



# SMALL PROJECT

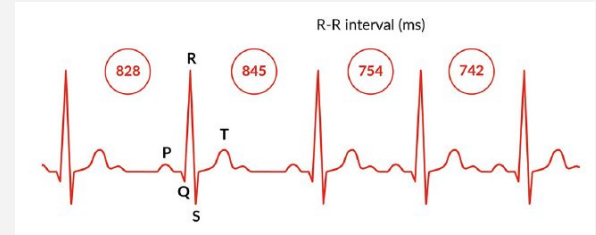
## CLASSIFICATION OF ECG SIGNALS

Biomedical Engineering - Biomedical Signals  
Sergi Marsol

# INTRODUCTION AND MOTIVATION

An important goal → automatic detection of arrhythmias

**Atrial Fibrillation** is defined as “Tachyarrhythmia characterized by predominantly uncoordinated atrial activation with consequent deterioration of atrial mechanical function”



One method for AF detection is to analyse the HRV signal, based on the detection of the R peak.  
Excessive HRV → correlated with atrial fibrillation

## GOALS AND METHODS OF THE PROJECT

### GOAL 1

Separate **CLEAN** ECG from **NOISY** ECG

- Read ECG files using Matlab
- Separate **CLEAN** from **NOISY** signals

### GOAL 2

Design and program our own R wave detector and extract main features

- Pre-Process ECG signals
- Find QRS peaks
- Extract features from original and RR signals
- Create a table of features

### GOAL 3

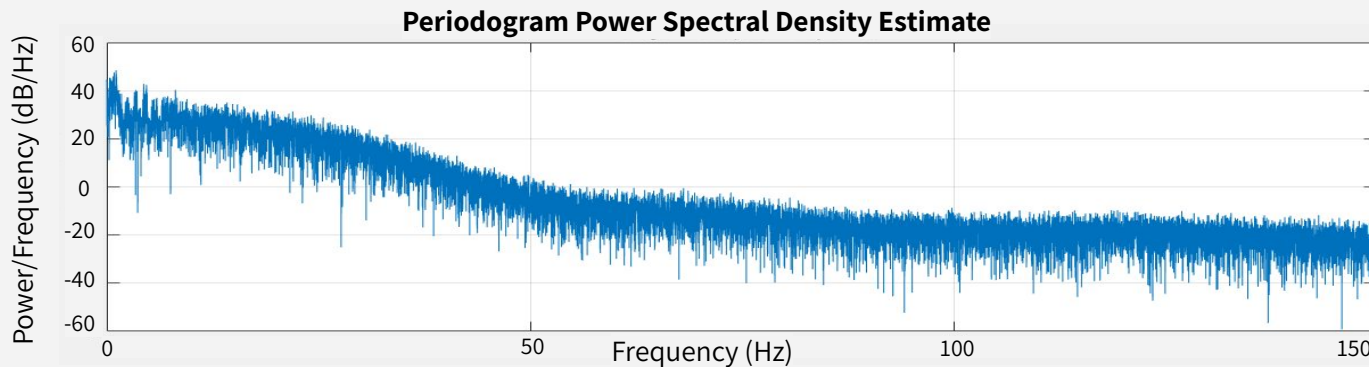
Provide a classifier for the 4 classes of the clean signal

- Build a classifier function
- Validate the predictions for the 4 classes and evaluate the results of the classifier

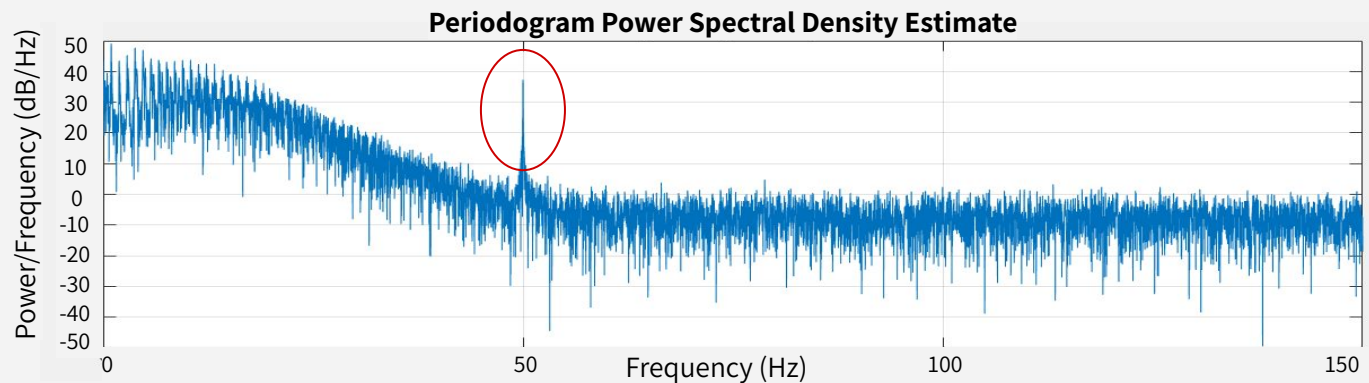
# GOAL 1: Separation of **clean** and **noisy** signals

## PERIDOGRAM

Clean  
signals



Noisy  
signals



# CLASSIFIERS

Clean and noisy signals separation using 1 feature → BANDPOWER at **frequency 50 Hz**

## 1. K-NEAREST NEIGHBORS

True Class	A	2484	68
	B	112	1394
		A	B
		Predicted Class	

Hyperparameter optimization

distance:  
*chebyshev*

$k = 17$

→ **ACCURACY = 0.9556**

Class	Precision	Recall
A	0.957	0.973
B	0.954	0.926

**Total F-Score = 0.952**

**Hyperparameter optimization**



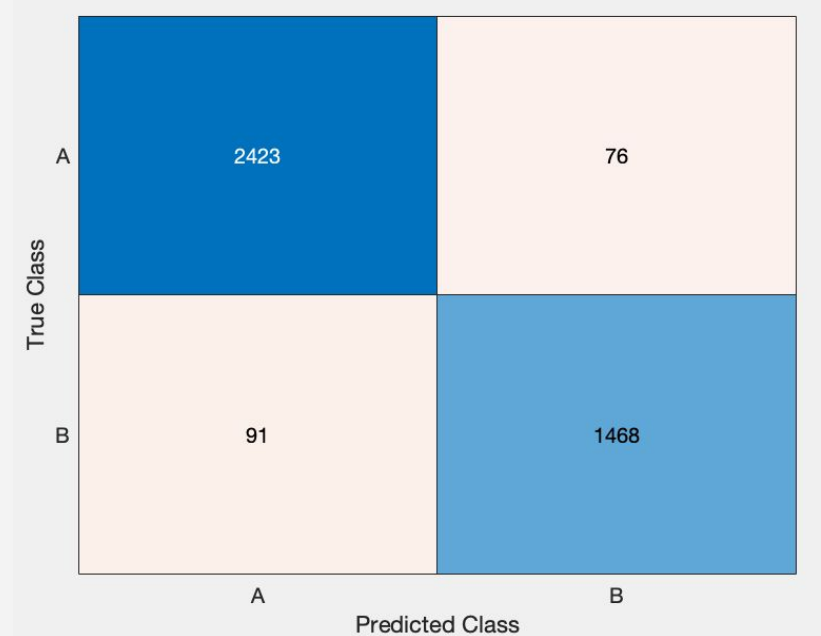
**MinLeafSize = 78**

**→ ACCURACY = 0.957**

Class	Precision	Recall
A	0.964	0.970
B	0.951	0.942

**Total F-Score = 0.957**

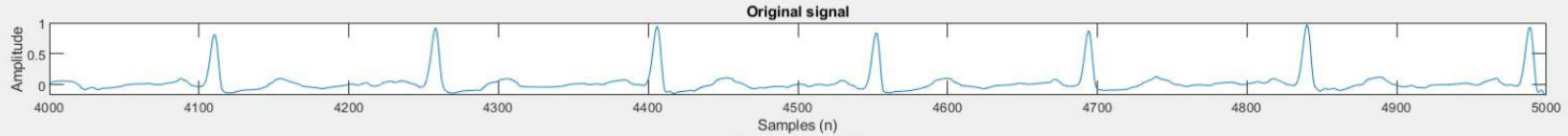
## 2. CLASSIFICATION TREE



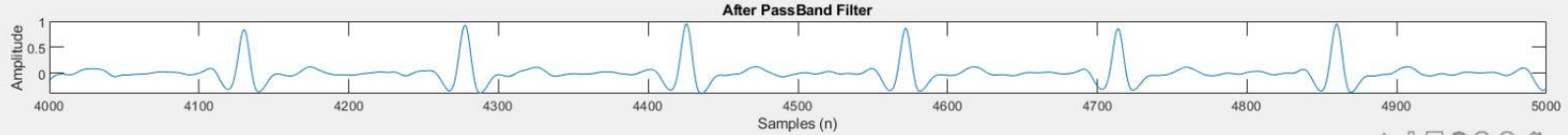
# GOAL 2: R wave detector and main features extraction

## PAN-TOMPKINS ALGORITHM

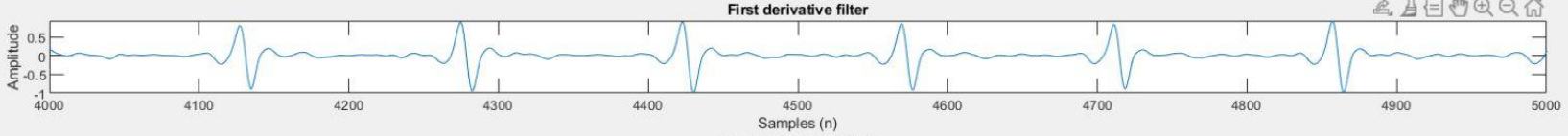
Original  
signal



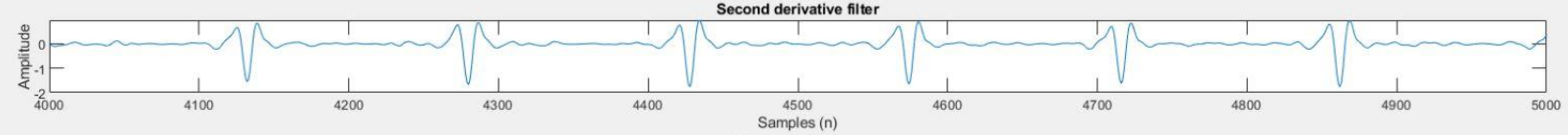
Bandpass  
filter



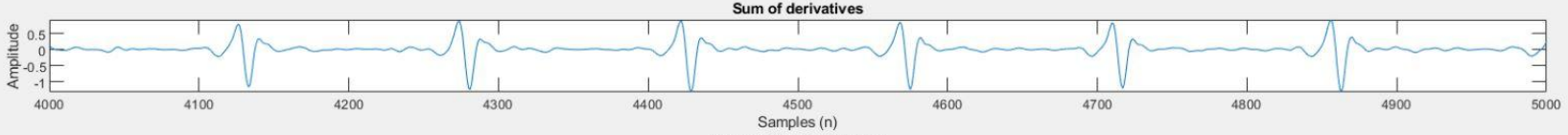
First  
derivative



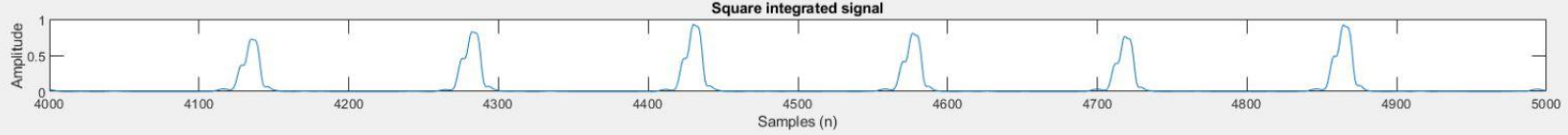
Second  
derivative



Sum of  
derivative



Square  
Integrator



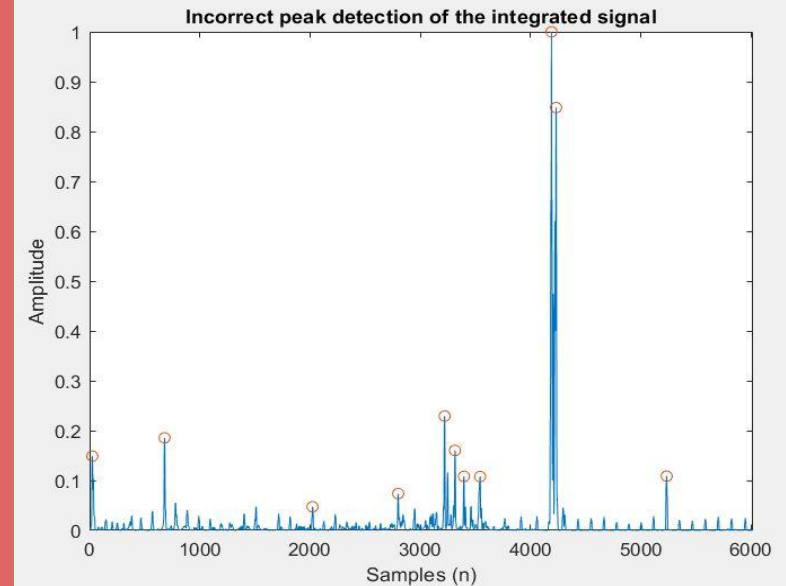
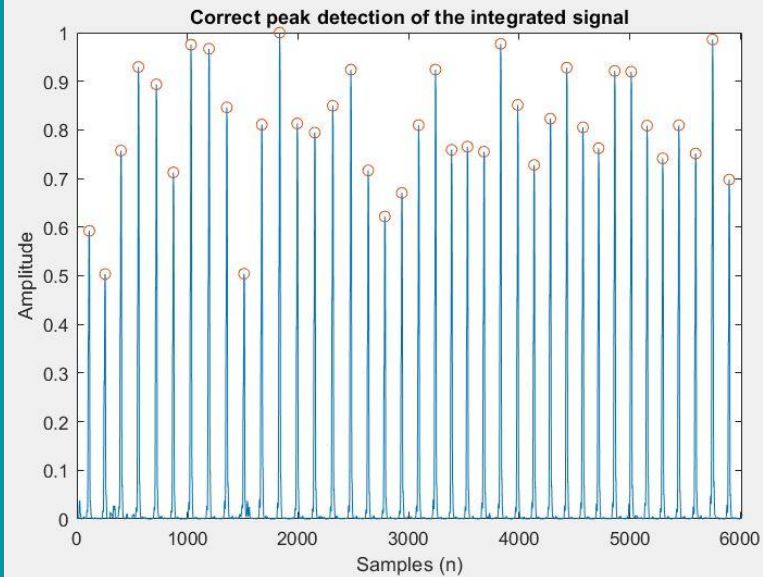
# QRS PEAK DETECTOR

$$\text{Thr1} = \text{npk1} + 0,25(\text{spk1} - \text{npk1})$$

$$\text{spk1} = 0,125\text{peak1} + 0,875\text{spk1}$$

$$\text{Thr2} = 0,5\text{Thr1}$$

$$\text{npk1} = 0,125\text{peak1} + 0,875\text{npk1}$$



# FEATURE EXTRACTION

## ECG signal

- Mean *mA*
- Standard deviation *sdA*
- Outlier fraction *outlierA*

## Autocorrelation

- Mean *autoA1*
- Standard deviation *autoA2*
- Maximum value *autoA3*

## Bandpower

- Bp [0,5-5]Hz *bp1*
- Bp [5-15]Hz *bp2*
- Bp [15-30]Hz *bp3*
- Bp [30-45]Hz *bp4*
- Number of Peaks *num\_pks*
- Peaks detected correctly *pks\_ok*

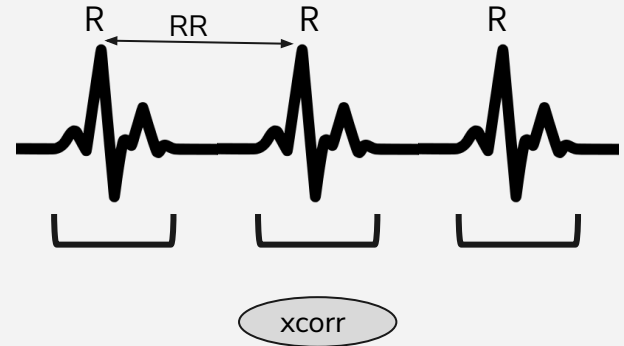
## RR signal

- Mean *mRR*
- Standard deviation *sdRR (HRV)*
- Outlier fraction *outlierRR*

- Standard deviation of successive differences *sdsd*

## Cross-correlation

- Correlation of the length of consecutive beats *corrlenNN*



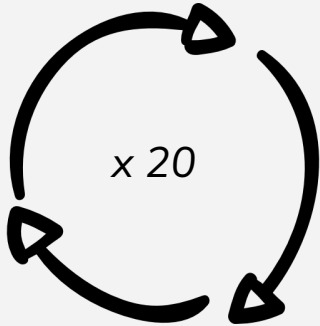


# FEATURE EXTRACTION

RR signal

➤ pNN20

➤ pNN50



*pNN10*

*pNN20*

*pNN30*

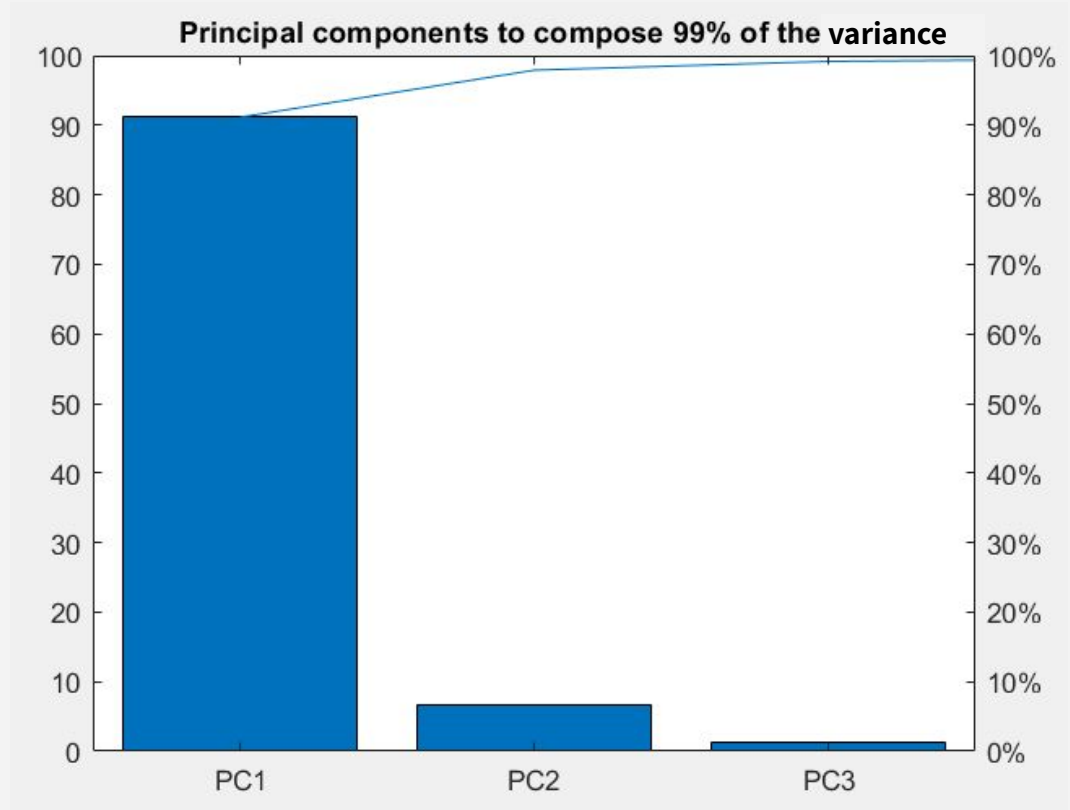
...

*pNN200*

➤ PC1 *PC1\_pNN*

➤ PC2 *PC2\_pNN*

➤ PC3 *PC3\_pNN*

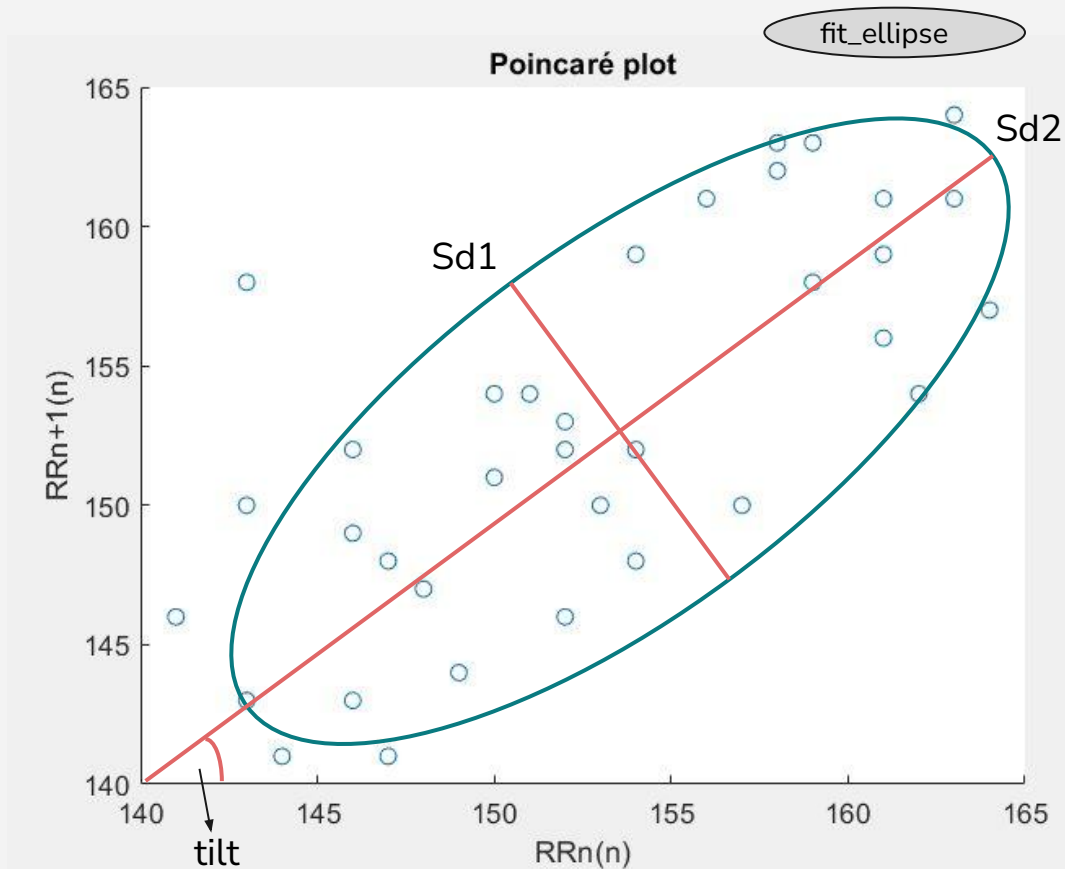
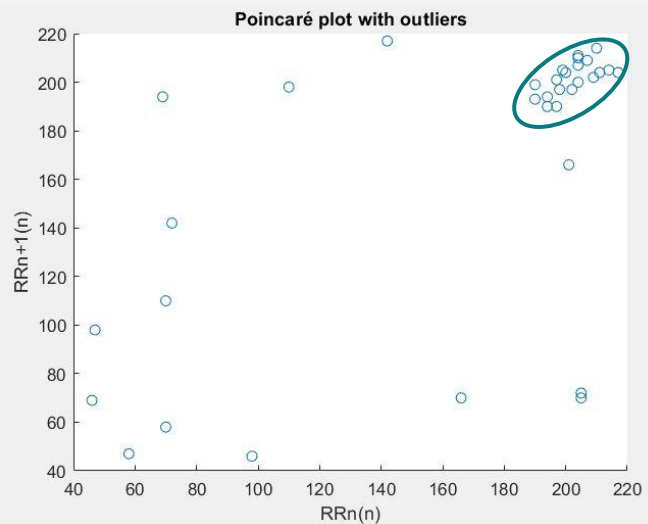


# FEATURE EXTRACTION

Poincaré

Scatter plot of  $nn-1$  vs  $nn$

➤ Sd1   ➤ Sd2   ➤ tilt



# FEATURE EXTRACTION

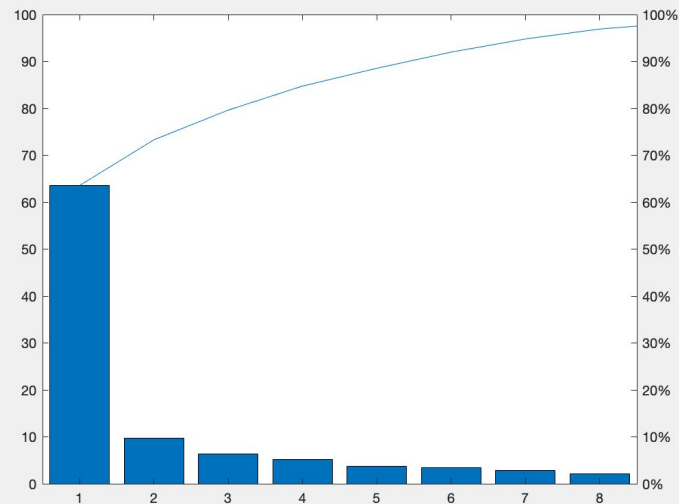
## ANOVA

Analysis of variance

Features	P-value	Features	P-value	Features	P-value
PC1	0	corrLenNN	1,8126e-40	bp3	0,0057
PC3	1,2338e-146	sd1	2,3819e-29	bp2	0,0088
PC2	3,0789e-116	mRR	2,1524e-17	tilt	0,0105
outlierA	3,6195e-97	bp1	1,1365e-11	autoA3	0,0450
sdsd	2,3383e-73	autoA1	1,5348e-07	bp4	0,8737
sdRR	1,1718e-65	autoA2	1,3441e-05	<div>Lower p-value ✓</div> <div>Higher p-value ✗</div>	
pks_ok	1,2913e-62	mA	4,2079e-04		
sd2	5,5304e-51	num_pks	4,7556e-04		
outlierRR	2,2740e-43	sdA	0,0056		

# GOAL 3: Build a classifier for all 4 clean ECG recordings

## 1. PRINCIPAL COMPONENT ANALYSIS (PCA)



Principal components explaining 95% of variance



We will build the classification model  
using **8 principal components**.

sdRR	0.09394	0.393	0.1623	0.2009	0.2767	0.09959	0.3168	0.1834
mRR	0.02409	0.3256	0.05352	0.2542	0.1455	0.02749	0.6684	0.5678
outlierA	0.02966	0.1095	0.5152	0.1353	0.4485	0.1623	0.2261	0.06005
outliersRR	0.06347	0.06883	0.4098	0.3146	0.27	0.05362	0.3814	0.5755
sd1	0.008086	0.4938	0.2611	0.02177	0.379	0.0741	0.07011	0.1027
sd2	0.03317	0.4894	0.2529	0.01976	0.381	0.07543	0.06667	0.1017
pc1 <sub>p</sub> NNw	0.9879	0.1016	0.01251	0.06679	0.07659	0.01792	0.03683	0.01503
pc2 <sub>p</sub> NNw	0.001749	0.2913	0.01764	0.8413	0.4087	0.04905	0.1694	0.08997
pc3 <sub>p</sub> NNw	0.0007253	0.01819	0.07301	0.05085	0.1028	0.008598	0.04209	0.09212
pks <sub>o</sub> k	0.007294	0.06365	0.01246	0.04512	0.01072	0.0003825	0.0431	0.07207
autoA1	0.007754	0.0131	0.04009	0.04842	0.2368	0.9658	0.06327	0.03076
bp1	0.0003137	0.0009984	0.001754	0.001086	0.005624	0.03301	0.002005	0.001351
corrienNN	0.003847	0.1171	0.588	0.1395	0.05594	0.02078	0.1912	0.513
sdsd	0.09225	0.355	0.2278	0.1954	0.3108	0.1089	0.4156	0.04673
	1	2	3	4	5	6	7	8

Principal Component

## 2. K-NEAREST-NEIGHBOR (KNN)

True Class	A	148	18	38	10
	N	20	1341	156	5
	O	70	343	313	11
	~	18	23	23	21
		A	N	O	~
		Predicted Class			

Hyperparameter optimization

distance:  
*cityblock*

$k = 14$

→ ACCURACY = 0.7217

Class	Precision	Recall
AF	0.578	0.692
Normal	0.777	0.881
Other	0.591	0.425
Noisy	0.447	0.247

Total F-Score = 0.567

### 3. TREE CLASSIFIER

Hyperparameter optimization

MinLeafSize = 49

→ ACCURACY = 0.6841

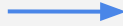
Class	Precision	Recall
AF	0.521	0.575
Normal	0.76	0.85
Other	0.543	0.43
Noisy	0.472	0.2

F-Score = 0.528

True Class	A	N	O	~
A	123	31	52	8
N	31	1293	194	4
O	67	346	317	7
~	15	32	21	17
Predicted Class				
	A	N	O	~

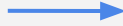
# CONCLUSIONS

## GOAL 1

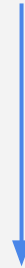


We have been able to separate clean from noisy ECG using the **BANDPOWER** as the main feature of the classifiers.

## GOAL 2

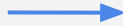


We have performed an analysis in amplitude and frequency of the ECG signals and have extracted **14 features**.



PCA has allowed dimensionality reduction from 14 features to 8 principal components.

## GOAL 3



We have built the classification models to **separate all clean ECG classes** using only 8 features and kNN is giving the best results.