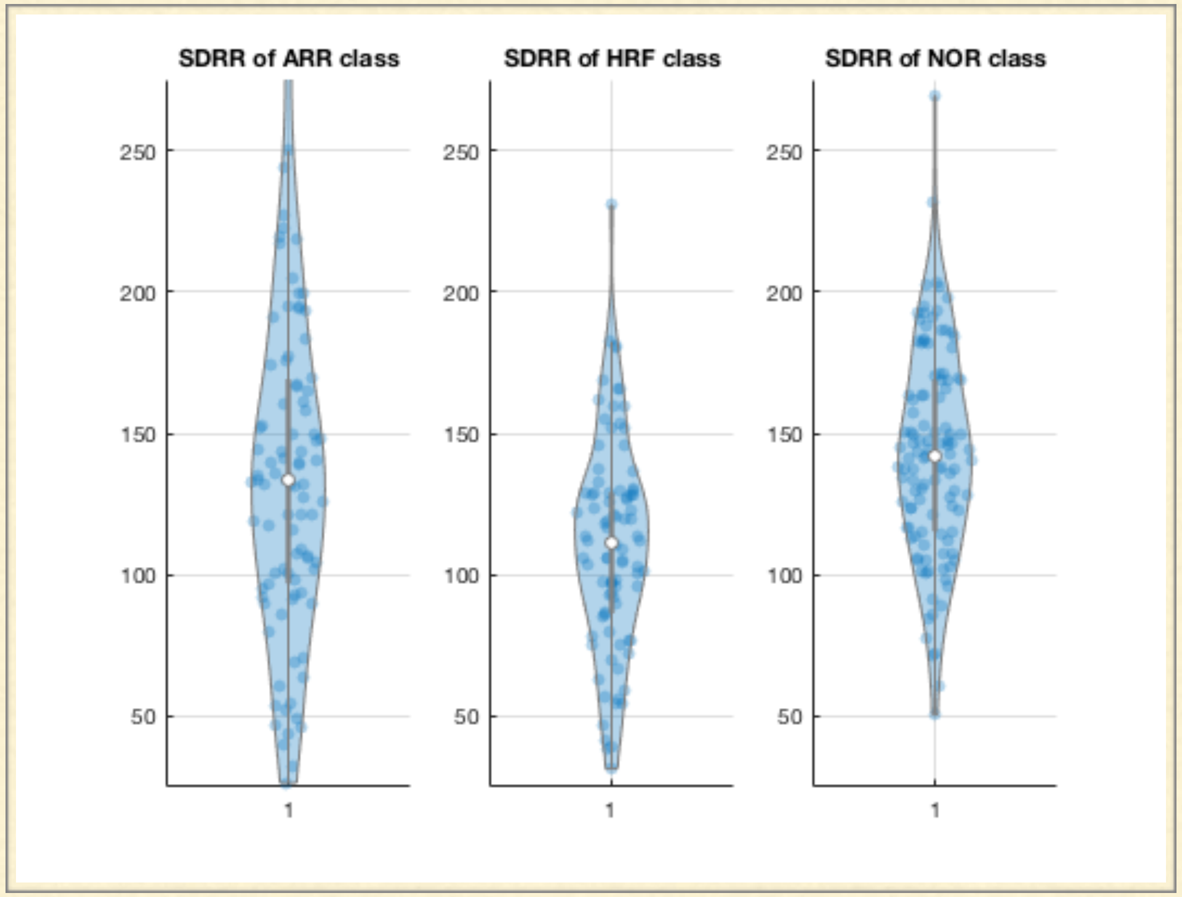
Violin Plots: cm20p and z3e20





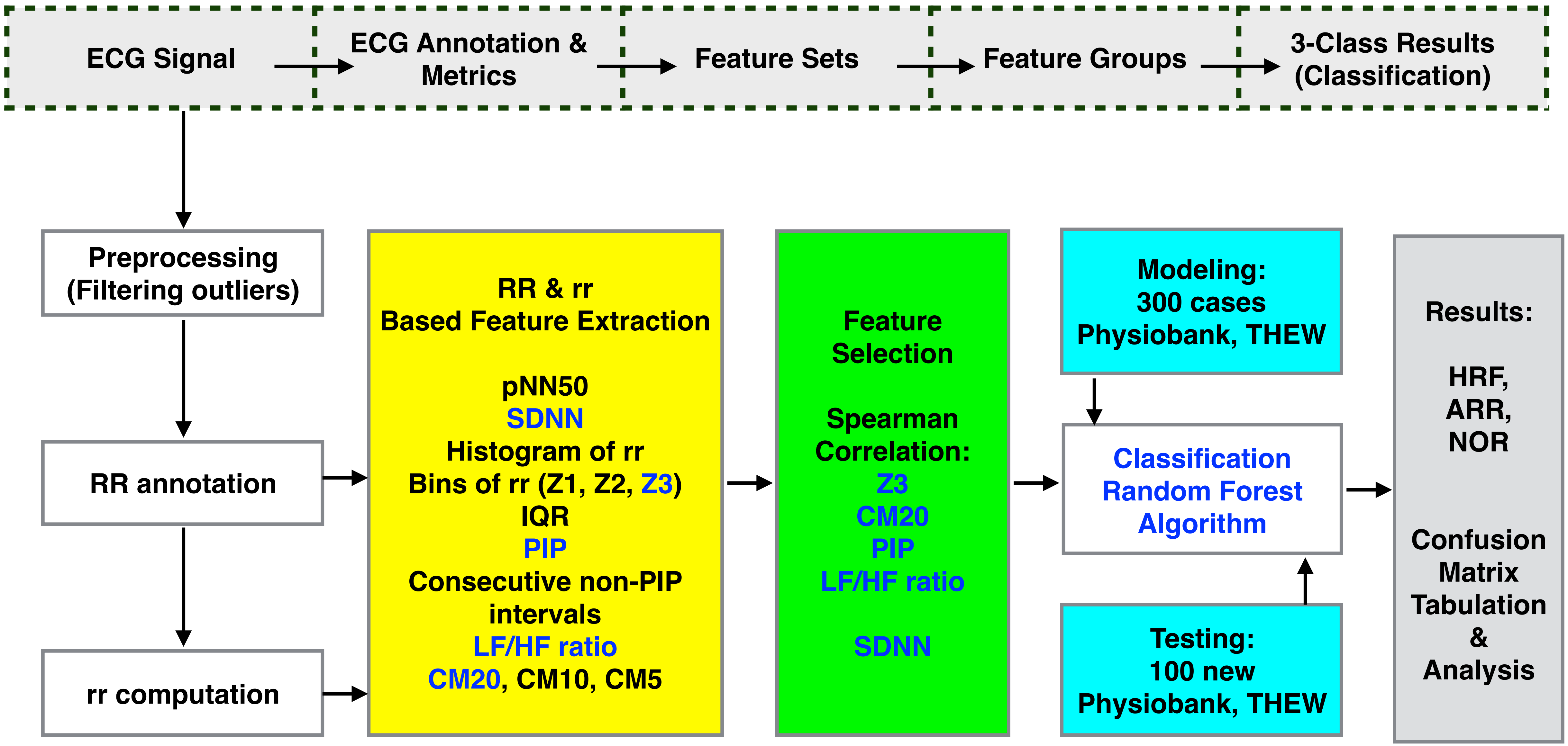
Class	Count	cm20p	z3e20	SDRR	sPIP	LFHF
Arrhythmic	95	1.91 [0.46, 9.26]	5.90 [1.96, 18.7]	133.17 [97, 169]	60.16 [55, 67]	0.50 [0.46, 0.68]
Heart Rate Fragmented	90	0.06 [5.1e-3, 0.16]	0.31 [0.09, 0.58]	111.19 [86, 129]	56.84 [52, 62]	0.82 [0.61, 1.03]
Normal	115	2.1e-3 [8e-4, 0.11]	0.14 [0.04, 0.38]	142.24 [115, 168]	47.78 [45, 50]	1.22 [0.93, 1.33]

Violin Plot: SDRR



Class	Count	cm20p	z3e20	SDRR	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 [0.46, 0.68]
Heart Rate Fragmented	90	0.06 [5.1e-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]

Workflow methodology: Normal - HRF - Arrhythmia Classification



Workflow implemented in HRVTool [Ref. 7], Excel & Matlab

Results of Random Forest Algorithm

Random Forest:
30 tree ensemble
(Out of Bag Error Minimization)

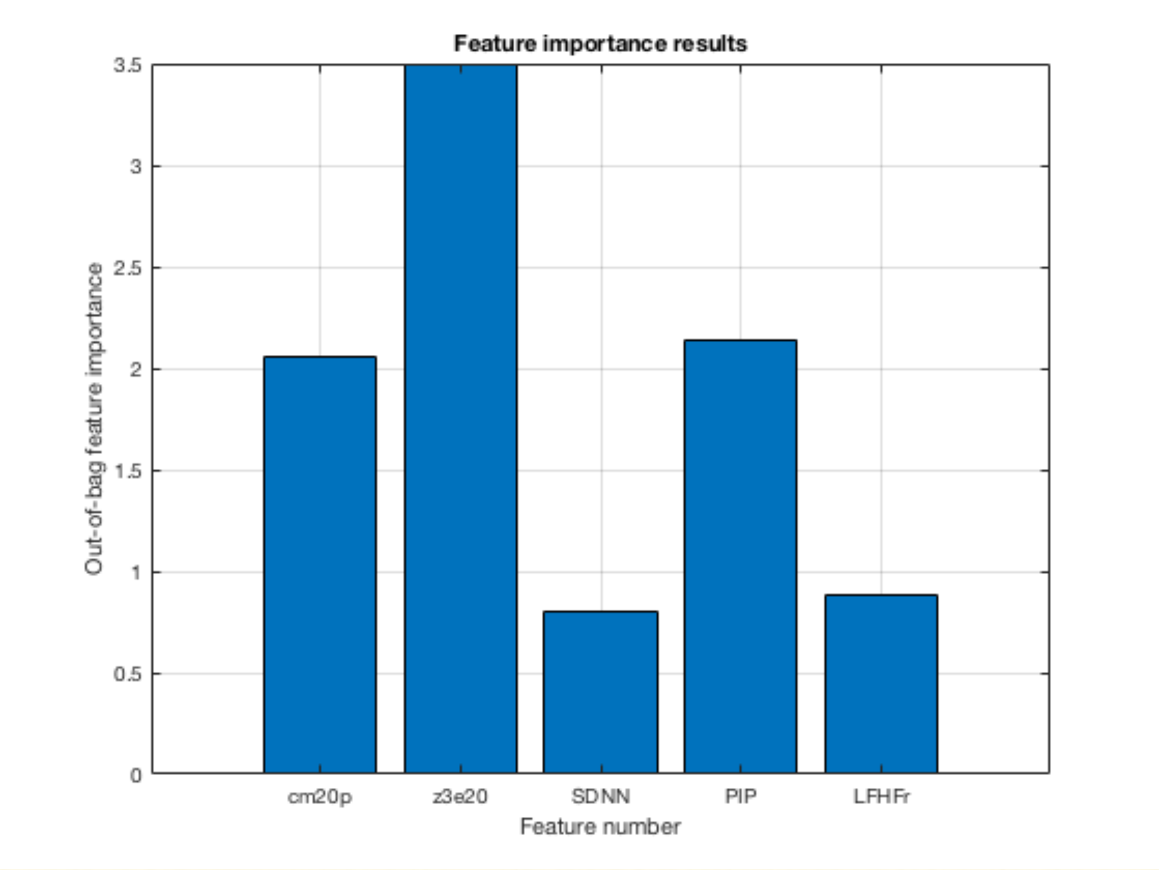
Classifier Testing:
Confusion Matrix

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

Classifier Performance

F1_ARR Score = 0.967
F1_CAD Score = 0.871
F1_NOR Score = 0.900
Averaged F1= = 0.911

Feature Pareto per Random Forest



Key Takeaway Messages

- RF results are conservative
- HRF discrimination
- Feature used: significant indicators
- How to improve performance?

Analysis of Results of Decision Tree

Insights

- Approx. Class Thresholds From Decision Tree:
  - **HRF:** sPIP>54% & LFHF<1 & z3e20p<1.1 (vs. Normal)
  - **HRF**: cm20p>0.02 & z3e20<0.97
  - **ARR**: z3e20p > 1.1

Analysis of Misclassifications

Patient & Feature	HLT 10090	HLT 10048	CAD 4198	AFDB 5261
cm20p	0.00	0.04	4.32	10.47
z3e20	0.01	0.34	6.64	16.56
sPIP	54.51	44.36	60.21	61.64
LFHF	1.47	1.34	0.42	0.48
SDRR	113.38	136.14	162.60	125.72
Actual Cat.	NOR	NOR	CAD	ARR
Pred. Cat.	NOR	HRF	ARR	ARR

cm20p,z3e20 weighting

Artifacts in data?

Probability Outputs of R-Forest

Case	Pred Class	Prob. ARR	Prob. HRF	Prob. NOR
HLT10090	NOR	0	0.433	0.567
HLT10048	CAD	0	0.567	0.433
HLT10094	NOR	0	0.133	0.867
CAD4189	CAD	0	0.833	0.167
CAD4185	CAD	0	0.7	0.3
CAD4198	ARR	1	0	0
AFDB0735	ARR	1	0	0
AFDB7162	ARR	0.967	0.033	0
AFDB5261	ARR	0.933	0.067	0

Key Learnings

- Noise filtering & artifact detection pre-processing of data is imperative
- Decision Tree thresholds provide intuitive understanding
- Probability outputs of R-Forest provide value to medical professionals in their final decision making

Summary and Next Steps

1. Cycle-by-cycle dynamics using rr-metric has been used for analysis

2. Heart Rate Fragementation detection using multiple variables yields promising results
3. The discrimination of HRF (as well as arrhythmias) using Random Forest seems to be a feasible path forward

4. The 5 features selected are useful in discriminating heart-rate dynamics for this purpose
5. Probability information from Random Forest algorithm is useful as per feedback from medical professionals (so that they can make the final decision with this information)

Next Steps:

7. Increase in ECG test cases for testing is planned
8. Incorporation of ECG noisy-segment determination and artifact detection into above workflow
9. Application/development of this methodology to short ECG segments (e.g. 5 mins etc.)

My sincere thanks to Seattle, WA based medical community researchers for their feedback
My sincere thanks to each of you for attending this presentation

References

[1] Madalena D. Costa, Roger B. Davis, Ary L. Goldberger, Heart Rate Fragmentation: A New Approach to the Analysis of Cardiac Interbeat Interval Dynamics, Frontiers in Physiology, 2017; 8: 255, https://doi.org/10.3389/fphys.2017.00255
[2] Madalena D. Costa, Roger B. Davis, Ary L. Goldberger, Heart Rate Fragmentation: A Symbolic Dynamical Approach, Frontiers in Physiology, Nov. 2017, Vol. 8, Article 827, doi: 10.3389/fphys.2017.00827
[3] Madalena D. Costa, Susan Redline, Roger B. Davis, Susan R. Heckbert, Elsayed Z. Soli-man, Ary L. Goldberger: Heart Rate Fragmentation as a Novel Biomarker of Adverse Cardiovascular Events: The Multi-Ethnic Study of Atherosclerosis, Front. Physiol., 03 Sep-tember 2018, https://doi.org/10.3389/fphys.2018.01117
[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 AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Compo-nents of a New Research Resource for Complex Physiologic Signals. Circula-tion 101(23):e215-e220. https://www.ahajournals.org/doi/10.1161/01.CIR.101.23.e215
[6] Marcus Vollmer, Ph.D Dissertation, https://d-nb.info/1124413723/34, pp. 63, Section 2.4.2