

# Relatório

## Trabalho Prático 2

Diogo Jesus, 2013137326

João Roque, 2013132242

**Informática Clínica e Sistemas de Tele-Saúde**

# 1 Introdução

Neste trabalho tinha-se como objectivos:

- Desenvolvimento de um programa para classificação de Contrações Ventriculares Prematuras;
- Desenvolvimento de um programa para a classificação de Fibrilação Atrial.

## 2 Desenvolvimento do Trabalho

### 2.1 PVC

#### Regras e Métodos Utilizados

Sendo um batimento PVC categorizado por um intervalo RR menor e por um complexo QRS maior, decidimos inicialmente explorar estas características para o nosso conjunto de regras. Tendo disponíveis os índices dos picos R, construímos um vector de features inicial que era constituído pelo valor ao último pico (Intervalo RR) e também pela distância do pico R ao pico S. Seleccionámos esta última distância como medida do tamanho do complexo QRS, que nos pareceu ser bastante distintiva numa primeira análise visual aos sinais. Para o seu cálculo, era procurado o mínimo numa janela de 120 samples (120 *ms*) à frente do pico R.

Ao analisar os boxplots das duas features exploradas observámos que apenas o tamanho do intervalo QS demonstrava capacidade de classificação para o problema. Isto pode ser explicado pelo simples facto de que nem todos os picos são identificados, logo realizar o cálculo do intervalo RR pelos indexes dos picos fornecidos é naturalmente sujeito a erro. Assim sendo, apenas este valor foi utilizado para a regra.

Para selecção do threshold utilizado analisámos o resultado da feature em todos os sinais, chegando à conclusão de que se o valor fosse acima de 50 *ms* considerava-se o batimento como sendo um PVC. De salientar que muitos dos sinais fornecidos apresentavam bastantes picos mal identificados, alguns dos quais estavam assinalados em mínimos dos sinais, logo impossibilitavam o correcto uso da nossa regra. Inicialmente construímos a regra com os 4 que *a priori* considerámos que estavam adequados, tendo apenas depois aplicado aos restantes.

#### Resultados

Aplicámos o nosso modelo a todos os dados fornecidos para a realização do trabalho, e os parâmetros de sensibilidade, especificidade e exactidão podem ser identificados na tabela

1. Os modelos com o melhor resultado foram os DPVC\_116, DPVC\_221, DPVC\_119 e DPVC\_228, nos quais a regra que utilizámos funciona bem.

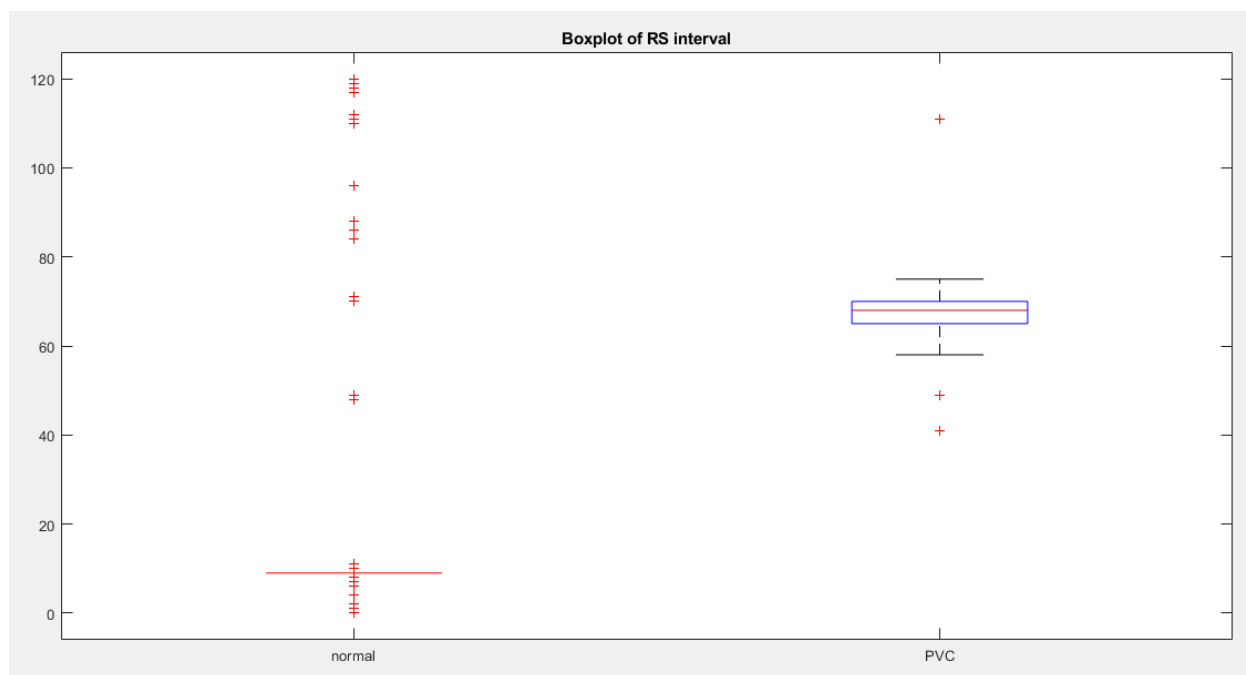


Figure 1: Boxplots por classe do intervalo RS (amostra DPVC116)

	Sensibilidade	Especificidade	Exactidão
DPVC_116	0,981651	0,992647	0,992152
DPVC_201	0	0,870722	0,78617
DPVC_221	0,967172	0,929332	0,935418
DPVC_233	0,021661	0,01508	0,016815
DPVC_119	1	0,969697	0,976122
DPVC_203	0,092342	0,896357	0,781461
DPVC_223	0,285412	0,340092	0,330306
DPVC_106	0,440385	0,968948	0,837941
DPVC_200	0,041162	0,906918	0,650788
DPVC_210	0,07732	0,659173	0,617132
DPVC_228	0,737569	0,759415	0,755722
Média	0,422243	0,755307	0,698184

Table 1: Resultados PVC (os 4 sinais mais adequados estão assinalados)

## 2.2 AF

### Regras e Métodos Utilizados

Um sinal ECG com Fibrilação Arterial distingue-se por não apresentar ondas P e também pela irregularidade ao nível dos intervalos RR. Decidimos explorar a característica referente à inexistência da onda P, pois novamente nos intervalos RR já esperávamos um grande erro associado que iria influenciar imenso os resultados. A identificação correcta de todos os picos é extremamente difícil, ainda mais quando o algoritmo Pan-Tompkins utilizado não é completo, e não possui um threshold adaptável para assim manter uma performance elevada ao longo do sinal. Ao utilizar uma contagem de ondas P em cada intervalo definido, podemos esquivar-nos um pouco a este problema.

De modo a averiguarmos a presença da onda P, iniciámos por aplicar um filtro passa-alto com frequência de corte de 0.5 Hz (para a eliminação de artefactos de baixa-frequência). De seguida, utilizámos os picos detectados e procurámos numa janela de 30 *ms* à esquerda do pico o mínimo d.o sinal. Numa janela de 100 *ms* à esquerda deste mínimo calculámos o desvio padrão, sendo nesta janela que esperamos encontrar a onda P. Em teoria, o desvio padrão seria menor para as janelas onde estivesse presente o ritmo AF, visto que o sinal deveria ser semelhante a uma flatline.

No entanto, ao analisarmos os resultados verificou-se o contrário, sendo os valores do desvio padrão distintos para janelas com AF mas apresentando valores maiores. Estes, pensamos nós, podem ser explicados pelo facto dos intervalos RR serem menores para AF sendo a janela de 100 *ms* suficiente para apanhar o batimento anterior, aumentando assim o valor de desvio padrão. Numa janela normal tal não se vai suceder com tanta frequência. Este pensamento está em concordância com a análise visual dos diferentes sinais fornecidos, onde verificámos que obtivemos melhores resultados, nos sinais onde a variação do ritmo era claramente visível.

Finalmente, para a detecção do AF, utilizámos uma janela de 5 segundos, que se sobrepõe à anterior em 2.5 segundos. Em cada janela fizémos uma contagem de pontos que estavam acima e abaixo de  $1.35 \times \text{Média de todos os desvios padrões}$ . Utilizámos este *threshold*, pois queríamos situar valores um pouco acima da média, e que fossem claramente discrepantes do resto do sinal. Calculámos a percentagem de pontos acima, e se esta fosse acima de 20%, considerámos a janela como AF. Algumas destas análises podem ser vistas nas imagens seguintes.

## Resultados

A regra que criámos para o AF apresentou resultados bastante satisfatórios. Na figura 2 pode observar-se os desvios padrões calculados ao longo do sinal, e o *threshold* ( $1.35 \times \text{Media}$  de todos os desvios padrões) utilizado. Comparando com o sinal ECG da figura 3 é possível visualizar a correlação entre a nossa feature e o sinal classificado como AF.

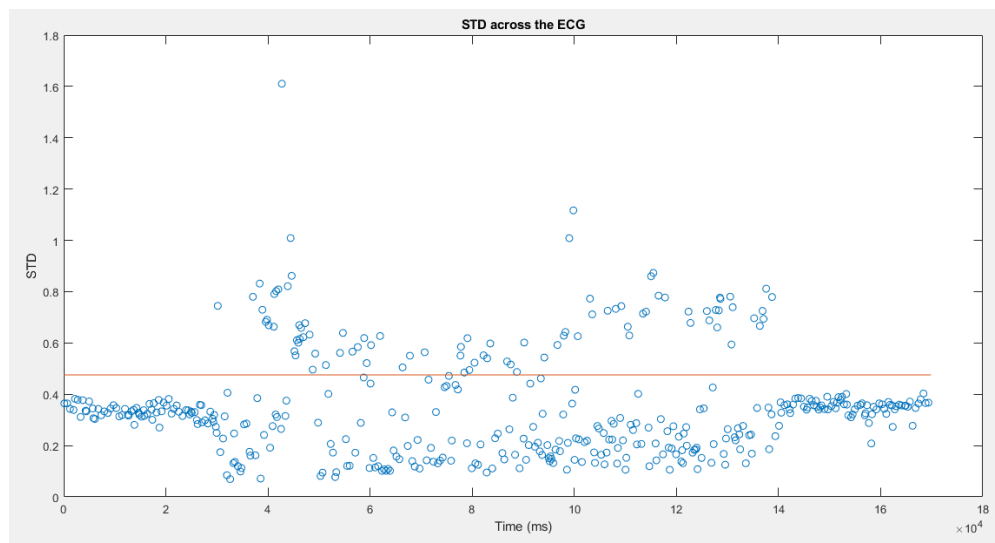


Figure 2: Desvio padrão ao longo do sinal ECG (amostra 04043-episode1)

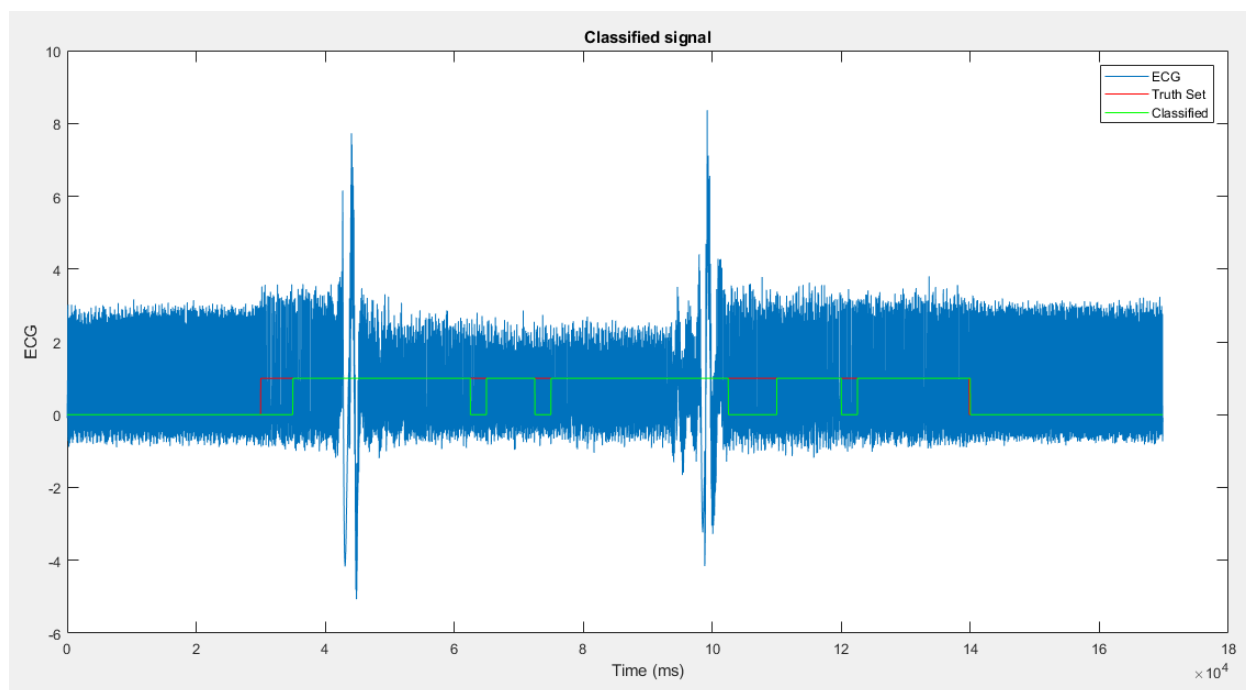


Figure 3: Classificação do sinal ECG (amostra 04043-episode1)

	Sensibilidade	Especificidade	Exactidão
afdb_file-04043_episode-1	0,8179	0,9972	0,8812
afdb_file-04043_episode-2	0,8549	0,8289	0,8347
afdb_file-04043_episode-3	0,9523	0,5417	0,8029
afdb_file-04043_episode-4	0,7821	0,4903	0,5967
afdb_file-04048_episode-1	1,0000	0,9049	0,9323
afdb_file-04048_episode-2	1,0000	0,9245	0,9333
afdb_file-04048_episode-3	1,0000	0,8838	0,8936
afdb_file-04746_episode-1	1,0000	0,8946	0,8966
afdb_file-04746_episode-2	1,0000	0,7500	0,7557
afdb_file-04746_episode-3	1,0000	0,8178	0,8358
afdb_file-05261_episode-1	1,0000	0,6392	0,6442
afdb_file-05261_episode-2	1,0000	0,8180	0,8296
afdb_file-05261_episode-3	1,0000	0,7994	0,8351
afdb_file-05261_episode-4	1,0000	0,7304	0,7465
afdb_file-08219_episode-1	0,9489	0,2917	0,5867
afdb_file-08219_episode-2	0,9454	0,6926	0,8453
afdb_file-08219_episode-3	0,7454	0,6980	0,7277
afdb_file-08219_episode-4	0,8748	0,7917	0,8447
Média	0,9401	0,7497	0,8013

Table 2: Resultados AF

### 3 Discussão e Conclusão

Quanto ao PVC, a regra de classificação que designámos mostrou-se muito boa em alguns casos, e má noutros. Como esta era extremamente baseada na boa identificação dos picos R, cremos que isto induziu um erro muito grande em casos que imensos picos não estavam bem identificados. Para melhorar a regra, será necessário criar uma regra que não esteja tão dependente deste factor, e que tenha em atenção as próprias limitações do dataset.

No caso do AF, tivémos mais atenção aos problemas da regra anterior, pois utilizámos contagens gerais de cada janela, o que faz com que alguns picos mal identificados não tenham tanto peso no resultado geral. Assim, embora este fosse um problema mais difícil, conseguimos obter resultados melhores. O desvio padrão revelou-se uma boa feature para utilizar na classificação pois o tempo entre picos R apresentou-se bastante variado quando existia AF. Como referido anteriormente, a nossa janela podia por vezes encontrar o pico anterior quando o tempo entre picos era menor (o que acontece sobretudo aquando da AF), que levava a uma variação maior. No entanto quando a janela não alcança o pico anterior, e não existe onda P, a variação será menor. Assim, como vemos na figura 2, na zona de AF os pontos do desvio padrão encontram-se tanto acima como abaixo da média. Assim, ao estabelecer-mos uma relação entre estes dois conjuntos de pontos, foi possível classificar a AF com sensibilidade, especificidade e exatidão todas bastante elevadas.

Assim, ao longo do trabalho, notámos um melhoramento das features que utilizámos, ao termos em atenção aos problemas que surgiram, obtendo resultados bastante razoáveis.