A Mathematical Diagnosis:

Classifying Electrocardiograms via Topological Time Series Analysis

Paul Samuel Ignacio
paulsamuel-ignacio@uiowa.edu
ppignacio@up.edu.ph

University of Iowa, USA
University of the Philippines Baguio, Philippines

V Escuela de Análisis Topológico de Datos

November 23, 2018

Background and Acknowledgements

- One of the projects undertaken at the 2018 MSRI-Undergraduate Program REU on "*Mathematics of Data Science*" held last summer.
- The project was undertaken by a team of three talented and (very) hardworking undergraduates led by the speaker and supervised by the MSRI-UP research director.
- A manuscript containing the results of this project is being prepared.
- We are grateful for the support of the following funding agencies:

Background and Acknowledgements

- One of the projects undertaken at the 2018 MSRI-Undergraduate Program REU on "*Mathematics of Data Science*" held last summer.
- The project was undertaken by a team of three talented and (very) hardworking undergraduates led by the speaker and supervised by the MSRI-UP research director.
- A manuscript containing the results of this project is being prepared.
- We are grateful for the support of the following funding agencies:

Background and Acknowledgements

- One of the projects undertaken at the 2018 MSRI-Undergraduate Program REU on "*Mathematics of Data Science*" held last summer.
- The project was undertaken by a team of three talented and (very) hardworking undergraduates led by the speaker and supervised by the MSRI-UP research director.
- A manuscript containing the results of this project is being prepared.
- We are grateful for the support of the following funding agencies:

Background and Acknowledgements

- One of the projects undertaken at the 2018 MSRI-Undergraduate Program REU on "*Mathematics of Data Science*" held last summer.
- The project was undertaken by a team of three talented and (very) hardworking undergraduates led by the speaker and supervised by the MSRI-UP research director.
- A manuscript containing the results of this project is being prepared.
- We are grateful for the support of the following funding agencies:

The 2018 MSRI-UP TDA Team

Christopher Dunstan

University of Maryland, Baltimore Country

Esteban Escobar

California Polytechnique University, Pomona

Luke Trujillo

Harvey Mudd College

Dr. David Uminsky

University of San Francisco

Given award for best poster in Mathematics at SACNAS Conference.

Outline

1. Introduction
   - Atrial Fibrillation and Electrocardiograms
   - Physionet Challenge 2017
2. Pipeline
3. Point Cloud Generation
4. Feature Extraction
5. Classification
6. Model Performance

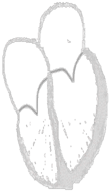
Introduction

Project Goal

Examine if TDA methods can be useful in diagnosis of heart conditions, specifically Atrial Fibrillation (AFib).

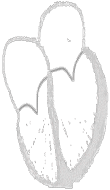
What is Atrial Fibrillation?

- A heart condition characterized by erratic heart beats caused by atrial spasms. It is estimated to affect 1% of the population.
- In a normal heartbeat, electric impulses are synergystically propagated from the sinoatrial node, located in the upper portion of the right atrium, towards the right ventricle.



What is Atrial Fibrillation?

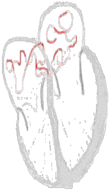
- A heart condition characterized by erratic heart beats caused by atrial spasms. It is estimated to affect 1% of the population.
- In a normal heartbeat, electric impulses are synergystically propagated from the sinoatrial node, located in the upper portion of the right atrium, towards the right ventricle.



A normal rhythm.
https://upload.wikimedia.org/wikipedia/commons/c/c3/Heart_conduct_sinus.gif

What is Atrial Fibrillation?

- In AFib, electrical impluses are in chaos in the atria, causing spasms and irregular opening of the valves leading to the ventricles.



AFib rhythm.
https://upload.wikimedia.org/wikipedia/commons/4/44/Heart_conduct_atrialfib.gif

Electrocardiograms

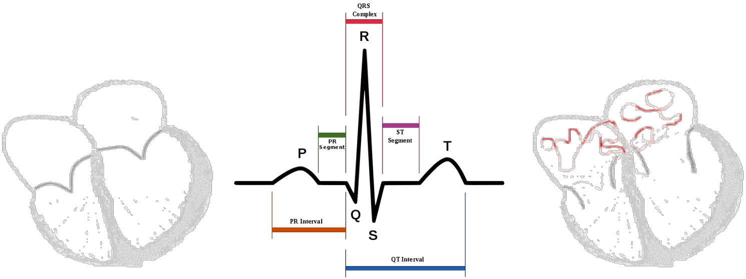


Image from `https://upload.wikimedia.org/wikipedia/commons/9/9e/SinusRhythmLabels.svg`.

Physionet Challenge

- Organized in conjunction with the annual Computing in Cardiology (CinC) conference.
- CinC is an international society of scientists and professionals in medicine, physics, engineering, and computer science.

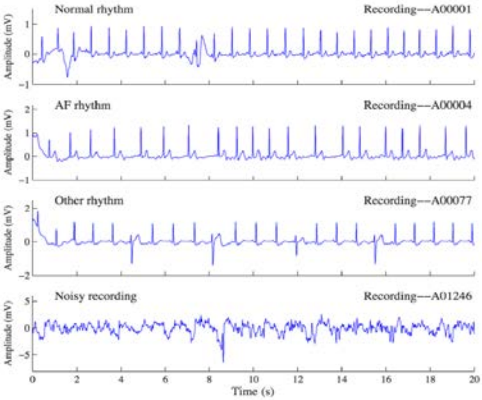
PhysioNet/Computing in Cardiology Challenges

In cooperation with the annual Computing in Cardiology conference, PhysioNet hosts a series of **challenges**, inviting participants to tackle clinically interesting problems that are either unsolved or not well-solved.

Year	*Topic*	*Papers*	*Contributed Software*
2000	Detecting Sleep Apnea from the ECG	13	1
2001	Predicting Paroxysmal Atrial Fibrillation	9	
2002	RR Interval Time Series Modeling	12	10
2003	Distinguishing Ischemic from Non-Ischemic ST Changes	3	1
2004	Spontaneous Termination of Atrial Fibrillation	12	1
2005	The First Five Challenges Revisited	5	
2006	QT Interval Measurement	20	8
2007	Electrocardiographic Imaging of Myocardial Infarction	6	
2008	Detecting and Quantifying T-Wave Alternans	19	5 • 1
2009	Predicting Acute Hypotensive Episodes	11	4
2010	Mind the Gap	13	5
2011	Improving the Quality of ECGs Collected using Mobile Phones	17	7
2012	Predicting Mortality of ICU Patients	17	58
2013	Noninvasive Fetal ECG	29	17
2014	Robust Detection of Heart Beats in Multimodal Data	15	35
2015	Reducing False Arrhythmia Alarms in the ICU	29	28
2016	Classification of Normal/Abnormal Heart Sound Recordings	11	48
2017	AF Classification from a Short Single Lead ECG Recording	57	64
2018	You Snooze, You Win	Ongoing	Ongoing

https://www.physionet.org/challenge/.

- Dataset consists of 8528 (training set) and 3658 (test set) ECG readings of length at most 60 seconds classified into four categories, **Normal**, **AF**, **Other**, and **Noisy**.



AF Classification from a Short Single Lead ECG Recording: the Physionet Computing in Cardiology Challenge 2017.

- Relabeling had to be done (twice) due to disagreements in classification of 1129 ECGs.

Type	# recordings (%)		
	V1	V2	V3
Training			
Normal	5154 (60.4)	5050 (59.2)	5076 (59.5)
AF	771 (9.0)	738 (8.7)	758 (8.9)
Other	2557 (30.0)	2456 (28.8)	2415 (28.3)
Noisy	46 (0.5)	284 (3.3)	279 (3.3)
Test			
Normal	2209 (60.4)	2195 (60.0)	2437 (66.6)
AF	331 (9.1)	315 (8.6)	286 (7.8)
Other	1097 (30.0)	1015 (27.8)	683 (18.7)
Noisy	21 (0.6)	133 (3.6)	252 (6.9)

Table 2. Data profile for the training/test set.

AF Classification from a Short Single Lead ECG Recording: the Physionet Computing in Cardiology Challenge 2017.

- For each of the four types, F_1 is defined as:

$$F_{1n} = \frac{2 \times Nn}{\Sigma N + \Sigma n}$$

$$F_{1a} = \frac{2 \times Aa}{\Sigma A + \Sigma a}$$

$$F_{1o} = \frac{2 \times Oo}{\Sigma O + \Sigma o}$$

$$F_{1p} = \frac{2 \times Pp}{\Sigma P + \Sigma p}$$

	Predicted Classification				
	Normal	AF	Other	Noisy	Total
Normal	Nn	Na	No	Np	$\sum N$
AF	An	Aa	Ao	Ap	$\sum A$
Other	On	Oa	Oo	Op	$\sum O$
Noisy	Pn	Pa	Po	Pp	$\sum P$
Total	$\sum n$	$\sum a$	$\sum o$	$\sum p$	

Table 3. Definition of parameters for scoring used in eq. 1.

AF Classification from a Short Single Lead ECG Recording: the Physionet Computing in Cardiology Challenge 2017.

The final score is:

$$\frac{F_{1n} + F_{1a} + F_{1o}}{3}$$

Out of 75 competing models... almost everyone won!

4 scored 83%

4 scored 82%

9 scored 81%

7 scored 80%

Next 28 scored above 70%

Out of 75 competing models... almost everyone won!

4 scored 83%
4 scored 82%
9 scored 81%
7 scored 80%
Next 28 scored above 70%

Out of 75 competing models... almost everyone won!
4 scored 83%
4 scored 82%
9 scored 81%
7 scored 80%
Next 28 scored above 70%

Out of 75 competing models... almost everyone won!
4 scored 83%
4 scored 82%
9 scored 81%
7 scored 80%
Next 28 scored above 70%

Out of 75 competing models... almost everyone won!

4 scored 83%

4 scored 82%

9 scored 81%

7 scored 80%

Next 28 scored above 70%

Out of 75 competing models... almost everyone won!
4 scored 83%
4 scored 82%
9 scored 81%
7 scored 80%
Next 28 scored above 70%

Out of 75 competing models... almost everyone won!
4 scored 83%
4 scored 82%
9 scored 81%
7 scored 80%
Next 28 scored above 70%

Model	Med./Alg[1]	# Feat.	Train	Validation	Test
Teijeiro et al.	Yes	86	0.893	0.912	0.831
Datta et al.	Yes	150	0.970	0.990	0.829
Zabihi et al.	Yes	150	0.951	0.968	0.826
Hong et al.	Yes	622	0.970	0.990	0.825

[1] Requires medical knowledge or other advanced algorithm to extract features.

Questions:

Is it possible to diagnose AF just from studying "features" of ECGs alone?

Is there signal from topological features of ECGs for detecting AF?

Questions:

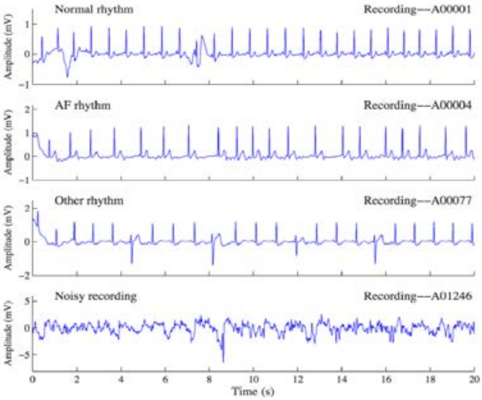
Is it possible to diagnose AF just from studying "features" of ECGs alone?

Is there signal from topological features of ECGs for detecting AF?

Tell Tale Signs

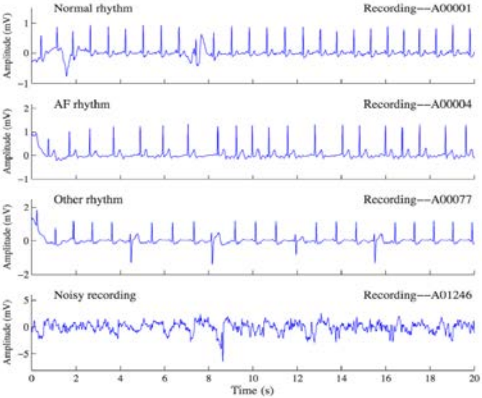
Two Recurrent Features

- RR intervals: AF rhythyms have irregular RR intervals.
- P waves: AF rhythyms are missing P waves.



AF Classification from a Short Single Lead ECG Recording: the Physionet Computing in Cardiology Challenge 2017.

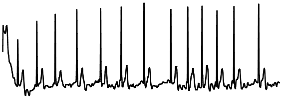
- RR intervals: AF rhythyms have irregular RR intervals.
- P waves: AF rhythyms are missing P waves.



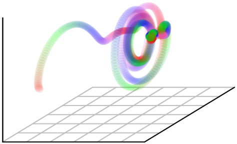
AF Classification from a Short Single Lead ECG Recording: the Physionet Computing in Cardiology Challenge 2017.

Pipeline

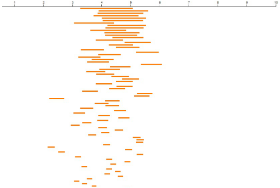
ECG



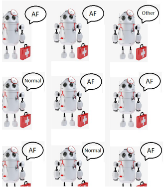
Embed to Hign Dimensional Space



Extract Features



Classify



Vote

Normal: 2
AF: 6
Other: 1
Noisy: 0

Diagnosis: AF

Pipeline

This approach is not original. A few references that use slight variation include Perea et al. (2015), Seversky et al. (2016), and Umeda (2017).

Sliding Windows

Idea: Store information about every part of the time series.
Steps

- Select a window length W.
- From time $t = 0$ to $t = W$, select D points on the time series and record it as a vector $v_0 \in \mathbb{R}^D$.
- Slide the window by incrementing t, and repeat the previous step for as long as possible. This generates a point clound in $\mathbb{R}^D$.

Sliding Windows

Idea: Store information about every part of the time series.

Steps

- Select a window length W.
- From time $t = 0$ to $t = W$, select D points on the time series and record it as a vector $v_0 \in \mathbb{R}^D$.
- Slide the window by incrementing t, and repeat the previous step for as long as possible. This generates a point clound in $\mathbb{R}^D$.

Sliding Windows

Idea: Store information about every part of the time series.
Steps

- Select a window length W.
- From time $t = 0$ to $t = W$, select D points on the time series and record it as a vector $v_0 \in \mathbb{R}^D$.
- Slide the window by incrementing t, and repeat the previous step for as long as possible. This generates a point clound in $\mathbb{R}^D$.

Sliding Windows

Idea: Store information about every part of the time series.
Steps

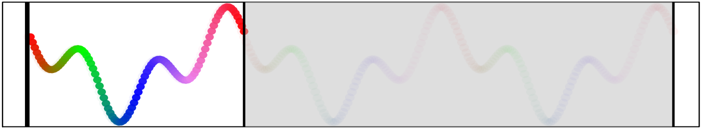
- Select a window length W.
- From time $t = 0$ to $t = W$, select D points on the time series and record it as a vector $v_0 \in \mathbb{R}^D$.
- Slide the window by incrementing t, and repeat the previous step for as long as possible. This generates a point clound in $\mathbb{R}^D$.

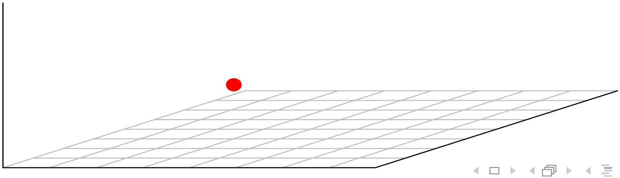
Sliding Windows

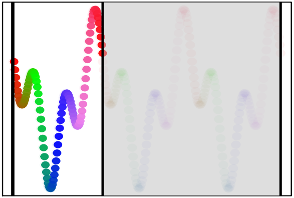
Idea: Store information about every part of the time series.
Steps

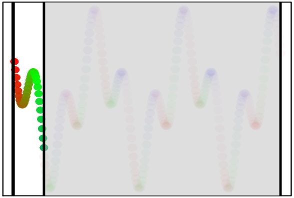
- Select a window length W.
- From time $t = 0$ to $t = W$, select D points on the time series and record it as a vector $v_0 \in \mathbb{R}^D$.
- Slide the window by incrementing t, and repeat the previous step for as long as possible. This generates a point clound in $\mathbb{R}^D$.

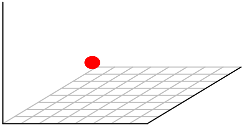
Sliding Windows
Example



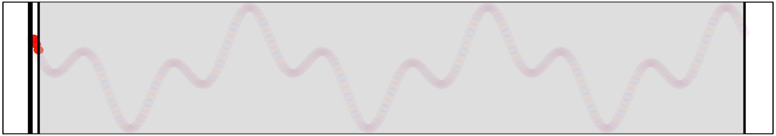


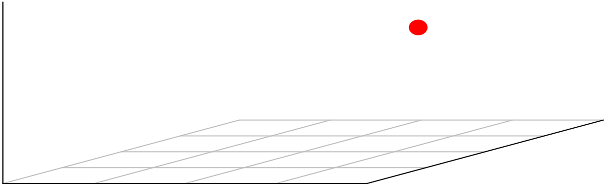




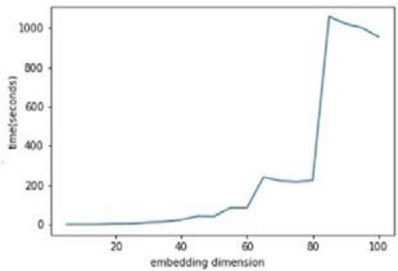


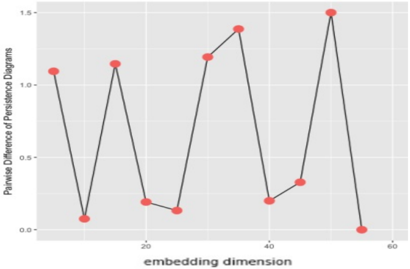
Sliding Windows
Embedding Dimension Dependence



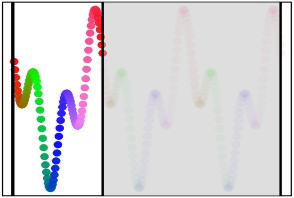


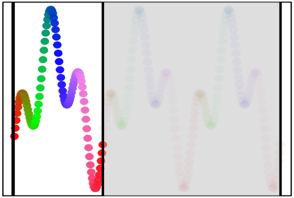
For our analysis, we chose the window size to be $250ms$ and $D = 50$.

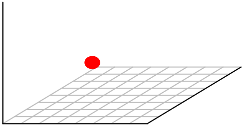


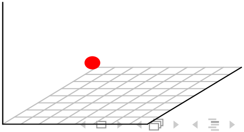


Sliding Windows
Inversion Invariance

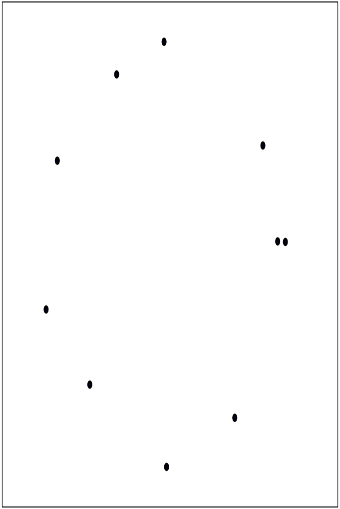


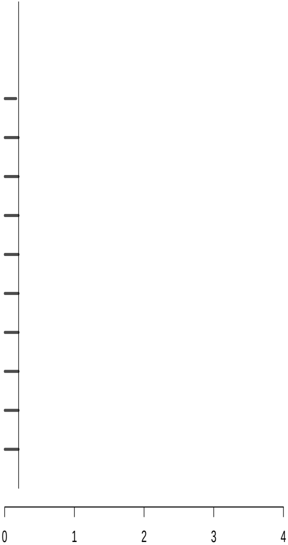






Persistent Homology





Persistent Homology

We computed the persistence of 1, and 2- dimensional cycles using *Ripser*.

BUT WAIT!!!!

WHAT DO THE HOLES MEAN?!?!?!?!?!

Persistent Homology

We computed the persistence of 1, and 2- dimensional cycles using *Ripser*.

BUT WAIT!!!!

WHAT DO THE HOLES MEAN?!?!?!?!?!

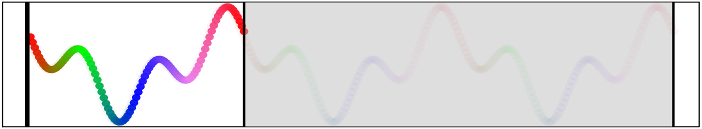
Persistent Homology

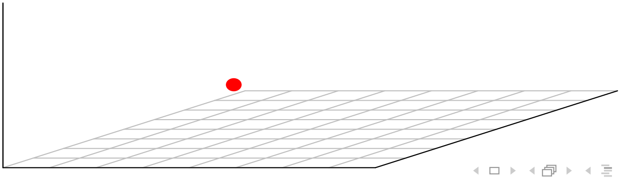
We computed the persistence of 1, and 2- dimensional cycles using *Ripser*.

BUT WAIT!!!!

WHAT DO THE HOLES MEAN?!?!?!?!?!

Sliding Windows
Revisited





Persistent Homology

We computed the persistence of 1, and 2- dimensional cycles using *Ripser*. We then computed statistics for the following features:

- Mean, sd, skewness, kurtosis of the birth, death, and persistence of features.
- Mean, sd, skewness, kurtosis of the birth, death, and persistence when most persistent features are removed.
- The number of features in each dimension.

We computed the persistence of 1, and 2- dimensional cycles using *Ripser*. We then computed statistics for the following features:

- Mean, sd, skewness, kurtosis of the birth, death, and persistence of features.
- Mean, sd, skewness, kurtosis of the birth, death, and persistence when most persistent features are removed.
- The number of features in each dimension.

Persistent Homology

We computed the persistence of 1, and 2- dimensional cycles using *Ripser*. We then computed statistics for the following features:

- Mean, sd, skewness, kurtosis of the birth, death, and persistence of features.
- Mean, sd, skewness, kurtosis of the birth, death, and persistence when most persistent features are removed.
- The number of features in each dimension.

Persistent Homology

We computed the persistence of 1, and 2- dimensional cycles using *Ripser*. We then computed statistics for the following features:

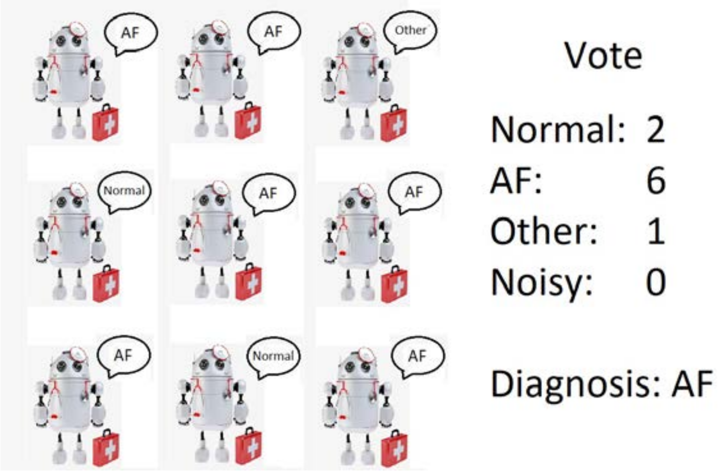
- Mean, sd, skewness, kurtosis of the birth, death, and persistence of features.
- Mean, sd, skewness, kurtosis of the birth, death, and persistence when most persistent features are removed.
- The number of features in each dimension.

Random Forest

We used a random forest to classify the ECGs.

Random Forest

We used a random forest to classify the ECGs.



Random Forest

We also examined feature importance and twelve were included in the final set of features. These are:

- Means and SD of birth and deaths of dimension 1 cycles.
- Accumulated persistence and skewness of persistence of dimension 1 cycles.
- Number, and SD of persistence of dimension 1 and 2 cycles.
- Mean of persistence for dimension 1 and 2 cycles.

Random Forest

We also examined feature importance and twelve were included in the final set of features. These are:

- Means and SD of birth and deaths of dimension 1 cycles.
- Accumulated persistence and skewness of persistence of dimension 1 cycles.
- Number, and SD of persistence of dimension 1 and 2 cycles.
- Mean of persistence for dimension 1 and 2 cycles.

Random Forest

We also examined feature importance and twelve were included in the final set of features. These are:

- Means and SD of birth and deaths of dimension 1 cycles.
- Accumulated persistence and skewness of persistence of dimension 1 cycles.
- Number, and SD of persistence of dimension 1 and 2 cycles.
- Mean of persistence for dimension 1 and 2 cycles.

Random Forest

We also examined feature importance and twelve were included in the final set of features. These are:

- Means and SD of birth and deaths of dimension 1 cycles.
- Accumulated persistence and skewness of persistence of dimension 1 cycles.
- Number, and SD of persistence of dimension 1 and 2 cycles.
- Mean of persistence for dimension 1 and 2 cycles.

Random Forest

We also examined feature importance and twelve were included in the final set of features. These are:

- Means and SD of birth and deaths of dimension 1 cycles.
- Accumulated persistence and skewness of persistence of dimension 1 cycles.
- Number, and SD of persistence of dimension 1 and 2 cycles.
- Mean of persistence for dimension 1 and 2 cycles.

Results

Model	Med./Alg[2]	# Feat.	Train	Validation	Test
Teijeiro et al.	Yes	86	0.893	0.912	0.831
Datta et al.	Yes	150	0.970	0.990	0.829
Zabihi et al.	Yes	150	0.951	0.968	0.826
Hong et al.	Yes	622	0.970	0.990	0.825
TDA	No	12	0.856	0.866	0.770
RR	Yes	3	0.809	0.856	0.700

[2] Requires medical inspection or other advanced algorithm to extract features.

Results

Model	Med./Alg[2]	# Feat.	Train	Validation	Test
Teijeiro et al.	Yes	86	0.893	0.912	0.831
Datta et al.	Yes	150	0.970	0.990	0.829
Zabihi et al.	Yes	150	0.951	0.968	0.826
Hong et al.	Yes	622	0.970	0.990	0.825
TDA	No	12	0.856	0.866	0.770
RR	Yes	3	0.809	0.856	0.700

[2] Requires medical inspection or other advanced algorithm to extract features.

Results

Model	Med./Alg[2]	# Feat.	Train	Validation	Test
Teijeiro et al.	Yes	86	0.893	0.912	0.831
Datta et al.	Yes	150	0.970	0.990	0.829
Zabihi et al.	Yes	150	0.951	0.968	0.826
Hong et al.	Yes	622	0.970	0.990	0.825
TDA	No	12	0.856	0.866	0.770
RR	Yes	3	0.809	0.856	0.700

[2] Requires medical inspection or other advanced algorithm to extract features.

Results

Model	f_{1a}	f_{1n}	f_{1o}	Final score
Teijeiro et al.	0.854	0.903	0.737	0.831
Datta et al.	0.823	0.916	0.750	0.829
Zabihi et al.	0.835	0.909	0.734	0.826
Hong et al.	0.823	0.912	0.751	0.825
TDA	0.590	0.973	0.749	0.770
RR	0.707	0.828	0.571	0.700
TDA + RR	0.716	0.976	0.842	0.844

Results

Model	f_{1a}	f_{1n}	f_{1o}	Final score
Teijeiro et al.	0.854	0.903	0.737	0.831
Datta et al.	0.823	0.916	0.750	0.829
Zabihi et al.	0.835	0.909	0.734	0.826
Hong et al.	0.823	0.912	0.751	0.825
TDA	0.590	0.973	0.749	0.770
RR	0.707	0.828	0.571	0.700
TDA + RR	0.716	0.976	0.842	0.844

Model	f_{1a}	f_{1n}	f_{1o}	Final score
Teijeiro et al.	0.854	0.903	0.737	0.831
Datta et al.	0.823	0.916	0.750	0.829
Zabihi et al.	0.835	0.909	0.734	0.826
Hong et al.	0.823	0.912	0.751	0.825
TDA	0.590	0.973	0.749	0.770
RR	0.707	0.828	0.571	0.700
TDA + RR	0.716	0.976	0.842	0.844

Model	f_{1a}	f_{1n}	f_{1o}	Final score
Teijeiro et al.	0.854	0.903	0.737	0.831
Datta et al.	0.823	0.916	0.750	0.829
Zabihi et al.	0.835	0.909	0.734	0.826
Hong et al.	0.823	0.912	0.751	0.825
TDA	0.590	0.973	0.749	0.770
RR	0.707	0.828	0.571	0.700
TDA + RR	0.716	0.976	0.842	0.844

Results

Model	Med./Alg[3]	# Feat.	Train	Validation	Test
Teijeiro et al.	Yes	86	0.893	0.912	0.831
Datta et al.	Yes	150	0.970	0.990	0.829
Zabihi et al.	Yes	150	0.951	0.968	0.826
Hong et al.	Yes	622	0.970	0.990	0.825
TDA	No	12	0.856	0.866	0.770
RR	Yes	3	0.809	0.856	0.700
TDA + RR	Yo (?)	15	0.921	0.957	0.844

[3] Requires medical inspection or other advanced algorithm to extract features.

Moving Forward.

- Use the hidden test set from the competition.
- Improve benchmarking on embedding dimension, and noise detection.
- Include P waves analysis.
- Apply to other time-series data.

References

- Umeda,Y. *Time Series Classification via Topological Data Analysis*,Transactions of the Japanese Society for Artificial Intelligence, 32 (3),1-12, 2017
- Seversky,L., Davis, S., Berger, M.,*On Time-Series Topological Data Analysis: New Data and Opportunities*,2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),1014-1022,2016
- Perea, J., Harer, J.,*Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis*, Foundations of Computational Mathematics, 15 (3), 799-838,2015

Maraming Salamat!!!