

Universidade Federal de Santa Maria - UFSM
Centro de Tecnologia - CT
Curso de Engenharia de Computação
ELC1147 - Inteligência Artificial Aplicada ao Processamento de Sinais
Biomédicos

IDENTIFICAÇÃO DE ATIVIDADE FÍSICA ATRAVÉS DE MODELOS DE INTELIGÊNCIA ARTIFICIAL NO PROCESSAMENTO DE SINAIS DE ECG

Luis Felipe de Deus – felipe.deus@ecomp.ufsm.br
Leonardo Ferreira - leonardoferreira@mail.ufsm.br
Tiago Knorst – tiago.knorst@ecomp.ufsm.br

Julho/2019

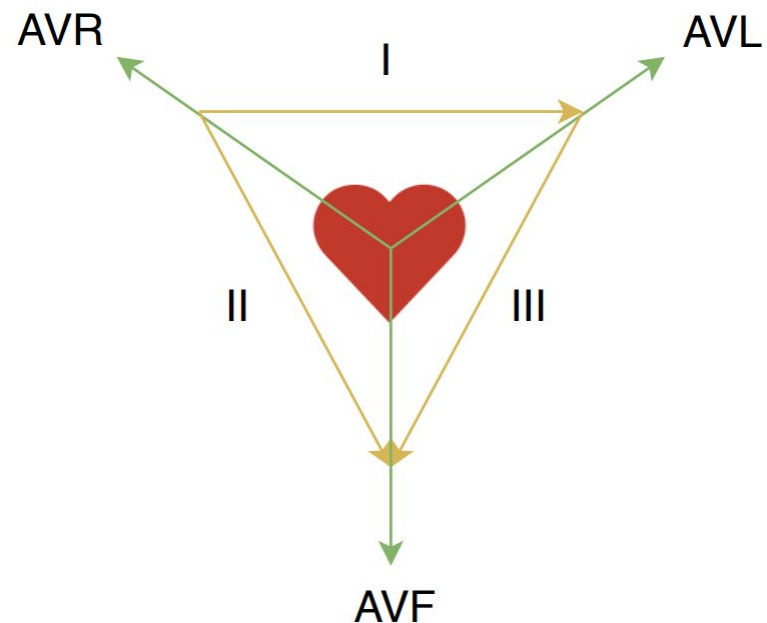
Agenda

- Introdução
- Objetivo
- Metodologia
- Desenvolvimento
- Resultados
- Conclusões

Introdução

Derivações Eletrocardiográficas

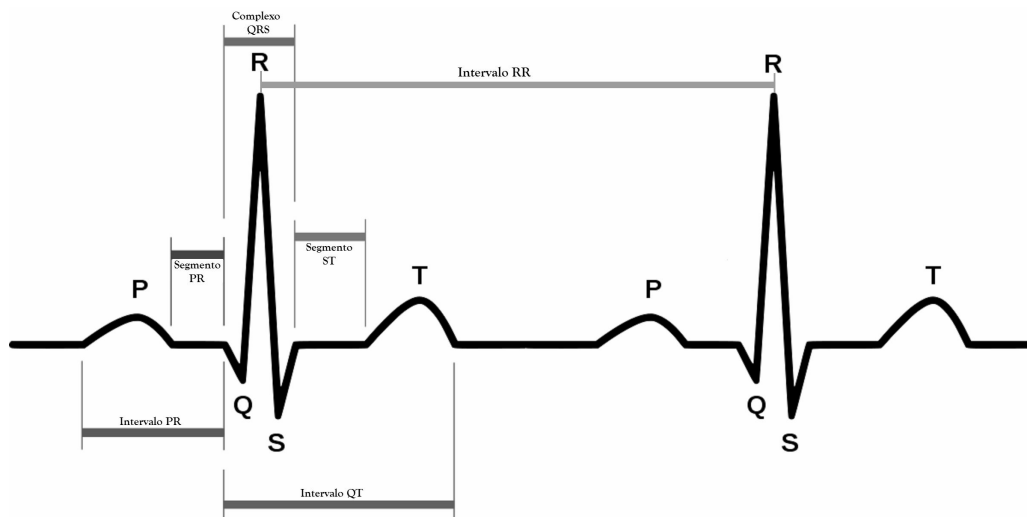
- Conforme o tipo de informação clínica necessária;
- Seis periféricas;
- Seis precordiais.



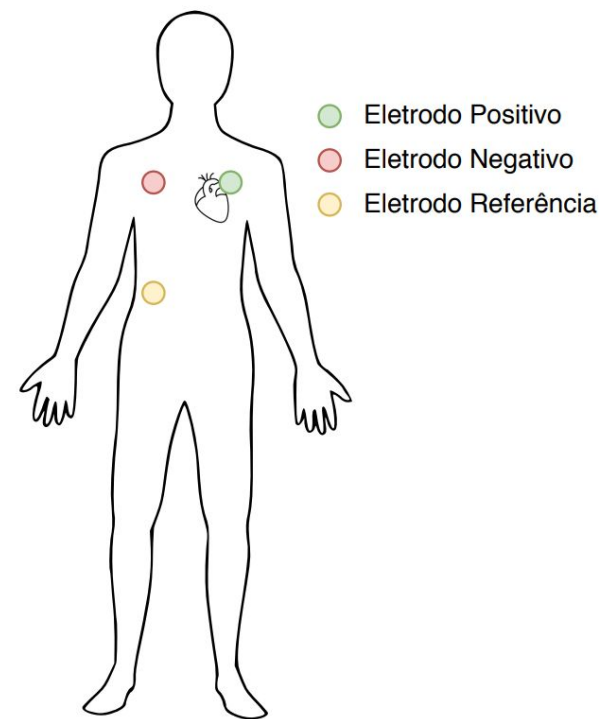
Introdução

Característica do sinal ECG

- Tipo de ondas;
- Intervalo R-R;
- Derivação I utilizada;
- Posicionamento dos eletrodos.



Fonte: Revista brasileira de cardiologia v17 n3.



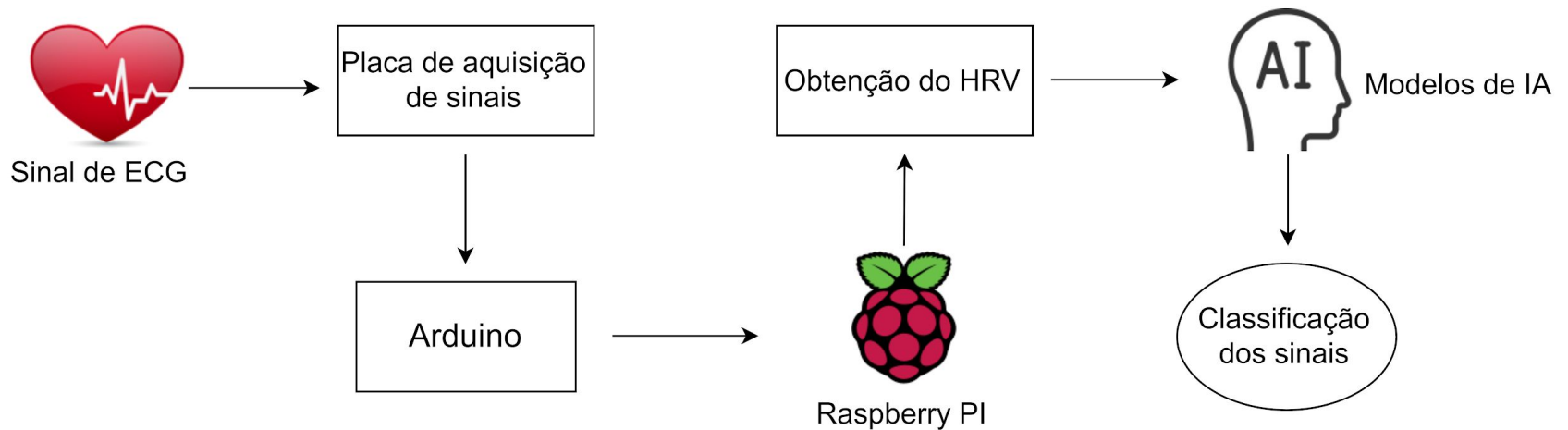
Objetivo

- Detectar se um indivíduo está, ou não, em atividade física;
- Classe 1: Atividade;
- Classe 0: Repouso.



Objetivo

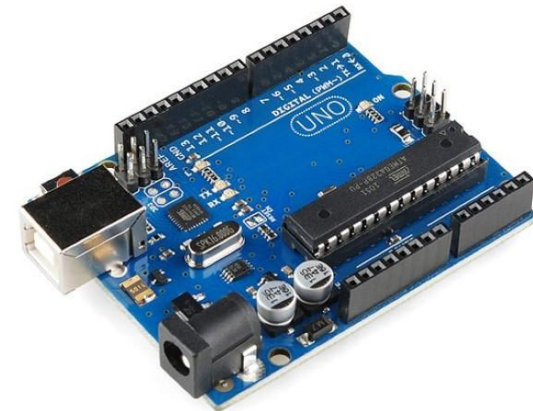
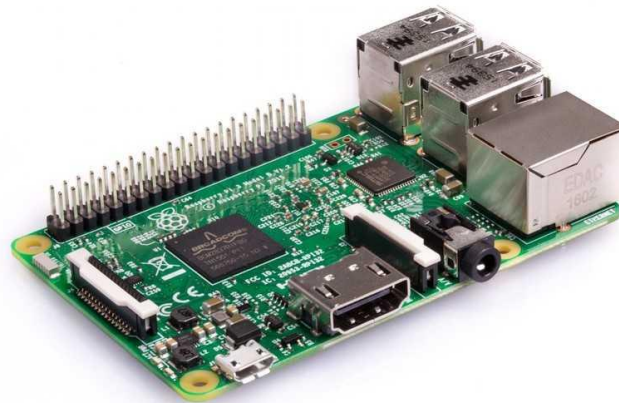
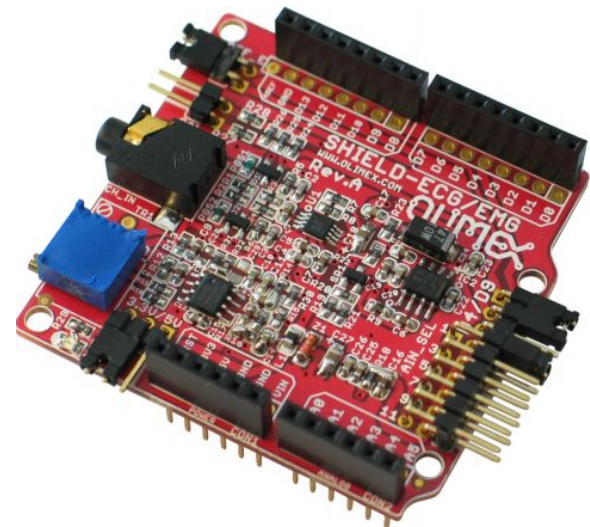
- Fluxograma de execução do projeto.



Metodologia

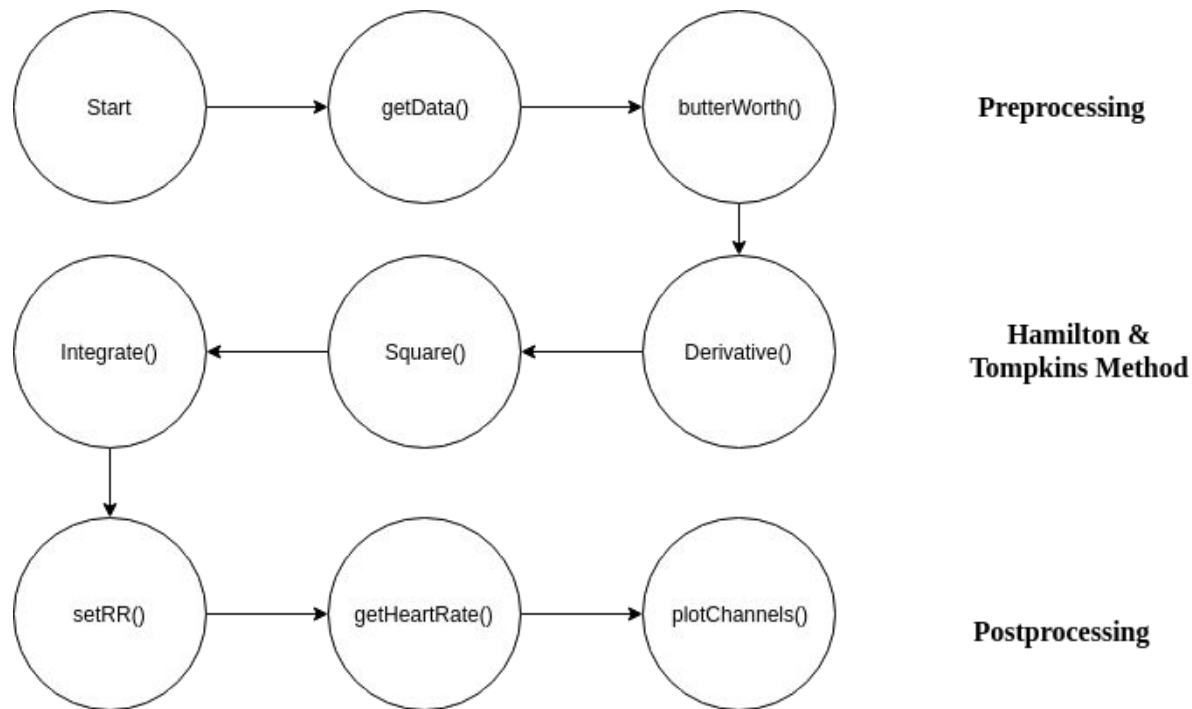
Hardware Utilizado

- *Shield Olimex;*
- *Arduino Uno;*
- *Raspberry PI 3 model B+.*



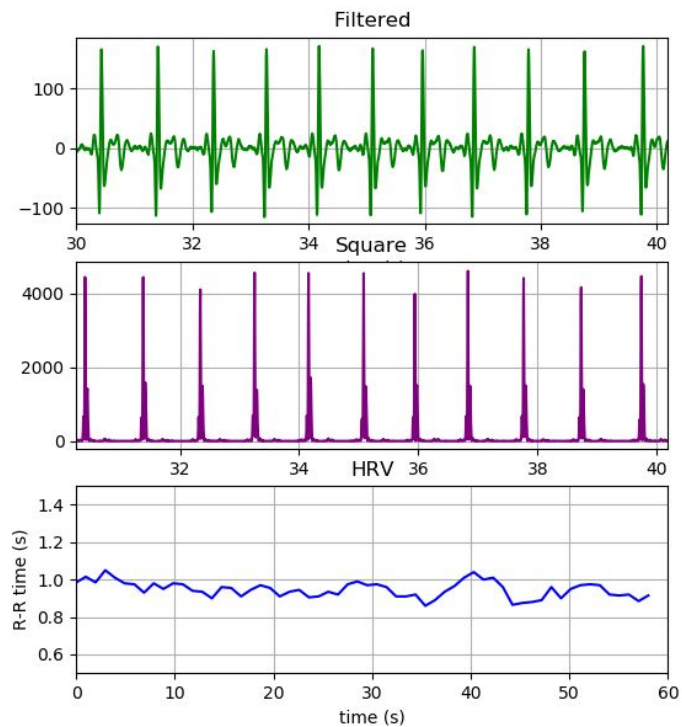
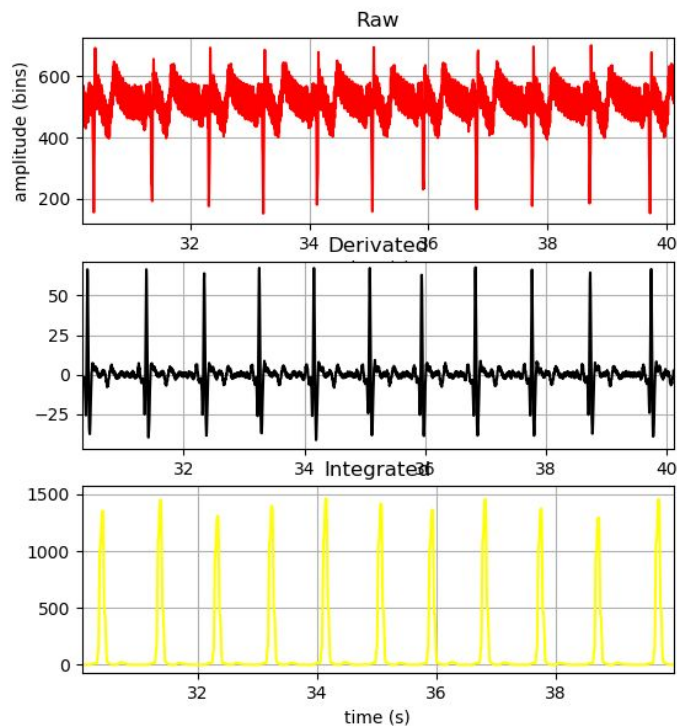
Metodologia

Método de Hamilton & Tompkins:



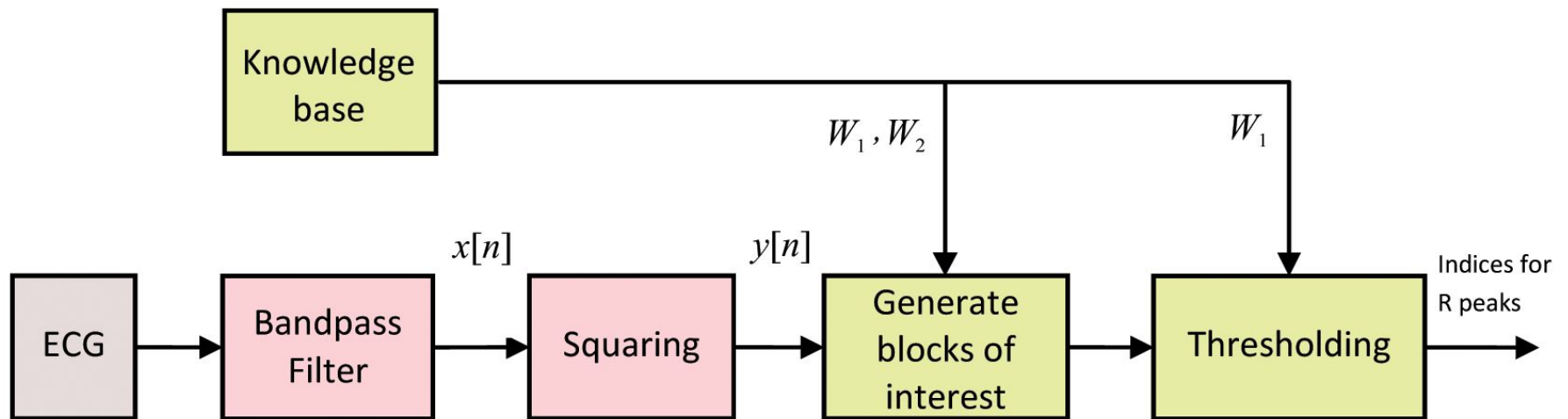
Metodologia

Método de Hamilton & Tompkins:



Metodologia

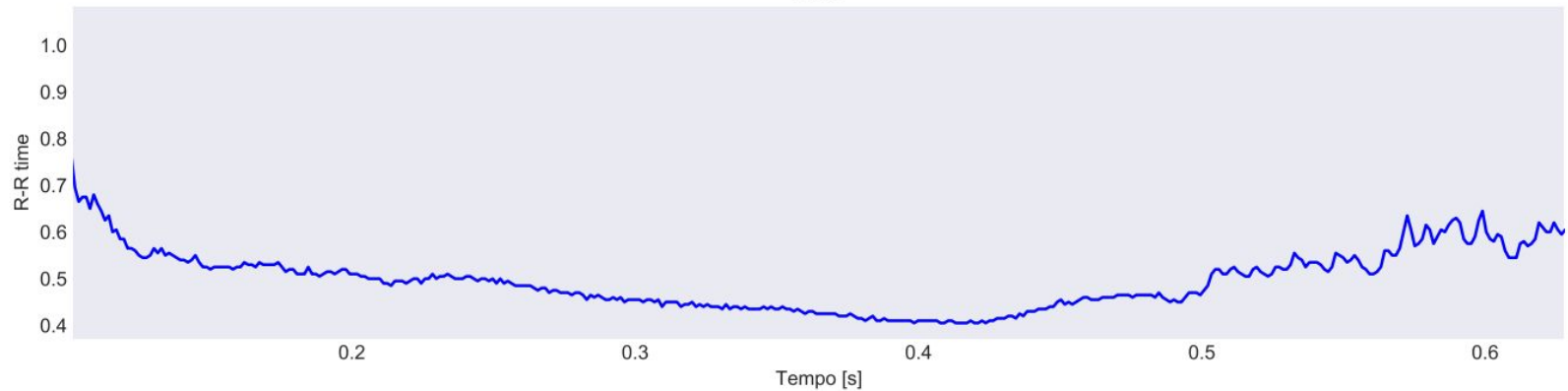
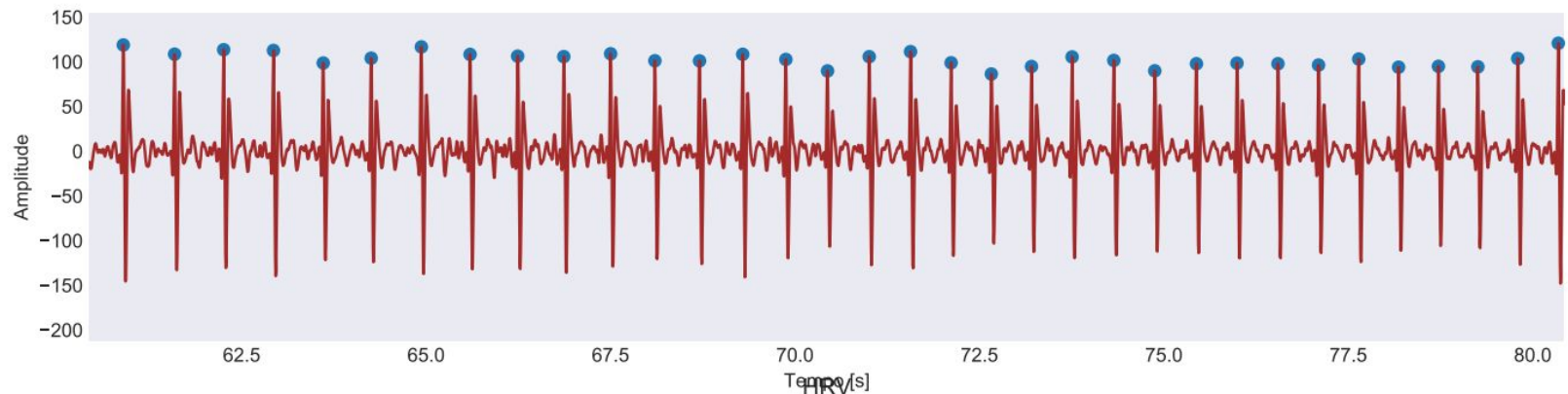
Método de Elgendi:



Fonte: (Elgendi, 2013).

Metodologia

Método de Elgendi:



Metodologia

Features utilizadas:

- Somente no domínio do tempo;
- Tempo de amostra;
- Custo computacional.

Tabela de <i>Features</i> Utilizadas	
Mean	Média do intervalo R-R
SDNN	Desvio Padrão do intervalo R-R
RMSSD	Raiz quadrada média do intervalo R-R
Median	Mediana do intervalo R-R
Var	Variância do intervalo R-R
Range	Maior variação do intervalo R-R
CVSD	Coefficiente de variação do intervalo R-R

Metodologia

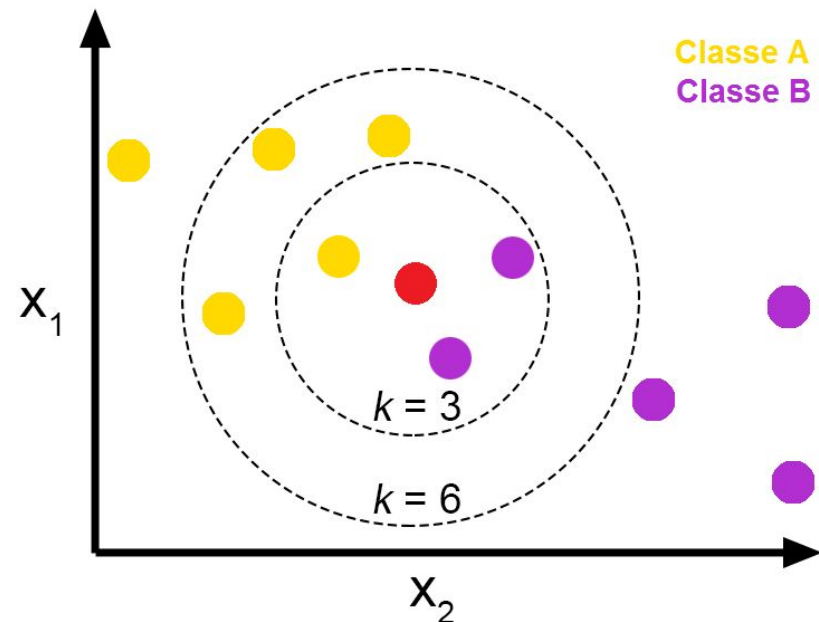
Algoritmos de Aprendizado de Máquina Supervisionado:

- *K-Nearest Neighbors* (KNN);
- *Random Forest*.

Desenvolvimento

K-Nearest Neighbors (KNN):

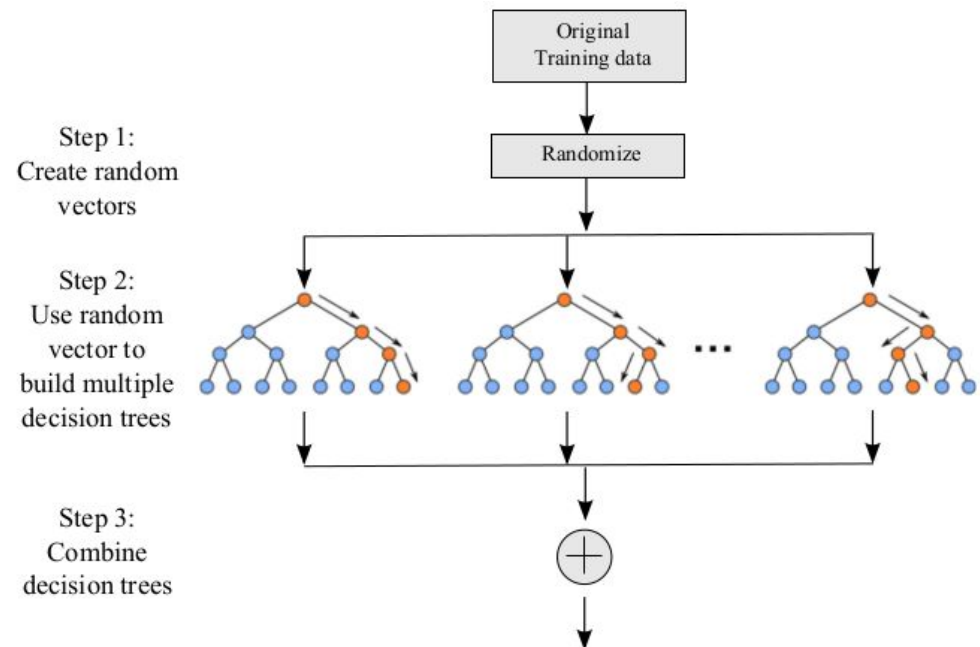
- Calcula para cada dado a ser classificado, a distância até os pontos do conjunto de treino;
- Elege a classe do novo dado por votação entre os K vizinhos;
- Utilizada a distância Euclidiana;
- $K = 3$ vizinhos.



Desenvolvimento

Random Forest:

- Conjunto (*Ensemble*) de árvores de decisão;
- Utilizadas 100 árvores;
- Adiciona aleatoriedade ao modelo;
- Evita ou minimiza *Overfitting*.



Desenvolvimento

Protocolo:

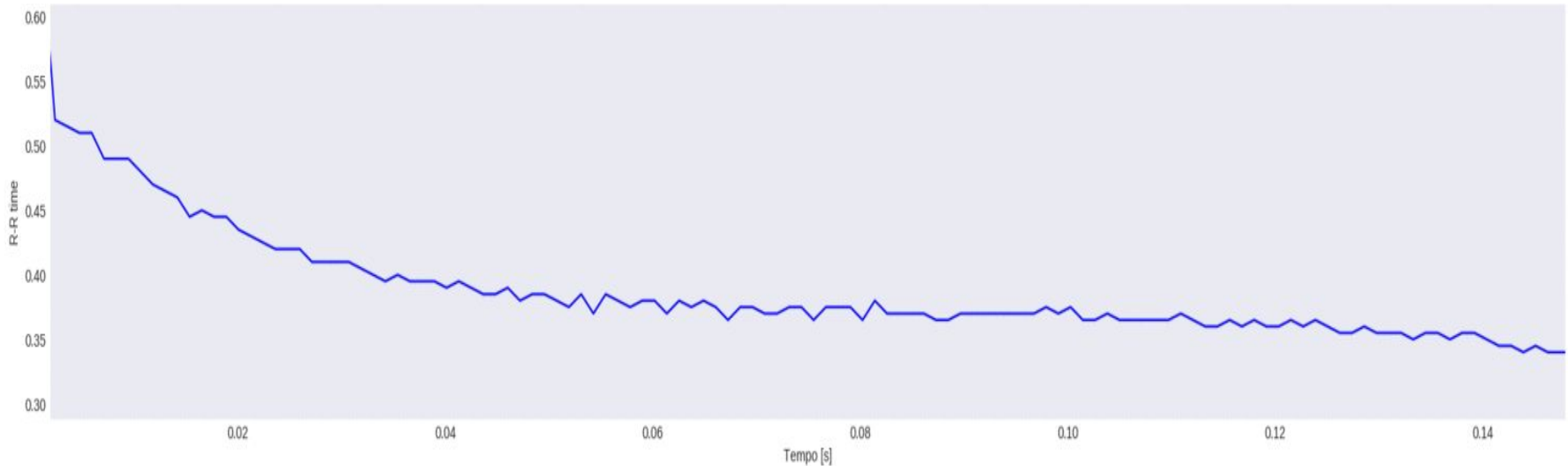
- 1 minuto de repouso;
- 2 minutos atividade física intensa;
- 2 minutos de recuperação.

Data Frame: Sete sinais de ECG obtidos dos autores

Metodologia

Extração de *features*:

- Janela móvel no sinal HRV;
- 1° ao 10° - s1 | 2° ao 11° - s2 | 3° ao 12° - s3 ...



Desenvolvimento

Frame de dados utilizado nos classificadores

<i>Features</i>	s1	s2	s3	sn	s3342
mean	0.777346	0.784846	0.785846	...	0.800180
sdnn	0.092493	0.081944	0.080988	...	0.027652
rmssd	0.074407	0.069519	0.067096	...	0.016400
sdsd	0.073906	0.069464	0.067039	...	0.016269
median	0.802513	0.802513	0.802513	...	0.802513
var	0.008555	0.006715	0.006559	...	0.000765
range	0.440007	0.345006	0.345006	...	0.100002
cvsd	0.095719	0.088577	0.085381	...	0.020496
activity	0	0	0	...	1

Resultados

Exemplo de execução do *Random Forest* em Raspberry PI

```
pi@raspberrypi: ~/Desktop/dev/processing
### RUNNING RANDOM FOREST ###
---- Accuracy: 0.980059820538
[[106  10]
 [ 10 877]]
True Positive:  877
False Positive:  10
True Negative: 106
False Negative:  10
```

	precision	recall	f1-score	support
0.0	0.91	0.91	0.91	116
1.0	0.99	0.99	0.99	887
accuracy			0.98	1003
macro avg	0.95	0.95	0.95	1003
weighted avg	0.98	0.98	0.98	1003

```
AUC: 0.996
```

Resultados

Exemplo de execução do KNN em Raspberry PI

```
pi@raspberrypi: ~/Desktop/dev/processing
### RUNNING KNN ###
--- Accuracy: 0.979062811565
[[126  9]
 [ 12 856]]
True Positive: 856
False Positive: 9
True Negative: 126
False Negative: 12

```

	precision	recall	f1-score	support
0.0	0.91	0.93	0.92	135
1.0	0.99	0.99	0.99	868
accuracy			0.98	1003
macro avg	0.95	0.96	0.96	1003
weighted avg	0.98	0.98	0.98	1003

```
AUC: 0.989
```

Resultados

Primeira análise:

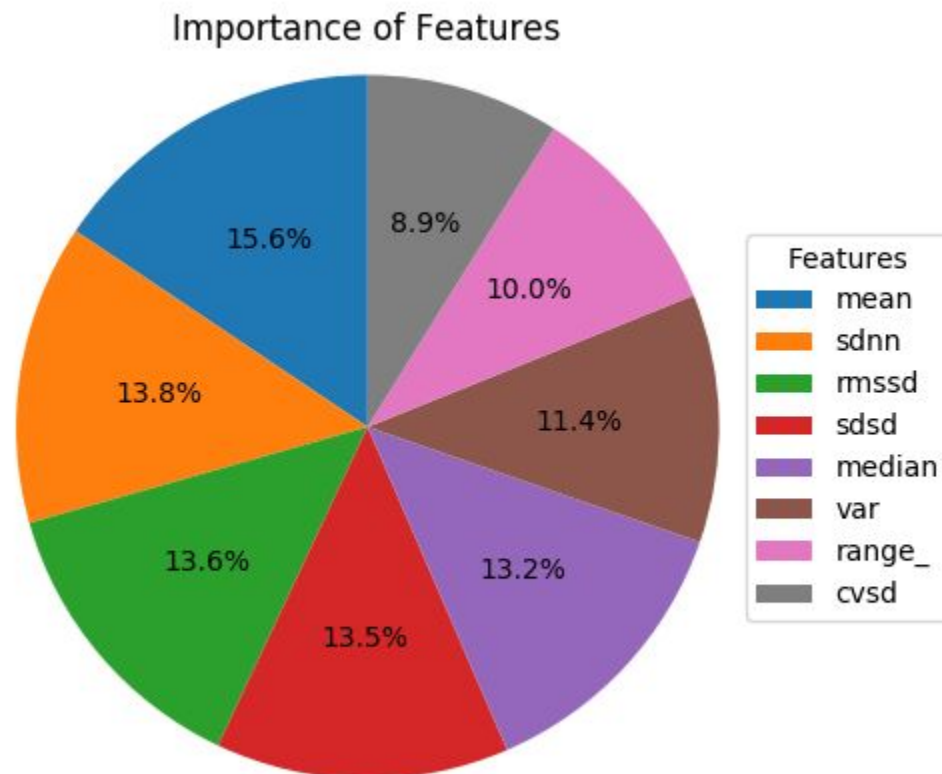
- Utilização de todas as *features*;
- Janela de 10 pontos por amostra.

Random Forest: Resultado da primeira análise								
Class	Precision	Recall	F1	Accuracy	TP	FP	TN	FN
0	0.80	0.79	0.80	0.9444	873	30	114	28
1	0.97	0.97	0.97					

KNN: Resultado da primeira análise								
Class	Precision	Recall	F1	Accuracy	TP	FP	TN	FN
0	0.88	0.78	0.82	0.9511	874	34	120	17
1	0.96	0.98	0.97					

Resultados

Importância das *features* usando como base a primeira análise:



Resultados

Segunda análise:

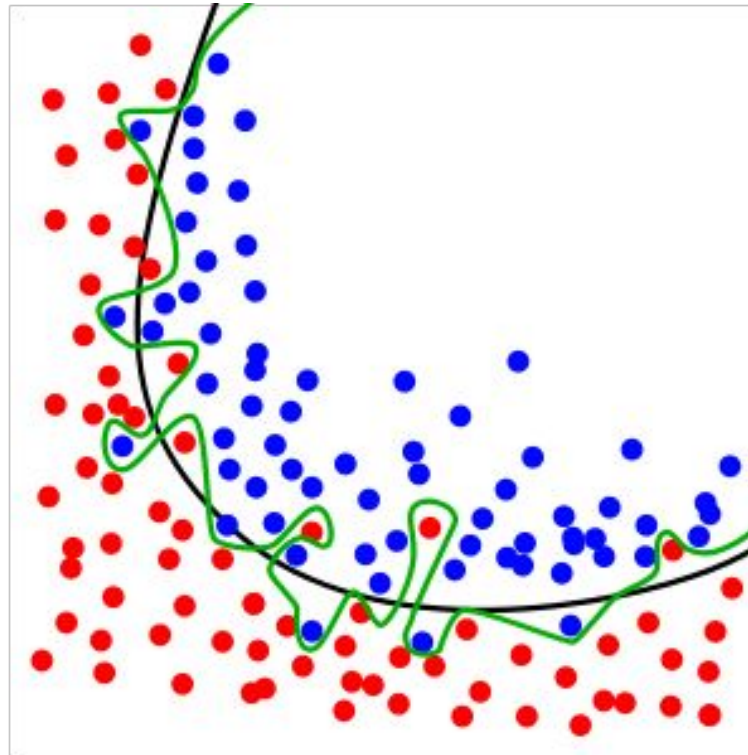
- Utilização de três *features* (Mean, SDNN, RMSSD);
- Janela de 10 pontos por amostra.

Random Forest: Resultado da segunda análise								
<i>Class</i>	<i>Precision</i>	<i>Recall</i>	F1	<i>Accuracy</i>	TP	FP	TN	FN
0	0.83	0.76	0.79	0.9425	870	36	115	24
1	0.96	0.97	0.97					

KNN: Resultado da segunda análise								
<i>Class</i>	<i>Precision</i>	<i>Recall</i>	F1	<i>Accuracy</i>	TP	FP	TN	FN
0	0.71	0.80	0.75	0.9301	863	28	109	45
1	0.97	0.95	0.96					

Resultados

- *Overfitting ?*



Resultados

Overfitting ?

- Novo sinal de ECG de um 4º indivíduo;
- Protocolo:
 - 3 minutos de repouso;
 - 12 minutos de atividade física moderada;
 - 3 minutos de recuperação.



Resultados

Terceira análise:

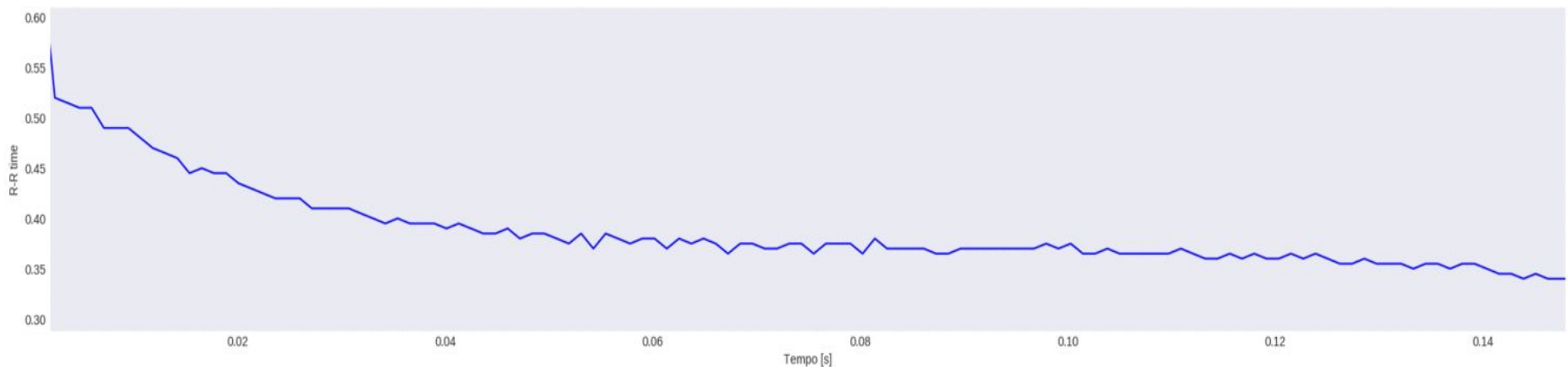
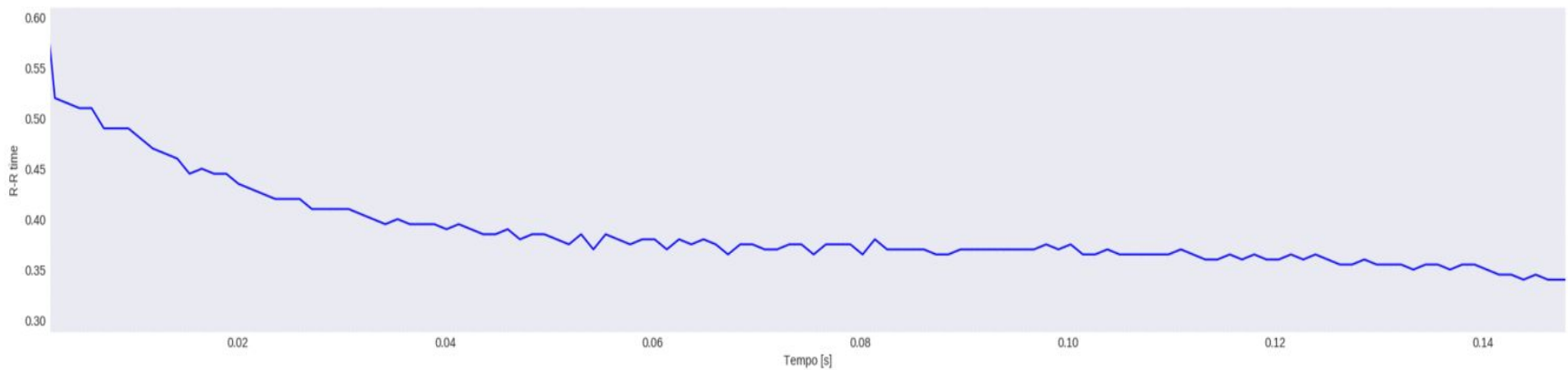
- Utilização de três *features* (Mean, SDNN, RMSSD);
- Janela de 10 pontos por amostra;
- Indivíduo de teste diferente dos de treino.

Random Forest: Resultado da terceira análise								
Class	Precision	Recall	F1	Accuracy	TP	FP	TN	FN
0	0.70	0.78	0.74	0.9365	1724	50	181	79
1	0.97	0.96	0.96					

KNN: Resultado da terceira análise								
Class	Precision	Recall	F1	Accuracy	TP	FP	TN	FN
0	0.54	0.73	0.62	0.8997	1661	62	169	142
1	0.96	0.92	0.94					

Resultados

- Tamanho da janela aumenta ou diminui o desempenho ?



Resultados

Quarta análise:

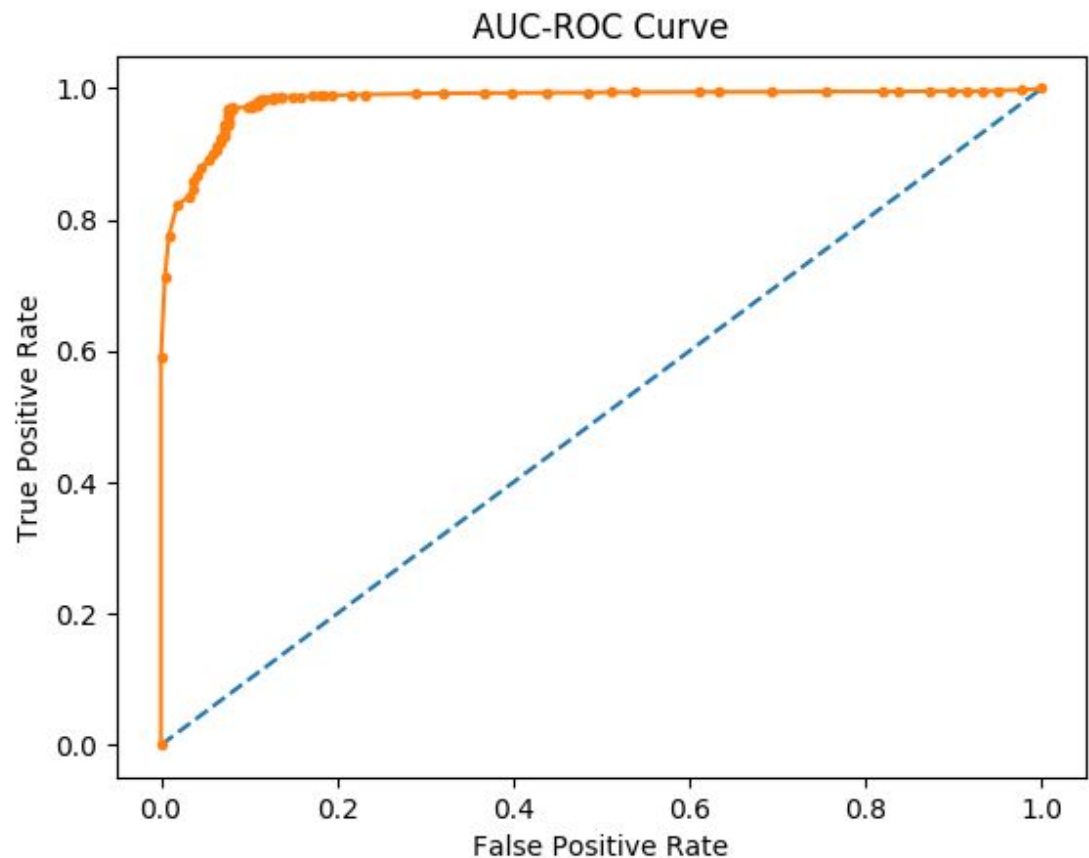
- Utilização de três *features* (Mean, SDNN, RMSSD);
- Janela de 20 pontos por amostra;
- Indivíduo de teste diferente dos de treino.

Random Forest: Resultado da quarta análise								
Class	Precision	Recall	F1	Accuracy	TP	FP	TN	FN
0	0.78	0.90	0.84	0.9614	1748	23	198	55
1	0.99	0.97	0.98					

KNN: Resultado da quarta análise								
Class	Precision	Recall	F1	Accuracy	TP	FP	TN	FN
0	0.70	0.88	0.78	0.9456	1720	27	194	83
1	0.98	0.95	0.97					

Resultados

- Curva AUC-ROC para *Random Forest*;
- AUC: 0.98.



Conclusões

- Trabalho propiciou o aprendizado sobre:
 - Área biomédica
 - Aquisição de sinais
 - Agrupamento de dados
 - Algoritmos de aprendizado de máquina
 - Métricas de desempenho
 - Metodologias de pesquisa

Conclusões

- Resultados obtidos:
 - Satisfatórios tanto para identificação de repouso quanto para a identificação de atividade
- Aplicabilidade em aplicações:
 - Desportivas
 - Médicas
- Trabalho futuro
 - Pesquisa por bases de dados que abrange pessoas de diferentes idades e condições de saúde

Referências

- [1] M. Elgendi, “Fast QRS Detection with an Optimized Knowledge-Based Method: Evaluation on 11 Standard ECG Databases,” PLoS ONE , 2013.
- [2] S. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” 2007.
- [3] G. D. Clifford, “Signal Processing Methods for Heart Rate Variability,” Ph.D. dissertation, 2002.
- [4] “Olimex,” Olimex LTD. [Online]. Available: www.olimex.com.
- [5] S. Narkhede, “Understanding AUC - ROC Curve - Towards Data Science,” 2018. [Online]. Available: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>.

Obrigado pela Atenção !

Perguntas?



Luis Felipe de Deus – felipe.deus@ecomp.ufsm.br

Leonardo Ferreira - leonardoferreira@mail.ufsm.br

Tiago Knorst – tiago.knorst@ecomp.ufsm.br