

1 Definição do problema

De acordo com a Organização Mundial de Saúde (OMS), doenças cardíacas são a principal causa de morte no planeta. E apesar de boa parte das doenças cardíacas poderem ser diagnosticadas com antecedência, o número de casos continuam a crescer muito. Estima-se que 80% dos casos aconteçam em países pobres e em desenvolvimento e, segundo a OMS, até o ano de 2030, 23,6 milhões de pessoas vão morrer de doenças cardíacas, pois as medidas de prevenção, aparentemente, não estão sendo adequadas.

O principal exame para diagnóstico de doenças cardíacas é chamado Eletrocardiograma (ECG). Segundo Cohen (1988), o ECG foi um dos primeiros sinais estudados pelas técnicas de processamento de sinais biomédicos e até os dias de hoje é a técnica mais utilizada no diagnóstico de doenças do coração. Isto se deve à simplicidade do exame, que é um processo não-invasivo.

Um dos problemas cardíacos detectável pelo exame de ECG são as arritmias. Arritmias são batimentos cardíacos anormais ou irregulares. Têm-se duas categorias de arritmias. Arritmias formadas por um único batimento cardíaco irregular, chamadas de arritmias morfológicas, ou um conjunto irregular de batimentos, denominadas arritmias rítmicas.

O processo de identificação e classificação de arritmias pode ser muito penoso para um ser humano e, ainda sujeito a erros, visto que seria necessário uma análise de batimento a batimento em horas ou até mesmo dias de registros de ECG capturados.

Uma alternativa em estudo consiste no uso de inteligência artificial para classificação automática destes padrões. Hoje em dia, métodos baseados em aprendizagem de máquina tem obtido resultados expressivos segundo Hannun et al. (2019), contudo, tem-se muito espaço para melhorias, em especial quando se considera uma avaliação mais rígida.

Quando tem-se como objetivo, a implementação de algoritmos e técnicas de inteligência artificial e a integração a equipamentos médicos, normas e protocolos de desenvolvimento de software foram

elaborados por ANSI/AAMI (2008). Na literatura, vários métodos publicados não levam em conta essas normas e práticas de avaliação e desenvolvimento, levando a reportar resultados super-estimados e assim, sem grande importância para fabricantes de equipamentos médicos.

Com isso, o trabalho consiste em estudar a aplicação de técnicas de mineração de dados na detecção de arritmias em ECG's. Como norma de desenvolvimento metodológico sugerido pela AAMI, o paradigma *inter-patient* será aplicado ao trabalho, onde o conjunto de teste será formado por dados de pacientes diferentes do conjunto de treino. O processo de classificação de arritmias consistirá em basicamente três etapas principais: 1) pré-processamento, 2) extração de características e 3) classificação.

Na etapa de pré-processamento será utilizado filtros com o intuito de reduzir ruídos de energia no sinal preservando as informações que caracterizam cada sinal. Os filtros a serem utilizados são: *Butterworth low pass filter*, *LOcally WEighted Scatterplot Smoothing (LOWESS) curve fitting*, *Non local means(NLM)*.

Na etapa de extração de características, será utilizado várias características disponibilizadas pelo dataset, como por exemplo, idade, gênero, taxa ventricular, taxa atrial, duração do QRS, intervalo QT, eixo R, eixo T, contador de QRS, Q onset, Q offset e T offset. A medida que testes forem sendo executados, o conjunto de variáveis podem sofrer alterações afim de melhorar o processo de classificação.

Por fim, na etapa de classificação, o vetor de características será utilizado para fornecer informação para o classificador *Support-vector machines - SVM's* bastante popular na literatura para problemas de classificação. A depender do desempenho do modelo, pode ser que outro modelo seja analisado para fins comparativos.

2 Trabalho relacionado

Zheng et al. (2020) realizaram o estudo da classificação de arritmias utilizando uma base de dados recente na literatura, produzida pela parceria entre Chapman University and Shaoxing People's Hospital. Esta base de dados contém mais de 10.000 registros. Logo, a aplicação e estudo desta base de dados se faz necessário, pois para a comunidade acadêmica, fontes públicas de dados são de extrema importância e fomentam cada vez mais estudos em diversas áreas. Os autores utilizaram o classificador *extreme gradient boosting tree* obtendo F1-score de 0.97, um valor considerável. Mais um motivo para estudo e

aplicações em outros métodos.

A relação deste trabalho tange o estudo de métodos para classificação de arritmias em ECG's. Neste trabalho, o objetivo é utilizar a mesma base dados aplicada em outro modelo classificador juntamente com a aplicação das normas da AAMI para geração de métodos padronizados de classificação de arritmias.

3 Conjunto de dados

O conjunto de dados mais utilizado na literatura para classificadores de arritmias é o MIT-BIH, dataset desenvolvido pelo Beth Israel Deaconess Medical Center em parceria com Massachusetts Institute of Technology. Porém, Zheng et al. (2020) produziram um novo dataset maior comparado com o MIT-BIH. Este dataset será analisado neste trabalho, permitindo assim uma análise de sua possível utilização em outros trabalhos acadêmicos. Abaixo segue uma breve descrição dos dados.

test_datasets

August 28, 2021

```
[1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import os
import openpyxl as op
import glob
```

```
[2]: files = os.listdir(os.path.abspath('ECGData'))
```

```
[3]: df_attrib = pd.read_excel(os.path.abspath('AttributesDictionary.xlsx'))
df_condnames = pd.read_excel(os.path.abspath('ConditionNames.xlsx'))
df_diagno = pd.read_excel(os.path.abspath('Diagnostics.xlsx'))
df_rhyt = pd.read_excel(os.path.abspath('RhythmNames.xlsx'))
```

```
[4]: df_attrib
```

```
[4]:
```

	Attributes	Type	ValueRange	\
0	FileName	String	NaN	
1	Rhythm	String	NaN	
2	Beat	String	NaN	
3	PatientAge	Numeric	0-999	
4	Gender	String	MALE/FEMAL	
5	VentricularRate	Numeric	0-999	
6	AtrialRate	Numeric	0-999	
7	QRSDuration	Numeric -	0-999	
8	QTInterval	Numeric	0-999	
9	QTCorrected	Numeric	0-999	
10	RAxis	Numeric	-179~180	
11	TAxis	Numeric	-179~181	
12	QRSCount	Numeric	0-254	
13	QOnset	Numeric	16 Bit Unsigned	
14	QOffset	Numeric	17 Bit Unsigned	
15	TOffset	Numeric	18 Bit Unsigned	

	Description
0	ECG data file name(unique ID)
1	Rhythm Label
2	Other conditions Label

```

3             Age
4             Gender
5     Ventricular rate in BPM
6             Atrial rate in BPM
7             QRS duration in msec
8             QT interval in msec
9     Corrected QT interval in msec
10            R axis
11            T axis
12            QRS count
13            Q onset(In samples)
14            Q offset(In samples)
15            T offset(In samples)

```

```
[5]: df_condnames
```

```

[5]:  Acronym Name                                Full Name
0      1AVB                                1 degree atrioventricular block
1      2AVB                                2 degree atrioventricular block
2      2AVB1  2 degree atrioventricular block(Type one)
3      2AVB2  2 degree atrioventricular block(Type two)
4      3AVB                                3 degree atrioventricular block
5      ABI                                atrial bigeminy
6      ALS                                Axis left shift
7      APB                                atrial premature beats
8      AQW                                abnormal Q wave
9      ARS                                Axis right shift
10     AVB                                atrioventricular block
11     CCR                                counterclockwise rotation
12     CR                                clockwise rotation
13     ERV    Early repolarization of the ventricles
14     FQRS                                fQRS Wave
15     IDC                                Interior differences conduction
16     IVB                                Intraventricular block
17     JEB                                junctional escape beat
18     JPS                                J point shift
19     JPT                                junctional premature beat
20     LBBB                                left bundle branch block
21     LBBBB                                left back bundle branch block
22     LFBBS                                left front bundle branch block
23     LRRI                                Long RR interval
24     LVH                                left ventricle hypertrophy
25     LVHV                                left ventricle high voltage
26     LVQRSAL                                lower voltage QRS in all lead
27     LVQRSCL                                lower voltage QRS in chest lead
28     LVQRSLL                                lower voltage QRS in limb lead
29     MI                                myocardial infarction

```

30	MIBW	myocardial infraction in back wall
31	MIFW	Myocardial infraction in the front wall
32	MILW	Myocardial infraction in the lower wall
33	MISW	Myocardial infraction in the side wall
34	PRIE	PR interval extension
35	PWC	P wave Change
36	QTIE	QT interval extension
37	RAH	right atrial hypertrophy
38	RAHV	right atrial high voltage
39	RBBB	right bundle branch block
40	RVH	right ventricle hypertrophy
41	STDD	ST drop down
42	STE	ST extension
43	STTC	ST-T Change
44	STTU	ST tilt up
45	TWC	T wave Change
46	TWO	T wave opposite
47	UW	U wave
48	VB	ventricular bigeminy
49	VEB	ventricular escape beat
50	VFW	ventricular fusion wave
51	VPB	ventricular premature beat
52	VPE	ventricular preexcitation
53	VET	ventricular escape trigeminy
54	WAVN	Wandering in the atrioventricular node
55	WPW	WPW

```
[7]: df_rhyt
```

[7]:	Acronym	Name	Full Name
0	SB		Sinus Bradycardia
1	SR		Sinus Rhythm
2	AFIB		Atrial Fibrillation
3	ST		Sinus Tachycardia
4	AF		Atrial Flutter
5	SI		Sinus Irregularity
6	SVT		Supraventricular Tachycardia
7	AT		Atrial Tachycardia
8	AVNRT	Atrioventricular Node Reentrant	Tachycardia
9	AVRT	Atrioventricular Reentrant	Tachycardia
10	SAAWR	Sinus Atrium to Atrial Wandering	Rhythm

```
[6]: df_diagno
```

[6]:	FileName	Rhythm	Beat	PatientAge	Gender	\
0	MUSE_20180113_171327_27000	AFIB	RBBB TWC	85	MALE	
1	MUSE_20180112_073319_29000	SB	TWC	59	FEMALE	

2	MUSE_20180111_165520_97000	SA	NONE	20	FEMALE
3	MUSE_20180113_121940_44000	SB	NONE	66	MALE
4	MUSE_20180112_122850_57000	AF	STDD STTC	73	FEMALE
...
10641	MUSE_20181222_204306_99000	SVT	NONE	80	FEMALE
10642	MUSE_20181222_204309_22000	SVT	NONE	81	FEMALE
10643	MUSE_20181222_204310_31000	SVT	NONE	39	MALE
10644	MUSE_20181222_204312_58000	SVT	NONE	76	MALE
10645	MUSE_20181222_204314_78000	SVT	NONE	75	MALE

	VentricularRate	AtrialRate	QRSDuration	QTInterval	QTCorrected	\
0	117	234	114	356	496	
1	52	52	92	432	401	
2	67	67	82	382	403	
3	53	53	96	456	427	
4	162	162	114	252	413	
...	
10641	196	73	168	284	513	
10642	162	81	162	294	482	
10643	152	92	152	340	540	
10644	175	178	128	310	529	
10645	117	104	140	312	435	

	RAxis	TAxis	QRSCount	QOnset	QOffset	TOffset
0	81	-27	19	208	265	386
1	76	42	8	215	261	431
2	88	20	11	224	265	415
3	34	3	9	219	267	447
4	68	-40	26	228	285	354
...
10641	258	244	32	177	261	319
10642	110	-75	27	173	254	320
10643	250	38	25	208	284	378
10644	98	-83	29	205	269	360
10645	263	144	19	208	278	364

[10646 rows x 16 columns]

```
[10]: patient1 = pd.read_csv('ECGData/'+ files[0])
      patient1
```

```
[10]:
```

	I	II	III	aVR	aVL	aVF	V1	V2	V3	\
0	-214.72	-229.36	-14.64	224.48	-102.48	-122.00	614.88	814.96	912.56	
1	-200.08	-209.84	-9.76	204.96	-97.60	-112.24	605.12	819.84	927.20	
2	-190.32	-195.20	-4.88	195.20	-92.72	-102.48	600.24	829.60	956.48	
3	-165.92	-165.92	0.00	165.92	-82.96	-82.96	590.48	844.24	971.12	
4	-161.04	-156.16	4.88	161.04	-82.96	-78.08	585.60	844.24	985.76	

```

...      ...      ...      ...      ...      ...      ...      ...      ...      ...
4995 -117.12      9.76  126.88  53.68 -122.00  68.32  439.20  634.40  800.32
4996 -102.48  14.64  117.12  43.92 -112.24  63.44  424.56  634.40  819.84
4997  -92.72  29.28  122.00  34.16 -107.36  73.20  424.56  644.16  829.60
4998  -92.72  34.16  126.88  29.28 -112.24  78.08  424.56  644.16  829.60
4999  -87.84  29.28  117.12  29.28 -102.48  73.20  424.56  653.92  834.48

```

```

      V4      V5      V6
0    126.88 -239.12 -507.52
1    165.92 -195.20 -463.60
2    204.96 -151.28 -429.44
3    239.12 -117.12 -400.16
4    273.28  -78.08 -370.88

```

```

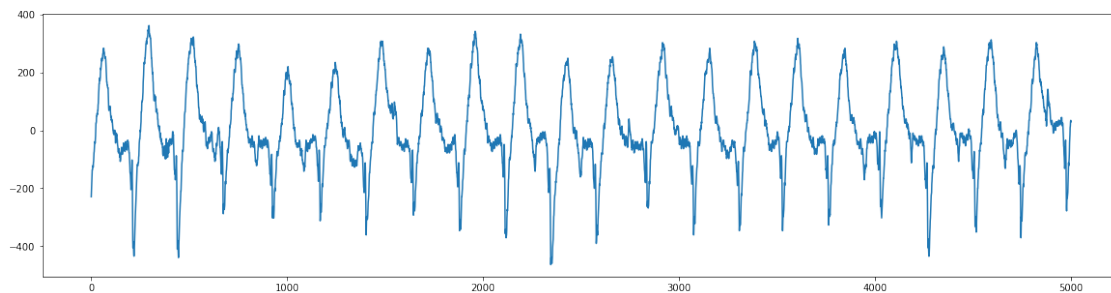
...      ...      ...      ...
4995  278.16      0.00 -273.28
4996  297.68  24.40 -248.88
4997  312.32  43.92 -229.36
4998  312.32  58.56 -204.96
4999  312.32  63.44 -200.08

```

[5000 rows x 12 columns]

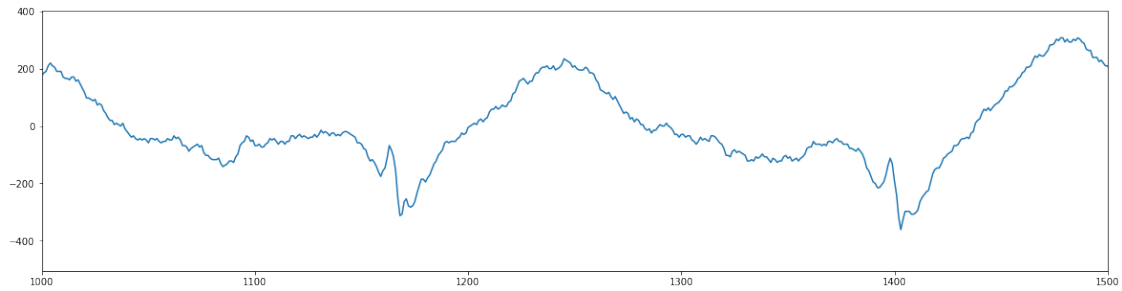
```
[13]: plt.figure(figsize=(20,5))
      patient1['II'].plot()
```

[13]: <AxesSubplot:>



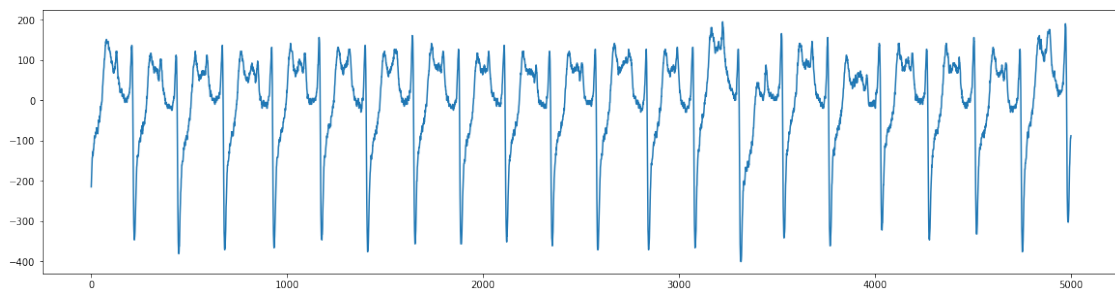
```
[14]: plt.figure(figsize=(20,5))
      plt.xlim(1000,1500)
      patient1['II'].plot()
```

[14]: <AxesSubplot:>



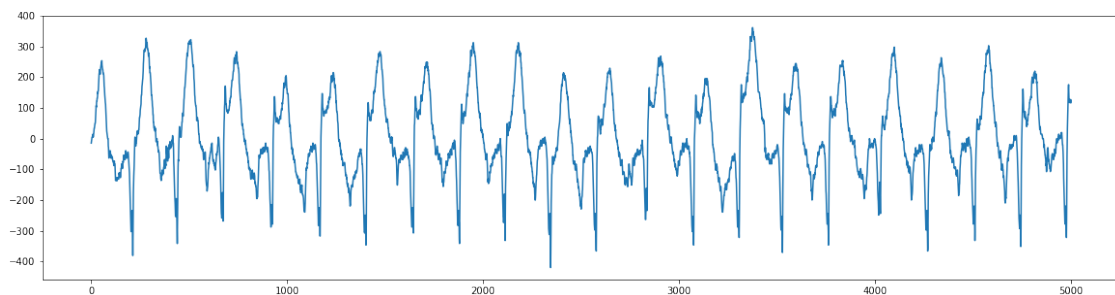
```
[15]: plt.figure(figsize=(20,5))
      patient1['I'].plot()
```

[15]: <AxesSubplot:>



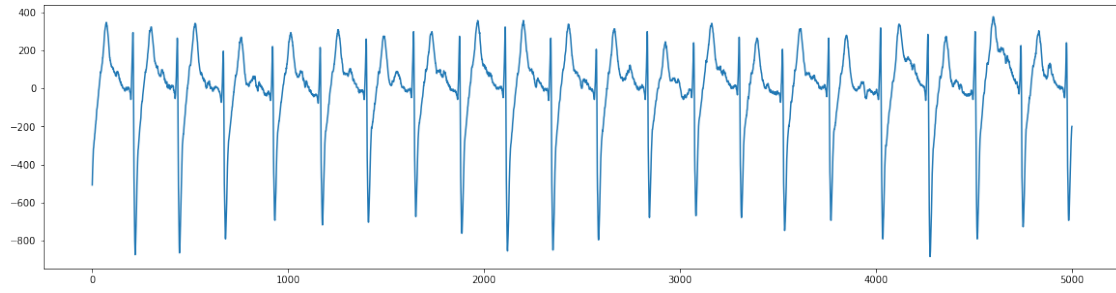
```
[16]: plt.figure(figsize=(20,5))
      patient1['III'].plot()
```

[16]: <AxesSubplot:>



```
[17]: plt.figure(figsize=(20,5))
      patient1['V6'].plot()
```

[17]: <AxesSubplot:>



References

ANSI/AAMI. *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. 2008. American National Standards Institute, Inc. (ANSI), Association for the Advancement of Medical Instrumentation (AAMI). ANSI/AAMI/ISO EC57, 1998-(R)2008.

COHEN, A. *Biomedical Signal Processing*. [S.l.]: CRC Press., 1988.

HANNUN, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, Nature Publishing Group, v. 25, n. 1, p. 65, 2019.

ZHENG, J. et al. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Scientific data*, Nature Publishing Group, v. 7, n. 1, p. 1–8, 2020.