



AVALIAÇÃO DO RITMO CARDÍACO EM ELETROCARDIOGRAMAS DE  
CURTA DURAÇÃO UTILIZANDO ANÁLISE DOS INTERVALOS RR E  
APRENDIZADO SUPERVISIONADO

Gabriel de Azevedo Limeira

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Biomédica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Biomédica.

Orientadores: Jurandir Nadal

Paulo Roberto Benchimol Barbosa

Rio de Janeiro

Março de 2019

AVALIAÇÃO DO RITMO CARDÍACO EM ELETROCARDIOGRAMAS DE  
CURTA DURAÇÃO UTILIZANDO ANÁLISE DOS INTERVALOS RR E  
APRENDIZADO SUPERVISIONADO

Gabriel de Azevedo Limeira

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO  
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)  
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM  
CIÊNCIAS EM ENGENHARIA BIOMÉDICA.

Examinada por:

---

Prof. Jurandir Nadal, D.Sc.

---

Prof. Antonio Mauricio Ferreira Leite Miranda de Sá, D.Sc.

---

Prof. José Manoel de Seixas, D.Sc.

RIO DE JANEIRO, RJ BRASIL

MARÇO DE 2019

Limeira, Gabriel de Azevedo.

Avaliação do Ritmo Cardíaco em Eletrocardiogramas de Curta Duração Utilizando Análise dos Intervalos RR e Aprendizado Supervisionado/ Gabriel de Azevedo Limeira. – Rio de Janeiro: UFRJ/COPPE, 2019.

IX, 108 p.: il.; 29,7 cm.

Orientadores: Jurandir Nadal

Paulo Roberto Benchimol Barbosa

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Biomédica, 2019.

Referências Bibliográficas: p. 104-108.

1. Ritmo cardíaco. 2. Fibrilação atrial. 3. Classificação supervisionada. I. Nadal, Jurandir *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Biomédica. III. Título.

## **Agradecimentos**

Agradeço à CAPES e ao CNPq pelo apoio financeiro a mim e ao meu programa, sem o qual este trabalho não poderia ter sido realizado.

Agradeço ao meu pai, à minha mãe, à minha vó e à minha irmã pelo apoio sempre presente em meio às dificuldades e pelos conselhos.

Agradeço aos meus orientadores, Jurandir Nadal e Paulo Roberto Benchimol Barbosa, pelos conselhos, pelas broncas e pela paciência.

Agradeço a Jesus por me proporcionar saúde, inteligência e por estar sempre presente em meio aos momentos mais difíceis.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

AVALIAÇÃO DO RITMO CARDÍACO EM ELETROCARDIOGRAMAS DE  
CURTA DURAÇÃO UTILIZANDO ANÁLISE DOS INTERVALOS RR E  
APRENDIZADO SUPERVISIONADO

Gabriel de Azevedo Limeira

Março/2019

Orientadores: Jurandir Nadal

Paulo Roberto Benchimol Barbosa

Programa: Engenharia Biomédica

Fibrilação atrial é uma condição que frequentemente não exibe sintomas e está fortemente relacionada com infarto e morte súbita cardíaca. O objetivo deste trabalho foi desenvolver um algoritmo que diferencia fibrilação atrial de ritmo normal e outros ritmos presentes em sinais de curta duração e de derivação única, coletados por um dispositivo móvel. Um total de 36 características foram coletadas, principalmente da sequência de intervalos batimento-a-batimento. Seleção de características por análise de componentes vizinhos (ACV) foi aplicada e quatro algoritmos de aprendizado supervisionado foram comparados e otimizados, utilizando-se treinamento com validação cruzada. A comparação dos desempenhos foi realizada por meio de um índice  $F_1$  que leva em conta sensibilidade e especificidade. A ACV permitiu selecionar 11 variáveis do conjunto inicial. O classificador que utilizou máquina de vetores de suporte forneceu o melhor índice geral ( $F_1 = 72,9$ ), onde o melhor desempenho ocorreu na detecção de fibrilação atrial ( $F_1 = 82,5$ ).

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

CARDIAC RHYTHM ASSESSMENT IN SHORT TERM ELECTROCARDIOGRAM  
USING RR FEATURES AND SUPERVISED LEARNING

Gabriel de Azevedo Limeira

March/2019

Advisors: Jurandir Nadal

Paulo Roberto Benchimol Barbosa

Department: Biomedical Engineering

Atrial fibrillation is a condition that often does not show itself through symptoms and is strongly related to infarction and sudden cardiac death. This work aims at developing an algorithm that differentiates atrial fibrillation rhythm from noise, normal and other rhythms in single short ECG leads collected by a mobile device. A total of 36 features were collected mostly from the sequence of beat-to-beat intervals. Neighborhood component analysis (NCA) feature selection technique was applied, and several supervised learning algorithms were compared and optimized using cross validation approach. Performances were compared with an index  $F_1$  that considers both sensitivity and specificity. NCA allowed selecting 11 features. The classifier based on support vector machines gave the best overall result ( $F_1 = 72,9\%$ ), were the best performance occurred for the atrial fibrillation class ( $F_1 = 82,5$ ).

# Sumário

1	Introdução.....	1
1.1	Objetivo geral .....	3
1.2	Objetivos específicos .....	3
2	Fundamentos teóricos.....	4
2.1	O sinal de ECG.....	4
2.2	Diretrizes para processamento.....	10
2.3	Pré-processamento do sinal .....	10
2.4	Detecção de picos do ECG .....	11
2.5	Fibrilação Atrial.....	17
2.6	Outras arritmias .....	19
2.7	Gráfico de Lorenz sobre RR e $\delta RR$ .....	20
2.8	Gráfico de Poincaré .....	25
2.9	Seleção de Características por Componentes Vizinhos.....	28
2.10	Métodos de classificação .....	33
2.10.1	O Classificador de Bayes.....	33
2.10.2	Árvore de decisão .....	34
2.10.3	K vizinhos mais próximos (KVP) .....	37
2.10.3.1	Relação do método com parâmetros do MATLAB® 2017a.....	41
2.10.4	Análise de discriminante linear .....	41
2.10.4.1	Introdução.....	41
2.10.4.1	Análise discriminante linear para um preditor ( $p = 1$ ) .....	42
2.10.4.2	Análise discriminante linear para mais de um preditor ( $p > 2$ ) .....	43
2.10.4.3	Análise Discriminante Quadrática.....	46
2.10.4.4	Relação do método com parâmetros do MATLAB® 2017a.....	47

2.10.5 Máquinas de Vetores de Suporte (MVS).....	48
2.10.5.1 Classificador de máxima margem .....	48
2.10.5.2 Classificador de vetores de suporte .....	51
2.10.5.3 Máquina de vetores de suporte .....	53
2.10.5.4 Extensão de MVS para mais de duas classes .....	55
2.10.5.5 Relação do método com parâmetros do MATLAB® 2017a.....	55
2.11 Validação cruzada.....	56
2.12 Códigos de saída para correção de erros (CSCE).....	57
3 Revisão da Literatura.....	59
4 Materiais e Métodos .....	63
4.1 Banco de dados .....	63
4.2 Anotação de dados.....	64
4.3 Ilustrações de sinais .....	65
4.4 Pré-processamento de sinal .....	68
4.5 Pontuação do desafio .....	69
4.6 Extração de características.....	70
4.7 Balanceamento do banco de dados por subamostragem .....	73
4.8 Seleção de características .....	74
4.9 Classificação .....	75
4.10 Índice de avaliação de desempenho.....	76
5 Resultados.....	77
5.1 Seleção de Características .....	77
5.2 Escolha de classificador e hiperparâmetro por Otimização.....	79
5.3 Índice de avaliação de desempenho do classificador .....	86
6 Discussão .....	88



6.1 Problemas na diferenciação entre outros ritmos e fibrilação atrial .....	88
6.2 Presença de ruído atrapalha diferenciação entre ritmo normal e atrial.....	90
6.3 Análise de variáveis.....	91
6.4 Comparação de desempenho dos competidores .....	97
6.5 Comparação do presente trabalho.....	100
6.6 Limitações do estudo .....	101
6.7 Propostas de trabalhos futuros .....	102
7 Conclusão .....	103
8 Referências bibliográficas .....	104

# 1 Introdução

A fibrilação atrial (FA) é a arritmia cardíaca sustentada mais frequente em seres humanos. Está associada ao envelhecimento e seu diagnóstico ainda necessita ser aprimorado. Nos Estados Unidos, pesquisadores estimaram, por cálculo retroativo, que o total de indivíduos idosos ( $\geq 65$  anos) acometidos seria de 3.873.900 (com intervalo de confiança IC 95%: 3.675.200-4.702.600), enquanto o total de indivíduos na população economicamente ativa seria de 1.457.100 (IC 95%: 1.218.500 – 1.695.800), compreendendo respectivamente 10% e 0,92% destas populações (TURAKHIA *et al.*, 2018). Outros estudos em diferentes países e populações apontam para prevalências entre 0,5 e 2,5% (GO AS *et al.*, 2001; STEWART *et al.*, 2001; CAMM *et al.*, 2010; CLAES *et al.*, 2012). Um estudo populacional feito em Rotterdam indica, em particular, os efeitos da idade na prevalência de fibrilação atrial: a prevalência geral ( $\geq 55$  anos) encontrada foi de 5,5%; no entanto, no grupo com idade entre 55 e 59 anos, a prevalência foi de 0,7%, enquanto naquele com mais de 85 anos a prevalência encontrada foi de 17,8%. A incidência de FA sobe de 1,1/1000 pessoas-ano no grupo de 55-59 anos para 20,7/1000 pessoas-ano no grupo de 80-84 anos (HEERINGA *et al.*, 2006).

Um estudo pioneiro realizado em Minas Gerais, com o auxílio da telemedicina, encontrou resultados de prevalência similares aos de países desenvolvidos: 1,8% de prevalência de FA em geral; 2,4% nos homens e 1,3% nas mulheres. Entre os homens, a prevalência variava de 0,001% (sujeitos com idade entre 5-19 anos) a 14,6%, (idade entre 90-100 anos). Entre as mulheres, a prevalência variava de 0,001% (sujeitos com idade entre 5-19 anos) a 8,7% (sujeitos com idade entre 90-100 anos) (MARCOLINO *et al.*, 2015).

Segundo as diretrizes sobre diagnóstico e tratamento de fibrilação atrial da Sociedade Europeia de Cardiologia, a FA está associada a uma taxa de mortalidade duas vezes maior, a um maior risco de acidente vascular encefálico (AVE) bem como a eventos com maior gravidade. Além disso, está associada a hospitalizações frequentes (um terço das hospitalizações por arritmias cardíacas ocorre devido a FA), disfunção cognitiva por eventos embólicos assintomáticos, perda no desempenho em atividades físicas e na qualidade de vida, além de perda na capacidade de bombeamento do ventrículo esquerdo.

A intensidade dos sintomas é bastante variável, gerando desde efeitos imperceptíveis até perda significativa na qualidade de vida (CAMM *et al.*, 2010).

Um estudo realizado na Finlândia mostra a associação entre fibrilação atrial e AVE (JAAKKOLA *et al.*, 2016). Um total de 3.623 pacientes com o primeiro AVE ou ataque isquêmico transiente entre os anos de 2003 até 2012 foram analisados. Destes, 753 (20,8%) apresentaram concomitantemente AVE e FA no primeiro diagnóstico. Para se certificar de que a FA era a causa do AVE e não o contrário, o estudo formou um grupo controle de 781 pacientes com hemorragia intracraniana. Destes, apenas 15 (1,5%) reportaram novo episódio de fibrilação atrial. O estudo mostrou ainda que existe uma forte correlação entre detecção de FA pela primeira vez durante o tratamento de um evento isquêmico e uma população mais jovem e saudável, sem histórico de doenças coronarianas ou outras patologias vasculares, insuficiência cardíaca e hipertensão.

Apesar de a associação entre FA e AVE ser conhecida e uma diretriz de tratamento estar bem estabelecida, um estudo retrospectivo baseado nos dados do programa *Get With The Guidelines-Stroke* com 94.744 pacientes com AVE isquêmico agudo e história pregressa conhecida de FA mostrou que 84% dos pacientes não estava recebendo tratamento de acordo com a recomendação da diretriz. O tratamento recomendado consiste em antagonista da vitamina K como a varfarina ou em anticoagulantes orais não antagonistas da vitamina K (NAOCS). Após ajustes para possíveis variáveis confundidoras, as chances de um AVE moderado ou grave são menores no grupo que fazia o tratamento em relação ao que não recebia tratamento antitrombótico (razão de chances ajustada [IC 95%] varfarina – (0,56 [0,51-0,60]); NAOCS- 0,65 [0,61-0,71]). As chances de mortalidades também são significativamente menores: (razão de chances ajustada [IC 95%] varfarina – (0,75 [0,67-0,85]); NAOCS- 0,79 [0,72-0,88]) (XIAN *et al.*, 2017). Embora exista uma diretriz de tratamento que melhora a condição de vida daqueles que tem FA, muitos casos de FA não são diagnosticados, principalmente os silenciosos. Para melhorar essa taxa de diagnóstico, um grande esforço tem sido feito por pesquisadores para buscar uma maneira de detectar a FA de maneira automática.

## 1.1 Objetivo geral

O objetivo deste trabalho é desenvolver um algoritmo que diferencie fibrilação atrial de ritmo normal e outros ritmos presentes em sinais curtos de derivação única que são coletados por um dispositivo móvel.

## 1.2 Objetivos específicos

Detectar as ondas R e extrair parâmetros a partir da sequência de intervalos RR.

Utilizar análise de componentes vizinhos para determinar, de maneira independente de métodos de classificação, qual o melhor conjunto de características que maximiza a acurácia da previsão de métodos de classificação.

Comparar diversos métodos de classificação supervisionada e determinar qual é o mais aplicável ao problema proposto.

## 2 Fundamentos teóricos

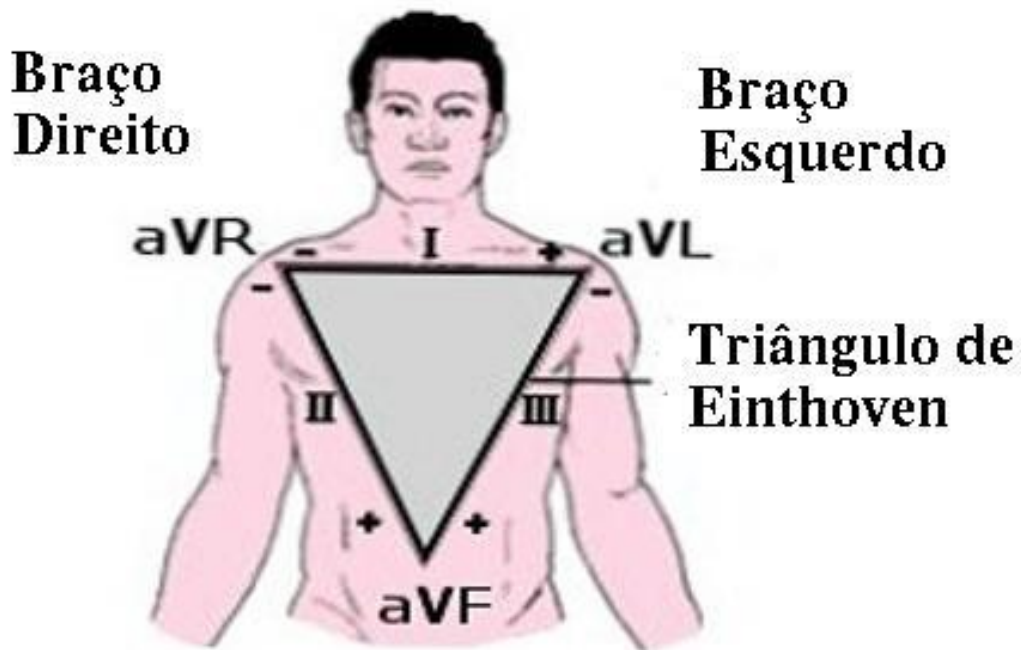
### 2.1 O sinal de ECG

O eletrocardiograma (ECG) constitui o registro da atividade bioelétrica cardíaca coletada na superfície do corpo. O nódulo sinusal se localiza na parede posterior do átrio direito e tem a propriedade de auto-excitação, ou seja, é capaz de se despolarizar de forma autônoma. A despolarização é a mudança do potencial transmembrana de uma célula excitável, em consequência a um estímulo forte o suficiente para causar tal mudança. A repolarização, por outro lado, é o retorno do potencial transmembrana ao nível basal ou pré-despolarização. A propagação do estímulo elétrico se faz de fibra cardíaca a fibra cardíaca, pois uma despolarização em um trecho da membrana causa despolarização no seguinte. Além disso, a despolarização em uma célula excitável é capaz de causar despolarização na seguinte, quando esta ocorre em um trecho de membrana que está próximo da membrana de célula seguinte. A diferença de potencial entre a célula despolarizada e a célula imediatamente vicinal ainda não despolarizada gera um dipolo, cuja diferença de potencial é registrada na superfície do corpo por meio de um eletrocardiograma. Einthoven criou um método preciso e reproduzível, capaz de registrar a variação de potencial entre células adjacentes (dipolos) ao longo do tempo. Considerando que um dipolo é uma grandeza vetorial, as projeções (derivações) criadas por Einthoven permitem registrar sinais elétricos do coração decorrente da propagação dos dipolos em diferentes perspectivas, interpretá-los e compará-los. O batimento normal possui uma representação morfológica característica no eletrocardiograma. No entanto, quando ocorrem distúrbios no ponto de origem autônomo da despolarização ou problemas diversos de condução, formas de onda anômalas são observadas no ECG (GUYTON *et al.*, 2006). A diferença de potencial observada na superfície do corpo é muito pequena, usualmente com magnitudes na ordem de microvolts, necessitando ser amplificada. O protocolo de colocação de eletrodos geralmente utilizado requer um mínimo de quatro eletrodos. Os eletrodos são posicionados nos pulsos direito e esquerdo e nas pernas direita e esquerda. O eletrodo da perna direita funciona como referência, enquanto os demais são considerados como equivalentes aos vértices do triângulo de

Einthoven e suas respectivas polaridades na construção de cada uma das derivações, as quais são exibidas na Figura 1 (GUYTON *et al.*, 2006).

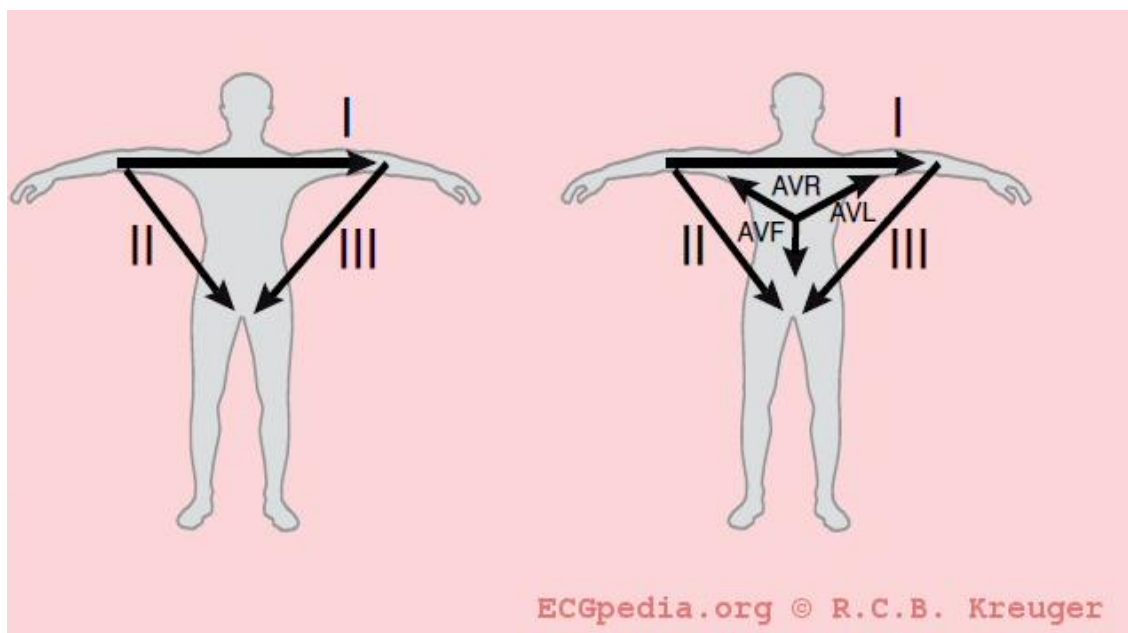
O protocolo de posicionamento dos eletrodos descrito anteriormente possibilita determinar qual a projeção do potencial elétrico instantâneo em diferentes direções. Essas direções são construídas a partir das diferenças de potencial entre combinações de eletrodos, em arranjos bipolares. Como observado na Figura 1, seis derivações podem ser construídas. Abaixo uma lista de como as derivações são construídas é apresentada:

- a) I = braço esquerdo – braço direito
- b) II = perna esquerda – braço direito
- c) III = perna esquerda – braço esquerdo
- d) AVR (*augmented vector right*) - braço direito – (“curto-circuito” entre braço esquerdo e perna esquerda)
- e) AVF (*augmented vector foot*) – perna esquerda – (“curto-circuito” entre braço esquerdo e braço direito)
- f) AVL (*augmented vector left*) – braço esquerdo – (“curto-circuito” entre braço direito e perna esquerda)



*Figura 1 - Triângulo de Einthoven. Os eletrodos colocados nos membros superiores direito (R) e esquerdo (L) e na perna esquerda (F) correspondem a cada vértice do triângulo. A polaridade das conexões também está indicada. Os lados do triângulo correspondem às derivações principais de Einthoven (I, II e III), enquanto os vértices correspondem às derivações aumentadas de Wilson (aVR, aVL e aVF), medidas em relação ao ponto médio das três derivações. Imagem traduzida da original disponível em [https://commons.wikimedia.org/wiki/File:Triangolo\\_di\\_Einthoven.jpg](https://commons.wikimedia.org/wiki/File:Triangolo_di_Einthoven.jpg), sob a licença Creative Commons Attribution-Share Alike 3.0 Unported.*

A distribuição do campo elétrico do coração no espaço tridimensional dada pelas derivações é mostrada na Figura 2. Com o uso de apenas três derivações, é possível determinar o campo elétrico resultante do coração no espaço tridimensional. Com duas derivações bipolares é possível determinar a direção do campo elétrico resultante no plano frontal (superior-inferior).



*Figura 2 - Projeções vetoriais dos três eletrodos do triângulo de Einthoven. Imagem publicada em [https://commons.wikimedia.org/wiki/File:De-Hart\\_axis\\_\(CardioNetworks\\_ECGpedia\).png](https://commons.wikimedia.org/wiki/File:De-Hart_axis_(CardioNetworks_ECGpedia).png), de acordo com a licença Creative Commons Attribution-Share Alike 3.0 Unported.*

Por sua vez, as denominadas derivações precordiais são utilizadas para determinar o campo elétrico resultante no plano transversal. O posicionamento dos respectivos eletrodos no corpo é exibido na Figura 3. A distribuição espacial dessas derivações no espaço tridimensional é exibida na Figura 4.



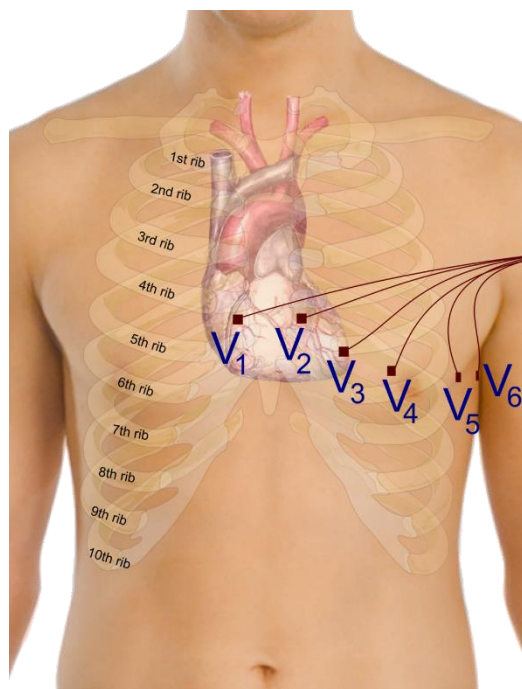


Figura 3 – Derivações precordiais. Imagem publicada em [https://commons.wikimedia.org/wiki/File:Precordial\\_leads\\_in\\_ECG.png](https://commons.wikimedia.org/wiki/File:Precordial_leads_in_ECG.png) de acordo com a licença Creative Commons CC0 1.0 Universal Public Domain Dedication.

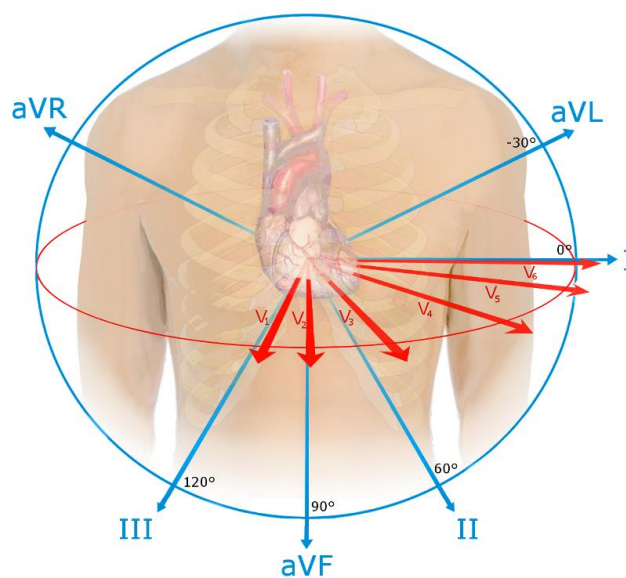
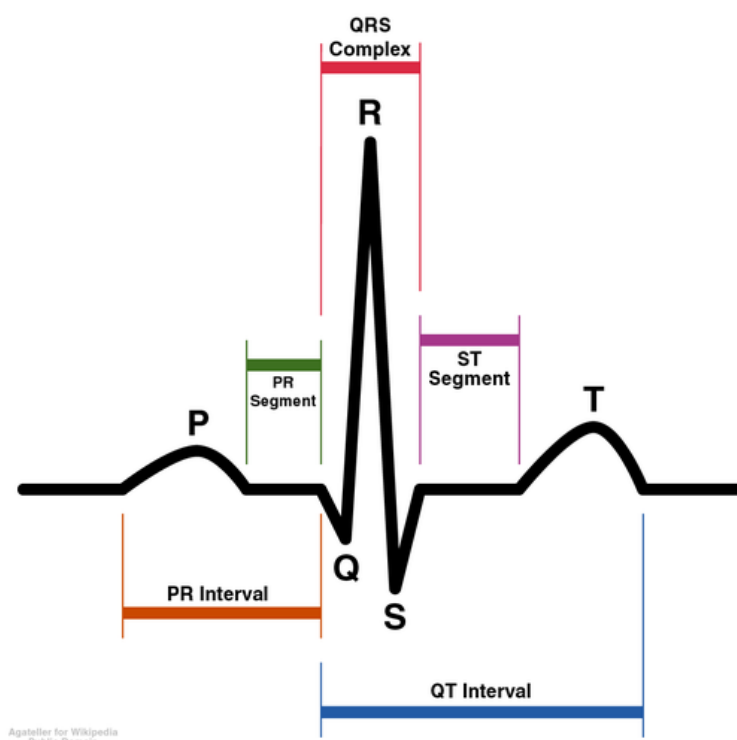


Figura 4 – Distribuição espacial das projeções do campo elétrico que são observadas através das derivações; Imagem publicada em [https://commons.wikimedia.org/wiki/File:EKG\\_leads.png](https://commons.wikimedia.org/wiki/File:EKG_leads.png) sob a licença Creative Commons Attribution-Share Alike 4.0 International.

Um batimento normal é visualizado no ECG pela onda P, complexo QRS e onda T. As distâncias entre esses pontos são denominadas como segmentos ou intervalos, conforme ilustrado na Figura 5. A onda P está relacionada com a despolarização dos átrios, o complexo QRS com a despolarização dos ventrículos e a onda T com a repolarização dos ventrículos. O intervalo RR é o intervalo de tempo entre duas ondas R consecutivas (GUYTON *et al.*, 2006).



*Figura 5 – Diagrama esquemático do padrão normal de um ECG. Imagem publicada em <https://commons.wikimedia.org/wiki/File:SinusRhythmLabels.png> sob domínio público*

A derivação II é a mais utilizada para diagnosticar doenças do coração. Ela realça todos os segmentos e ondas, enquanto V5 é mais adequada para avaliar arritmias ventriculares, visto que ela fica bem próxima do ventrículo esquerdo. Alguns trabalhos mostram que para efeitos de classificação o uso da combinação das derivações II e V5 é melhor (LLAMEDO *et al.*, 2012; LUZ *et al.*, 2016). É possível sintetizar a informação das 12 derivações com o uso de menos derivações, através de técnicas de síntese (TOMAŠIĆ *et al.*, 2014).

## 2.2 Diretrizes para processamento

O trabalho de classificação automática de batimentos pode ser dividido em cinco etapas:

- a) Pré-processamento do ECG
- b) Detecção de picos no ECG
- c) Extração de características
- d) Seleção de características
- e) Classificação

## 2.3 Pré-processamento do sinal

A abordagem mais largamente utilizada é o emprego de filtros tipo *finite impulse response (FIR)*, com filtros digitais lineares para atenuação de ruídos de medição, flutuações de linha de base e interferências elétrica e eletromiográfica (LYNN, 1971). Esta abordagem pode ser satisfatória para muitas situações. Outras abordagens têm sido utilizadas e podem ter desempenho superior, no entanto, deve ser considerado o custo benefício de implementar uma técnica específica de processamento em relação ao método canônico de pré-processamento (LUZ *et al.*, 2016). Dentre as alternativas utilizadas, podem-se citar filtros adaptativos baseados em redes neurais (XUE *et al.*, 1992), transformadas *wavelet* (SAYADI *et al.*, 2007), filtros Bayesianos não-lineares (SAMENI *et al.*, 2007) e Filtros de Kalman estendidos (SAYADI *et al.*, 2008). Para este trabalho, considerou-se que o uso de um pré-processamento simples era suficiente, visto que apenas as ondas R, que têm uma boa relação sinal-ruído, são utilizadas para extração de características.

## 2.4 Detecção de picos do ECG

O método escolhido para detecção de ondas R foi o baseado no método DOM (“diferença de operação”) (YEH *et al.*, 2008). Este método, além de ser de simples implementação, é de fácil compreensão e possui desempenho superior ao proposto por Pan e Tompkins (PT) (PAN *et al.*, 1985) e semelhante ao baseado em Transformada Wavelet (WT) (LI *et al.*, 1995).

Resumidamente, os passos do detector são como se segue: (YEH *et al.*, 2008)

- 1) Obter a diferença de sinal discreta (Figura 6a,b):

$$x_d[n] = x[n] - x[n - 1] \quad (1)$$

- 2) Aplicar filtro passa baixas de 100 Hz em  $x_d$  para eliminar ruídos de alta frequência. O sinal resultante do filtro é chamado de  $x_{df}$  (Figura 6c)
- 3) Passar o sinal  $x_{df}$  por dois filtros: o primeiro filtro elimina sinais positivos que estiverem abaixo de  $T_1$ ; o segundo filtro elimina sinais negativos que estiverem acima de  $T_2$ .  $T_1 = 2$  MVp e  $T_2 = 2$  MVn, onde MVp representa a média dos valores positivos do sinal, enquanto MVn representa a média dos valores negativos de sinal. Segue adiante uma representação desse processo na fórmula (Figura 6d,e):

$$\hat{x}_{df} = \begin{cases} 0, & \text{se } 0 < x_{df} < T_1, \text{ ou } T_2 < x_{df} < 0 \\ x_{df}, & \text{se } x_{df} \geq T_1, \text{ ou } x_{df} \leq T_2 \end{cases} \quad (2)$$

- 4) O sinal  $\hat{x}_{df}$  deve ser separado em um sinal apenas positivo  $\hat{x}_{df}^+$  e outro negativo  $\hat{x}_{df}^-$ . Para cada sinal, devem ser encontrados os máximos e mínimos locais em janelas consecutivas de 0,14 s (tempo máximo considerado para intervalo QRS). Para ser considerado um máximo/mínimo local, adiciona-se um segundo requisito, de que o valor deve ser diferente de zero (Figura 7a)
- 5) Para cada par de máximos, se a diferença entre dois máximos é menor do que 0,14 s, escolhe-se o máximo com maior valor. Se a diferença é maior que esse intervalo, ambos são aceitos. O mesmo processo é executado para a sequência de mínimos, mas com o critério de escolha como sendo o de menor valor (Figura 7b,c e Figura 8)

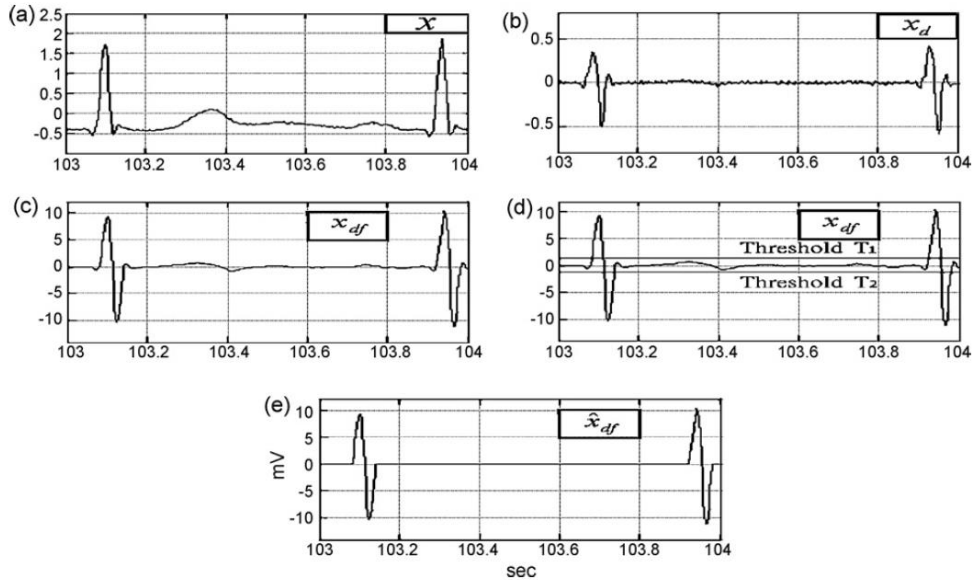


Figura 6 - (a)  $x$ , o sinal #103 do banco de dados MIT-BIH Arrhythmia Database. (b)  $x_d$ , o sinal de ECG após aplicar a operação de diferença; (c)  $x_{df}$ , o sinal após o uso do filtro passa-baixas; (d) uso de valores de limiar (e)  $\hat{x}_{df}$ , o sinal depois de passar por limiares. Reproduzido a partir de (YEH et al., 2008), com permissão da Elsevier.

- 6) As sequências separadas de máximos e mínimos aceitos na etapa anterior são então ajuntadas e reordenadas. Nessa nova sequência, composta pelos máximos e mínimos locais, se a diferença entre um máximo local e um mínimo local é menor do que 0,14, então esse par é aceito. Caso a diferença entre os pares seja maior do que 0,14 s, o par é rejeitado (Figura 7d)
- 7) No método original, para cada par de máximo e mínimo local do sinal diferencial encontrado, busca-se no sinal original o máximo local no intervalo delimitado pelos marcadores encontrados anteriormente. Como no banco de dados utilizado por esta pesquisa (seção 4.1) os sinais poderiam vir com polaridade invertida, esse passo foi adaptado. Caso o mínimo local de derivada viesse depois do máximo local de derivada, considerava-se que a onda R era

positiva e buscava-se no sinal original pelo máximo local. Caso o máximo local de derivada viesse depois do mínimo local de derivada, considerava-se que a polaridade do sinal estava invertida e buscava-se no sinal original pelo mínimo local. Essa abordagem trazia o benefício adicional de detectar eventos correspondentes a picos de eventos não causados por batimentos normais, como por exemplo uma extrassístole. Ou seja, todo pico relacionado à ocorrência de um batimento é detectado, inclusive a onda R (Figura 7e)

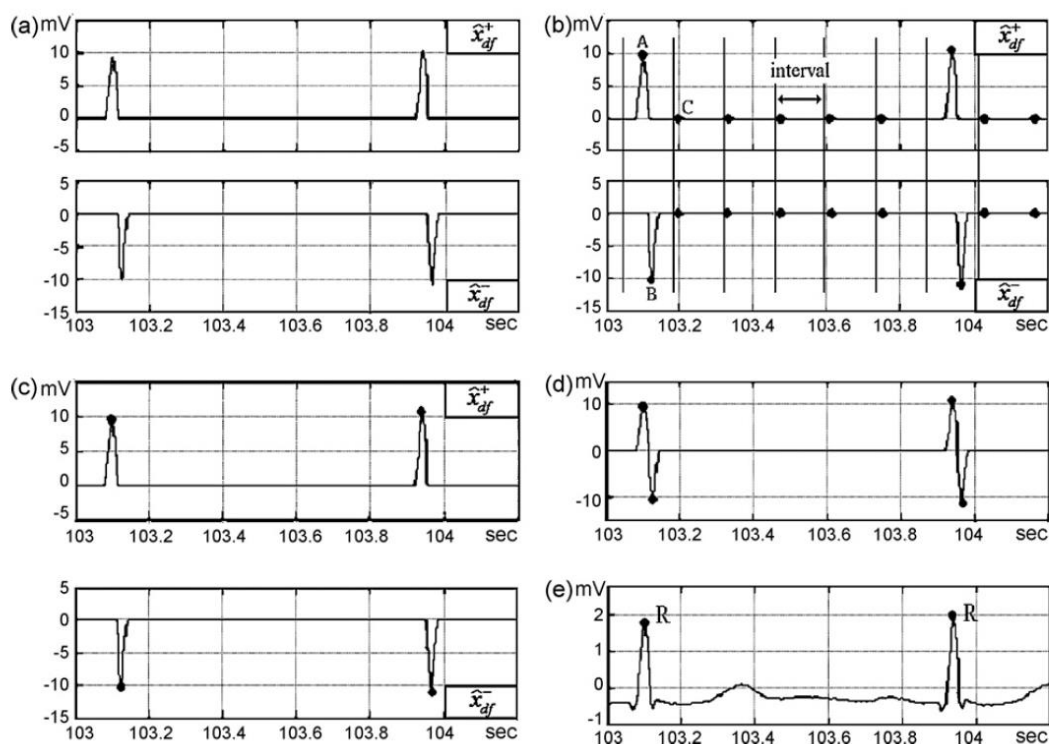


Figura 7 – (a)  $\hat{x}_{df}^+$  e  $\hat{x}_{df}^-$  são separados do sinal diferencial  $\hat{x}_{df}$ ; (b) selecionar os valores extremos (onde os marcadores em formato de círculo representam os extremos); (c) os pontos de valores extremos selecionados (d) os pontos de valores positivos e negativos no sinal de ECG original; (e) a posição do valor máximo positivo é o ponto R. Reproduzido a partir de (YEH et al., 2008), com permissão da Elsevier.

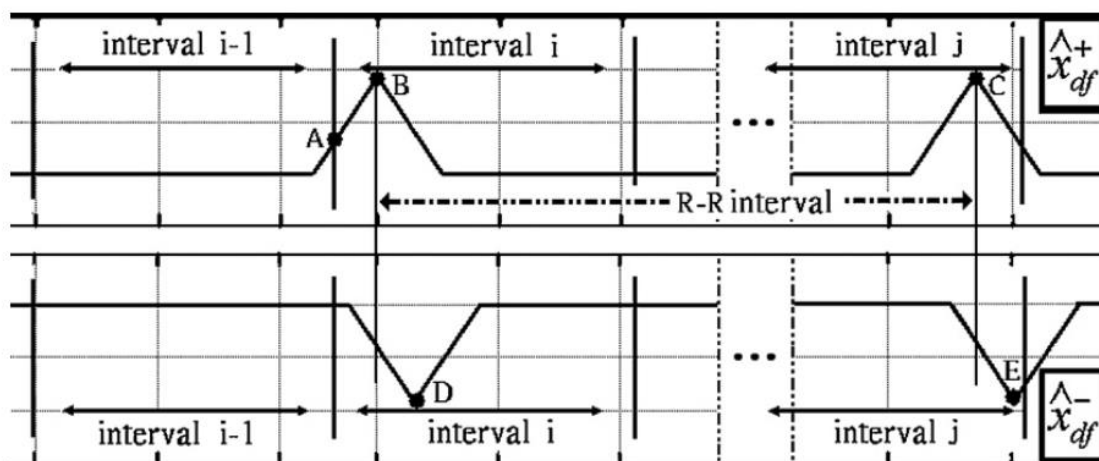
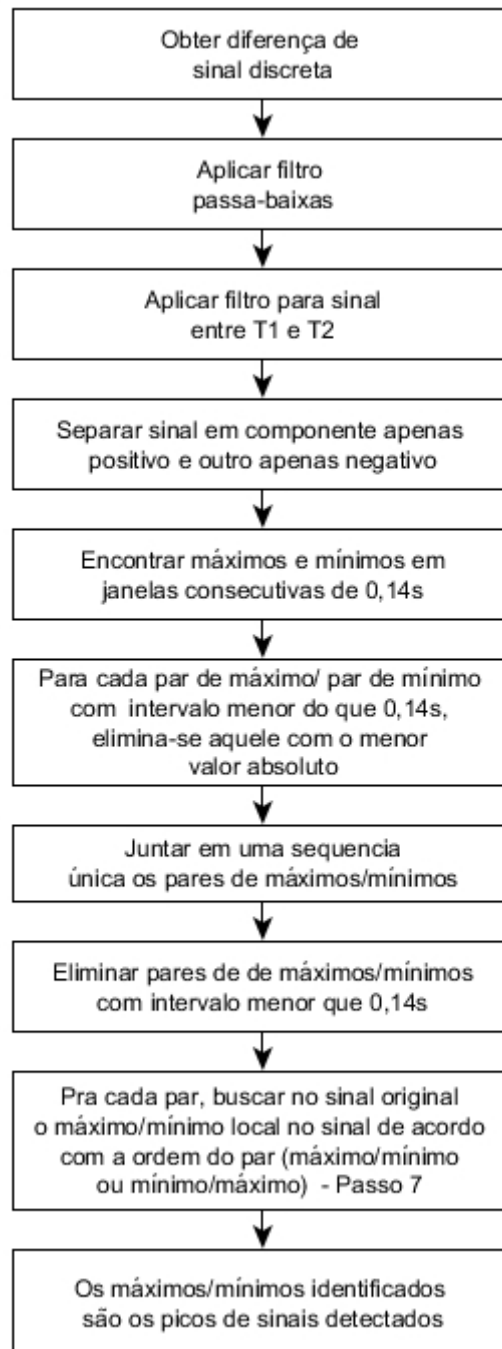


Figura 8 - Exemplo de seleção dos valores corretos extremos na análise individual de sinal positivo/negativo. Como o intervalo entre A e B é próximo e B é maior do que A, apenas B é aceito. Como o intervalo entre D e E é maior que 0,14 segundos, ambos são aceitos. Reproduzido a partir de (YEH et al., 2008), com permissão da Elsevier.

O processo como um todo é representado pelo fluxograma da Figura 9. O resultado do detector é ilustrado para alguns sinais do banco de dados MIT-BIH Arrhythmia Database (MOODY *et al.*, 2001). A Figura 10 ilustra a detecção dos pontos Q e S também, no entanto este trabalho se preocupou apenas em detectar o evento pico relacionado a um evento cardíaco, não se preocupando, portanto, em delinear o complexo QRS. Observa-se um desempenho satisfatório para os quatro sinais ilustrados, que possuem morfologias distintas.



*Figura 9 -Fluxograma do detector de picos utilizado, seguindo a proposta de (YEH et al., 2008).*



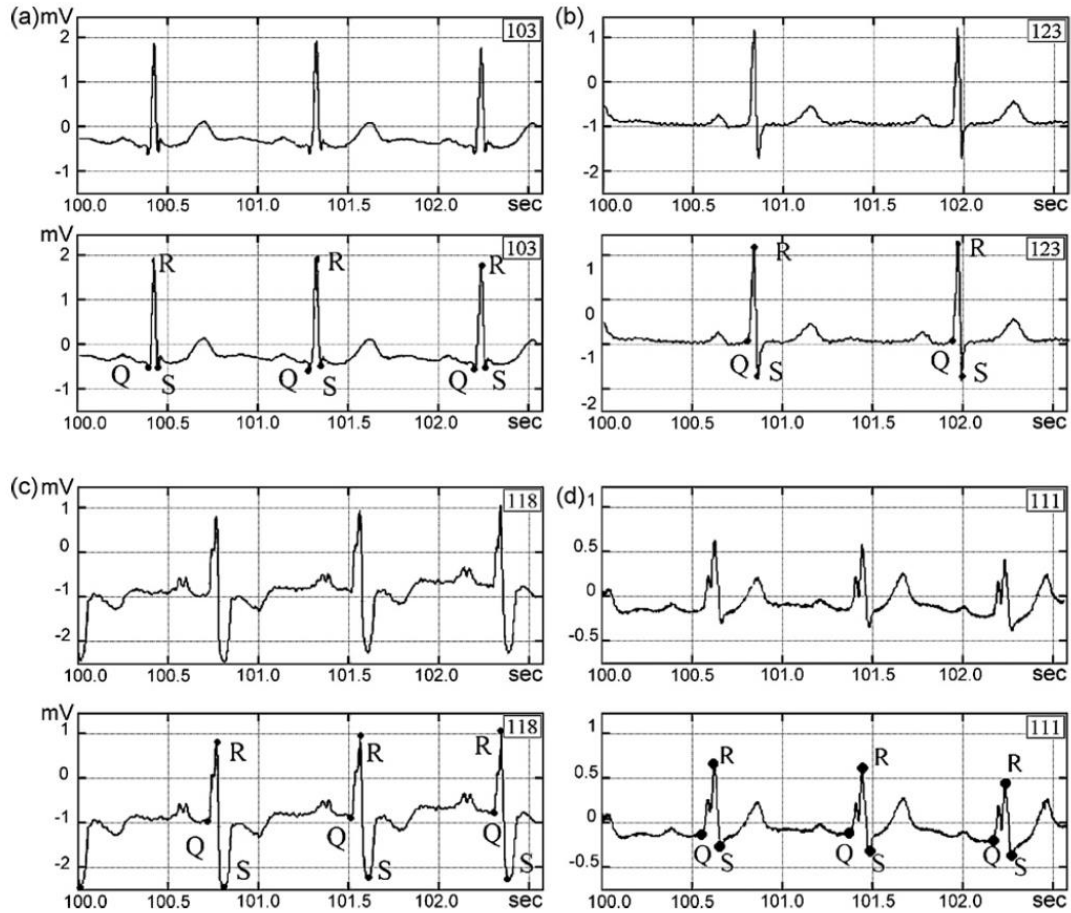


Figura 10 - Para cada par de figuras, o sinal original é mostrado na figura superior e os complexos QRS detectados são mostrados na figura inferior. O detector implementado não efetua detecção de pontos R e S. Reproduzido a partir de (YEH et al., 2008), com permissão da Elsevier.

O desempenho do detector foi comparado originalmente usando todos os batimentos presentes no banco de dados do MIT-BIH Arrhythmia Database em termos de falsos positivos (FP) e falsos negativos (FN). A marcação do detector foi comparada com o padrão ouro determinado pela marcação manual dada pelos especialistas dos organizadores do banco de dados, conforme descrito em (MOODY *et al.*, 2001). O índice  $F_d$  é utilizado como índice geral de erro, sendo dado por:

$$F_d = \frac{FP+FN}{total\ de\ batimentos} \quad (3)$$

Os resultados da comparação realizada pelo autor do artigo estão exibidos na Tabela 1, reproduzida a partir do artigo original do método DOM.

*Tabela 1 - Comparação de desempenho entre três métodos em termos de FP e FN.  $F_d$  é o índice indicado na equação 3. Pan e Tompkins (PT) (PAN et al., 1985); Transformada Wavelet (WT) (LI et al., 1995). (Reproduzido e adaptado a partir de (YEH et al., 2008), com permissão da Elsevier)*

<i>Método</i>	<i>Total de batimentos</i>	<i>Batimentos FP</i>	<i>Batimentos FN</i>	<i><math>F_d</math></i>
<i>DOM</i>	116.137	58 (0,05%)	166 (0,14%)	0,19
<i>PT</i>	116.137	507 (0,43%)	277 (0,24%)	0,68
<i>WT</i>	116.137	65 (0,06%)	112 (0,10%)	0,15

Obs: 100 ms de tolerância entre o ponto detectado e a marcação do *database*

## 2.5 Fibrilação Atrial

A fibrilação atrial (FA) é a arritmia cardíaca sustentada mais comum em seres humanos e sua característica mais marcante no eletrocardiograma é o ritmo irregularmente irregular dos intervalos entre os batimentos cardíacos. As ondas P frequentemente desaparecem e são substituídas por um sinal de baixa tensão e alta frequência. Um ECG com fibrilação atrial é exibido na Figura 11. A propagação da despolarização ocorre de forma descoordenada, com conseqüente decréscimo da eficiência no bombeamento de sangue para os ventrículos (RUFFY, 1995; GUYTON *et al.*, 2006; BERNE *et al.*, 2008). Seu mecanismo fisiológico é bastante controverso, de modo que não será discutido neste trabalho.

A fibrilação atrial é rara em adolescentes e crianças, mas sua prevalência aumenta conforme a idade, especialmente em idosos. O estudo Framingham indica que a mortalidade em geral é dobrada, e a mortalidade por doenças cardiovasculares mais do

que dobra. Os fatores de risco para FA incluem idade avançada, sexo masculino, regurgitação mitral e insuficiência cardíaca congestiva. (RUFFY, 1995)

A irregularidade no ritmo cardíaco e a perda da função atrial pode causar diferentes sintomas, devido à perda da performance cardíaca. A variação de sintomas depende do estado do paciente e de quais medicamentos ele utiliza. Idosos sedentários podem ser assintomáticos, enquanto jovens em plena atividade física podem reclamar de perda da capacidade em fazer exercício. (RUFFY, 1995)

A fibrilação atrial pode ser tratada com drogas antiarrítmicas ou tratamento elétrico. Tanto em uma abordagem como na outra, é necessário utilizar concomitantemente tratamento de anticoagulação, pois a associação entre eventos tromboembólicos e a FA é bem conhecida. (RUFFY, 1995)



*Figura 11 – (Superior) ECG com fibrilação atrial; (Inferior) ECG Normal. A frequência cardíaca do ECG com fibrilação atrial é irregular, como pode ser observado pelo espaçamento irregular entre os picos. Além disso, observa-se no período entre os picos um sinal com aspecto ruidoso, efeito da estimulação randômica que ocorre no interior do átrio. Essa estimulação gera a chamada onda f. Imagem traduzida da versão original disponível em [https://en.wikipedia.org/wiki/File:Afib\\_ecg.jpg](https://en.wikipedia.org/wiki/File:Afib_ecg.jpg), sob a licença Creative Commons Attribution-Share Alike 3.0 Unported.*

## 2.6 Outras arritmias

Para os propósitos deste trabalho, alguns breves comentários sobre alguns tipos simples de arritmias serão feitos, a título de contextualizar alguns algoritmos simples que serão aplicados sobre o intervalo RR.

A taquicardia significa frequência cardíaca alta, em geral considerada como acima de 100 batimentos por minuto (bpm). Uma bradicardia significa frequência cardíaca lenta, em geral menos de 60 bpm (GUYTON *et al.*, 2006).

Algumas substâncias como cafeína, fármacos ou condições como isquemia e calcificação de áreas do coração podem causar inícios de propagação de despolarização cardíaca não comandadas pelo nodo sinusal, ocorrendo a chamada extrassístole, ou batimento prematuro. Se o ponto de estímulo inicial se situar no átrio, este provavelmente atingirá o nodo AV antes do esperado, causando uma despolarização com complexo QRS de morfologia normal, porém prematura, acompanhada de uma onda P com segmento P-R diminuído. No caso de uma extrassístole ventricular, como na Figura 12, o complexo QRS é alargado, pois a propagação de despolarização ocorre por vias de condução mais lentas que o feixe de His, e a despolarização do ventrículo oposto ocorre com atraso. O potencial elétrico também pode ficar mais elevado ou atenuado, em relação ao complexo QRS normal, em função da mudança do eixo elétrico da despolarização. Finalmente, a onda T também tem polaridade invertida porque as fibras que se repolarizam primeiro são as que se despolarizaram primeiro, ao contrário do que ocorre no batimento normal (GUYTON *et al.*, 2006).

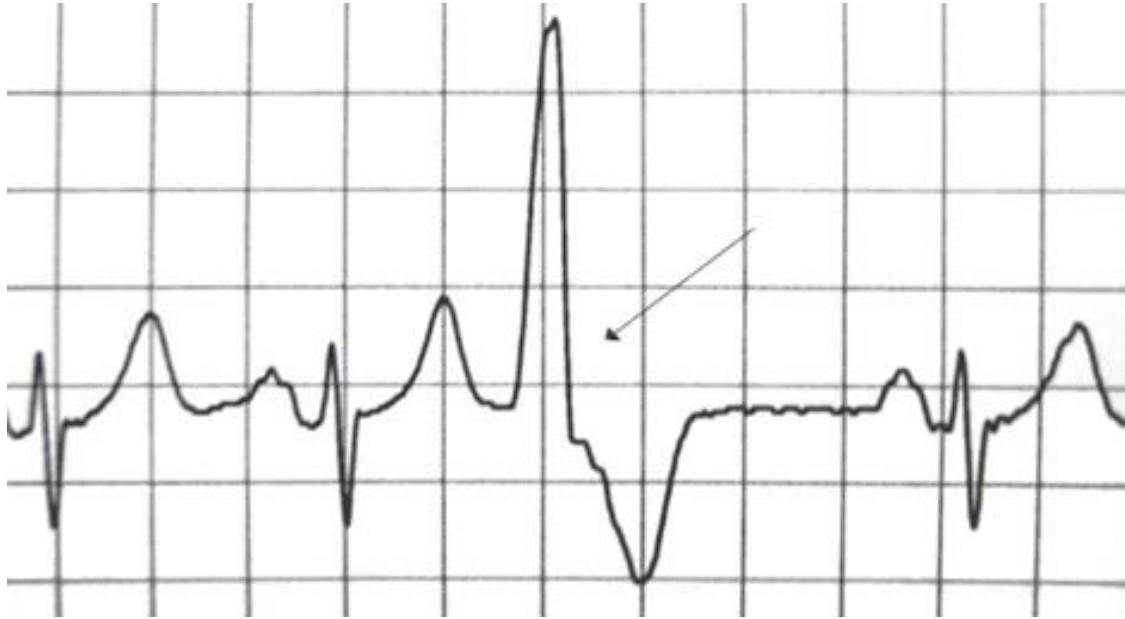
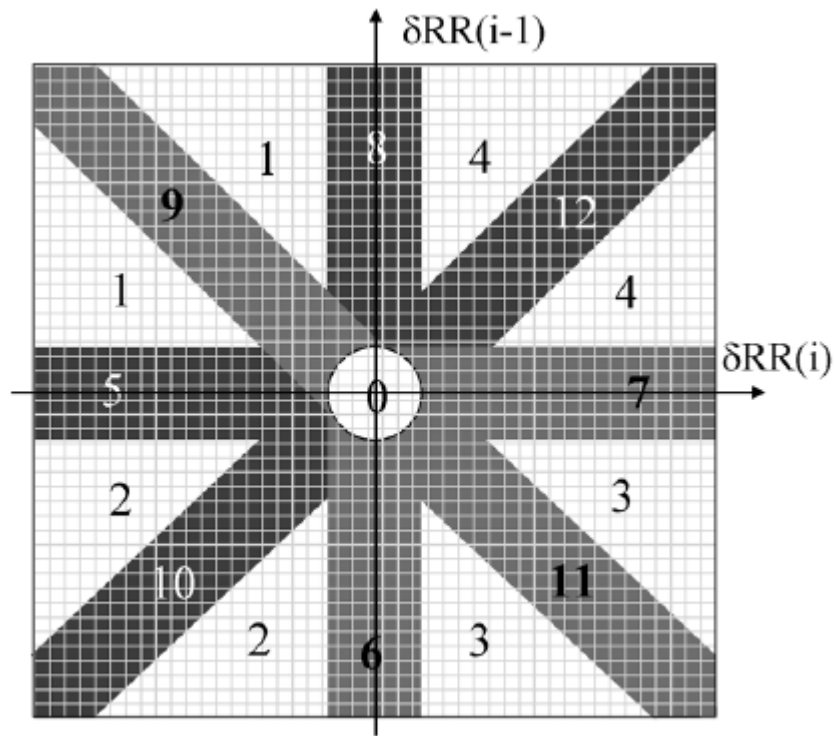


Figura 12 – Extrassístole ventricular. Conteúdo disponibilizado por (HEILMAN, , p. 10), sob a licença Creative Commons Attribution Share-Alike 3.0 Unported (“Creative Commons — Attribution-ShareAlike 3.0 Unported — CC BY-SA 3.0”, ).

## 2.7 Gráfico de Lorenz sobre RR e $\delta RR$

O trabalho desenvolvido por Sarkar e seus colaboradores utilizou o gráfico de Lorenz para identificar padrões de aceleração e desaceleração do ritmo cardíaco, para dessa maneira distinguir fibrilação atrial do ritmo normal e de outras arritmias. O gráfico de Lorenz é um gráfico de dispersão de  $\delta RR(i)$  contra  $\delta RR(i-1)$  codifica em um único ponto as direções de mudança dos intervalos RR de três batimentos consecutivos. (SARKAR *et al.*, 2008)

O gráfico de Lorenz é exemplificado na Figura 13, com as suas respectivas regiões, e o mapeamento de diferentes tipos de batimentos cardíacos é resumido na Tabela 2.



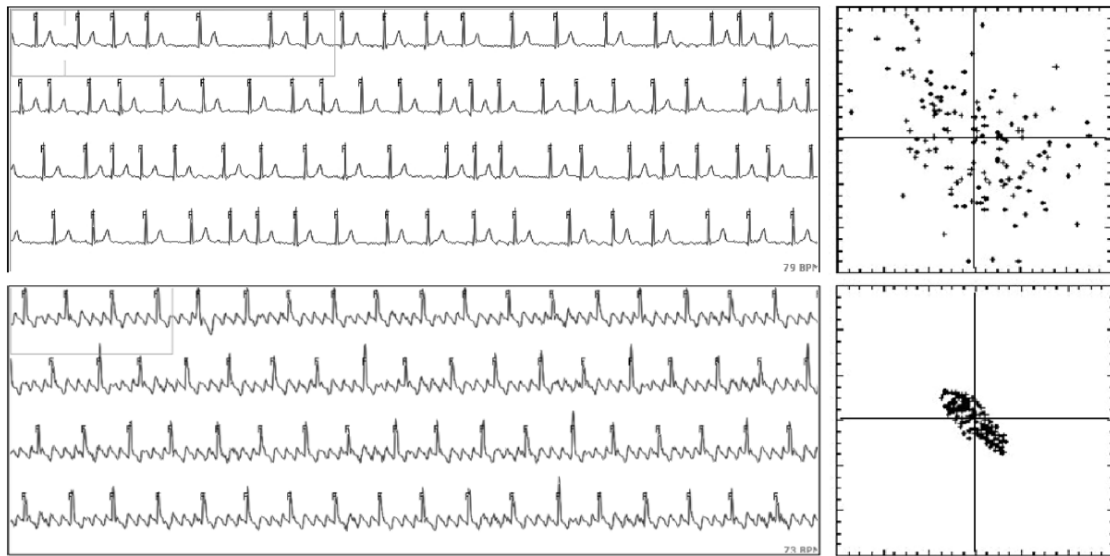
*Figura 13 – Histograma 2-D, representação numérica do diagrama de Lorenz, com dispersão de  $\delta RR(i)$  contra  $\delta RR(i-1)$ . O significado das regiões numeradas é melhor explicada na tabela a seguir. Reproduzido a partir do artigo de (SARKAR et al., 2008). A IEEE, editora deste artigo, não exige solicitação formal de autorização de uso para trabalhos em teses e dissertações. IEEE © 2008.*

Para exemplificar a análise desse gráfico, algumas regiões serão tomadas como exemplo. A região 0 é a ocupada majoritariamente pelo ritmo sinusal normal e quando existe uma taquicardia com pouca variação dos intervalos RR. Por sua vez, o segmento 9 é ocupado quando existe um batimento curto, seguido de um batimento longo seguido de outro batimento curto, ou seja, é uma sequência de aceleração e desaceleração. Na Figura 14 segue uma comparação de como o gráfico muda diante de uma fibrilação atrial e de uma taquicardia.

Tabela 2 – Mapeamento de regiões do gráfico de Lorenz

Segmentos	Sequência	Valores de $\delta RR$
0	P-P-P/G-G-G	$ \delta RR  < máscaraRSN$
1	P-G-P	$ \delta RR(i - 1)  \neq  \delta RR(i) $
2	G-M-P	$ \delta RR(i - 1)  \neq  \delta RR(i) $
3	G-P-G	$ \delta RR(i - 1)  \neq  \delta RR(i) $
4	P-M-G	Um dos $ \delta RR $ menor que máscaraRSN
5	G-G-P	Um dos $ \delta RR $ menor que máscaraRSN
6	G-P-P	Um dos $ \delta RR $ menor que máscaraRSN
7	P-P-G	Um dos $ \delta RR $ menor que máscaraRSN
8	P-G-G	$ \delta RR(i - 1)  \cong - \delta RR(i) $
9	P-G-P	$ \delta RR(i - 1)  \cong - \delta RR(i) $
10	G-M-P	$ \delta RR(i - 1)  \cong  \delta RR(i) $
11	G-P-G	$ \delta RR(i - 1)  \cong - \delta RR(i) $
12	P-M-G	$ \delta RR(i - 1)  \cong  \delta RR(i) $

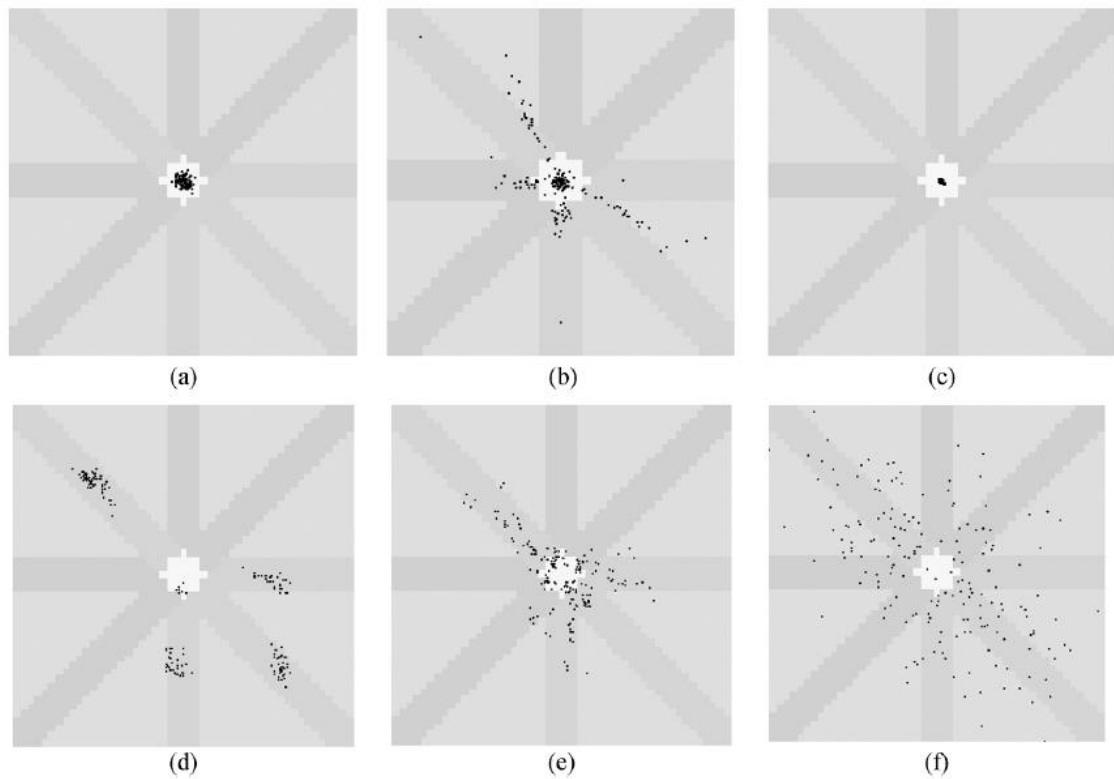
A numeração dos segmentos corresponde à mostrada na Figura 13. A coluna sequência faz referência ao tamanho de cada intervalo RR de uma sequência de três batimentos representada por um ponto no gráfico de Lorenz. Pequeno é abreviado por P, médio é abreviado por M e grande é abreviado por G. MáscaraRSN (Ritmo Sinusal Normal) corresponde ao raio da região 0 do gráfico de Lorenz, para a qual foi utilizado o valor 80 ms. Adaptado e traduzido a partir do artigo de (SARKAR et al., 2008) A IEEE, editora deste artigo, não exige solicitação formal de autorização de uso para trabalhos em teses e dissertações. IEEE © 2008.



*Figura 14 – Exemplo de 2 minutos de sinal de fibrilação atrial (parte superior) e 2 minutos de taquicardia atrial com flutter (parte inferior) com o corresponde gráfico de Lorenz. As variações da fibrilação atrial são dispersas e não correlacionadas, enquanto as da taquicardia se concentram bem ao redor da origem. Reproduzido a partir do artigo de (SARKAR et al., 2008). A IEEE, editora deste artigo, não exige solicitação formal de autorização de uso para trabalhos em teses e dissertações. IEEE © 2008.*

A Figura 15 mostra a distribuição da ocupação de áreas do histograma de acordo com o tipo de ritmo cardíaco apresentado.





*Figura 15 – Gráfico de Lorenz para diferentes tipos de ritmos cardíacos: a) Ritmo sinusal normal b) séries de contrações atriais prematuras c) taquicardia atrial com resposta ventricular regular d) Taquicardia Atrial com batimentos agrupados (bigeminia, trigeminia) e) Taquicardia Atrial com resposta ventricular irregular. f) fibrilação atrial. O intervalo do gráfico é de -600ms a 600ms. A máscaraRSN (definida na Tabela 2) é de 80 ms. Reproduzido a partir do artigo de (SARKAR et al., 2008) A IEEE, editora deste artigo, não exige solicitação formal de autorização de uso para trabalhos em teses e dissertações. IEEE © 2008.*

A partir da observação destes padrões, Sarkar e seus colaboradores desenvolveram algumas características que “evidenciam” diferentes ritmos cardíacos. Os indicadores são:

$$Irregularidade = \sum_{i=1}^{12} Q_{Caixas_n} \quad (4)$$

$$Densidade = \sum_{i=5}^{12} (Q_{Pontos_n} - Q_{Caixas_n}) \quad (5)$$

$$Anisotropia = \left| \sum_{i=9,11}^{12} Q_{Pontos_n} - \sum_{i=10,12}^{12} Q_{Pontos_n} \right| + \left| \sum_{i=6,7}^{12} Q_{Pontos_n} - \sum_{i=5,8}^{12} Q_{Pontos_n} \right| \quad (6)$$

$$PACO = \sum_{i=1}^4 (Q_{Pontos_n} - Q_{Caixas_n}) + \sum_{i=5,6,10} (Q_{Pontos_n} - Q_{Caixas_n}) - \sum_{i=7,8,12} (Q_{Pontos_n} - Q_{Caixas_n}) \quad (7)$$

onde  $Q_{Pontos_n}$  é o número de pontos que ocuparam as caixas da região n e  $Q_{Caixas_n}$  é a quantidade de caixas com pelo menos um ponto ocupado na região n.

Outro indicador denominado Quantidade\_Origem conta o número de pontos dentro das caixas que contém a origem. Irregularidade mede a dispersão da ocupação de diferentes regiões que não a origem. Densidade mede o quanto os pontos se concentram em torno das mesmas caixas. Anisotropia mede a orientação da distribuição. PACO é uma medida que evidencia pausas compensatórias. Todas estas características foram usadas posteriormente no trabalho desenvolvido.

## 2.8 Gráfico de Poincaré

O gráfico de Poincaré foi proposto por Park e seus colaboradores. O princípio de visualização é semelhante ao do Lorenz explicitado anteriormente, porém trabalha sobre um conjunto diferente de características. Os eixos desse gráfico estão baseados no intervalo RR de um batimento e no seu consecutivo, logo cada ponto traz informação sobre dois batimentos. Diferentes padrões de ritmos cardíacos se refletem em diferentes

padrões visuais (PARK *et al.*, 2009), conforme exemplificado na Figura 16, Figura 17 e Figura 18.

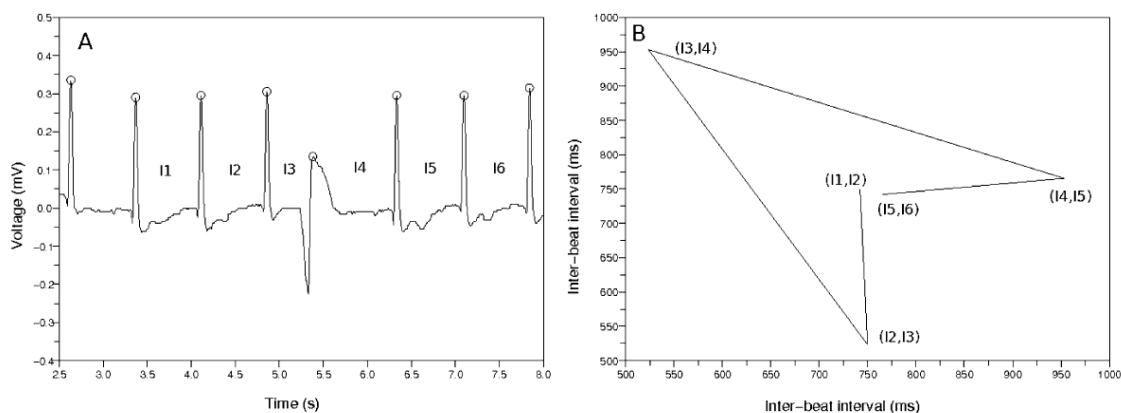


Figura 16 – Exemplo de gráfico de Poincaré – a) No gráfico da esquerda é exibida uma amostra de ECG que exibe contração ventricular prematura b) O gráfico da direita é do tipo Poincaré e foi montado a partir dos intervalos indicados no gráfico à esquerda. Observar que é formado um padrão visual que se assemelha a uma cunha. Figura copiada de um trabalho publicado de acordo com a licença livre Attribution 2.0 Generic (<https://creativecommons.org/licenses/by/2.0/>) (PARK *et al.*, 2009).

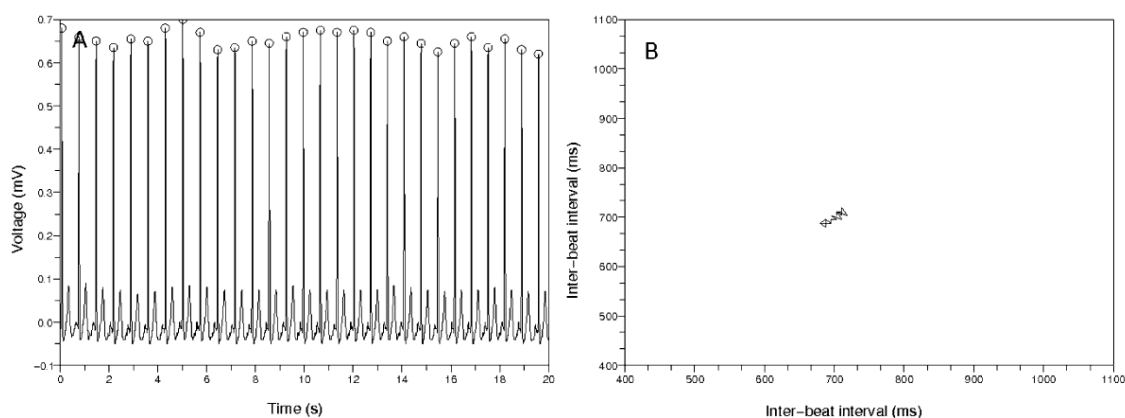


Figura 17 – Exemplo de gráfico de Poincaré para batimentos sem fibrilação atrial. – a) O gráfico da esquerda exibe uma amostra de ECG com batimentos com intervalo uniforme. b) O gráfico da direita do tipo Poincaré mostra que os pontos nesse caso se reúnem em torno de um único ponto. Figura copiada de um trabalho publicado de acordo com a licença livre Attribution 2.0 Generic (<https://creativecommons.org/licenses/by/2.0/>) (PARK *et al.*, 2009).

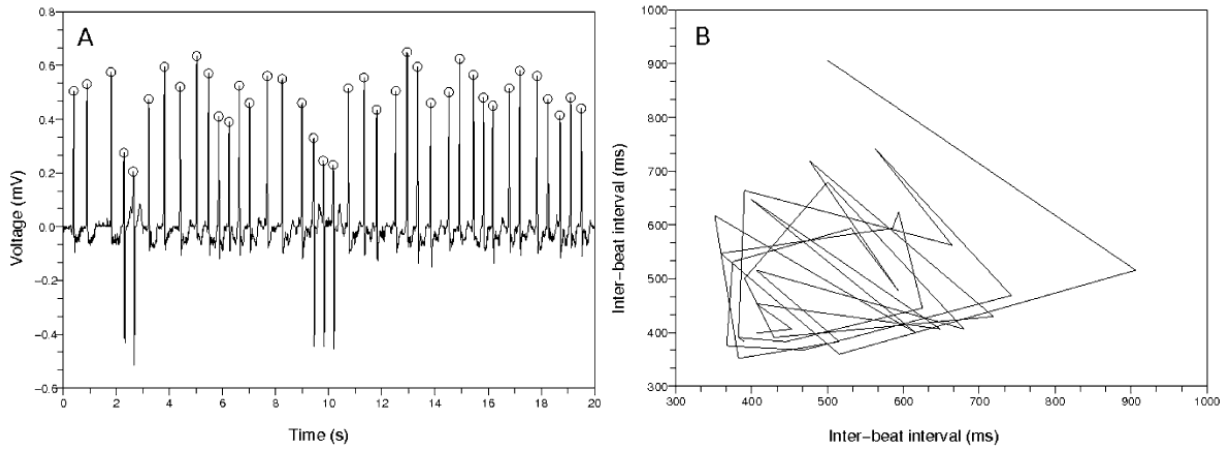


Figura 18 – Exemplo de gráfico de Poincaré para ECG exibindo fibrilação atrial. a) No gráfico da esquerda é exibido um sinal com batimentos com intervalos irregularmente distribuídos. b) Observar que não há padrão no gráfico da direita. Os batimentos frequentemente atravessam a linha diagonal. Figura copiada de um trabalho publicado de acordo com a licença livre Attribution 2.0 Generic (<https://creativecommons.org/licenses/by/2.0/>) (PARK et al., 2009).

A partir da observação dos padrões exemplificados, duas características foram propostas por Park e seus colaboradores. A primeira é o incremento médio do intervalo entre batimentos, que, considerando dois pontos consecutivos  $(I_j, I_{j+1})$  e  $(I_{j+1}, I_{j+2})$  conforme a Figura 16, é dado pela fórmula:

$$IM = \frac{\frac{1}{n-2} \sum_{j=1}^{n-2} \sqrt{(I_j - I_{j+1})^2 + (I_{j+1} - I_{j+2})^2}}{\frac{1}{n} \sum_{j=1}^n I_j} \quad (8)$$

Para encontrar a segunda característica, é necessário encontrar o ponto central  $(x, x)$  que minimiza a sua soma de distância quadrática em relação a todos os pontos do gráfico. Este ponto central  $(a, a)$  é dado por:

$$a = \frac{1}{2(n-1)} (-I_1 - I_n + 2 \sum_{j=1}^{n-1} I_j) \quad (9)$$

O desvio padrão da distância entre os pontos do gráfico e um gráfico de linha reta  $y = x$  é dado por:

$$\sqrt{\frac{1}{2(n-1)} \sum_{j=1}^{n-1} (I_j - I_{j-1})^2 - \left( \frac{1}{(n-1)\sqrt{2}} \sum_{j=1}^{n-1} |I_j - I_{j+1}| \right)^2} \quad (10)$$

A característica utilizada para representar a dispersão normalizada dos pontos é dada então por:

$$dispersão = \frac{\sqrt{\frac{1}{2(n-1)} \sum_{j=1}^{n-1} (I_j - I_{j-1})^2 - \left( \frac{1}{(n-1)\sqrt{2}} \sum_{j=1}^{n-1} |I_j - I_{j+1}| \right)^2}}{\frac{1}{2(n-1)} (-I_1 - I_n + 2 \sum_{j=1}^{n-1} I_j)} \quad (11)$$

Estas duas características também são utilizadas no trabalho desenvolvido.

## 2.9 Seleção de Características por Componentes

### Vizinhos

A técnica utilizada para selecionar o conjunto de características mais relevantes para o problema de classificação é a Seleção de Características por Componentes Vizinhos (SCCV). Essa é uma técnica baseada em componentes vizinhos que busca maximizar a acurácia de classificação determinando pesos associados a cada característica, sendo tais pesos obtidos por otimização (YANG *et al.*, 2012).

Segundo o autor desta técnica, existem três tipos básicos de seletores de características: filtro, empacotador e embutido. Seletores do tipo filtro buscam avaliar características intrínsecas como distância interclasse (Fisher Score). Seletores do tipo empacotador se baseiam na performance de classificação de uma técnica de classificação específica para avaliar o conjunto de características a cada iteração. Por outro lado, os seletores do tipo embutido utilizam um tipo de “classificador” no qual a seleção de características esteja embutida e numa fase seguinte métodos de otimização são utilizados para encontrar os pesos para cada característica. O seletor do tipo filtro tem geralmente um desempenho menor do que os do tipo empacotador e embutido. Por sua vez, o método embutido geralmente é bem mais eficiente porque não precisa que um processo completo de classificação seja feito para cada subconjunto de características, mas determina um conjunto único de pesos para as características em um único fluxo. O método embutido possui ainda a vantagem de não estar intrinsecamente ligado a um método particular, ou seja, o mesmo conjunto de características selecionadas pode ser usado em diversos classificadores (YANG *et al.*, 2012).

A técnica pode ser descrita da seguinte maneira:

Seja  $T = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\}$  um conjunto de treinamento, onde  $x_i$  é um vetor de características  $d$ -dimensional e  $y_i$  é seu rótulo correspondente. Deve-se encontrar um vetor de pesos  $w$  que otimiza a acurácia de classificação baseada em classificação de vizinhos mais próximos. A distância ponderada entre duas amostras  $x_i$  e  $x_j$  é dada por:

$$D_w(x_i, x_j) = \sum_{l=1}^d w_l^2 |x_{il} - x_{jl}| \quad (12)$$

onde  $w_l$  é o peso da característica  $l$ .

Uma forma eficaz de determinar a acurácia do método de vizinho mais próximo seria executar a classificação utilizando a validação cruzada, deixando uma amostra por vez para teste e usando o resto do banco como treinamento. Essa proposta é comumente denominada como validação “*leave-one-out*”. O processo de seleção do vizinho mais próximo, no entanto, é um processo não diferenciável, portanto se optou por determinar que o ponto de referência é dado por uma distribuição de probabilidade que é inversa à distância relativa:

$$p_{ij} = \begin{cases} \frac{\kappa(D_w(x_i, x_j))}{\sum_{k \neq i} \kappa(D_w(x_i, x_k))}, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (13)$$

onde  $\kappa(z) = \exp(-z/\sigma)$  é uma função de *kernel* e o comprimento de *kernel*  $\sigma$  é um importante parâmetro que influencia a escolha do ponto de referência. Se  $\sigma$  tende a zero, apenas o vizinho mais próximo tende a ser escolhido. Se  $\sigma$  tende ao infinito, todos os pontos têm a mesma probabilidade de serem escolhidos. Sendo assim, a probabilidade de se escolher corretamente o ponto  $x_i$  é:

$$p_i = \sum_j y_{ij} p_{ij} \quad (14)$$

onde  $y_{ij} = 1$  se e apenas se  $y_i = y_j$  e 0 caso contrário. A acurácia de classificação *leave-one-out* é dada por:

$$\xi(w) = \frac{1}{N} \sum_i p_i = \frac{1}{N} \sum_i \sum_j y_{ij} p_{ij} \quad (15)$$

Os autores da técnica introduzem um peso de regularização ao termo para penalizar o uso de pesos com alto valor.

$$\xi(w) = \sum_i \sum_j y_{ij} p_{ij} - \lambda \sum_{l=1}^d w_l^2 \quad (16)$$

Esse hiperparâmetro pode ser estimado por validação cruzada. O coeficiente  $1/N$  é ignorado porque sua presença apenas muda o valor ótimo do hiperparâmetro sem alterar o resultado final.

Substituindo a função de *kernel* na equação 13, obtém-se:

$$p_{ij} = \frac{e^{-\frac{D_w(x_i, x_j)}{\sigma}}}{\sum_{k \neq i} e^{-\frac{D_w(x_i, x_k)}{\sigma}}} \quad (17)$$

Substituindo equação 12 na equação 17:

$$p_{ij} = \frac{e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{jl}|}{\sigma}}}{\sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}{\sigma}}} \quad (18)$$

Note que se  $j = k$ , tem-se:

$$p_{ik} = \frac{e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}{\sigma}}}{\sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}{\sigma}}} \quad (19)$$

Derivando  $p_{ij}$  por  $w_l$  com o uso da regra da cadeia e do quociente, tem-se:

$$\begin{aligned} \left( \frac{\partial p_{ij}}{\partial w_l} \right) &= \frac{\left( \sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}{\sigma}} \right) \left( -\frac{2w_l |x_{il} - x_{jl}|}{\sigma} \right) \left( e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{jl}|}{\sigma}} \right)}{\left( \sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}{\sigma}} \right)^2} \\ &\quad - \frac{\left( e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{jl}|}{\sigma}} \right) \left( -\sum_{k \neq i} \frac{2w_l |x_{il} - x_{kl}|}{\sigma} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}{\sigma}} \right)}{\left( \sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}{\sigma}} \right)^2} \end{aligned} \quad (20)$$

Percebe-se que a derivada de  $p_{ij}$  em ambos os termos possui um componente igual a  $p_{ij}$  conforme a equação 18:

$$\left(\frac{\partial p_{ij}}{\partial w_l}\right) = p_{ij} \left\{ \frac{\left( \sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}} \right) \left( -\frac{2w_l |x_{il} - x_{jl}|}{\sigma} \right)}{\left( \sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}} \right)} - \frac{\left( \sum_{k \neq i} e^{-\frac{2w_l |x_{il} - x_{kl}|}{\sigma}} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}} \right)}{\left( \sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}} \right)} \right\} \quad (21)$$

No primeiro termo dentro dos colchetes há um termo no numerador que é igual ao do denominador, portanto pode ser cortado. No segundo termo uma reagrupação de termos é feita para proporcionar uma substituição, colocando o denominador dentro do somatório do denominador. Assim,

$$\left(\frac{\partial p_{ij}}{\partial w_l}\right) = p_{ij} \left\{ \left( -\frac{2w_l |x_{il} - x_{jl}|}{\sigma} \right) - \left( \sum_{k \neq i} e^{-\frac{2w_l |x_{il} - x_{kl}|}{\sigma}} \frac{e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}}}{\left( \sum_{k \neq i} e^{-\frac{\sum_{l=1}^d w_l^2 |x_{il} - x_{kl}|}} \right)} \right) \right\} \quad (22)$$

No segundo termo é encontrado um componente igual a  $p_{ik}$  conforme a equação 19.

$$\left(\frac{\partial p_{ij}}{\partial w_l}\right) = p_{ij} \left\{ \left( -\frac{2w_l |x_{il} - x_{jl}|}{\sigma} \right) + \left( \sum_{k \neq i} \frac{2w_l |x_{il} - x_{kl}|}{\sigma} p_{ik} \right) \right\} \quad (23)$$

A derivada de  $p_{ij}$  pode ser descrita mais sucintamente, então, como:

$$\left(\frac{\partial p_{ij}}{\partial w_l}\right) = \frac{2p_{ij}w_l}{\sigma} \left\{ -|x_{il} - x_{jl}| + \left( \sum_{k \neq i} |x_{il} - x_{kl}| p_{ik} \right) \right\} \quad (24)$$

Voltando à equação 16, a sua derivada em relação a  $w_l$  é igual a:

$$\frac{\partial \xi(w)}{\partial w_l} = \sum_i \sum_j y_{ij} \frac{\partial p_{ij}}{\partial w_l} - 2w_l \lambda \quad (25)$$

Substituindo  $\frac{\partial p_{ij}}{\partial w_l}$  pela equação 24, tem-se:

$$\frac{\partial \xi(w)}{\partial w_l} = \sum_i \sum_j y_{ij} \left[ \frac{2p_{ij}w_l}{\sigma} \left\{ -|x_{il} - x_{jl}| + \left( \sum_{k \neq i} |x_{il} - x_{kl}| p_{ik} \right) \right\} \right] - 2w_l \lambda \quad (26)$$

Usando equação 14 para encontrar  $p_i$ , tem-se:

$$\frac{\partial \xi(w)}{\partial w_l} = \frac{2w_l}{\sigma} \sum_i \left[ -\sum_j y_{ij} p_{ij} |x_{il} - x_{jl}| + p_i \left( \sum_{k \neq i} |x_{il} - x_{kl}| p_{ik} \right) \right] - 2w_l \lambda \quad (27)$$

Finalmente, os resultados podem ser resumidos em:



$$\frac{\partial \xi(w)}{\partial w_l} = 2w_l \left\{ \frac{1}{\sigma} \sum_i [-\sum_j y_{ij} p_{ij} |x_{il} - x_{jl}| + p_i (\sum_{j \neq i} |x_{il} x_{jl}| p_{ij})] - \lambda \right\} \quad (28)$$

Esta equação da derivada de acurácia é utilizada por um método numérico denominado gradiente ascendente. O algoritmo proposto e denominado de seleção de características por componentes vizinhos é dado por:

T – conjunto de treinamento;  $\alpha$ : comprimento de passo inicial;  $\sigma$ : comprimento de kernel.  $\Lambda$  - parâmetro de regularização;  $\eta$  – pequena constante positiva.

Inicialização:  $w^{(0)} = (1, 1, \dots, 1)$ ,  $\epsilon^{(0)} = -\infty$ ,  $t=0$

Enquanto  $|\epsilon^{(t)} - \epsilon^{(t-1)}| < \eta$

Para  $i = 1, \dots, N$  (número de amostras)

Calcular  $p_{ij}$  e  $p_i$  usando  $w^{(t)}$  de acordo com as equações 13 e 14

Para  $l = 1, \dots, d$

$$\Delta_l = 2w_l^{(t)} \left\{ \frac{1}{\sigma} \sum_i [-\sum_j y_{ij} p_{ij} |x_{il} - x_{jl}| + p_i (\sum_{j \neq i} |x_{il} - x_{jl}| p_{ij})] - \lambda \right\}$$

$t=t+1$

$$w_l^{(t)} = w_l^{(t-1)} + \alpha \Delta$$

$$\epsilon^{(t)} = \xi(w^{(t)})$$

Se  $\epsilon^{(t)} > \epsilon^{(t-1)}$  então:

$$\alpha = 1,01\alpha$$

Caso contrário:

$$\alpha = 0,4\alpha$$

$w = w^{(t)}$

Resultado final =  $w$

O autor da técnica de SCCV comparou seu método com o de outros três autores: Simba (RANB *et al.*, 2004), FSSUN (SUN *et al.*, 2010) e LMFV (CHEN *et al.*, 2009), todas baseadas em K vizinhos mais próximos (método que será descrito na seção 2.10.3). Testes comparativos foram feitos em oito grupos de testes largamente utilizados (ALON *et al.*, 1999; ALIZADEH *et al.*, 2000; STATNIKOV *et al.*, 2005), incluindo *BrainTumor2*, *Colon\_Cancer*, *DLBCL*, *Leukemia1*, *Prostate\_Tumor* e *SRBCT*. Nesse subconjunto, quatro problemas de duas classes e quatro problemas de multiclasse foram

criados. Em todos os problemas, o número de genes analisados é muito maior do que o número de amostras. Busca-se correlacionar a presença do gene com a condição clínica. Utilizou-se uma mesma técnica de K vizinhos mais próximos com a distância tipo Manhattan associado a uma validação cruzada *leave-one-out* (deixar um para fora) para estimar o erro de classificação de acordo com o número de características escolhidas.

Dos oito *arrays* testados, o SCCV obteve o menor erro de classificação em quatro casos, empatou com outros métodos em três casos e não obteve o menor erro de classificação em um caso.

## 2.10 Métodos de classificação

### 2.10.1 O Classificador de Bayes

O menor erro de teste é obtido, em média, quando se designa cada observação para sua classe mais provável, dados seus preditores (JAMES *et al.*, 2013a). Ou seja, uma observação de teste com vetor de predição  $x_0$  deve ser designado à classe  $j$  para a qual

$$P(Y = j|X = x_0) \quad (29)$$

é maior. Essa probabilidade é uma probabilidade condicional: a probabilidade de  $Y$  ser igual a  $j$  dado que  $X = x_0$ .

O conjunto de pontos para uma classe  $j$  para a qual a probabilidade indicada na equação 29 é maior forma uma região para esta classe. Nos pontos em que essa probabilidade condicional de uma classe  $j$  é igual a de outra classe diferente de  $j$ , pode-se escolher tanto uma classe quanto a outra, formando assim uma fronteira de Bayes.

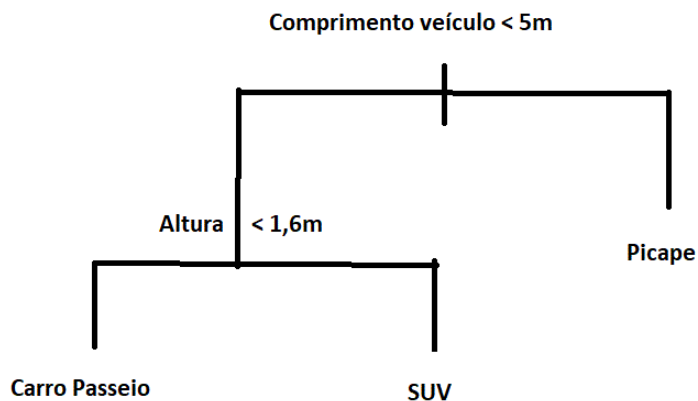
O teorema de Bayes diz que (ALBUQUERQUE *et al.*, 2008):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (30)$$

## 2.10.2 Árvore de decisão

As árvores de decisão especificam uma série de decisões previamente definidas para cada amostra. As decisões partem de uma “raiz” (o nó inicial) até as folhas, que são os nós finais. Árvores de classificação dão respostas categóricas, enquanto árvores de regressão dão respostas numéricas. Para este trabalho, apenas árvores de classificação foram utilizadas.

Cada decisão da árvore de decisão é binária e atua sobre um único preditor por vez conforme ilustrado na Figura 19.



*Figura 19 -Exemplo hipotético de passo de árvore de decisão. O modelo deve escolher se um veículo pertence à categoria Carro Passeio, Picape ou SUV. No nó raiz, o inicial, se a comprimento veículo maior do que 5m, diz-se que o veículo é uma picape. Se menor do que 5m, a variável altura é avaliada. Se a altura é menor do que 1,6m, diz-se que o veículo é um carro de passeio; se maior do que 1,6m, diz-se que o veículo é um SUV.*

O algoritmo utilizado é baseado no algoritmo CART, (BREIMAN *et al.*, 1993) que consiste nos seguintes passos: (“Decision Trees - MATLAB & Simulink”, )

- No nó raiz, usar todo o conjunto de treinamento, comparando todo tipo possível de divisão binária em todos os preditores.
- Comparar o desempenho das divisões segundo algum critério.
- Efetuar a divisão que for escolhida.

- d) Repetir recursivamente o procedimento para os nós filhos até que se alcance uma das condições de parada.

As regras de parada, na versão utilizada do MATLAB® (2017a), são:

- O nó é puro, ou seja, só contém observações de uma das classes.
- Há menos que TamanhoMínimoNóPai de observações no nó que faria a divisão
- Qualquer divisão que seria feita gera um nó filho com menos do que TamanhoMínimoFolha
- O procedimento excede o número NumeroMáximoDivisões de nós-ramo

Os possíveis critérios de otimização para a melhor divisão são: (“Fit binary classification decision tree for multiclass classification - MATLAB fitctree”, )

- Índice de diversidade de Gini (gdi) – O índice de Gini de um nó é

$$1 - \sum_i p^2(i) \quad (31)$$

onde o índice  $i$  se refere às classes de resultados envolvidas no processo, e  $p(i)$  é a fração da classe  $i$  presente no nó em relação a todas as outras. Um nó com apenas uma classe (nó puro) possui algum  $p(i) = 1$ , conseqüentemente seu gdi é igual a zero. Logo essa é uma medida de “impureza” do nó.

- Desvio – Definindo-se  $p(i)$  da mesma maneira que antes, tem-se:

$$- \sum_i p(i) \log p(i) \quad (32)$$

Um nó puro também tem desvio zero. Qualquer outra proporção gera um valor positivo, que também é uma medida de impureza.

- Regra de Twoing – Ao contrário dos dois exemplos anteriores, esta medida não representa a impureza do nó filho. No entanto, ela indica se uma determinada divisão é melhor do que outra. Uma divisão é melhor quando esta leva a ramos filhos mais “puros”. Quanto maior o valor, melhor é a divisão. O cálculo é dado por:

$$P(E)P(D)(\sum_i |E(i) - D(i)|)^2 \quad (33)$$

onde  $P(E)$  e  $P(D)$  são as frações do nó pai que se dividem para os nós filhos da esquerda e direita respectivamente.  $E(i)$  e  $D(i)$  são as frações de observações da classe(i) no nó filho, respectivamente o esquerdo e o direito.

No presente estudo não se trabalhou com amostras faltantes, portanto o tratamento para este caso não está sendo abordado. Além disso, ele tem como objetivo classificar amostras de sinal em variáveis categóricas, logo o processo para determinar o valor de variáveis contínuas também não será abordado. O processo exato para determinar as divisões é dado por (para cada preditor  $x_j$ ): (“Fit binary classification decision tree for multiclass classification - MATLAB fitctree”, )

- 1) O erro médio quadrático ponderado das respostas no nó  $t$  é dado por:

$$\varepsilon_t = \sum_{j \in T} w_j (y_j - \bar{y}_t)^2 \quad (34)$$

onde  $T$  representa todas as informações do nó  $t$ ,  $w_j$  é o peso da observação  $j$ ,  $\bar{y}_t$  é o valor verdadeiro da variável resposta e  $y_j$  é o valor predito da variável resposta.

- 2) A probabilidade de que um determinado conjunto de observações esteja no nó  $t$  é ponderada de acordo com pesos atribuídos a cada observação e é dada por:

$$P(T) = \sum_{j \in T} w_j \quad (35)$$

onde  $T$  é o conjunto de todas as observações presentes no nó  $t$ .

- 3) Os preditores  $x_i$  são ordenados de forma crescente, sendo que cada elemento deste preditor pode ser considerado como um candidato a um ponto de corte.
- 4) É necessário escolher qual candidato é o melhor ponto de corte para cortar o nó  $T$  usando  $x_i$ . O algoritmo se utiliza para este fim do ganho de impureza ( $\Delta I$ ). Este cálculo é dado por:
  - a. O nó é dividido nos nós filhos esquerdos e direitos no ponto de corte candidato ( $t_L$  e  $t_R$ ).

- b. Considerar  $T_L$  e  $T_R$  os conjuntos de amostras contidos em  $t_L$  e  $t_R$ , e  $i$  a impureza do nó indicado no seu respectivo índice subscrito. Calcular  $\Delta I$ , que é dado por:

$$\Delta I = P(T)i_t - P(T_L)i_{t_L} - P(T_R)i_{t_R} \quad (36)$$

- 5) O candidato escolhido é aquele com maior  $\Delta I$ .

Para um determinado nó, o processo descrito é repetido para todos os preditores. A divisão é feita no preditor que apresentar o maior  $\Delta I$ . O procedimento é então repetido para cada nó filho até que este atinja uma das condições de parada.

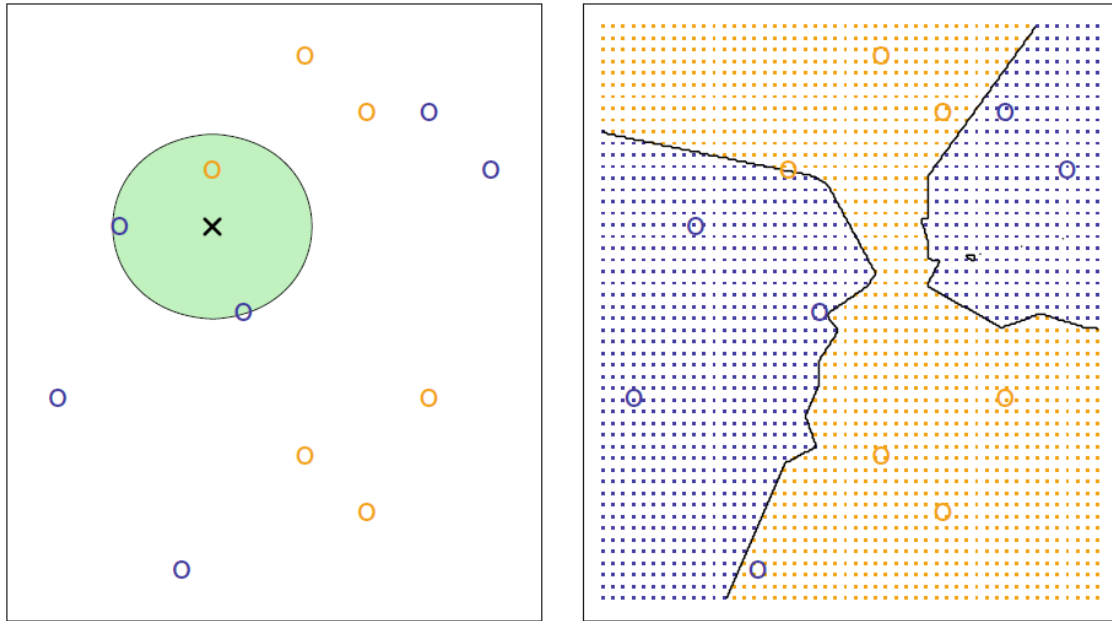
### 2.10.3 K vizinhos mais próximos (KVP)

O classificador KVP é um dos tipos mais simples. Dada uma observação de teste  $x_0$ , o classificador procura os  $K$  pontos no conjunto de treinamento que mais estão próximos de  $x_0$ , representados por  $\mathcal{N}_0$ . Então o KVP estima a probabilidade condicional para a classe  $j$  como a fração de pontos em  $\mathcal{N}_0$  cuja resposta seja igual a  $j$  (JAMES *et al.*, 2013b):

$$P(Y = j|X = x_0) = \frac{1}{K} \sum_{i \in \mathcal{N}_0} I(y_i = j) \quad (37)$$

A aplicação da regra de Bayes descrita na seção 2.10.1 determina então que a observação de teste seja designada à classe com maior probabilidade. Uma ilustração desse processo se encontra na Figura 20.

O valor escolhido para  $K$  determina o nível de flexibilidade do classificador. Se  $K = 1$ , a fronteira será bem flexível e variável. Se  $K$  for 100, será bem mais rígida e inflexível. Esse conceito é ilustrado na Figura 21. A escolha do  $K$  ideal de maneira objetiva é dependente de conhecer o erro de teste para várias escolhas de  $K$ . A maneira correta de determinar o erro de teste será explicitada de forma melhor mais adiante.



*Figura 20 – Ilustração de classificação KVP. Esquerda: Os três pontos mais próximos da observação de teste (cruz preta) são identificados pelo círculo verde. Como o tipo de observação mais comum entre os mais próximos é o azul, a observação de teste é predita como azul. Direita: Ilustração das fronteiras e áreas de decisão para cada classe. Adaptado com permissão de Springer Nature (JAMES et al., 2013a).*

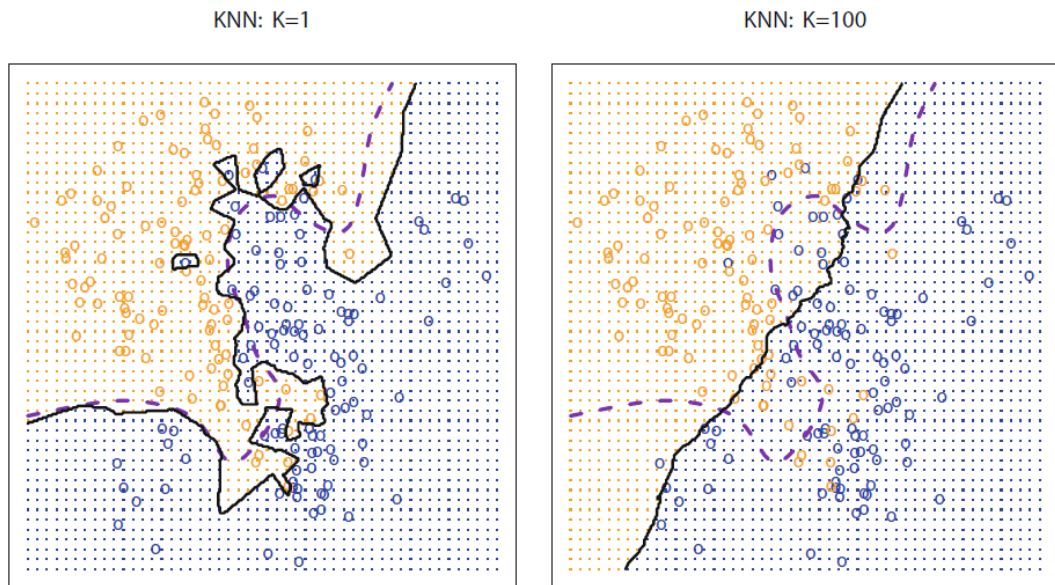


Figura 21 – Ilustração do efeito do número de  $K$  vizinhos na fronteira de classificação. Quando  $K=1$  (gráfico da esquerda), a fronteira de região é bem flexível e instável. Quando  $K=100$ , a fronteira é bem rígida e não acompanha a não linearidade do problema. Adaptado com permissão de Springer Nature (JAMES et al., 2013a).

Para implementação, será considerado que se têm duas matrizes, a de treinamento e a de teste. A matriz de treinamento ( $m_x$  por  $n$ ) é composta de  $m_x$  vetores (1 por  $n$ )  $x_1, x_2, \dots, x_{m_x}$  e a matriz de teste ( $m_y$  por  $n$ ) é composta de  $m_y$  vetores (1 por  $n$ )  $y_1, y_2, \dots, y_{m_y}$ . Com essas definições, as medidas de distância entre pontos disponíveis no programa utilizado na pesquisa (MATLAB® 2017®) são (“Classification Using Nearest Neighbors - MATLAB & Simulink”,):

- Medida de Minowski. Para  $p = 1$ , essa medida fornece a medida de quarteirão. Para  $p = 2$ , fornece a distância Euclidiana. Para  $p = \infty$ , fornece a distância de Chebychev.

$$d_{st} = \sqrt[p]{\sum_{j=1}^n |x_{sj} - y_{tj}|^p} \quad (38)$$

- Distância de quarteirão

$$d_{st} = \sum_{j=1}^n |x_{sj} - y_{tj}| \quad (39)$$



- Distância Euclidiana.

$$d_{st}^2 = (x_s - y_t)(x_s - y_t)' \quad (40)$$

- Distância Euclidiana padronizada, onde V é uma matriz diagonal n por n cujo elemento diagonal j é  $S(j)^2$ , onde S é o vetor que contém os pesos inversos.

$$d_{st}^2 = (x_s - y_t)V^{-1}(x_s - y_t)' \quad (41)$$

- Distância de Chebychev

$$d_{st} = \max_j \{|x_{sj} - y_{tj}|\} \quad (42)$$

- Distância de Mahalanobilis, onde C é a matriz de covariância

$$d_{st}^2 = (x_s - y_t)C^{-1}(x_s - y_t)' \quad (43)$$

- Distância de cosseno

$$d_{st} = \left( 1 - \frac{x_s y_t'}{\sqrt{(x_s x_s')(y_t y_t')}} \right) \quad (44)$$

- Distância de correlação

$$d_{st} = \left( 1 - \frac{(x_s - \bar{x}_s)(y_t - \bar{y}_t)'}{\sqrt{(x_s - \bar{x}_s)(x_s - \bar{x}_s)'} \sqrt{(y_t - \bar{y}_t)(y_t - \bar{y}_t)'}} \right) \quad (45)$$

Sendo

$$\bar{x}_s = \frac{1}{n} \sum_j x_{sj} \quad (46)$$

e

$$\bar{y}_t = \frac{1}{n} \sum_j y_{tj} \quad (47)$$

- Distância de Hamming

$$d_{st} = \left( \frac{\#(x_{sj} \neq y_{tj})}{n} \right) \quad (48)$$

- Distância de Jaccard

$$d_{st} = \left( \frac{\#[(x_{sj} \neq y_{tj}) \cap ((x_{sj} \neq 0) \cup (y_{tj} \neq 0))]}{\#[(x_{sj} \neq 0) \cup (y_{tj} \neq 0)]} \right) \quad (49)$$

- Distância de Spearman é outra utilizada pelo MATLAB® 2017®, mas sua descrição foge ao escopo do trabalho.

### 2.10.3.1 Relação do método com parâmetros do MATLAB® 2017a

O MATLAB® define um parâmetro para o tipo de distância conforme explicitado anteriormente e no caso específico da distância de Minowski um parâmetro expoente que define qual o valor “p” a ser usado para exponenciação/radiciação. Além disso, há um parâmetro peso de distância que pondera, para as K amostras escolhidas para uma determinada observação de teste, o peso para determinar sua classificação final. Nesse caso, “igual” implica em não utilização de pesos, “inversa” implica em um peso igual  $1/\text{distância}$  e “inversaoquadrado” implica em  $1/\text{distância}^2$ .

## 2.10.4 Análise de discriminante linear

### 2.10.4.1 Introdução

Na análise discriminante linear, modela-se o comportamento de X em função da resposta Y e então se usa o teorema de Bayes para inverter a relação e obter Y em função de X. Suponha-se que uma determinada observação possa ser alocada em K classes diferentes. Nesse exemplo,  $\pi_k$  representa a probabilidade a priori de que uma observação escolhida randomicamente a partir da população seja da classe k. A expressão  $f_k(X) = P(X = x|Y = k)$  é uma função de densidade de probabilidade dos preditores, dado que a observação pertença à classe k. Denota-se  $p_k = P(Y = k|X = x)$  como a probabilidade a posteriori de que uma observação com preditores x seja de uma classe k.

A partir do teorema de Bayes e das definições acima pode-se escrever (JAMES *et al.*, 2013b):

$$P(Y = k|X = x) = \frac{P(X=x|Y=k)P(Y=k)}{P(X=x)} \quad (50)$$

$$p_k = \frac{f_k(x) \pi_k}{\sum_{l=1}^K f_l(x) \pi_l} \quad (51)$$

Se fosse possível obter  $\pi_k$  e  $f_k(X)$  verdadeiras, o classificador resultante teria o menor erro de teste possível. Como geralmente não é possível ter esses parâmetros de antemão, é necessário estimá-los. Estimar  $\pi_k$  normalmente significa usar a proporção de observações da amostra de treinamento, dado que esta tenha sido coletada randomicamente da população. Estima-se  $f_k(X)$  assumindo que esta possui uma forma simples. O modelo vai ser mais assertivo, evidentemente, na medida em que a  $f_k(X)$  verdadeira estiver próxima da  $f_k(X)$  suposta.

#### 2.10.4.1 Análise discriminante linear para um preditor ( $p = 1$ )

Este método assume que a função de densidade  $f_k(X)$  é uma Gaussiana. No caso de um preditor, esta função assume a forma:

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right)} \quad (52)$$

onde  $\mu_k$  e  $\sigma_k$  são a média e a variância para a classe k. Também se assume que todos os  $\sigma_k$  são iguais, denotando-se assim simplesmente como  $\sigma$ . Usando 52 em 51, tem-se:

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}}{\sum_{l=1}^K \pi_l \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}} \quad (53)$$

Ao se tomar o log de 53 e se preocupando apenas com os termos do numerador que são diferentes para cada classe, tem-se que a classe escolhida será aquela que apresentar o maior  $\delta_k(x)$  tal que (JAMES *et al.*, 2013b):

$$\delta_k(x) = x \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k) \quad (54)$$

O nome linear para a análise discriminante vem do fato de que a função discriminante descrita em 54 é linear em relação a x. No caso particular em que  $K = 2$  e

$\pi_1 = \pi_2$ , este classificador designa para a classe 1 se  $2x(\mu_1 - \mu_2) > \mu_1^2 - \mu_2^2$ . Caso contrário, designa para a classe 2. Então o equivalente da fronteira de Bayes seria:

$$x = \frac{\mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} = \frac{\mu_1 + \mu_2}{2} \quad (55)$$

Os parâmetros  $\mu_k$  e  $\sigma_k$  são estimados por (JAMES *et al.*, 2013b):

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i:y_i=k} x_i \quad (56)$$

$$\hat{\sigma}^2 = \frac{1}{n-K} \sum_{k=1}^K \sum_{y_i=k} (x_i - \hat{\mu}_k)^2 \quad (57)$$

$$\hat{\pi}_k = \frac{n_k}{n} \quad (58)$$

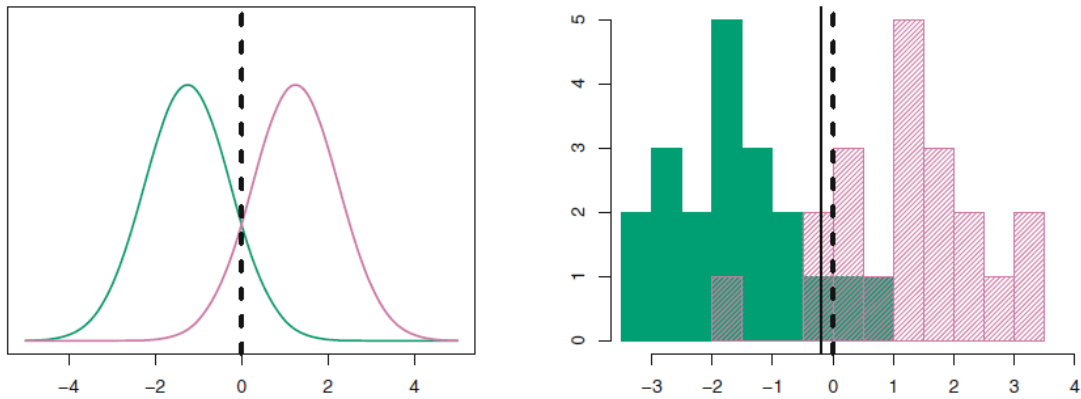


Figura 22 – Esquerda: Neste exemplo hipotético, duas funções gaussianas hipotéticas de uma dimensão representativas de uma distribuição de parâmetros para cada classe são mostradas. A linha tracejada representa a fronteira da decisão de Bayes, considerando pleno conhecimento das distribuições subjacentes. Direita: 20 observações retiradas de amostras das duas classes, mostradas como histogramas. A linha sólida representa a fronteira de decisão de Bayes como a linha tracejada. A linha vertical sólida representa a fronteira de decisão ADL (Análise de Discriminantes Lineares) feita a partir de estimativas dos dados de treinamento. Adaptado com permissão de Springer Nature (JAMES *et al.*, 2013b).

#### 2.10.4.2 Análise discriminante linear para mais de um preditor ( $p > 2$ )

Neste caso,  $X$  é um vetor  $(X_1, X_2, \dots, X_p)$ , pois mais de um preditor é usado para prever a classe. A ADL assume que  $X$  é amostrado por uma distribuição multivariada

Gaussiana, com médias e variâncias possivelmente distintas para cada classe, mas sem correlação entre os preditores. A Figura 23 ilustra o efeito das médias e variâncias distintas.

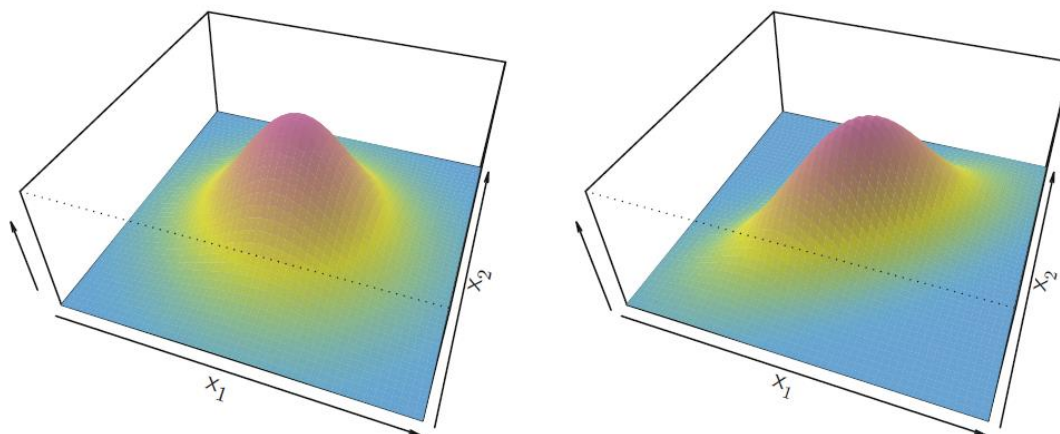


Figura 23 -Duas funções Gaussianas multivariadas são mostradas, com  $p = 2$ . A altura da superfície em qualquer ponto representa a probabilidade de que  $X_1$  e  $X_2$  venham a cair em uma região pequena em torno do ponto. Esquerda: Preditores não correlacionados e com mesma variância (Forma em sino.) Direita: As duas variáveis têm uma correlação de 0,7 (Forma elíptica). Adaptado com permissão de Springer Nature (JAMES et al., 2013b).

A notação  $X \sim N(\mu, \Sigma)$  indica que a variável aleatória  $X$  vem de uma distribuição Gaussiana multivariável.  $E(X) = \mu$  é a média de  $X$  e  $\text{Cov}(X) = \Sigma$  é a matriz  $p \times p$  de covariância de  $X$ . A função de densidade multivariada Gaussiana é definida por:

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)} \quad (59)$$

A ADL assume que as observações de cada classe  $k$  são retiradas de uma distribuição variável  $X \sim N(\mu_k, \Sigma)$ , com uma média específica  $\mu_k$  para cada classe e covariância em comum  $\Sigma$ . Ao se usar a equação 59 em 51, é possível verificar que, usando a mesma lógica de função discriminantes apresentada na seção anterior, o classificador irá atribuir à observação com preditores  $X = x$  a classe com maior valor para a função (JAMES et al., 2013b):

$$\delta_k = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k \quad (60)$$

A equação logo acima é a versão vetorial da equação 54. Esta também é dependente linearmente de  $x$ . A Figura 24 ilustra uma função gaussiana multivariável hipotética e sua aplicação com ADL.

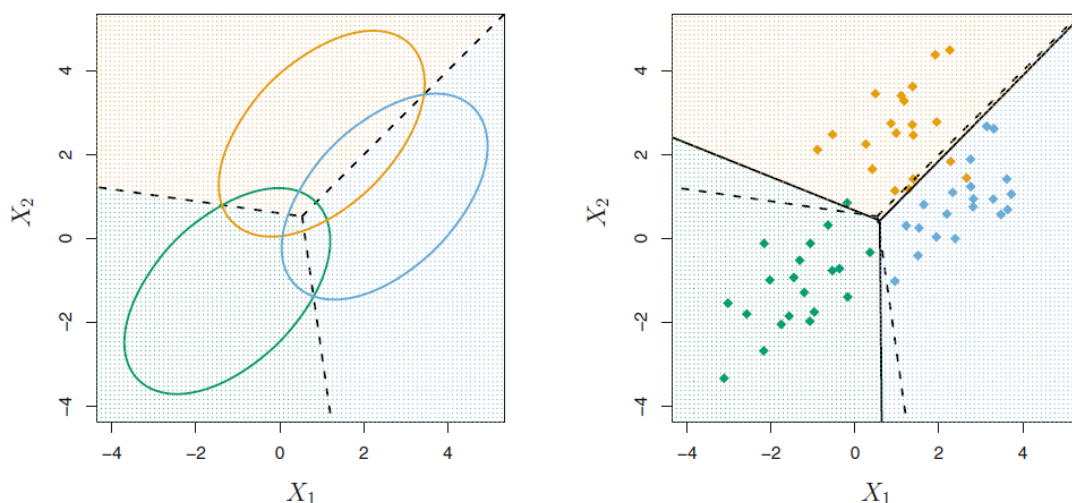


Figura 24 -Exemplo com três classes. As observações de cada classe são retiradas de distribuições multivariadas hipotéticas, com médias distintas e matrizes de covariância única. Esquerda: elipses que contém 95% de probabilidade para cada uma das três classes. A linha tracejada corresponde à fronteira de Bayes, construída baseado no conhecimento prévio da distribuição verdadeira. Direita: 20 observações são geradas de cada classe, e as fronteiras baseadas no classificador da ADL a partir da estimativa dos parâmetros da distribuição é indicado por linhas sólidas. Para comparação, a linha tracejada corresponde ao classificador de Bayes. Adaptado com permissão de Springer Nature (JAMES et al., 2013b).

As fronteiras entre as classes na hipótese de todos os  $\pi_k$  serem iguais, conforme exibido na Figura 24, correspondem aos pontos em que  $\delta_k(x) = \delta_l(x)$  para  $k \neq l$ , ou seja (JAMES et al., 2013b):

$$x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k = x^T \Sigma^{-1} \mu_l - \frac{1}{2} \mu_l^T \Sigma^{-1} \mu_l \quad (61)$$

Para estimar os parâmetros  $\mu_1, \mu_2, \dots, \mu_k, \pi_1, \pi_2, \dots, \pi_k$  e  $\Sigma$  é necessário usar as formas vetoriais dos estimadores das equações 56, 57 e 58. No caso da matriz de covariância, considera-se as amostras de todas as classes no mesmo cálculo. Os termos da diagonal se referem às variâncias de cada preditor e os termos de fora da diagonal

expressam a covariância entre um par de preditores, uma medida do quanto aqueles preditores estão correlacionados. Observa-se pelo exemplo da Figura 24 que a fronteira de separação da ADL é próxima da teórica que seria estipulada pelo classificador de Bayes, quando a função de densidade de probabilidade verdadeira é de fato uma Gaussiana de acordo com as hipóteses levantadas pela ADL.

### 2.10.4.3 Análise Discriminante Quadrática

Foi visto na seção 2.10.4.2 que a ADL assume que todas as observações vêm de uma Gaussiana multivariada com médias específicas para cada classe e uma matriz de covariância compartilhada. A Análise Discriminante Quadrática (ADQ) difere da ADL apenas por assumir que cada classe possui sua própria matriz de covariância. Em outras palavras, a observação da classe  $k$  é da forma  $X \sim N(\mu_k, \Sigma_k)$ , onde  $\Sigma_k$  é a matriz de covariância específica. Usando a mesma lógica de função discriminante explicada nas seções 2.10.4.1 e 2.10.4.2, o classificador designará a observação  $X = x$  para a classe  $k$  que obtiver o maior valor da função (JAMES *et al.*, 2013b):

$$\begin{aligned} \delta_k &= -\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) - \frac{1}{2} \log |\Sigma_k| + \log \pi_k \\ \delta_k &= -\frac{1}{2} x^T \Sigma_k^{-1} x + x^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \log |\Sigma_k| + \log \pi_k \end{aligned} \quad (62)$$

É notório que esta função depende quadraticamente de  $X$ , o que justifica o nome ADQ para o classificador. É necessário fazer estimativas de  $\pi_k$ ,  $\Sigma_k$  e  $\mu_k$ .

A diferença fundamental entre este modelo e o seu antecessor é a flexibilidade. Por supor uma matriz de covariância diferente para cada classe, é criado um modelo com maior flexibilidade de mudança, de modo que se pode ajustar a um modelo apresentando comportamento não-linear mais intenso. Um modelo linear, nesse caso, pode não se ajustar bem a um problema não-linear, revelando assim um problema de tendência. Essa flexibilidade, no entanto, tem um preço: o modelo pode se ajustar demais ao modelo de treinamento e ser muito dependente deste, sofrendo grande variância caso ocorra treinamento com uma nova amostra da população. Ou seja, é possível que o ADQ se ajuste bem ao grupo de treinamento, mas falhe ao avaliar um grupo de teste. Este é o compromisso entre tendência e variância, que deve sempre ser considerado pelo pesquisador e é ilustrado na Figura 25.

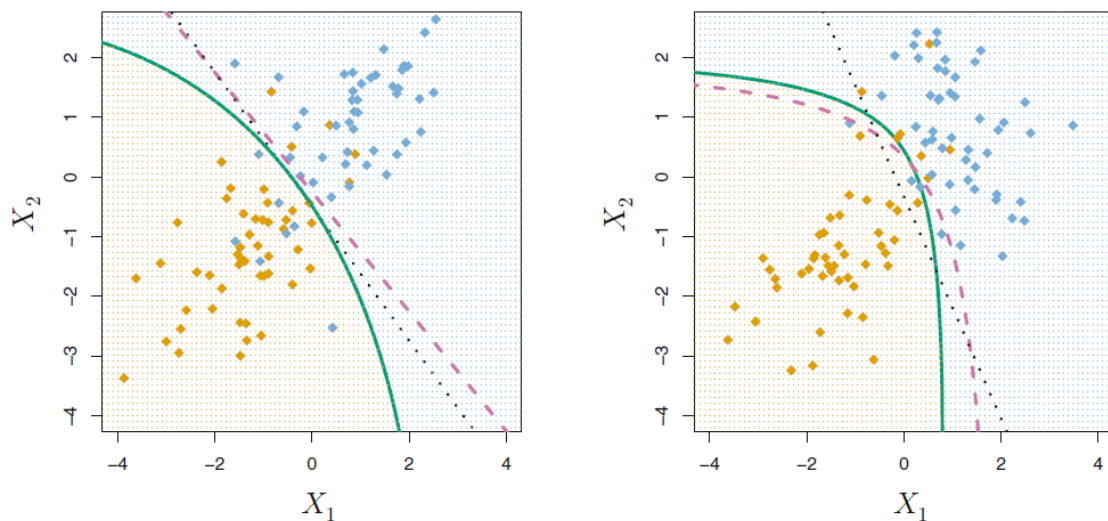


Figura 25 – Roxo tracejado– Fronteira de Bayes; Preto pontilhado – ADL; Verde sólido – ADQ. Esquerda: Fronteiras de decisão para um problema em que sabidamente os dados vem de uma gaussiana multivariável com  $\Sigma_1 = \Sigma_2$ . Nesse caso, a fronteira da ADL é mais próxima da verdadeira. Direita: Fronteiras de decisão para um problema em que sabidamente os dados vem de uma gaussiana multivariável com  $\Sigma_1 \neq \Sigma_2$ . Nesse caso, a fronteira da ADQ é mais próxima da verdadeira. Adaptado com permissão de Springer Nature (JAMES et al., 2013b).

#### 2.10.4.4 Relação do método com parâmetros do MATLAB® 2017a

Para a análise discriminante linear, os preditores cujo coeficiente linear seja menor do que o parâmetro *delta* são eliminados. O coeficiente a que se refere aqui é o de separação de planos como o da Figura 24. Por sua vez, o parâmetro *gamma* permite um ajuste fino da matriz de covariância. Se este for igual a zero, nenhuma modificação é feita, exceto se for computacionalmente impossível de se obter uma matriz inversa da covariância (ou seja, se esta for singular). Se este parâmetro for igual a 1, apenas matrizes de covariância diagonais são aceitas (alternativamente, pode-se usar o tipo “*diaglinear*” - ADL linear diagonal ou “*diagquadratic*” – ADL quadrática diagonal). Se estiver entre 0 e 1, uma regularização é feita sobre a matriz de covariância para deixá-la menos singular. O software não especifica como é feito esse processo de regularização.



## 2.10.5 Máquinas de Vetores de Suporte (MVS)

A MVS é uma generalização de um classificador simples denominado classificador de máxima margem. Este possui a limitação de exigir que os dados de treinamento sejam perfeitamente separáveis por uma fronteira linear. Uma extensão deste modelo, o classificador de suporte a vetores, pode ser aplicado em casos que não sejam perfeitamente separáveis, além de poder ser parametrizado para evitar ajuste exagerado aos dados de treinamento. A extensão deste segundo modelo, a máquina de vetores de suporte, possibilita o uso de fronteiras não-lineares. Por fim, é possível estender a máquina de vetores de suporte para problemas de multi-classe (JAMES *et al.*, 2013c).

### 2.10.5.1 Classificador de máxima margem

Em um espaço p-dimensional, um hiperplano é subespaço “plano” de dimensão p-1 que não precisa passar pela origem. Em duas dimensões é uma linha e em três dimensões é um plano. O hiperplano é definido matematicamente como:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = 0 \quad (63)$$

Qualquer ponto que satisfaça a equação 63 está no plano. Por outro lado, se o lado esquerdo da equação 63 for maior que zero, então o ponto está em um lado do plano. Se estiver menor do que zero, está do outro lado do plano.

Suponha-se que se tem uma matriz X que possua n observações de treinamento em um espaço p dimensional, com n vetores  $x_n = (x_{n1} \dots x_{np})^T$  e n respostas  $y_1, \dots, y_n$  que pertençam ao conjunto (1,-1), sendo que 1 representa pertencer a uma classe e -1 à outra.

Caso as observações de uma classe sejam perfeitamente separáveis por um plano, qualquer plano que tenha a propriedade de:

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} > 0 \text{ se } y_i = 1 \quad (64)$$

e

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} < 0 \text{ se } y_i = -1 \quad (65)$$

para todo  $i$  é um hiperplano de separação viável. De forma equivalente, as duas equações anteriores podem ser descritas unicamente como:

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) > 0 \quad (66)$$

Cria-se então uma fronteira  $f(x) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$ . Para uma observação de teste  $x^*$ , se  $f(x^*) > 0$ , a observação é dita como da classe 1. Caso contrário, é dita como da classe -1. A magnitude de  $f(x^*)$  é uma medida do quão afastado está a observação do plano de separação. Quanto maior a distância, mais segurança se tem na atribuição de classe indicada pelo modelo.

No caso de ser possível separar perfeitamente duas classes com um hiperplano, existem infinitas possibilidades de escolha, visto que este pode ser transladado e rotacionado. Uma possível escolha é a que apresenta a maior margem. A margem neste contexto é definida como a menor distância perpendicular entre todos os pontos de treinamento até o hiperplano. Busca-se, portanto, maximizar essa margem. Este conceito é ilustrado na Figura 26.

Os pontos cujas distâncias até o hiperplano de separação são iguais à margem são denominados vetores de suporte. Qualquer alteração na posição desses vetores alteraria o hiperplano de separação. Por outro lado, qualquer alteração em outro vetor que não seja o de suporte e que não atinja ou toque na fronteira determinada pela margem não altera o hiperplano de separação.

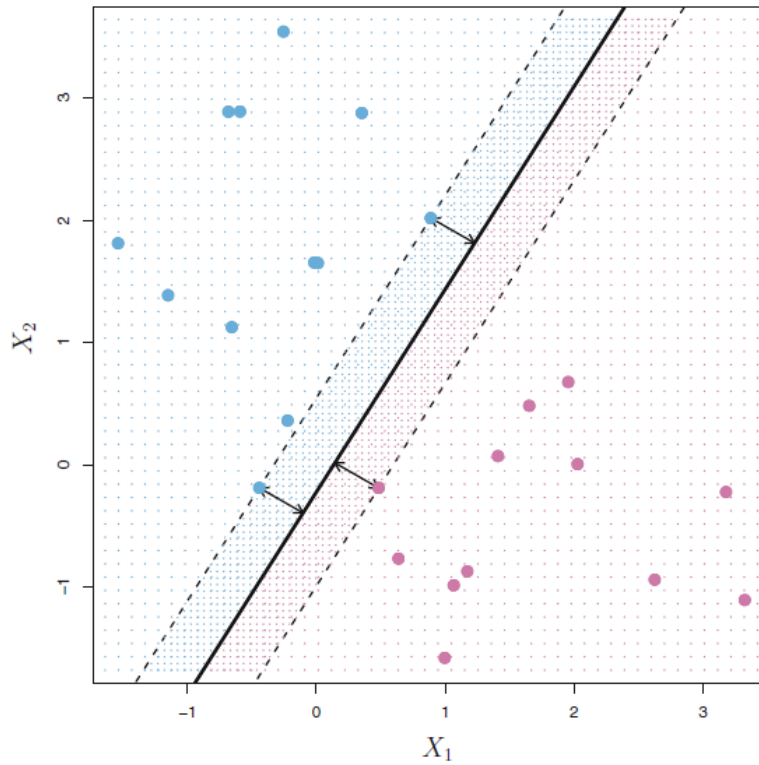


Figura 26 -O hiperplano de máxima margem é mostrado em linha contínua. A margem é a distância entre as linhas pontilhadas e a linha contínua, indicada pela seta. Os pontos que estão sendo cortados pelas linhas pontilhadas são os denominados de vetores de suporte. Adaptado com permissão de Springer Nature (JAMES et al., 2013c).

Formalmente, o plano de separação por margem máxima é construído como o seguinte problema de otimização (JAMES et al., 2013c):

Maximizar  $M$  com os parâmetros  $\beta_1, \beta_2$  e  $\beta_p$

$$\text{Sujeito a } \sum_{j=1}^p \beta_j^2 = 1 \quad (67)$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M \quad \forall i = 1, \dots, n \quad (68)$$

A equação 68 força que todos os pontos estejam no hiperplano correto, caso  $M$  seja maior que zero. Quanto maior ficar  $M$ , maior a margem e conseqüentemente a segurança de que o hiperplano calculador é um bom separador de classes. A restrição 67 apenas permite que o termo à esquerda da equação 68 signifique diretamente a distância perpendicular do ponto  $i$  até o hiperplano (JAMES et al., 2013c).

### 2.10.5.2 Classificador de vetores de suporte

Quando as duas classes de um problema não são separáveis por um hiperplano, o classificador de máxima margem não pode ser construído. Porém um conceito de margem flexível que não separa completamente as classes pode ser estendido a partir do conceito anterior. Além disso, observações que possam ser adicionadas a um conjunto de treinamento que estejam dentro da margem do classificador de máxima margem podem ter um efeito bem forte de alteração do hiperplano de separação, levando assim a uma variabilidade não desejada que pode levar a um excessivo ajuste aos dados de treinamento. No classificador de suporte a vetores, é permitido que observações estejam do lado incorreto da margem ou do hiperplano de separação.

Formalmente, esse classificador é obtido como segue (JAMES *et al.*, 2013c):

$$\begin{aligned} & \text{maximizar } M \text{ com os parâmetros } \beta_0, \beta_1, \dots, \beta_n \text{ e } \epsilon_0, \epsilon_1, \dots, \epsilon_n \\ & \text{sujeito a } \sum_{j=1}^p \beta_j^2 = 1 \end{aligned} \quad (69)$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i) \quad (70)$$

$$\epsilon_i \geq 0, \sum_{i=1}^n \epsilon_i \leq C \quad (71)$$

onde  $C$  é um parâmetro de ajuste não negativo e  $M$  continua sendo uma medida do tamanho da margem. As variáveis de relaxamento  $\epsilon_0, \epsilon_1, \dots, \epsilon_n$  permitem que a observação fique fora do limite da margem. Quando  $\epsilon_n$  é igual a 1, a observação está exatamente no hiperplano de separação. Quando  $\epsilon_n$  é maior do que 1, a observação se localiza no lado incorreto do hiperplano. O parâmetro  $C$ , portanto, indica quanto se permite que haja relaxamento das restrições do classificador de máxima margem. Se  $C$  for igual a zero, este classificador fica igual ao anterior.

Quando  $C$  é pequeno, a margem se torna menor e o erro de classificação no conjunto de treinamento também fica menor, porém há maior variância (a superfície de separação pode estar muito ajustada ao conjunto de treinamento, sofrendo fortes variações com novos conjuntos de treinamento). Quando  $C$  é grande, a margem se torna maior, o modelo se ajusta bem menos ao conjunto de treinamento, podendo em exagero perder muitos dos padrões apresentados neste conjunto. Por outro lado, ocorre menor variância,

ou seja, diferentes amostras de conjuntos de treinamentos geram hiperplanos de separação parecidos. O efeito do parâmetro  $C$  na margem e no hiperplano de separação é ilustrado na Figura 27.

As observações que estão no lado correto da margem, da mesma forma que no classificador anterior, não afetam o hiperplano gerado. Dessa forma, este classificador é bastante insensível a outliers. Por outro lado, todas as observações que estão na margem ou no lado incorreto da margem são chamadas de vetores de suporte e influenciam a localização do hiperplano de separação.

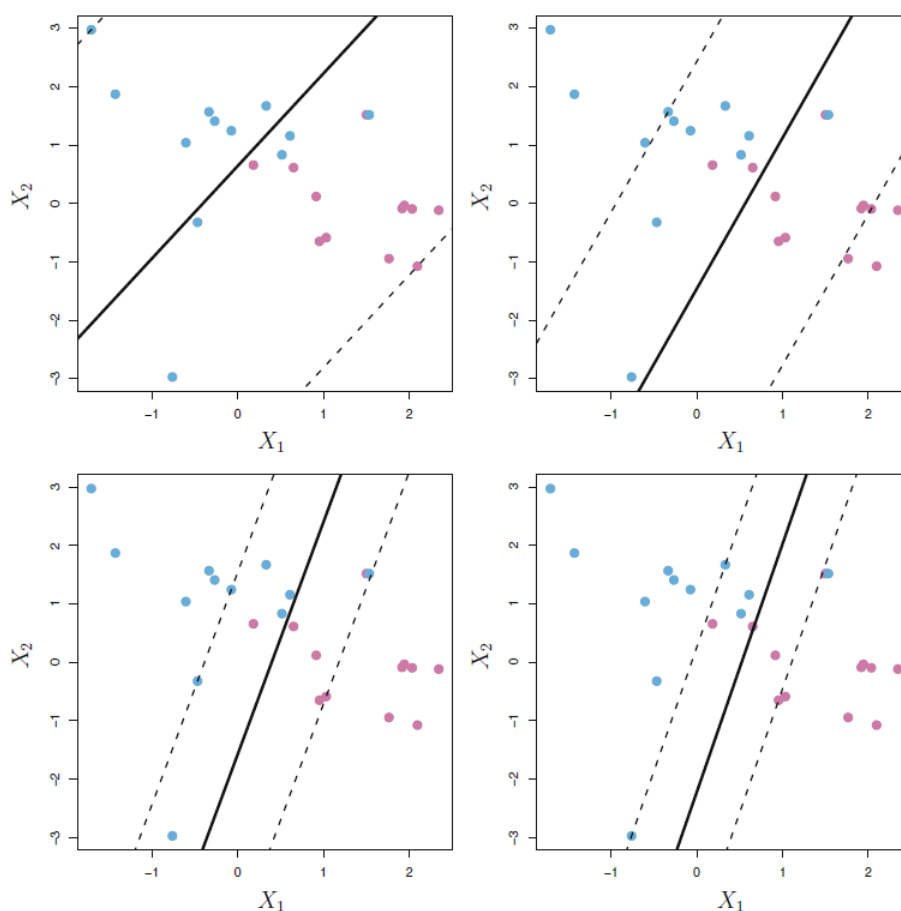


Figura 27 – Efeito do parâmetro  $C$  na máquina de vetores de suporte.  $C$  decresce na seguinte ordem (superior esquerdo, superior direito, inferior esquerdo e inferior direito). Um alto valor de  $C$  implica em muita tolerância para as observações estarem longe da margem, o que permite que esta esteja grande. O efeito inverso é observado ao decrescer  $C$ . Adaptado com permissão de Springer Nature (JAMES et al., 2013c).

### 2.10.5.3 Máquina de vetores de suporte

O classificador de suporte a vetores só pode gerar no espaço utilizado fronteiras lineares, no entanto alguns problemas claramente possuem fronteiras não-lineares, como o exibido na Figura 28. Uma possível solução, utilizada por exemplo em regressão linear, é a adição de termos não lineares no espaço de características, como funções polinomiais.

Dito de outro modo, em vez de usar  $p$  preditores  $X_1, X_2, \dots, X_p$ , utiliza-se por exemplo  $2p$  preditores  $X_1, X_1^2, X_2, X_2^2, \dots, X_p, X_p^2$ . No espaço de características aumentado, a fronteira continua linear, mas o mapeamento desse espaço ao de características original leva a uma fronteira não-linear. No entanto, essa abordagem pode levar ao uso de muitas características, o que pode tornar o problema computacionalmente difícil de ser tratado (JAMES *et al.*, 2013c).

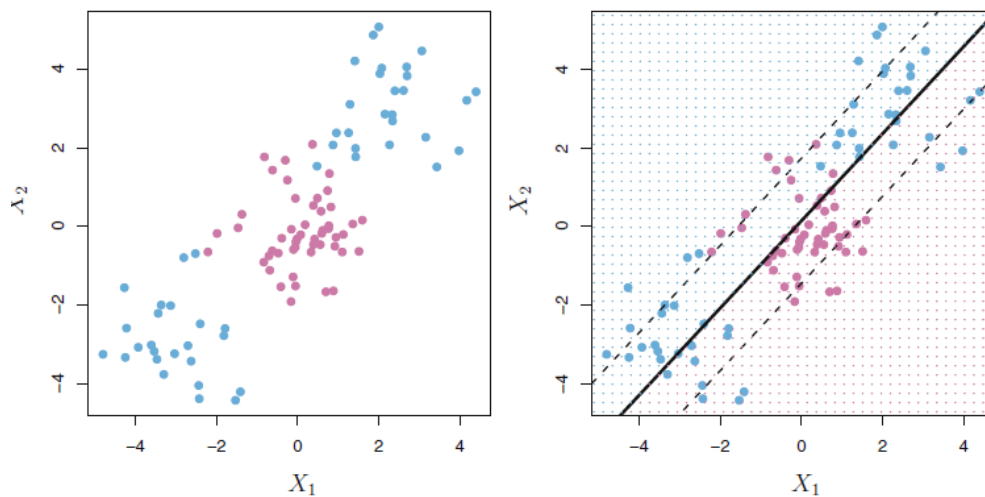


Figura 28 – Esquerda: Classes apresentam fronteiras não lineares. Direita: Fronteira linear não consegue separar classes. Adaptado com permissão de Springer Nature (JAMES *et al.*, 2013c).

A máquina de vetores de suporte se utiliza de um artifício para fazer computações eficientes: o produto interno:

$$\langle x_i, x_{i'} \rangle = \sum_{j=1}^p x_{ij} x_{i'j} \quad (72)$$

É possível mostrar que o hiperplano de separação de um classificador de suporte a vetores com  $n$  observações também pode ser escrito como: (JAMES *et al.*, 2013c)

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i \langle x, x_i \rangle \quad (73)$$

Notar que é necessário calcular apenas os  $n(n - 1)/2$  pares de produtos internos nessa configuração para calcular esta função para um dado ponto de observação  $x_{i'}$ .

Como, no entanto, apenas os vetores de suporte têm  $\alpha_i$  diferente de zero, só eles podem afetar o hiperplano de separação. Considerando então  $\mathcal{S}$  como o conjunto que possui os vetores de suporte,  $f(x)$  pode ser escrito como:

$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle \quad (74)$$

A máquina de vetores de suporte estende o classificador de suporte a vetores ao generalizar o conceito de produto interno com uma função  $K$ , referida como *kernel*, que quantifica a similaridade de duas observações. O *kernel* pode inclusive ser igual ao produto interno, neste caso a máquina de vetores de suporte gera uma superfície linear da mesma maneira como foi visto para o classificador de suporte a vetores. Um exemplo de *kernel* polinomial de ordem  $d$  seria (JAMES *et al.*, 2013c):

$$K(x_i, x_{i'}) = \left(1 + \sum_{j=1}^p x_{ij} x_{i'j}\right)^d \quad (75)$$

Outro exemplo de *kernel* é o radial, que é representado por:

$$K(x_i, x_{i'}) = e^{\left(-\gamma \sum_{j=1}^p (x_{ij} x_{i'j})^2\right)} \quad (76)$$

onde  $\gamma$  é uma constante positiva.

O *kernel* radial decai rapidamente conforme a observação de treinamento se afasta da de teste, implicando que apenas observações de treinamento próximas a de teste terão efeito relevante na composição da fronteira de separação, levando assim a um comportamento local. Dois exemplos dos *kernels* anteriormente citados podem ser vistos na Figura 29.

O *kernel*, portanto, é capaz de representar o espaço de características aumentado de uma forma computacionalmente eficiente. Além disso, é capaz de representar espaços que não poderiam ser representados por um espaço de características finitos, como é o caso do *kernel* radial (JAMES *et al.*, 2013c).

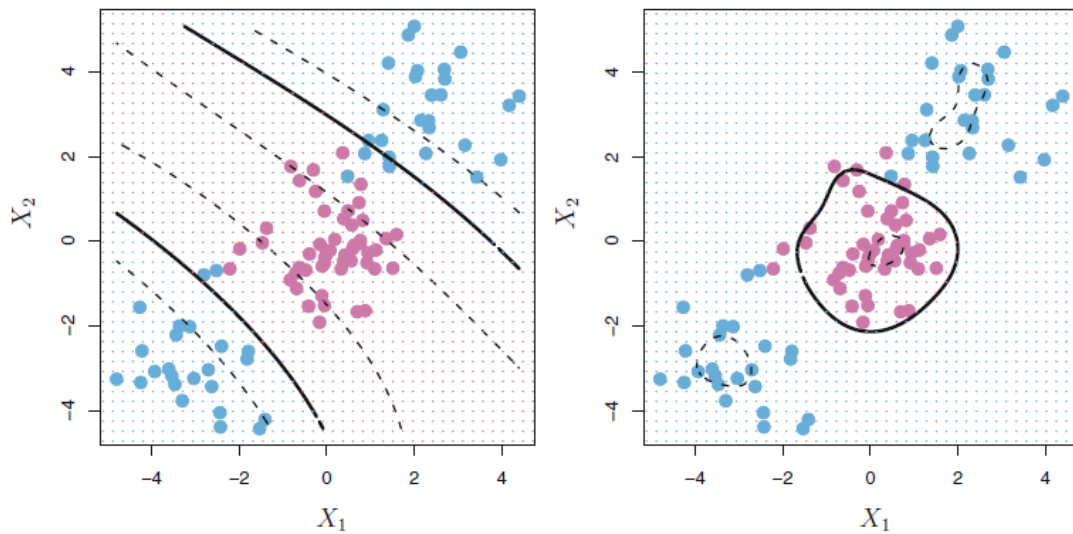


Figura 29 – Esquerda: Aplicação de um MVS com kernel polinomial de ordem 3. Direita: Aplicação de kernel radial. Ambos conseguem separar bem as classes. Adaptado com permissão de Springer Nature (JAMES et al., 2013c).

#### 2.10.5.4 Extensão de MVS para mais de duas classes

Para estender a técnica de MVS para mais de duas classes, utilizam-se comparações entre vários pares de classes. As duas estratégias mais comuns são um-contra-um e um-contra-todos. Na estratégia um-contra-um para  $K$  classes, os  $(k(k-1))/2$  pares de classes são comparados para cada observação de teste, e a predição é dada para a classe que aparece mais vezes como “vencedora” entre os pares (sistema de votação). Na estratégia um-contra-todos, cada classe é comparada contra todas as outras ao mesmo tempo. A observação de teste é designada à classe que no seu respectivo MVS apresentar a maior distância até o hiperplano de separação, pois quanto maior essa distância, maior a segurança em se afirmar que uma dada observação pertença àquela classe.

#### 2.10.5.5 Relação do método com parâmetros do MATLAB® 2017a

O parâmetro “*box constraint*” (restrição de caixa) utilizado pelo programa desenvolvido neste trabalho cumpre o mesmo papel da constante  $C$  descrita na seção 2.10.5.3. O parâmetro função do *kernel* possibilita usar funções de *kernel* pré-definidas pelo programa ou criar uma nova customizada. A matriz de dados de entrada pode ser



dividida pelo parâmetro escalar “escala do *kernel*” antes de ser executado o procedimento de produto interno, de modo a gerar uma pré-normalização dos dados.

## 2.11 Validação cruzada

Quando se tem disponível um banco de dados grande o suficiente, é possível dividir uma parte do banco para treinamento de um modelo e outra para teste, na técnica conhecida como *hold-out* ou reserva. Na maioria dos casos, no entanto, o número de amostras é restrito em relação ao ideal. Em função disso, a estimativa de erro de teste obtida pode ser muito diferente do obtida com a aplicação do método na população em geral.

Com uma amostragem da população mais reduzida, a estimativa de erro produzida a partir do grupo de teste pode sofrer alta variabilidade, pois apresenta dependência de quais amostras são separadas para teste. Para um dado conjunto de amostras, escolher diferentes subconjuntos para teste e treinamento comumente leva a estimativas distintas (JAMES *et al.*, 2013d).

Para melhorar a estimativa de erro, uma abordagem é a divisão da amostra de forma aleatória em  $k$  grupos. Escolhe-se um grupo para teste e treina-se um modelo com os outros  $k-1$  grupos. Em seguida outro grupo não escolhido anteriormente é utilizado para teste e treina-se outro modelo com os restantes, sendo o processo repetido até que todos os grupos tenham participado como teste.

No caso de problemas de classificação, considere-se  $\text{Err}_j$  como  $I(y_j \neq \hat{y}_j)$ , ou seja, igual a 1 quando a estimativa do modelo para o grupo  $i$  for distinta da verdadeira, e 0 caso contrário. Considere-se também  $n_i$  como o número de observações de teste no grupo  $i$ . Nessas condições, o erro de validação cruzada para  $n$  observações  $CV_{(k)}$  é dado por:

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^k \left( \frac{1}{n_i} \sum_{j=1}^{n_i} \text{Err}_j \right) \quad (77)$$

A escolha do número  $k$  é importante. Se  $k$  for muito alto, o grupo de teste será tão pequeno que, apesar de observações de teste diferentes serem escolhidas a cada passo, os grupos de treinamento estarão altamente correlacionados, pois terão muitas amostras em

comum. Nesse caso, a estimativa de erro de teste sofrerá maior variabilidade ao se comparar com estimativas obtidas pela repetição de todo o processo de validação cruzada (ou seja, novas estimativas a partir da geração aleatória de outros conjuntos de k-grupos). Por isso o valor de k é normalmente um valor intermediário, tipicamente igual a 5 ou 10 (JAMES *et al.*, 2013d).

## 2.12 Códigos de saída para correção de erros (CSCE)

O CSCE trata um problema multiclasse como uma série de problemas de classificação binária. Um *design* de códigos indica de que maneira se deve combinar os pares ou grupos para prever as observações. Para muitas classes e para um problema com muitos dados, algumas combinações podem ser computacionalmente intratáveis. Um *design* simples e efetivo, especialmente para um problema com poucas classes, é o um-contra-um. Um exemplo será utilizado para ilustrar como funciona.

Suponha que se tenham três classes, que se use um índice g para erro e que os classificadores binários sejam SVMs. Nesse caso, tem-se a seguinte codificação: (“Fit multiclass models for support vector machines or other classifiers - MATLAB fitcecoc”, )

Tabela 3 – Exemplo um-contra-um de CSCE

	Classificador 1	Classificador 2	Classificador 3
Classe 1	1	1	0
Classe 2	-1	0	1
Classe 3	0	-1	-1

Ou seja, cada classificador compara apenas duas classes dentre todas, ignorando as restantes. Seja um problema de k classes e w classificadores. Considere  $m_{kw}$  como os elementos da matriz de CSCE conforme a Tabela 3 e  $s_w$  o score que o classificador w dá para uma observação, sendo a classe positiva aquela marcada como 1.

Nessas condições busca-se a classe  $\hat{k}$  que minimiza a agregação de perdas para os  $L$  classificadores binários, conforme a fórmula a seguir (“Fit multiclass models for support vector machines or other classifiers - MATLAB fitcecoc”, ):

$$\hat{k} = \underset{k}{\operatorname{argmin}} \frac{\sum_{w=1}^L |m_{kw}| g(m_{kw}, s_w)}{\sum_{w=1}^L |m_{kw}|} \quad (78)$$

onde  $g$  é uma função de perda, que para a solução de problemas de classificação pode ser usada a de Hamming:

$$g(y_j, s_j) = \frac{[1 - \operatorname{sign}(y_j s_j)]}{2} \quad (79)$$

onde  $y$  é o rótulo atribuído a uma observação  $j$  que pertence a uma determinada classe conforme ilustrado na Tabela 3 e  $s$  é a pontuação que o classificador binário atribui a essa observação  $j$ . A função sinal é 0 quando seu argumento é 0, 1 quando seu argumento é maior do que zero e -1 quando seu argumento é menor do que 0.

Essa abordagem pode render resultados melhores do que abordar esse problema diretamente como multiclasse (FURNKRANZ, 2002).

### 3 Revisão da Literatura

Duas principais estratégias de detecção de fibrilação atrial têm sido usadas: análise da atividade atrial e análise da atividade ventricular no eletrocardiograma (ECG). Um exemplo de análise do primeiro tipo consiste em detectar sequências de ondas f no intervalo TQ, que são geradas pela FA. Essa técnica funciona satisfatoriamente quando o sinal possui uma boa relação sinal-ruído (DU *et al.*, 2014).

Outra abordagem consiste em fixar, por inspeção visual, 10 min de sinal com ritmo sinusal extraído de cada sinal de longa duração, com ocorrências de FA. Observa-se empiricamente quantos pontos amostrais, em média, a onda P antecede a onda R. Fixa-se também um intervalo de duração da onda P e divide-se o mesmo em seis segmentos equidistantes (determinação empírica). De cada segmento se retira a média. Deste conjunto de seis valores, calcula-se variância, curtose e assimetria. Para cada batimento, portanto, é formado um conjunto de nove características da onda P. Esse conjunto de valores em um espaço de dimensão nove é utilizado para treinar um modelo Gaussiano multivariável de misturas. Através de um processo iterativo, é determinada a quantidade de distribuições Gaussianas do modelo, o peso relativo de cada distribuição, o vetor médio e a matriz de covariância que maximiza a verossimilhança. Após o treinamento do modelo, para cada batimento de teste é determinada uma pontuação que depende de uma distância de Mahalanobis (dependente dos vetores médios e de covariância) que é escalada por uma função de *kernel* radial. Caso esta pontuação esteja abaixo de um limiar determinado por treinamento experimental, o batimento é considerado normal. Caso esteja acima, é entendido como batimento FA. Um algoritmo baseado em votação por maioria determina se o sinal como um todo é FA ou não, a partir dos resultados individuais dos batimentos do sinal (LADAVICH *et al.*, 2015).

Uma alternativa é utilizar transformadas tempo-frequência para analisar a atividade atrial. Um estudo separou intervalos TQ de duração variável, fixando o fim a 60 ms antes da onda R subsequente e a duração a um quarto da mediana do intervalo RR associado aos últimos 10 batimentos. Sobre cada segmento, foi calculado o *wavelet entropy*, uma medida de dispersão ou “caos” que, acima de um determinado limiar, indicava a presença de ruído. Dos trechos não ruidosos, uma média coerente de um

número L de batimentos anteriores foi feita. A demarcação de cada segmento TQ utilizava o mesmo critério de fim anteriormente descrito, mas a duração foi de um quarto da mediana do intervalo RR dos últimos L batimentos. Sobre a média coerente calculada, foi efetuada a transformada *wavelet*, de modo a calcular as energias contidas em cada escala e a *wavelet entropy*. Este último marcador provou ser o melhor discriminante, sendo a entropia presente nos casos com fibrilação atrial maior do que nos casos com ritmo sinusal normal. Como na fibrilação atrial ocorre propagação desordenada de despolarização no átrio durante todo o ciclo cardíaco, esse resultado era o esperado. Em virtude da média coerente, esta técnica apresenta um atraso até o início da detecção de um evento de fibrilação atrial; no entanto, consegue detectar sinais com fibrilação atrial mesmo com pouca alteração na frequência cardíaca (RÓDENAS *et al.*, 2015).

Sobre a análise de intervalos RR, outro grupo de pesquisadores detectou fibrilação atrial baseado em um histograma, que foi construído a partir da diferença entre um histograma de variações do intervalo RR antes do batimento e em um histograma das diferenças entre intervalos RR consecutivos ( $\delta RR$ ) após o batimento. Um índice construído a partir desse histograma foi usado para detectar possíveis casos de FA. Uma regra com limiar adaptativo minimizava falsos positivos. Para os casos que passavam para a próxima etapa, uma série de regras heurísticas foi utilizada para determinar se realmente era o caso de um FA ou se o distúrbio foi causado por outro tipo de arritmia. Dentre essas regras, havia análise de histograma dos batimentos após o evento de FA, análise de desvio padrão e comparação de padrões de batimentos com padrões conhecidos como bigeminismo e trigeminismo. Ao final, foi realizada uma comparação de teste estatístico entre a distribuição cumulativa das razões dos intervalos RRs sucessivos ( $RR[i]/RR[i-1]$ ) com uma distribuição esperada para um caso de FA. A estratégia teve bom desempenho geral, mas falhou em diferenciar alguns tipos de arritmias cardíacas de FA (CHAO HUANG *et al.*, 2011).

Sarkar e seus colaboradores utilizaram histogramas 2-D da diferença entre RRs consecutivos  $\delta RR(i) \times \delta RR(i-1)$  para detectar FA no ECG e diferenciá-lo de taquicardia. Cada *bin* do referido histograma trazia informação sobre aceleração e desaceleração referente a três batimentos consecutivos. Cada tipo de ritmo cardíaco possui um padrão diferente de dispersão neste histograma, e diversas índices envolvendo a quantidade de

pontos e/ou quadrados ocupados por região foram calculados, de modo a detectar fibrilação atrial e taquicardia (SARKAR *et al.*, 2008).

Park e seus colaboradores propuseram o uso de um gráfico de Poincaré que representa em cada ponto dois intervalos RR consecutivos. Um ritmo irregular específico como uma contração ventricular prematura apresenta padrões que determinam a formação de diferentes *clusters* de pontos. Ritmos normais tendem a ter um cluster ou poucos *clusters*, pois os batimentos ficam concentrados. No caso de FA, muitos *clusters* podem ser encontrados ou eventualmente apenas um grande cluster. Para fazer a diferenciação entre FA e outros ritmos, medidas de dispersão foram criadas e utilizadas com base em inferências obtidas a partir da visualização dos gráficos de Poincaré (PARK *et al.*, 2009).

Para o diagnóstico de doenças cardíacas, é comum se analisar o ECG. Tradicionalmente, os estudantes e médicos generalistas são treinados em cursos, aprendem com a vivência prática da profissão e racionalmente passam a identificar padrões morfológicos. No entanto, aprender esta habilidade de identificação demanda esforço e prática. Em muitas unidades de saúde, não existe disponibilidade de cardiologistas durante todo o tempo para atender às emergências médicas. Além disso, para a análise de um banco de dados de muitos pacientes com muitas horas de gravação, torna-se trabalhoso e dispendioso que cardiologistas vasculhem as formas de onda em busca de anomalias.

Com o objetivo de automatizar esse trabalho, há décadas pesquisadores buscam formas de realizar detecção e classificação automática de batimentos (LUZ *et al.*, 2016). Para permitir a avaliação de algoritmos e a comparação do desempenho de diferentes sistemas automáticos, bancos de dados anotados por especialistas têm sido divulgados para fomentar a cooperação científica. O banco de dados mais utilizado como referência para avaliação do desempenho dos detectores e classificadores de batimentos cardíacos é o *MIT-BIH Arrhythmia Database* (MOODY *et al.*, 2001). Esse banco de dados constitui um padrão de referência confiável, tendo sofrido um número muito pequeno de revisões ao longo de quatro décadas de existência; está disponível publicamente e vem sendo extensivamente utilizado.

Um banco de dados para identificar FA foi proposto pelo desafio da conferência *Computers in Cardiology 2017* e do site *Physionet*, descrito com detalhes na seção 4.1.

Este banco possui a característica de ter muitos registros eletrocardiográficos (aproximadamente 11 mil para treinamento + teste) de curta duração (10-60 s), coletados a partir de um dispositivo portátil que o paciente leva consigo em sua rotina diária, cujos dados são transmitidos para o celular e deste para um banco de dados na “nuvem”, que é acessível pelo médico acompanhante.

Em função das limitações da instrumentação, os sinais coletados apresentam maior ruído que o usual e representam um desafio em relação aos utilizados por trabalhos anteriores, que eram mais longos e menos ruidosos. Por outro lado, como dito anteriormente, a fibrilação atrial pode se apresentar sem sintomas perceptíveis, por isso a aproximação do exame de ECG ao paciente é positiva, pois facilita o diagnóstico precoce.

Este cenário com muitos dados e com uma qualidade de dados inferior aponta que uma abordagem efetiva seria a utilização de muitas características do sinal associada a técnicas de redução por otimização da sua quantidade, de modo que o banco de dados pudesse ser utilizado para determinar quais são as características relevantes para separar os diferentes ritmos cardíacos.

Como apontado, muitos pesquisadores propuseram diferentes características para diferenciar FA de outras arritmias. A proposta deste trabalho, portanto, é reunir algumas destas características que foram apresentadas, propor novas e utilizar o método de seleção por componentes vizinhos (seção 2.9) para definir um conjunto ótimo, que é então utilizado em diversas técnicas de classificação, cujos resultados são comparados por validação cruzada.

## 4 Materiais e Métodos

### 4.1 Banco de dados

O banco de dados utilizado foi fornecido pelo desafio da conferência *Computers in Cardiology 2017*, que pode ser baixado gratuitamente em <https://physionet.org/challenge/2017/>. Segundo o artigo publicado pelos organizadores do desafio (CLIFFORD *et al.*, 2017), um total de 12.186 gravações foram doadas pela empresa AliveCor. Cada gravação foi realizada por sujeitos que tinham os dispositivos da AliveCor. Estes consistem de um amplificador de ECG, ilustrado na Figura 30, em que os sujeitos colocam os seus polegares em um par de eletrodos. O sinal padrão gerado é equivalente ao da derivação I do ECG. Era possível que o sujeito usasse o dispositivo com orientação invertida, pois isso não era validado pelo dispositivo. Nesse caso, o sinal obtido era de polaridade invertida (CLIFFORD *et al.*, 2017).



*Figura 30 – Foto do dispositivo de coleta de eletrocardiograma da AliveCor®. Imagem gentilmente cedida por Paulo Roberto Benchimol Barbosa.*

As gravações tinham média de 30 s. Os dados eram transmitidos dos eletrodos para o microfone do celular por meio de uma frequência portadora de 19 KHz e um índice de modulação de 200 Hz/mV. Os dados foram digitalizados com uma frequência de



44,1 KHz e resolução de 24 bits, com posterior demodulação digital. O resultado desta operação foi armazenado com uma amostragem de 300 Hz, com 16 bits de resolução, com uma banda de frequência de 0,5-40 Hz e com faixa dinâmica de  $\pm 5$  mV (CLIFFORD *et al.*, 2017).

Os dados foram disponibilizados em formato MATLAB® v4 junto com arquivo de cabeçalho descrevendo o formato de dados. O conjunto de treinamento ficou disponível publicamente, consistindo de 8.528 gravações com durações que variavam entre 9 e 61 s. O conjunto de teste continha 3.658 gravações, mas não foi disponibilizado publicamente. Como o autor não conseguiu se inscrever e participar do desafio, não teve acesso à sua pontuação baseada no grupo de teste, tendo apenas acesso ao grupo de treinamento (CLIFFORD *et al.*, 2017).

## 4.2 Anotação de dados

As marcações iniciais dos ritmos foram cedidas pela AliveCor. Quatro classes de ritmos foram utilizadas: ritmo normal, ritmo de fibrilação atrial, outros ritmos e sinais ruidosos. Na versão de anotação final, o banco de treinamento consistiu de 5.076 gravações normais, 758 gravações com fibrilação atrial, 2.415 gravações com outros ritmos e 279 gravações ruidosas (CLIFFORD *et al.*, 2017).

A partir do conjunto de anotação inicial (versão 1), alguns sinais foram visualmente reverificados de acordo com uma metodologia não especificada, de forma que alguns sinais foram reclassificados, em geral remarcando sinais de difícil detecção como ruidosos. Essa remarcação gerou uma nova anotação (versão 2) para as classificações (CLIFFORD *et al.*, 2017).

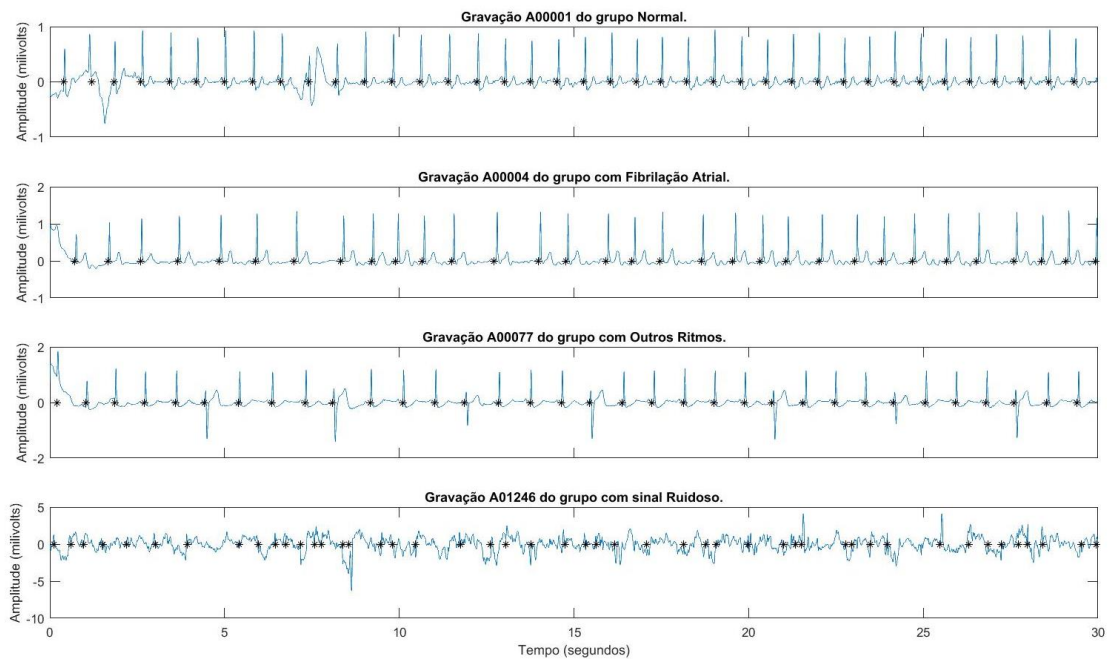
Os organizadores perceberam que deveriam fazer uma análise mais aprofundada das anotações, porém, por restrições de tempo e dinheiro, decidiram fazer uma análise direcionada pelos algoritmos com melhores performances até então. Os organizadores supuseram que algoritmos que estivessem com um desempenho abaixo de um índice de referência não trariam novas informações úteis e que utilizar um esquema de votação entre os melhores algoritmos melhoraria o padrão ouro utilizado. Supuseram que os sinais

que causaram pouca discordância de classificação eram sinais fáceis de classificar, enquanto os que causaram muita discordância deveriam ser reanalisados por especialistas (CLIFFORD *et al.*, 2017).

Os organizadores utilizaram um índice de confiabilidade de concordância entre avaliadores denominado Fleisskappa. Do conjunto de testes, os 1.129 sinais com maior diferença de concordância foram reavaliados por um conjunto de especialistas sem conhecimento da marcação dada anteriormente, sendo no mínimo três especialistas para cada gravação. Esse processo de correção ocorreu principalmente sobre o grupo de teste, que não foi disponibilizado publicamente como dito anteriormente. Algumas poucas marcações do grupo de treinamento foram também revisadas, mas apenas por um único especialista, gerando a anotação definitiva (versão 3) do banco de treinamento (CLIFFORD *et al.*, 2017).

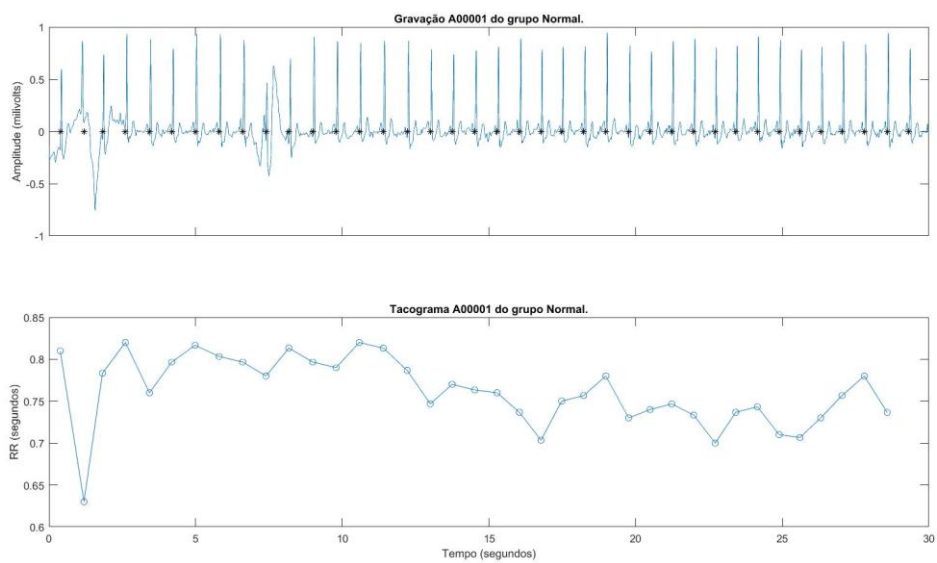
### **4.3 Ilustrações de sinais**

Exemplos de sinais com diferentes ritmos cardíacos são exibidos na Figura 31. Observa-se que o sinal de fibrilação atrial apresenta um ritmo desordenado, enquanto o sinal com outros ritmos tem um padrão consistente de alteração. Por outro lado, em caso de ruído, o detector apresenta dificuldade para detectar picos.

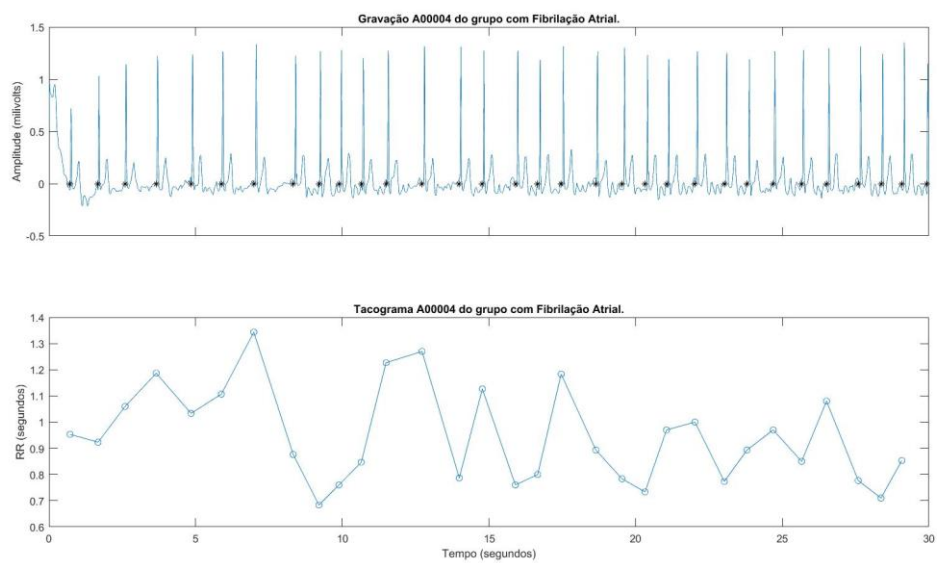


*Figura 31 – Gráfico comparativo das formas de onda dos diversos tipos de sinais. Os pontos pretos denotam detecção de onda R.*

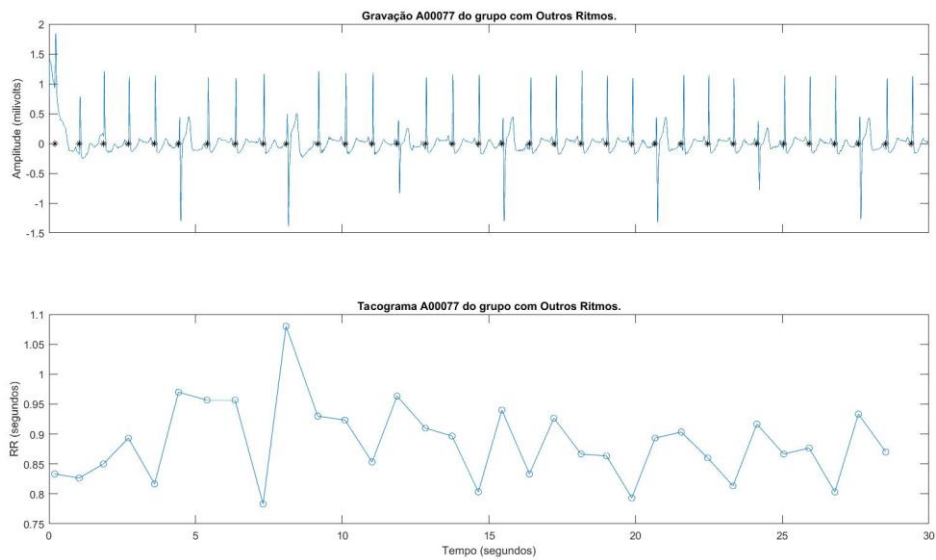
Comparações ilustrativas de cada sinal com o seu respectivo tacograma são realizadas na Figura 32, na Figura 33, na Figura 34 e na Figura 35. O tacograma é uma representação que dispõe no eixo da abscissa o tempo transcorrido até o início do batimento e no eixo da ordenada o intervalo de tempo entre batimentos. Essa visualização expõe a variabilidade dos intervalos RRs ao longo do tempo.



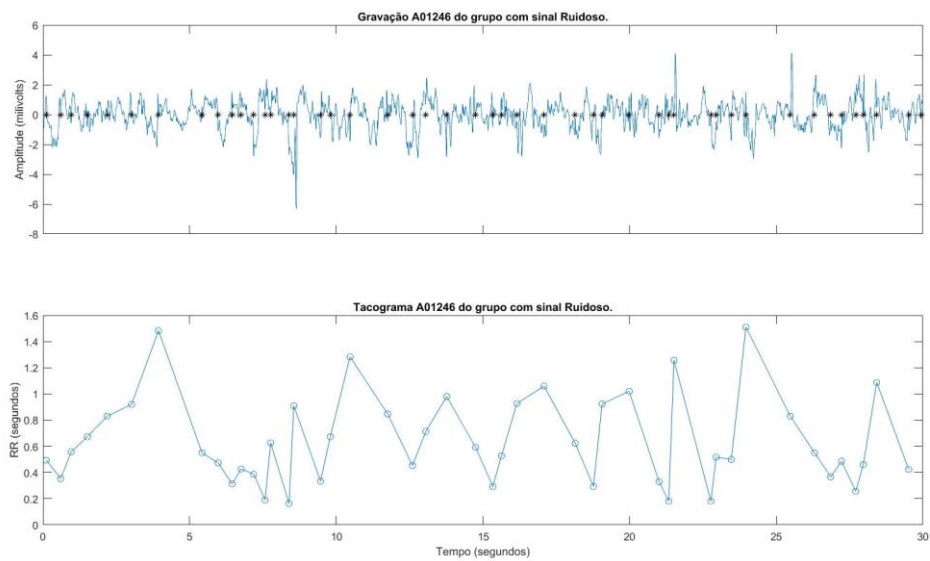
*Figura 32 – Tacograma de um sinal normal*



*Figura 33 -Tacograma de um sinal com fibrilação atrial*



*Figura 34 – Tacograma com sinal de outros ritmos*



*Figura 35 – Tacograma de sinal ruidoso*

## 4.4 Pré-processamento de sinal

Sobre o sinal bruto, foi aplicado um pré-processamento clássico com filtros digitais para remoção de interferência elétrica, linha de base e ruído de alta frequência. O

filtro escolhido foi do tipo *Butterworth* passa banda de segunda ordem com filtragem direta e reversa. Utilizando a função `filtfilt` do MATLAB®, o sinal é filtrado, a sequência do sinal filtrado é invertida e esta é submetida novamente ao filtro. Dessa maneira, a distorção de fase é zero e a ordem do filtro efetivo é dobrada em relação ao filtro original. As frequências de corte foram estabelecidas em 10 e 18 Hz.

## 4.5 Pontuação do desafio

Os organizadores do CinC inicialmente propuseram uma pontuação de avaliação que foi publicada no site do desafio:

		Classificação Preditada				
		Normal	FA	Outros	Ruidoso	Total
Classificação de Referência	Normal	$N_n$	$N_a$	$N_o$	$N_p$	$\sum N$
	FA	$A_n$	$A_a$	$A_o$	$A_p$	$\sum A$
	Outros	$O_n$	$A_o$	$O_o$	$O_p$	$\sum O$
	Ruidoso	$P_n$	$P_a$	$P_o$	$P_p$	$\sum P$
	Total	$\sum n$	$\sum a$	$\sum o$	$\sum p$	

Para cada tipo, tem-se uma medida  $F_1$ :

- Ritmo normal

$$F_{1n} = \frac{2Nn}{\Sigma N + \Sigma n} \quad (80)$$

- Ritmo de fibrilação atrial

$$F_{1a} = \frac{2Aa}{\Sigma A + \Sigma a} \quad (81)$$

- Outro tipo de Ritmo

$$F_{1o} = \frac{2Po}{\Sigma O + \Sigma o} \quad (82)$$

- Ruidoso

$$F_{1p} = \frac{2Pp}{\Sigma P + \Sigma p} \quad (83)$$

A pontuação  $F_1$  foi inicialmente igual a:

$$F_1 = \frac{F_{1n} + F_{1a} + F_{1o} + F_{1p}}{4} \quad (84)$$

A maioria dos competidores de desafio não conseguiu detectar corretamente os sinais ruidosos, o que levou a comissão organizadora a modificar a pontuação para:

$$F_1 = \frac{F_{1n} + F_{1a} + F_{1o}}{3} \quad (85)$$

Ou seja, os sinais ruidosos passaram a ser ignorados.

## 4.6 Extração de características

Foi implementado pelo autor desta pesquisa um detector de picos que segue as instruções dadas por (YEH *et al.*, 2008). Esta técnica é detalhada na seção 2.4. A partir da detecção dos instantes dos picos complexo QRS, foram calculados os intervalos entre esses picos (intervalo entre picos R, denominado de RR) e a diferença entre esses intervalos ( $\delta RR$ ), uma medida de aceleração/desaceleração cardíaca.  $\delta RR$  positivo implica em desaceleração cardíaca, enquanto  $\delta RR$  negativo implica aceleração cardíaca.

Trinta e cinco características foram propostas para avaliar a existência de fibrilação atrial. Quase todas envolvem análises derivadas do intervalo RR, portanto é esperado que essas características estejam altamente correlacionadas. As seguintes características foram utilizadas:

- 1) Média dos intervalos RR
- 2) Mediana dos intervalos RR
- 3) Desvio médio absoluto dos intervalos RR
- 4) Desvio mediano absoluto dos intervalos RR
- 5) Desvio padrão dos intervalos RR
- 6) Média geométrica dos intervalos RR
- 7) Média harmônica dos intervalos RR
- 8) Média “aparada” dos intervalos RR
- 9) Intervalo interquartil dos intervalos RR
- 10) Assimetria dos intervalos RR
- 11) Curtose dos intervalos RR
- 12) Média do  $\delta RR$
- 13) Mediana do  $\delta RR$
- 14) Desvio médio absoluto do  $\delta RR$
- 15) Desvio mediano absoluto do  $\delta RR$
- 16) Desvio padrão do  $\delta RR$
- 17) Média geométrica do  $\delta RR$
- 18) Média harmônica do  $\delta RR$
- 19) Média “aparada” do  $\delta RR$
- 20) Intervalo interquartil do  $\delta RR$
- 21) Assimetria do  $\delta RR$
- 22) Curtose do  $\delta RR$



- 23) Quantidade\_Origem (seção 2.7)
- 24) Irregularidade (seção 2.7)
- 25) PACO (seção 2.7)
- 26) Densidade (seção 2.7)
- 27) Anisotropia (seção 2.7)
- 28) Dispersão (seção 2.8)
- 29) RRs Curtos – frequência de intervalos RR menores do que 0,25 s

Além dessas características, outro conjunto de características discretas foram utilizadas, da seguinte forma:

- 30) Sinalizador de Bradicardia – 1 se for encontrado mais do que cinco intervalos RR superiores a 1,2 s (50 batimentos por minuto); 0 caso contrário
- 31) Sinalizador de Taquicardia – 1 se for encontrado mais do que cinco intervalos RR inferiores a 0,6 s (100 batimentos por minuto); 0 caso contrário
- 32) Sinalizador de Extra-sístole – Usando uma média móvel dos últimos cinco batimentos, 1 se observar uma queda de 15% no RR em relação a essa média seguida de um batimento com aumento de 15% em relação à mesma média. Ou seja, foi observado um batimento com intervalo menor do que a média seguido de uma pausa compensatória. Além disso, dois batimentos seguidos com menos do que 15% da média móvel dos intervalos RR (ou seja, sem pausa compensatória). 0 caso não entre nas condições anteriores.
- 33) Sinalizador de Prematuro – 1 se for encontrado um intervalo RR menor do que 0,5 s, seguido de outro na mesma condição ou então com um intervalo RR maior do que 1 s. Busca capturar o mesmo evento da característica anterior, porém com valores absolutos.
- 34) Sinalizador de Outro Ritmo - 1 se forem encontradas as condições descritas nas características 32 e 33, ou se a mediana dos intervalos RR é menor do que 0,6 (taquicardia), ou se a mediana dos intervalos RR é maior do que 1 (bradicardia). 0 caso contrário.

Além dessas características, outras duas características foram utilizadas com o objetivo de detectar sinais que fossem muito ruidosos. É importante destacar que o sinal, por ser transportado por sinais de ultrassom, possui características de ruído peculiar, apresentando alta potência de sinal tanto em altas frequências como em baixas frequências. Isso foi observado empiricamente, mas condiz com o fato de o ultrassom operar em altas frequências, gerando assim *aliasing* com frequências próximas do zero.

35) Frequência média baixa frequência – Frequência normalizada média do sinal bruto, não filtrado, dentro da faixa de 0 a 30 Hz. Foi utilizada a função `meanfreq` do Matlab 2017a ©.

36) Potência média alta frequência – Potência média do sinal do sinal bruto, não filtrado, em alta frequência para a faixa de frequência de 50 Hz a 145 Hz. Foi utilizado a função `bandpower` do Matlab 2017a ©.

## 4.7 Balanceamento do banco de dados por subamostragem

O banco de dados utilizado é consideravelmente desbalanceado. Como dito na seção 4.2, o banco de treinamento possuía 5.076 gravações normais, 758 gravações com fibrilação atrial, 2.415 gravações com outros ritmos e 279 gravações ruidosas.

Conforme especificado na seção 4.5, o sinal ruidoso passou a ser desconsiderado na análise. A solução utilizada para abordar o desbalanceamento foi fazer uma subamostragem (eliminação de gravações) das classes com maior quantidade de observações para que todas ficassem com a mesma quantidade de observações da classe com fibrilação atrial. Esta subamostragem foi feita de forma aleatória até a obtenção da quantidade desejada para cada classe.

## 4.8 Seleção de características

Conforme explicitado na seção 2.9, é necessário explicitar um fator de regularização  $\lambda$  para o SCCV. Este parâmetro penaliza o comprimento do vetor de características, de modo a direcionar algumas das características para zero ou aproximadamente zero. Este valor pode ser otimizado por validação cruzada.

O seguinte procedimento foi adotado para determinar o parâmetro  $\lambda$  do SCVV: (“Feature selection using neighborhood component analysis for classification - MATLAB fscnca”, )

- 1) Uma lista de 30 elementos que com intervalo constante iniciando com 0 e terminando com 30 foi posteriormente dividida pela quantidade de observações do banco de dados. Esta lista foi usada como possíveis valores para  $\lambda$ , conforme sugestão de boa prática dada recomendada pela documentação do MATLAB® 2017a.
- 2) A totalidade do banco de dados foi dividida em 10 grupos de treinamento/validação para validação cruzada, conforme descrito na seção 2.11.
- 3) Para cada  $\lambda$  e para cada grupo, foi executado todo o procedimento de SCVV e determinado qual seria o erro de classificação utilizando o classificador análogo ao 1-KNN conforme explicado na seção 2.9,
- 4) Para cada  $\lambda$ , foi calculada a média dos erros de classificação calculados no passo anterior.
- 5) Por fim, o  $\lambda$  com menor erro de classificação foi escolhido. Com o parâmetro de regularização  $\lambda$  escolhido, foi aplicado o método SCVV em todo o banco de dados e extraído o vetor de pesos para as características. Foi determinado que os parâmetros escolhidos seriam aqueles com pelo menos 25% do peso da característica do maior peso. Essa escolha de limiar fica ao critério do pesquisador, que pode preferir usar mais ou menos características.

## 4.9 Classificação

Utilizando as características selecionadas pelo procedimento descrito na seção 4.8, quatro classificadores diferentes foram treinados com validação cruzada (seção 2.11) no banco de dados, para posterior comparação de desempenho.

Os classificadores utilizados foram:

- 1) Árvore de decisão (seção 2.10.2)
- 2) K-vizinhos mais próximos (seção 2.10.3)
- 3) Análise discriminante (seção 2.10.4)
- 4) Máquina de vetores de suporte (MVS) (seção 2.10.5)

Um mesmo conjunto de 10 grupos de validação/treinamento foi utilizado para comparar os classificadores, pois grupos diferentes poderiam dificultar a comparação de desempenho dos grupos. Todas as variáveis sofreram padronização antes de entrarem no classificador, pois as características utilizadas possuem escalas bem distintas.

O problema multiclasse foi tratado como um conjunto de pares de classes, de acordo com a estratégia um-contra-um descrita na seção 2.12.

Para cada classificador, diferentes configurações de hiperparâmetros foram testadas e comparadas por um processo de otimização que objetivava minimizar o percentual de observações classificadas incorretamente. Uma busca em grade foi feita com cinco pontos em cada dimensão de hiperparâmetro, testando todas as possíveis combinações de forma aleatória até um limite de 100 avaliações por classificador.

A configuração de classificador escolhida foi aquela que apresentava o menor percentual de erro de classificação.

As configurações de hiperparâmetro que foram variadas pelo processo de otimização, todas relacionadas à configuração de algoritmo do MATLAB® 2017a, foram:

- 1) Máquina de vetores de suporte
  - a. Função *kernel* (Gaussiana e polinomial)

- b. Escala do *kernel*
  - c. *Box constraint*
- 2) Análise discriminante
  - a. Delta
  - b. Gamma
- 3) K vizinhos mais próximos
  - a. Expoente
  - b. Tipo de distância
- 4) Árvore de decisão
  - a. Tamanho mínimo da folha
  - b. Número máximo de divisões
  - c. Critério para divisão

## 4.10 Índice de avaliação de desempenho

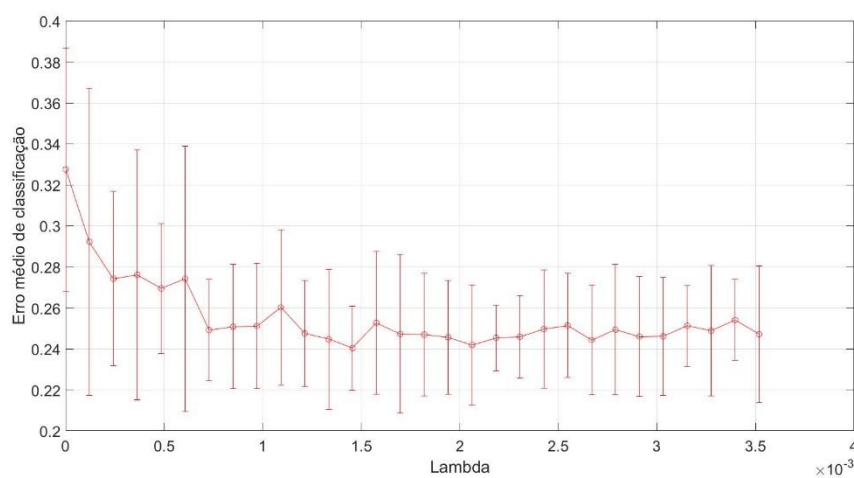
Após a escolha das melhores características, do método de classificação e dos hiperparâmetros, é necessário avaliar o desempenho do modelo. Uma maneira de se alcançar esse objetivo é realizando a validação cruzada descrita na seção 2.11. Divide-se o banco de treinamento em 10 grupos, treina-se com nove e faz-se a predição sobre o que foi separado. Repete-se então o processo separando outro grupo, até que todos os grupos tenham sido separados uma vez. Para utilizar o índice de comparação proposto sobre o desafio, foi realizada predição de classificação utilizando validação cruzada.

Com esta predição e a anotação da classe de referência, foi montada uma matriz, conforme especificado na seção 4.5 e os índices  $F_1$  foram então calculados.

# 5 Resultados

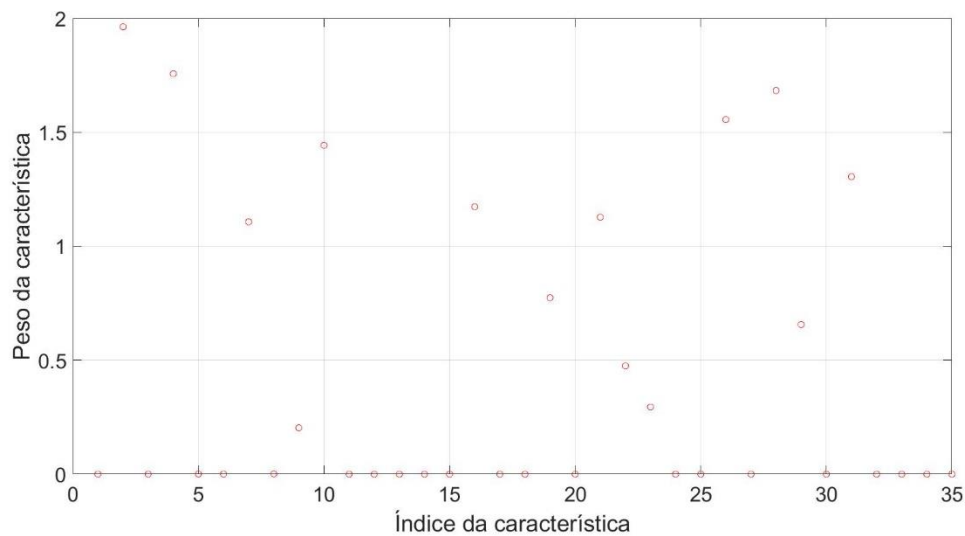
## 5.1 Seleção de Características

Utilizando-se do procedimento descrito na seção 4.8 sobre o banco de dados subamostrado conforme o procedimento descrito na seção 4.7, a otimização do  $\lambda$  pode ser representada graficamente pela Figura 36.



*Figura 36 – Variação do erro de classificação em função de  $\lambda$  utilizando validação cruzada. O valor de  $\lambda$  escolhido entre 0,001 e 0,003 é satisfatório. As barras de erro representam intervalo de confiança de 95%.*

Com o  $\lambda$  escolhido pela otimização igual 0,04, a seleção de características gerou seus pesos, exibidos na Figura 37.



*Figura 37 – Peso de cada variável para cada característica. É notório identificar que muitas características ficaram com peso igual ou próximo de zero.*

As características selecionadas pelo método, dentre as apresentadas anteriormente na seção 4.6, são:

- a) Mediana dos intervalos RR
- b) Desvio mediano absoluto dos intervalos RR
- c) Média harmônica dos intervalos RR
- d) Assimetria dos intervalos RR
- e) Desvio padrão do  $\delta RR$
- f) Intervalo interquartil do  $\delta RR$
- g) Curtose do  $\delta RR$
- h) Evidência de densidade
- i) Dispersão
- j) Frequência média
- k) Sinalizador de outro ritmo

## 5.2 Escolha de classificador e hiperparâmetro por Otimização

Para a escolha do classificador, foi utilizado o procedimento descrito na seção 4.9. O resultado completo da otimização de cada classificador e hiperparâmetro pode ser visto na Tabela 4, conforme exposto pelo log da função fitcecoc do MATLAB® 2017a. Cada hiperparâmetro é explicado na sua respectiva seção de revisão da literatura.

*Tabela 4 - Resultado completo da otimização de cada classificador testado com análise discriminante*

Iter	Eval	Objective	Objective	BestSoFar	Delta	Gamma	DiscrimType
	result		runtime	(observed)			
1	Best	0.35884	4.7289	0.35884	0.00017783	0.5	diagLinear
2	Accept	0.37467	0.84113	0.35884	0.00017783	1	pseudoLinear
3	Accept	0.35928	1.0094	0.35884	-	0	pseudoQuadra
4	Accept	0.66667	0.81344	0.35884	1000	1	diagLinear
5	Accept	0.35928	0.87116	0.35884	-	0	quadratic
6	Accept	0.35928	0.8302	0.35884	-	0	pseudoQuadra
7	Accept	0.35928	0.82919	0.35884	-	0	quadratic
8	Best	0.32938	0.95798	0.32938	1e-06	0	pseudoLinear
9	Accept	0.32938	0.81857	0.32938	0.00017783	0	linear
10	Accept	0.35928	0.83643	0.32938	-	0	quadratic
11	Accept	0.35928	0.84741	0.32938	-	0	quadratic
12	Accept	0.35884	0.78566	0.32938	1e-06	0.5	diagLinear
13	Accept	0.34609	0.83998	0.32938	1e-06	0.25	pseudoLinear
14	Accept	0.35928	0.84385	0.32938	-	0	pseudoQuadra
15	Accept	0.35928	0.8279	0.32938	-	0	pseudoQuadra
16	Accept	0.66667	0.84676	0.32938	5.6234	0	linear
17	Accept	0.35928	0.9976	0.32938	-	0	pseudoQuadra
18	Accept	0.37467	0.82049	0.32938	0.00017783	1	diagLinear
19	Accept	0.34609	0.777	0.32938	0.00017783	0.25	pseudoLinear
20	Accept	0.66667	0.80233	0.32938	5.6234	0.75	linear

Depois de algumas iterações, chega-se ao resultado da otimização (Tabela 5).



Tabela 5 - Resultado da otimização da análise discriminante

	90	Accept		0.38742		0.80071		0.3285		-		0		diagQuadrati	
	91	Accept		0.35928		0.83333		0.3285		-		0		pseudoQuadra	
	92	Accept		0.38742		0.84959		0.3285		-		0		diagQuadrati	
	93	Accept		0.38742		0.8732		0.3285		-		0		diagQuadrati	
	94	Accept		0.35928		0.83826		0.3285		-		0		quadratic	
	95	Accept		0.35928		0.80514		0.3285		-		0		quadratic	
	96	Accept		0.37467		0.75294		0.3285		0.031623		0		diagLinear	
	97	Accept		0.66667		0.75716		0.3285		5.6234		0.5		linear	
	98	Accept		0.35884		0.75287		0.3285		1e-06		0.5		pseudoLinear	
	99	Accept		0.66667		0.75441		0.3285		5.6234		1		linear	
	100	Accept		0.36324		0.97028		0.3285		0.031623		0.75		linear	

Optimization completed.  
 MaxObjectiveEvaluations of 100 reached.  
 Total function evaluations: 100  
 Total elapsed time: 100.8824 seconds.  
 Total objective function evaluation time: 86.6471

Best observed feasible point:

Delta	Gamma	DiscrimType
0.031623	0	linear

Observed objective function value = 0.3285  
 Function evaluation time = 0.7753

O resultado dessa otimização pode ser descrito da seguinte maneira: após 100 iterações, para o classificador do tipo análise discriminante, a menor função objetivo (que neste caso é o menor erro de classificação) foi encontrada com o valor de 32,85% para os hiperparâmetros  $\delta = 0,031623$ ,  $\gamma = 0$  e usando a análise do tipo linear. Além disso, o tempo para treinar este modelo foi de 0,77 s em um notebook com processador Intel® Core™ i7-3632QM CPU 2,20 GHz e 8 GB de memória RAM.

Para o classificador SVM, tem-se o resultado apresentado na Tabela 6.

Tabela 6 - Resultado utilizando máquina de vetores de suporte

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BoxConstraint t	KernelScale	KernelFunction
1	Best	0.28188	282.04	0.28188	1	-	polynomial
2	Accept	0.66799	2.7652	0.28188	1	0.031623	gaussian
3	Accept	0.66799	2.6317	0.28188	1000	0.031623	gaussian
4	Accept	0.30783	312.02	0.28188	31.623	-	polynomial
5	Accept	0.2832	28.552	0.28188	0.001	-	polynomial
6	Best	0.26517	106.28	0.26517	0.031623	-	polynomial
7	Accept	0.66535	1.7587	0.26517	1	1000	gaussian
8	Accept	0.27089	4.1828	0.26517	1000	31.623	gaussian
9	Accept	0.30783	312.22	0.26517	31.623	-	polynomial
10	Accept	0.33641	3.6051	0.26517	1000	1	gaussian
90	Accept	0.2832	28.519	0.26517	0.001	-	polynomial
91	Accept	0.32454	2.7532	0.26517	31.623	1	gaussian
92	Accept	0.26517	106.81	0.26517	0.031623	-	polynomial
93	Accept	0.30783	311.7	0.26517	31.623	-	polynomial
94	Accept	0.51143	1.6609	0.26517	0.031623	1	gaussian
95	Accept	0.50528	310.46	0.26517	1000	-	polynomial
96	Accept	0.66799	2.4493	0.26517	0.031623	0.031623	gaussian
97	Accept	0.50528	310.18	0.26517	1000	-	polynomial
98	Accept	0.2832	28.574	0.26517	0.001	-	polynomial
99	Accept	0.30783	312.55	0.26517	31.623	-	polynomial
100	Accept	0.2832	28.663	0.26517	0.001	-	polynomial

Optimization completed.  
 MaxObjectiveEvaluations of 100 reached.  
 Total function evaluations: 100  
 Total elapsed time: 10264.1754 seconds.  
 Total objective function evaluation time: 10251.5365

Best observed feasible point:

BoxConstraint	KernelScale	KernelFunction
0.031623	NaN	polynomial

Observed objective function value = 0.26517  
 Function evaluation time = 106.2764

O resultado dessa otimização pode ser descrito da seguinte maneira: após 100 iterações, para o classificador do tipo máquina de vetores de suporte, a menor função objetivo (que neste caso é o menor erro de classificação) foi encontrada com o valor de 26,5% para os hiperparâmetros  $box\ constraint = 0,031623$  e usando a análise do tipo polinomial (neste caso, a escala do *kernel* não se aplica). Além disso, o tempo para treinar este modelo foi de 106,274 s.

Para o classificador de k-vizinhos mais próximos, tem-se o resultado apresentado na Tabela 7.

Tabela 7 - Resultado utilizando k-vizinhos mais próximos

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	NumNeighbors	Distance	DistanceWeight	Exponent
1	Best	0.32498	2.9003	0.32498	196	spearman	squaredinver	-
2	Accept	0.59894	0.97491	0.32498	6	hamming	inverse	-
3	Best	0.31398	0.93601	0.31398	6	cosine	squaredinver	-
4	Accept	0.66667	1.8461	0.31398	1137	hamming	equal	-
5	Best	0.29991	1.5871	0.29991	34	mahalanobis	equal	-
6	Accept	0.42392	1.6876	0.29991	1137	chebychev	inverse	-
7	Best	0.28628	0.86503	0.28628	34	cityblock	squaredinver	-
8	Accept	0.28892	0.956	0.28628	6	seuclidean	squaredinver	-
9	Accept	0.58135	1.7306	0.28628	1137	jaccard	inverse	-
10	Accept	0.35048	2.2666	0.28628	1	minkowski	equal	3
90	Accept	0.57784	1.7161	0.281	1137	jaccard	squaredinver	-
91	Accept	0.28892	0.84018	0.281	6	correlation	equal	-
92	Accept	0.58135	1.7013	0.281	1137	jaccard	inverse	-
93	Accept	0.36939	1.5874	0.281	1137	cityblock	squaredinver	-
94	Accept	0.34081	2.0499	0.281	196	spearman	inverse	-
95	Accept	0.66667	1.6777	0.281	1137	jaccard	equal	-
96	Accept	0.29551	2.3135	0.281	34	minkowski	inverse	3
97	Accept	0.33729	1.8938	0.281	6	spearman	squaredinver	-
98	Accept	0.28584	0.87458	0.281	34	seuclidean	squaredinver	-
99	Accept	0.29639	0.89514	0.281	6	chebychev	equal	-
100	Accept	0.55717	0.96897	0.281	1	jaccard	squaredinver	-

Optimization completed.  
MaxObjectiveEvaluations of 100 reached.  
Total function evaluations: 100  
Total elapsed time: 152.7404 seconds.  
Total objective function evaluation time: 141.9119

NumNeighbors	Distance	DistanceWeight	Exponent
6	seuclidean	equal	NaN

Observed objective function value = 0.281  
Function evaluation time = 0.83788

O resultado dessa otimização pode ser descrito da seguinte maneira: após 100 iterações, para o classificador do tipo k vizinhos mais próximos, a menor função objetivo (que neste caso é o menor erro de classificação) foi encontrada com o valor de 28,1% para os hiperparâmetros “Número de Vizinhos” = 6, tipo de distância = “EuclidianaPadronizada” (para a qual o parâmetro expoente não se aplica) e PonderamentoDeDistância= “igual” (ou seja, sem diferenciação). Além disso, o tempo para treinar este modelo foi de 0,83 s.

Para o classificador de árvore de decisão, tem-se o resultado apresentado na Tabela 8.

Tabela 8 - Resultado utilizando árvore de decisão

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	MinLeafSize	MaxNumSplits	SplitCriterion
1	Best	0.38391	4.3071	0.38391	196	1	twoing
2	Best	0.32674	0.87124	0.32674	6	7	deviance
3	Accept	0.33509	0.94028	0.32674	1	329	gdi
4	Accept	0.66843	0.74603	0.32674	1137	1	gdi
5	Best	0.31442	0.85726	0.31442	34	329	deviance
6	Best	0.30299	0.92558	0.30299	6	48	deviance
7	Accept	0.66843	0.72628	0.30299	1137	329	gdi
8	Accept	0.31662	0.86479	0.30299	1	7	twoing
9	Accept	0.38083	0.77817	0.30299	196	7	twoing
10	Accept	0.33553	0.9753	0.30299	1	2273	twoing
90	Accept	0.32718	0.8245	0.29727	1	7	deviance
91	Accept	0.31794	0.81912	0.29727	34	329	gdi
92	Accept	0.32498	0.89397	0.29727	6	2273	gdi
93	Accept	0.31486	0.95961	0.29727	1	48	deviance
94	Accept	0.66843	0.71158	0.29727	1137	7	deviance
95	Accept	0.66843	0.70817	0.29727	1137	1	twoing
96	Accept	0.66843	0.73824	0.29727	1137	2273	deviance
97	Accept	0.38347	0.83893	0.29727	196	48	deviance
98	Accept	0.38083	0.76987	0.29727	196	329	twoing
99	Accept	0.33509	0.92925	0.29727	1	2273	gdi
100	Accept	0.38083	0.80087	0.29727	196	2273	gdi

Optimization completed.  
MaxObjectiveEvaluations of 100 reached.  
Total function evaluations: 100  
Total elapsed time: 94.2558 seconds.  
Total objective function evaluation time: 84.6137

Best observed feasible point:

MinLeafSize	MaxNumSplits	SplitCriterion
6	48	gdi

Observed objective function value = 0.29727  
Function evaluation time = 0.86352

O resultado dessa otimização pode ser descrito da seguinte maneira: após 100 iterações, para o classificador do tipo k vizinhos mais próximos, a menor função objetivo (que neste caso é o menor erro de classificação) foi encontrada com o valor de 29,7% para os hiperparâmetros “TamanhoMínimoFolha” = 6, NúmeroMáximoDivisões = 48 e CritérioDeDivisão = “gdi”. Além disso, o tempo para treinar este modelo foi de 0,86 s.

Com o objetivo de determinar se o número de iterações escolhido por tipo de classificação foi adequado, foi criado um gráfico que exibiu o menor erro de classificação encontrado por quantidade de iterações (Figura 38, Figura 39, Figura 40 e Figura 41). Um comportamento com quedas abruptas seguido de estabilidade indicaria que se chegou ao mínimo, enquanto quedas mais frequentes e incrementais indicariam que um valor menor poderia ser encontrado.

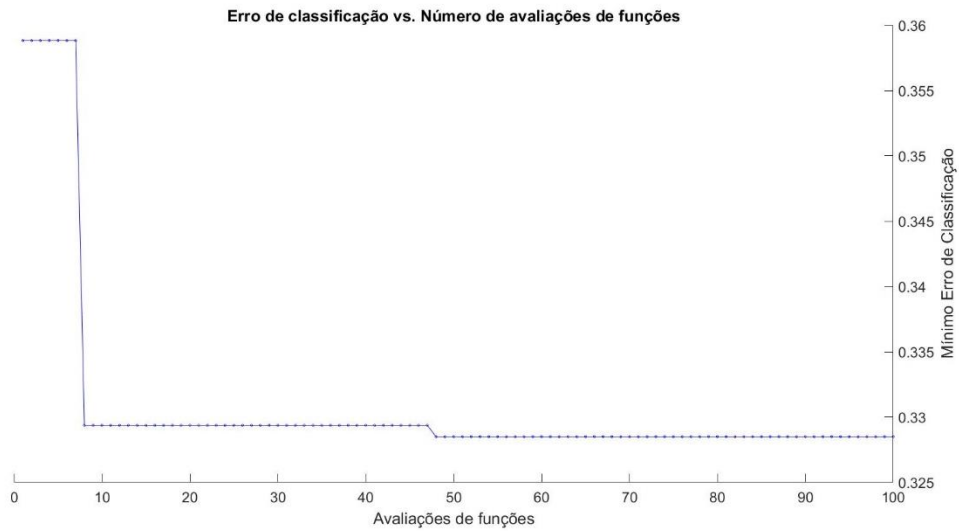


Figura 38 – Evolução do mínimo erro de classificação conforme o número de avaliações de modelos para o classificador análise discriminante.

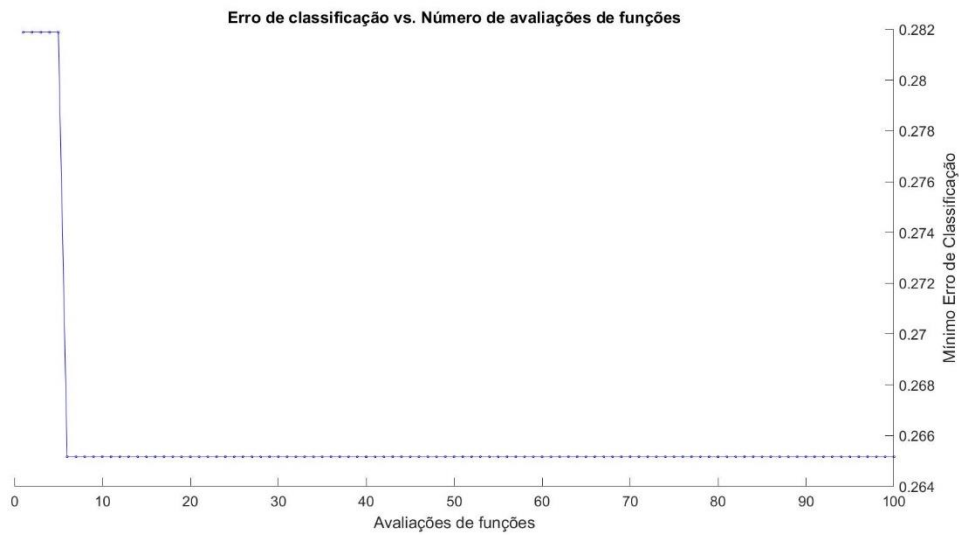
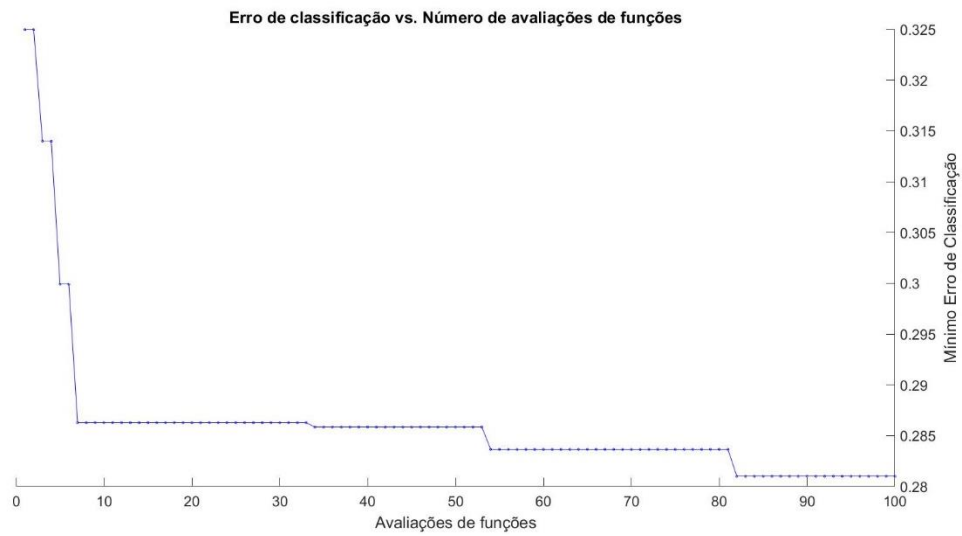
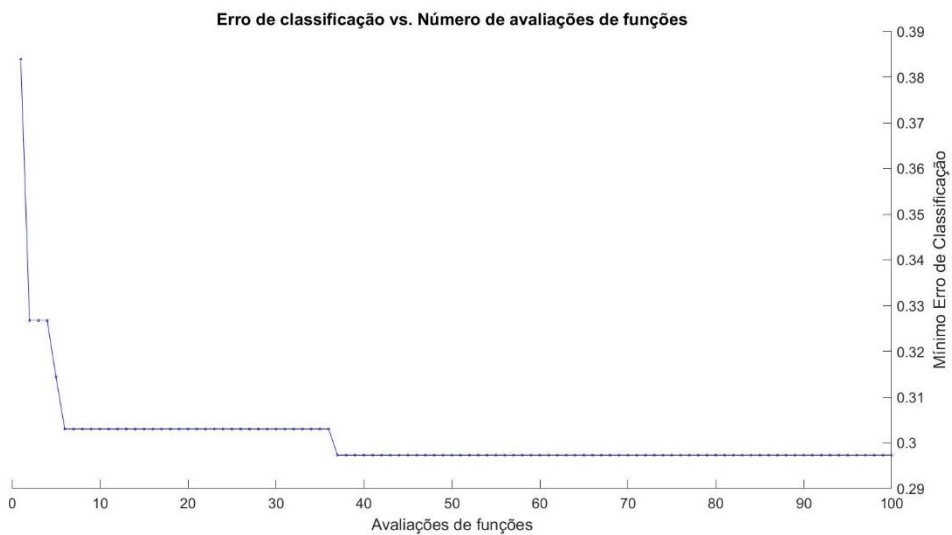


Figura 39 - Evolução do mínimo erro de classificação conforme o número de avaliações de modelos para o classificador máquina de vetores de suporte.



*Figura 40 - Evolução do mínimo erro de classificação conforme o número de avaliações de modelos para o classificador K Vizinhos mais próximos.*



*Figura 41 - Evolução do mínimo erro de classificação conforme o número de avaliações de modelos para o classificador árvore de decisão.*

O resultado pode ser sumarizado na Tabela 9.

*Tabela 9 - Sumário dos resultados utilizando diferentes métodos*

	Análise Discriminante	Máquina de vetores de suporte	K Vizinhos Mais Próximos	Árvore de Decisão
Menor erro de classificação	32,85%	26,5%	28,1%	29,7%
Tempo de execução (segundos)	0,78	106,27	0,88	0,86

O classificador escolhido foi o de máquina de vetores de suporte, que apresentava o menor erro.

### **5.3 Índice de avaliação de desempenho do classificador**

Escolhido o classificador e seu conjunto de hiperparâmetros, foi aplicado o procedimento descrito na seção 4.10.

A matriz de contingência encontrada para o classificador é apresentada na Tabela 10. Os índices  $F_1$  foram calculados e estão exibidos na Tabela 11:

Tabela 10 - Matriz de contingência para o classificador SVM

		Classificação Predita			
		Normal	FA	Outros	Total
Classificação de Referência	Normal	629	27	111	$\sum N$
	FA	28	590	203	$\sum A$
	Outros	101	141	444	$\sum O$
	Total	$\sum n$	$\sum a$	$\sum o$	

Tabela 11 - Índices  $F_1$  para o classificador escolhido

	Normal	FA	Outros	Global
Índice F1	0,7473	0,8249	0,6150	0,7291

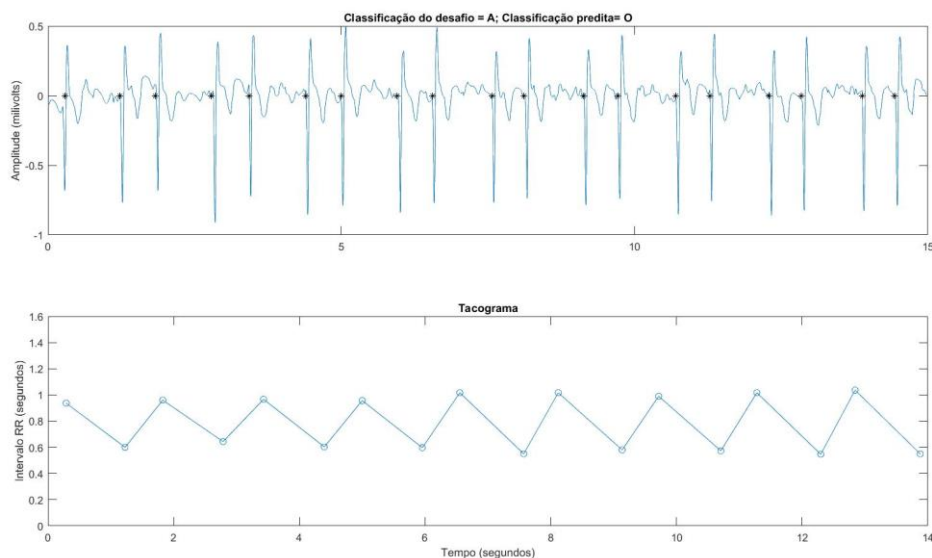


## 6 Discussão

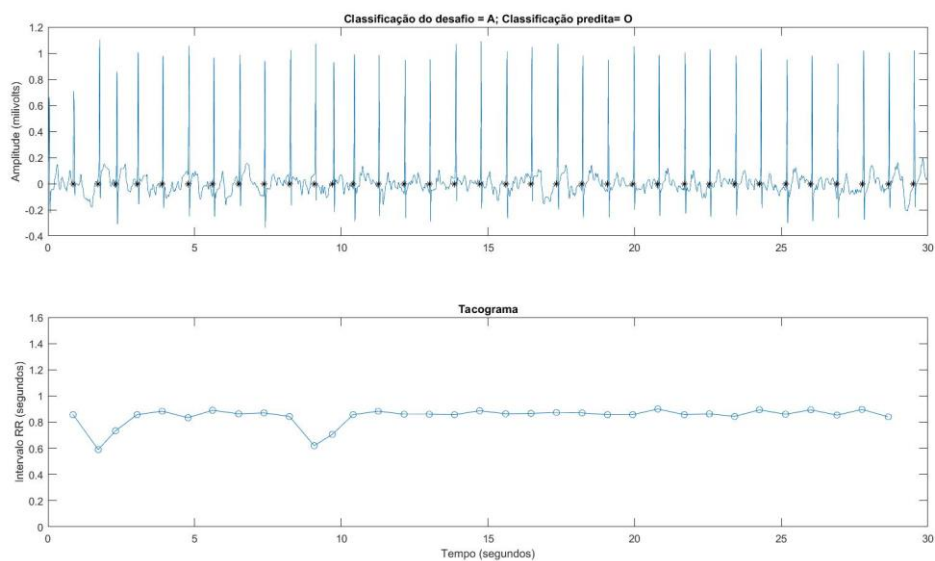
### 6.1 Problemas na diferenciação entre outros ritmos e fibrilação atrial

O método proposto teve baixo desempenho na diferenciação entre outros ritmos e fibrilação atrial. No entanto, a baixa qualidade das anotações de referência foi reconhecida na publicação que apresenta o desafio (CLIFFORD *et al.*, 2017). Como o método proposto utiliza menos características do que o de outros pesquisadores, ele é menos flexível e mais resistente a erros no banco de treinamento. Ou seja, ele será menos afetado por idiosincrasias da avaliação das marcações.

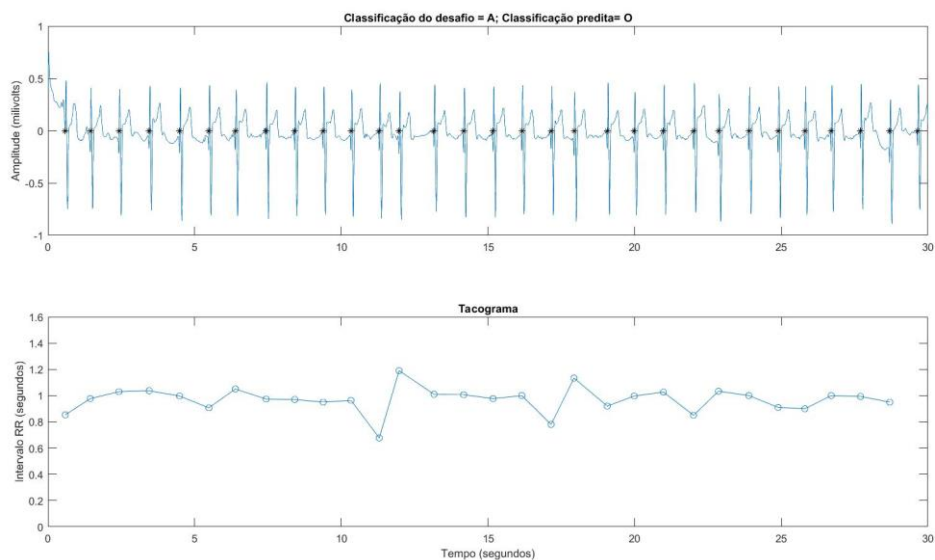
Alguns exemplos de sinais com marcação incorreta serão exibidos a seguir. Na parte superior o sinal será exibido e na parte inferior seu respectivo tacograma.



*Figura 42 – Sinal A00438 do banco de treinamento. O tacograma revela padrão de zigue-zague com valor fixo para o intervalo RR encurtado e alargado. O sinal é marcado pelo desafio incorretamente como fibrilação atrial, porém é um outro ritmo.*



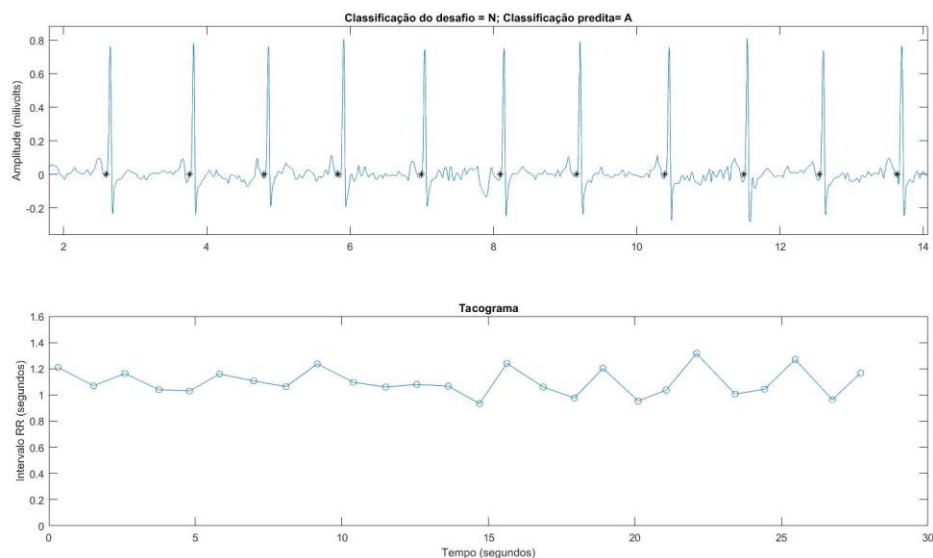
*Figura 43 - Sinal A00551 do banco de treinamento. O tacograma não exibe grande variação nos intervalos entre batimentos, e é observado aspecto de sinal serrilhado monomórfico, indicativo de flutter atrial. O sinal é marcado pelo desafio incorretamente como fibrilação atrial, porém é do tipo outro ritmo.*



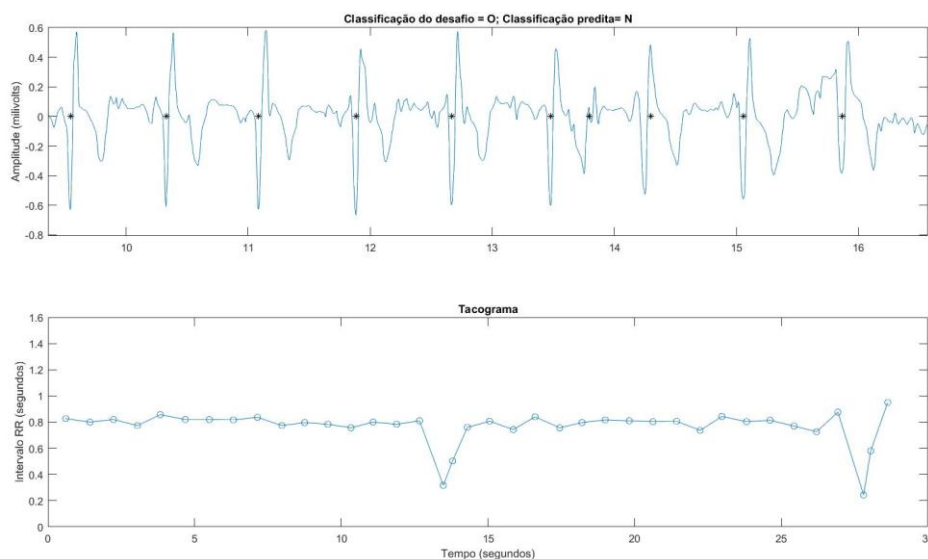
*Figura 44 - Sinal A00680 do banco de treinamento. O tacograma apenas exibe um encurtamento seguido de um alongamento do intervalo entre batimentos. Uma inspeção mais detalhada do sinal entre as ondas R exibe um sinal limpo. O sinal é marcado pelo desafio incorretamente como fibrilação atrial, porém é do tipo outro ritmo.*

## 6.2 Presença de ruído atrapalha diferenciação entre ritmo normal e atrial

Um dos marcadores da presença da fibrilação atrial, além da irregularidade na onda R, é a presença das ondas f, com aspecto de serrilhado polimórfico. O diagnóstico diferencial da fibrilação atrial da FA é o *flutter* atrial, que apresenta serrilhado monomórfico. No entanto, utilizar alguma técnica que detecte este tipo de morfologia exige um sinal limpo, pois o ruído pode ser facilmente confundido com a onda f. Dois conjuntos de sinais serão exibidos na Figura 45 e na Figura 46 para exemplificar a dificuldade na detecção.



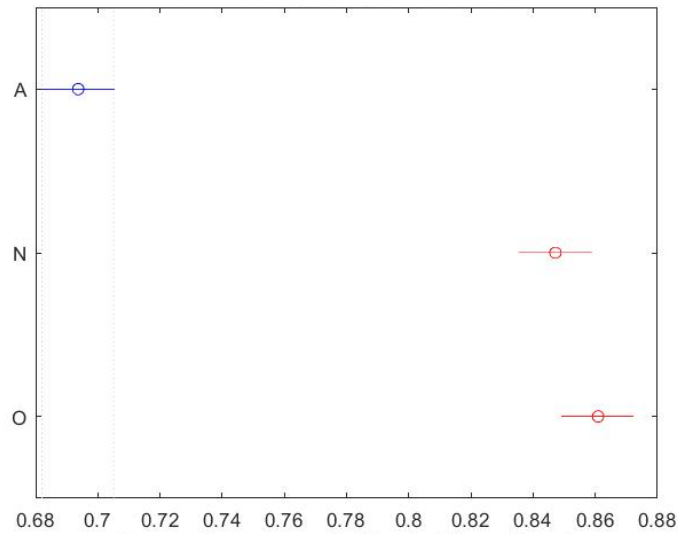
*Figura 45- Na parte superior da figura é exibido um gráfico de detalhe do sinal, enquanto na parte inferior um tacograma do sinal inteiro é exibido. O sinal apresenta bradicardia (RR médio > 1 segundo), então apenas por esse critério o sinal não poderia ter sido classificado como normal, como foi feito pelo avaliador do desafio. É observada uma certa variabilidade aleatória de intervalos de batimentos no sinal, além de um serrilhado polimórfico entre os batimentos, indicando o ritmo de fibrilação atrial, que foi a classificação gerada pelo modelo.*



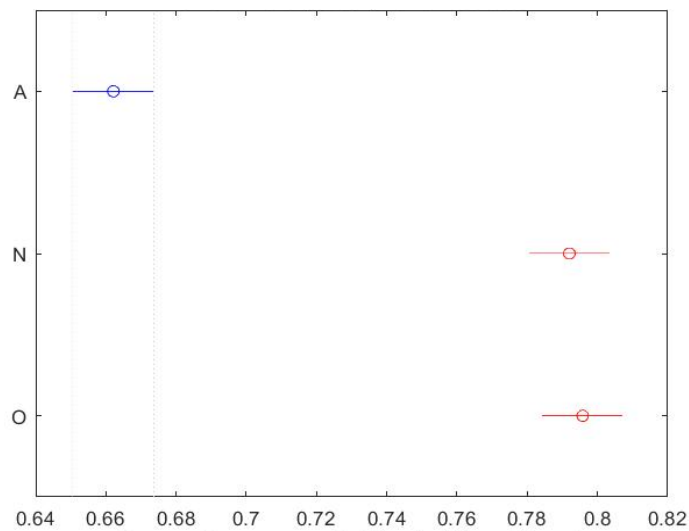
*Figura 46 - Na parte superior da figura é exibido um gráfico de detalhe do sinal, enquanto na parte inferior um tacograma do sinal inteiro é exibido. O sinal apresenta apenas um erro na detecção do pico em torno dos 14 segundos. A variabilidade do sinal é baixa, indicando um sinal normal, como predito pelo modelo deste trabalho. A classificação do desafio, no entanto, indica este sinal como sendo de outro ritmo. O sinal apresenta ruído de base forte, que é bem similar ao da onda f.*

### 6.3 Análise de variáveis

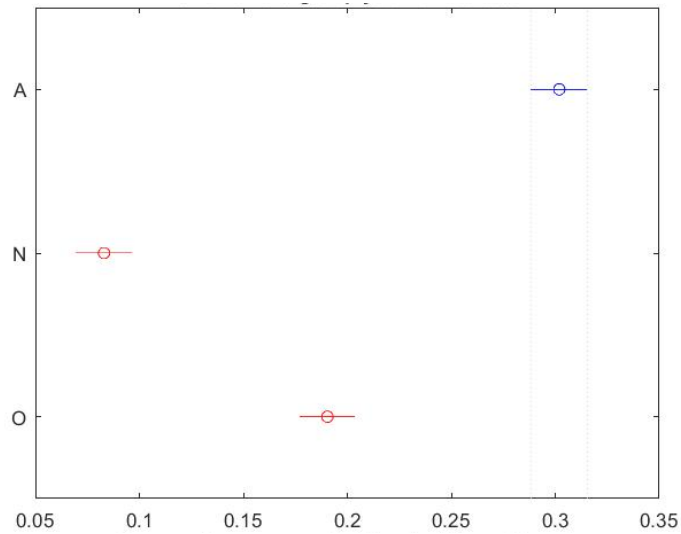
Com o objetivo de avaliar se há diferença significativa nas médias entre-grupos das variáveis escolhidas pelo método de seleção, uma análise de comparação múltipla baseada na técnica de Tukey-Kramer foi realizada (Figura 47 a Figura 56). Com esta técnica, é possível verificar se uma variável é capaz de diferenciar os grupos e em qual grupo ela apresenta média significativamente diferente das demais.



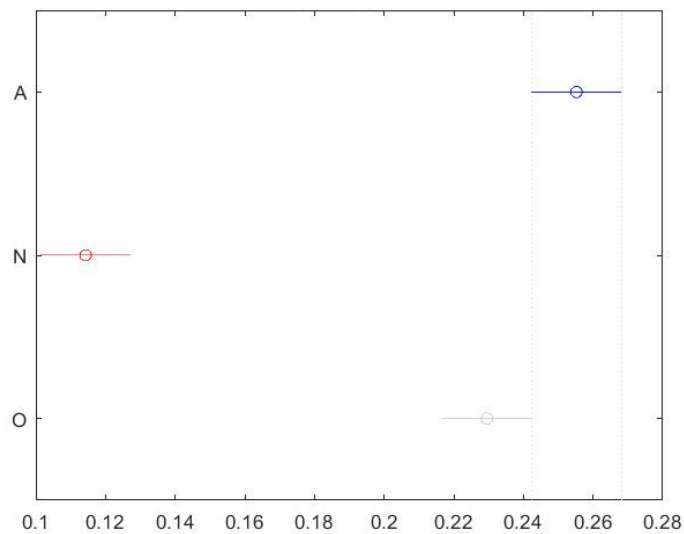
*Figura 47 - Comparação múltipla de mediana de RR. A mediana do RR é significativamente menor no grupo com fibrilação atrial. Ou seja, em média, a frequência cardíaca desse ritmo é maior, conforme esperado.*



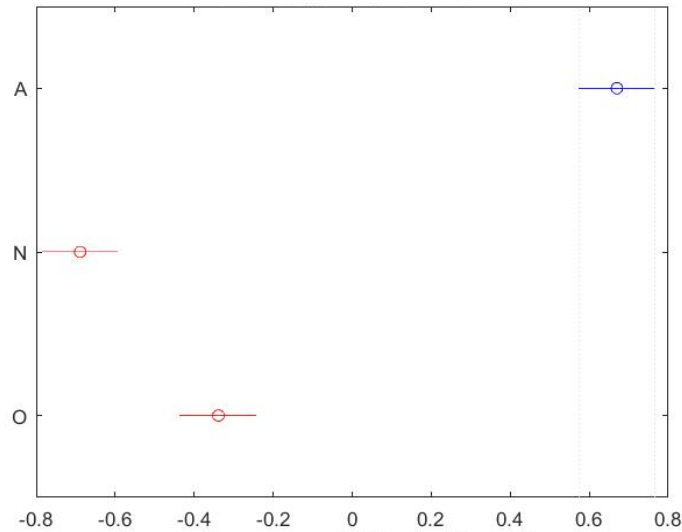
*Figura 48 - Comparação múltipla de média harmônica. A média harmônica do RR é significativamente menor no grupo com fibrilação atrial. Ou seja, em média, a frequência cardíaca desse ritmo é maior, conforme esperado. A média harmônica atribui mais peso a valores menores.*



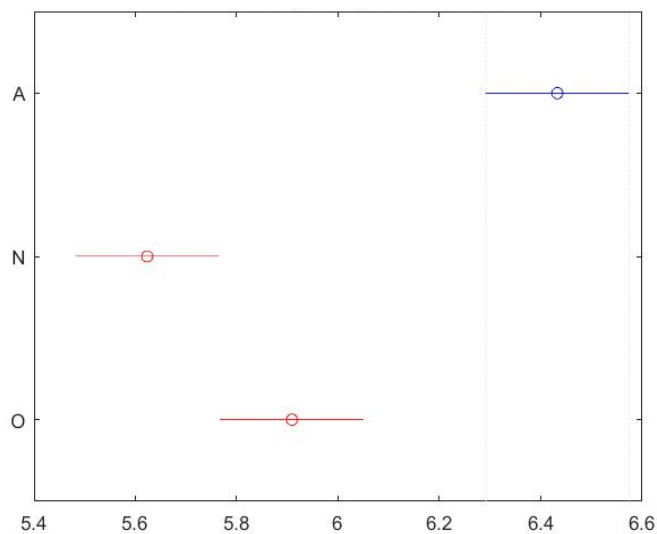
*Figura 49 - Comparação múltipla de intervalo interquartil. O intervalo interquartil é significativamente diferente entre os grupos. O ritmo normal apresenta pouca irregularidade de frequência cardíaca, o “outros ritmos” apresenta um nível de irregularidade intermediário e o ritmo fibrilação atrial apresenta o maior nível de irregularidade, conforme esperado.*



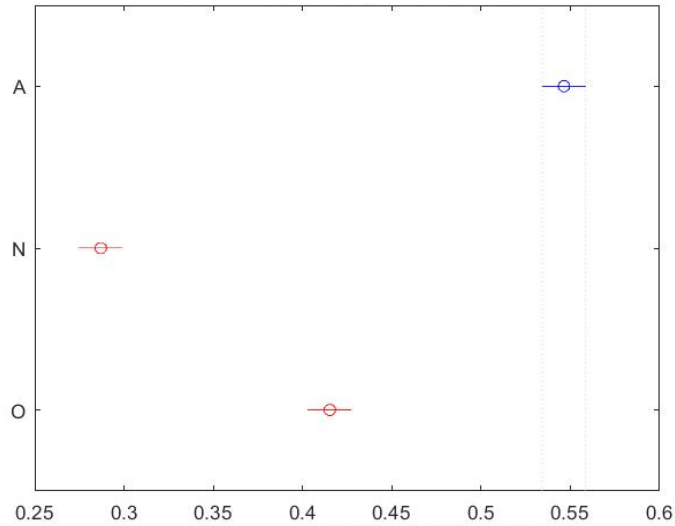
*Figura 50 - Comparação múltipla de desvio padrão. O desvio padrão é significativamente diferente entre fibrilação atrial e os outros grupos. O ritmo normal apresenta pouca irregularidade de frequência cardíaca, no entanto o desvio padrão não é uma medida tão boa quanto o intervalo interquartil para diferenciar fibrilação atrial de outros ritmos, pois é mais sensível a valores extremos.*



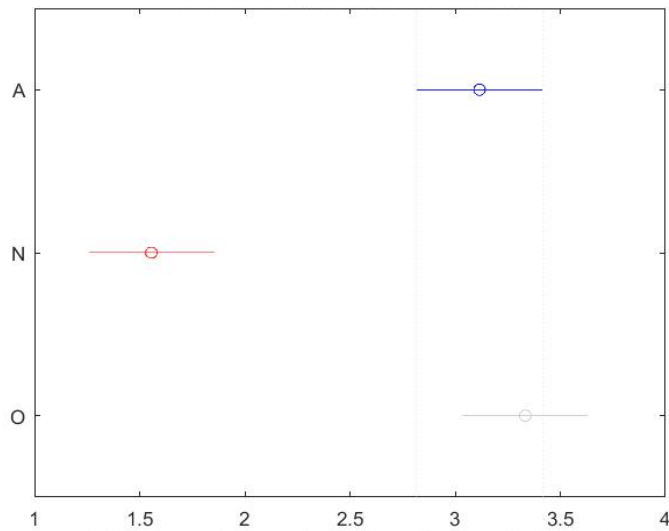
*Figura 51 - Comparação múltipla de desvio médio absoluto. A assimetria é significativamente diferente entre os grupos. O ritmo normal apresenta uma assimetria menor, “outros ritmos” valores intermediários e fibrilação atrial maior assimetria, conforme esperado. A fibrilação atrial é caracterizada por acelerações e desacelerações descoordenadas, logo a distribuição de RRs pode ser mais assimétrica.*



*Figura 52 - Comparação múltipla de frequência média em baixa frequência. A frequência média é significativamente diferente entre os grupos. O ritmo normal apresenta uma frequência menor, o “outros ritmos” apresenta uma frequência média maior e o ritmo fibrilação atrial apresenta o maior nível de frequência média, conforme esperado. O maior nível de frequência média nos sinais com fibrilação atrial pode estar relacionado as ondas f que são visíveis entre os complexos QRS.*

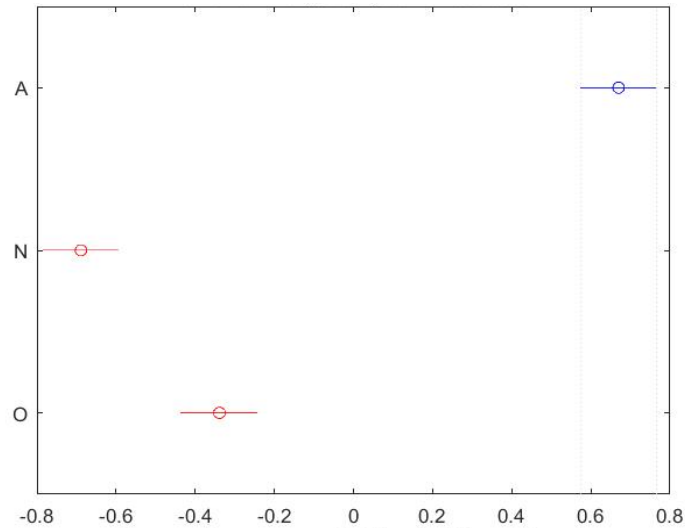


*Figura 53 - Comparação múltipla de dispersão. A dispersão é significativamente diferente entre os grupos. O ritmo normal apresenta uma dispersão menor, o “outros ritmos” apresenta uma dispersão maior e o ritmo fibrilação atrial apresenta o maior nível de dispersão, conforme esperado. A dispersão está relacionada com a variabilidade do RR em relação ao ponto central, logo é mais uma variável com tendência semelhante a de outras que capturam a irregularidade do RR.*

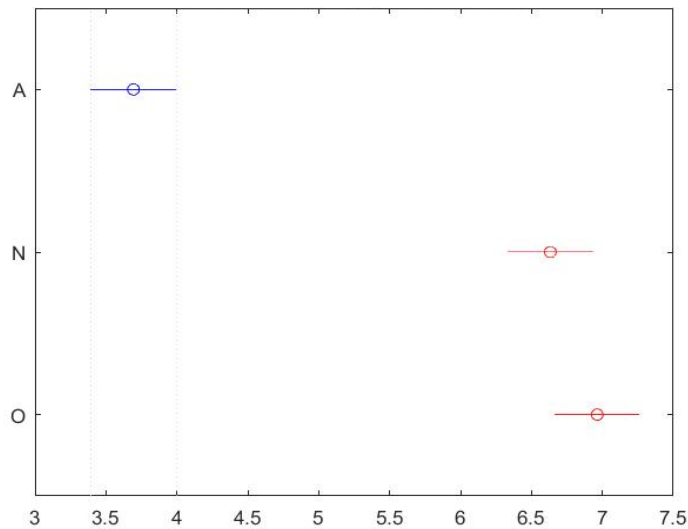


*Figura 54 - Comparação múltipla da evidência de densidade. Quanto maior esse valor, mais dispersos estão os  $\delta RR$ s nos compartimentos do histograma 2-D. É esperado que o ritmo normal tenha menor densidade, conforme observado na análise.*





*Figura 55 - Comparação múltipla da assimetria. A assimetria é diferente significativamente em cada um dos grupos, indicando que cada ritmo tem uma tendência maior para acelerações ou desacelerações em relação à média.*



*Figura 56 - Comparação múltipla da curtose. A curtose é maior na fibrilação atrial do que nos outros grupos e mede o achatamento da curva de distribuição de probabilidade. O ritmo com fibrilação atrial apresenta mais valores extremos do que as outras duas classificações de ritmos.*

## 6.4 Comparação de desempenho dos competidores

Os desempenhos dos quatro primeiros colocados no desafio e de dois outros utilizados para comparação, utilizando o conjunto de teste secreto, é resumido na Tabela 12.

*Tabela 12 – Pontuação dos quatro primeiros colocados no desafio e de dois outros utilizados para comparação. Reproduzido parcialmente de (CLIFFORD et al., 2017), trabalho licenciado sob os termos da Creative Commons attribution 3.0 unported (<https://creativecommons.org/licenses/by/3.0/>)*

Grupo de pesquisa	$F_1$ Oficial	$F_1$ Treinamento
Teijeiro <i>et al.</i>	0,83	0,89
Datta <i>et al.</i>	0,83	0,97
Zabihi <i>et al.</i>	0,83	0,95
Hong <i>et al.</i>	0,83	0,97
Liu <i>et al.</i>	0,80	0,85
Athif <i>et al.</i>	0,78	0,90

Como notado pelo trabalho publicado pela organização do desafio, os principais competidores obtiveram bom desempenho de classificação, porém tiveram ajuste excessivo ao banco de treinamento, mesmo com validação cruzada (CLIFFORD *et al.*, 2017).

O grupo de Teijeiro marcou alguns sinais do grupo de treinamento segundo critério próprio, pré-processou os sinais com um detector de sinal invertido e usou uma técnica de abdução interpretativa chamada *Construe* para identificar características relevantes para o problema. Usou um conjunto de 79 características, divididas em características morfológicas, de ritmo e de qualidade de sinal. Este conjunto de características foi treinado com um classificador Xboost. Em paralelo, características

baseadas em sequências específicas de batimento que caracterizam anomalias foram detectadas por uma segunda rede neural especializada em sequências. Um terceiro classificador, denominado classificador de empilhamento gerava a classificação final de acordo com o resultado dos classificadores anteriores. Esse foi o grupo que mais buscou selecionar melhor as características relacionadas ao problema, por isso também obteve a menor diferença entre o grupo de treinamento e o teste final. Ainda assim, o ajuste ao banco de treinamento ainda foi superior ao de teste (TEIJEIRO *et al.*, 2017).

O grupo de Datta utilizou para pré-processamento do sinal um detector de altas frequências baseado em espectrograma, eliminando trechos de sinais ruidosos. Se 80% do sinal estivesse ruidoso, este era descartado como sinal da classe ruidoso. O grupo utilizou inicialmente mais de 150 características do sinal, baseados em pontos fiduciais do ECG, características propostas por outros pesquisadores, características de HRV, análise na frequência, variáveis de entropia e estatísticas de RR, que serviram como entrada para um classificador *Adaboost* que se utiliza de uma combinação interativa de outros classificadores, com parâmetros ajustados por otimização Bayesiana (DATTA *et al.*, 2017).

O grupo de Zabihi partiu de um conjunto de 491 características e usou uma técnica de *random forest* para ordenar as características por ordem de importância de modo a selecionar as 150 mais importantes. As características se baseiam no domínio do tempo, em características morfológicas, de ritmo, de frequência, de entropia, de fase de sinal e em características “meta”. As características “meta” são descritores estatísticos da predição de características base de segmentos de sinal. O treinamento para as características “meta” foi realizado em um subconjunto separado de 20% de total. O conjunto de características base e “meta” serviram como entrada para um outro classificador do tipo *random forest* (ZABIHI *et al.*, 2017).

O grupo de Hong utilizou como características estatísticas de RR, dados baseados em frequência (escalas wavelet, potência em bandas, entropia Shannon) e características baseadas em conhecimento médico com as da variabilidade de frequência cardíaca. Além disso, o grupo usou um procedimento de análise de clusters em cada sinal para identificar qual o trecho de sinal mais representativo, denominado de *centerwave*. Ademais, o grupo usou uma técnica de *oversampling* para expandir os sinais originais, gerando maior

expansão para as classes com menos observações. Sobre os dados com *oversampling*, foi aplicado uma rede neural DNN e extraído a última camada escondida da rede como conjunto de características. Sobre o sinal *centerwave*, foi aplicado uma rede neural RNN e foi extraído outro conjunto de características. Sobre todos os conjuntos de características citados anteriormente, foi aplicado um *Ensemble* de cinco classificadores *Xboost*, cada um com 3.000 árvores de decisão (HONG *et al.*, 2017).

Por outro lado, o grupo de pesquisadores de Liu usou uma proposta parecida com a do presente trabalho e alcançou performance no grupo de teste apenas ligeiramente inferior. Utilizou um conjunto bem mais enxuto de 33 variáveis, composto de potência de sinal, análise na frequência para detectar ondas f, cálculo de entropia adaptado, estatísticas de RR, estatísticas relacionadas com ondas P, características de trabalhos anteriores como o *IrregularEvidence* e técnicas de *P-wave absence*. No entanto, não usou nenhuma técnica de seleção de características nem fez uma parametrização mais fina dos hiperparâmetros do modelo (LIU *et al.*, 2018).

O grupo de Athif usou um esquema que separa a classificação em dois fluxos diferentes de acordo com a frequência cardíaca (bradicardia e taquicardia em um grupo, o restante em outro). Para o grupo com frequência normal, usou uma configuração de SVM parecida com a utilizada nesse trabalho, de acordo com o algoritmo CSCE do MATLAB® (seção 2.12). Como características, utilizou características baseadas em autovetores, de histograma de RR, de qualidade de sinal, de estatística de RR e de morfologia de onda P (ATHIF *et al.*, 2018). Apesar da sua pontuação de treinamento ter sido superior à de Liu *et al.* (2018), sua pontuação no teste foi inferior.

A diferença da pontuação de treinamento e do conjunto secreto de teste para o grupo de Liu foi consideravelmente menor do que a dos outros casos. A comparação mostra que, apesar dos modelos mais complexos terem apresentado um modelo ligeiramente superior, é provável que um modelo mais simples como o de Liu possa ter desempenho igual ou superior, caso seja melhor parametrizado.

## 6.5 Comparação do presente trabalho

Como a pesquisa não foi concluída a tempo de participar do desafio e o banco de dados de teste não foi divulgado publicamente, não foi possível obter a pontuação no banco de teste para o presente trabalho. No entanto, este trabalho será comparado com o de Liu *et al.* (2018) e Athif *et al.* (2018), pois estes tiveram metodologias mais parecidas com a do presente trabalho, além de usar menos características e terem um desempenho razoável. A comparação é exibida na Tabela 13.

*Tabela 13 – Comparação de resultados usando validação cruzada no banco de treinamento do desafio*

	$F_{1N}$	$F_{1A}$	$F_{10}$	$F_1$
Liu <i>et al.</i>	0,907	0,867	0,765	0,846
Athif <i>et al.</i>	0,938	0,902	0,854	0,898
Este trabalho	0,747	0,825	0,615	0,729

O presente trabalho teve desempenho inferior ao dos outros competidores. Os outros competidores exploraram um conjunto de características mais abrangente, envolvendo características morfológicas do complexo QRS, no domínio da frequência do sinal, além de características morfológicas e estatísticas da onda P. Além disso, usaram técnicas de pré-processamento mais avançadas, baseadas em transformada wavelet, permitindo assim que um sinal mais limpo seja entregue ao detector de ondas P.

Por outro lado, os pesquisadores não reportaram o uso de métodos de seleção de características ou de otimização de hiperparâmetros, indicando que talvez eles ainda estejam usando características demais ou não estejam usando o melhor classificador que se adapta ao problema.

Uma inspeção visual dos sinais classificados incorretamente pelo presente trabalho frequentemente revela que a classificação dada pelo modelo estava de fato correta. Este tipo de problema foi frequentemente relatado, inclusive pelo próprio

organizador do desafio (CLIFFORD *et al.*, 2017; TEIJEIRO *et al.*, 2017; LIU *et al.*, 2018). Como o *range* de tempo do sinal variava muito (9 a 60 s), a padronização de características era dificultada.

Além disso, como poucas e relevantes características foram utilizadas, a possibilidade de este trabalho ter incorrido em *overfitting* é bem menor, o que pode ajudar a explicar o pior resultado. O puro acréscimo de características, mesmo que fracamente relacionadas com o problema subjacente, é suficiente para aumentar o desempenho dos índices de avaliação, no entanto sua aplicação no mundo real pode vir a não ter o desempenho esperado.

## 6.6 Limitações do estudo

Este trabalho teve diversas dificuldades no seu desenvolvimento. Muito tempo foi gasto tentando usar o detector de complexos QRS do WFDB toolbox, sem sucesso. Outros pesquisadores reportaram sucesso com o uso dessas ferramentas *open source*, possibilitando assim o rápido acesso a muitas características do ECG (DATTA *et al.*, 2017; ATHIF *et al.*, 2018; LIU *et al.*, 2018) Tardiamente foi descoberto que isto se devia a problemas de compatibilidade do Windows 10 com o WFDB. Apenas o sistema Linux funciona bem com esse pacote, associado a versões mais antigas do MATLAB®, que possuem uma versão específica de Java. Em função desses problemas, foi necessário implementar um detector de QRS.

Houve ao mesmo tempo um foco em melhorar o classificador com a parametrização de hiperparâmetros e seleção de características. Enquanto essas técnicas são importantes e úteis, foi visto que Liu *et al.* (2018) conseguiram um desempenho superior mesmo sem usar esses recursos.

Foram feitas algumas tentativas de implementação de leitores de onda P tanto no domínio da frequência como no tempo, no entanto nenhuma delas parecia ter funcionado adequadamente, especialmente considerando o caráter mais ruidoso do banco de dados, quando se comparava sinais de exemplo com o resultado que poderia ser esperado. Como

foi visto na seção 6.5, no entanto, alguns pesquisadores reportaram sucesso no uso dessas características.

## 6.7 Propostas de trabalhos futuros

Como proposta de futuro trabalho, propõe-se separar o ritmo “outros ritmos” em um conjunto de ritmos diferentes, cada um com sua especificidade e característica.

Propõe-se também usar técnicas de pré-processamento baseadas em transformadas de sinal mais complexas, complementada por técnicas de detecção de inversão de eletrodo. Seria interessante também usar ferramentas prontas como o detector de onda P do WFDB e usar o procedimento completo proposto por (YEH *et al.*, 2008) para obter os pontos Q e S. Com mais pontos fiduciais do ECG, seria possível obter um conjunto de características morfológicas e temporais mais completo.

É necessário avaliar também características no domínio na frequência, como espectrograma, transformadas *wavelet* e entropia. Além disso, seria importante adicionar características baseadas em estatísticas da onda P.

## 7 Conclusão

A aplicação desenvolvida separa sinais de ECG com fibrilação atrial de sinais com ritmo sinusal normal e com outras arritmias. Os sinais eram de curta duração e apresentavam alta presença de ruído, apresentando assim um problema desafiador e ainda não resolvido de forma definitiva nem mesmo por grandes grupos internacionais dedicados a esse problema.

O método de seleção de características por componentes vizinhos foi efetivo para a obtenção de um classificador com menos características, descartando variáveis com alta correlação cruzada. Além disso, o método de otimização empregado para ajuste automático de hiperparâmetros de modelos supervisionados levou a um incremento relevante no desempenho inicial de cada método testado.

Potencialmente, modelos com poucas variáveis permitem uma melhor interpretação dos fatores que mais contribuem para a identificação do ritmo. Por outro lado, não foi possível igualar o desempenho obtidos na literatura com o uso de muitas características e de técnicas complexas.

A combinação da arquitetura de classificação descrita neste trabalho com algumas características relevantes utilizadas por outros pesquisadores pode levar a um avanço no desenvolvimento de classificação.



## 8 Referências bibliográficas

- ALBUQUERQUE, J. P. de A.; FORTES, J. M. P.; FINAMORE, W. A. **Probabilidade, variável aleatória e processos estocásticos**. 1. ed. Rio de Janeiro: Interciência, 2008.
- ALIZADEH, A. A.; EISEN, M. B.; DAVIS, R. E.; *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. (cover story). **Nature**, v. 403, n. 6769, p. 503, 3 fev. 2000.
- ALON, U.; BARKAI, N.; NOTTERMAN, D. A.; *et al.* Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays. **Proceedings of the National Academy of Sciences**, v. 96, n. 12, p. 6745–6750, 8 jun. 1999.
- ATHIF, M.; YASAWARDENE, P. C.; DALUWATTE, C. Detecting Atrial Fibrillation from Short Single Lead ECGs Using Statistical and Morphological Features. **Physiological Measurement**, v. 39, n. 6, p. 064002, 2018.
- BERNE; LEVY. **Fisiologia**. 6. ed. Rio de Janeiro: Elsevier, 2008.
- BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; *et al.* **Classification and Regression Trees**. Boca Raton: Chapman & Hall, 1993.
- CAMM, A. J.; KIRCHHOF, P.; LIP, G. Y. H.; *et al.* Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). **Europace: European Pacing, Arrhythmias, And Cardiac Electrophysiology: Journal Of The Working Groups On Cardiac Pacing, Arrhythmias, And Cardiac Cellular Electrophysiology Of The European Society Of Cardiology**, v. 12, n. 10, p. 1360–1420, out. 2010.
- CHAO HUANG; SHUMING YE; HANG CHEN; *et al.* A Novel Method for Detection of the Transition Between Atrial Fibrillation and Sinus Rhythm. **IEEE Transactions on Biomedical Engineering**, v. 58, n. 4, p. 1113–1119, abr. 2011.
- CHEN, B.; LIU, H.; CHAI, J.; *et al.* Large Margin Feature Weighting Method via Linear Programming. **IEEE Transactions on Knowledge and Data Engineering**, v. 21, n. 10, p. 1475–1488, out. 2009.
- CLAES, N.; VAN LAETHEM, C.; GOETHALS, M.; *et al.* Prevalence of atrial fibrillation in adults participating in a large-scale voluntary screening programme in Belgium. **Acta Cardiologica**, v. 67, n. 3, p. 273–278, jun. 2012.
- Classification Using Nearest Neighbors - MATLAB & Simulink**. Disponível em: <<https://www.mathworks.com/help/releases/R2017a/stats/classification-using-nearest-neighbors.html>>. Acesso em: 18 jul. 2018.

CLIFFORD, G.; LIU, C.; MOODY, B.; *et al.* AF classification from a short single lead ECG recording: The Physionet Computing in Cardiology Challenge 2017. **Computing in Cardiology**, v. 44, 2017.

**Creative Commons — Attribution-ShareAlike 3.0 Unported — CC BY-SA 3.0.**  
Disponível em: <<https://creativecommons.org/licenses/by-sa/3.0/deed.en>>. Acesso em: 18 ago. 2018.

DATTA, S.; PURI, C.; MUKHERJEE, A.; *et al.* Identifying Normal, AF and other Abnormal ECG Rhythms using a Cascaded Binary Classifier. In: **Anais... In: 2017 COMPUTING IN CARDIOLOGY CONFERENCE**. 14 set. 2017. Disponível em: <<http://www.cinc.org/archives/2017/pdf/173-154.pdf>>. Acesso em: 20 ago. 2018.

**Decision Trees - MATLAB & Simulink.** Disponível em: <<https://www.mathworks.com/help/releases/R2017a/stats/classification-trees-and-regression-trees.html>>. Acesso em: 3 jun. 2018.

DU, X.; RAO, N.; QIAN, M.; *et al.* A Novel Method for Real-Time Atrial Fibrillation Detection in Electrocardiograms Using Multiple Parameters. **Annals of Noninvasive Electrocardiology**, v. 19, n. 3, p. 217–225, 1 maio 2014.

**Feature selection using neighborhood component analysis for classification - MATLAB fscna.** Disponível em: <<https://www.mathworks.com/help/releases/R2017a/stats/fscna.html#zmw57dd0e377598>>. Acesso em: 19 ago. 2018.

**Fit binary classification decision tree for multiclass classification - MATLAB fitctree.** Disponível em: <<https://www.mathworks.com/help/releases/R2017a/stats/fitctree.html>>. Acesso em: 16 jun. 2018.

**Fit multiclass models for support vector machines or other classifiers - MATLAB fitcecoc.** Disponível em: <<https://www.mathworks.com/help/releases/R2017a/stats/fitcecoc.html#bufm0tv>>. Acesso em: 18 ago. 2018.

FURNKRANZ, J. Round Robin Classification. v. 2, p. 27, 3 jan. 2002.

GO AS; HYLEK EM; PHILLIPS KA; *et al.* Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. **JAMA: Journal of the American Medical Association**, v. 285, n. 18, p. 2370–2394, 9 maio 2001.

GUYTON, A. C.; HALL, J. E. **Tratado de Fisiologia Médica**. Rio de Janeiro: Elsevier, 2006.

- HEERINGA, J.; KUIP, V. D.; A.M, D.; *et al.* Prevalence, Incidence and Lifetime Risk of Atrial Fibrillation: The Rotterdam Study. **European Heart Journal**, v. 27, n. 8, p. 949–953, 1 abr. 2006.
- HEILMAN, J. **File:PVC10.JPG - Wikimedia Commons**. Disponível em: <<https://commons.wikimedia.org/wiki/File:PVC10.JPG>>. Acesso em: 18 ago. 2018.
- HONG, S.; WU, M.; ZHOU, Y.; *et al.* ENCASE: an ENsemble CIASSifiEr for ECG Classification Using Expert Features and Deep Neural Networks. In: **Anais... In: 2017 COMPUTING IN CARDIOLOGY CONFERENCE**. 14 set. 2017. Disponível em: <<http://www.cinc.org/archives/2017/pdf/178-245.pdf>>. Acesso em: 20 ago. 2018.
- JAAKKOLA, J.; MUSTONEN, P.; KIVINIEMI, T.; *et al.* Stroke as the First Manifestation of Atrial Fibrillation. **PLoS ONE**, v. 11, n. 12, p. 1–9, 9 dez. 2016.
- JAMES, G.; WITTEN, D.; HASTIE, T.; *et al.* Statistical Learning. In: JAMES, G.; WITTEN, D.; HASTIE, T.; *et al.* (Ed.). **An Introduction to Statistical Learning: with Applications in R**. New York, NY: Springer New York, 2013a. p. 15–57.
- JAMES, G.; WITTEN, D.; HASTIE, T.; *et al.* Classification. In: JAMES, G.; WITTEN, D.; HASTIE, T.; *et al.* (Ed.). **An Introduction to Statistical Learning: with Applications in R**. New York, NY: Springer New York, 2013b. p. 127–173.
- JAMES, G.; WITTEN, D.; HASTIE, T.; *et al.* Support Vector Machines. In: JAMES, G.; WITTEN, D.; HASTIE, T.; *et al.* (Ed.). **An Introduction to Statistical Learning: with Applications in R**. New York, NY: Springer New York, 2013c. p. 337–372.
- JAMES, G.; WITTEN, D.; HASTIE, T.; *et al.* Resampling Methods. In: JAMES, G.; WITTEN, D.; HASTIE, T.; *et al.* (Ed.). **An Introduction to Statistical Learning: with Applications in R**. New York, NY: Springer New York, 2013d. p. 175–201.
- LADAVICH, S.; GHORAANI, B. Rate-Independent Detection of Atrial Fibrillation by Statistical Modeling of Atrial Activity. **Biomedical Signal Processing and Control**, v. 18, p. 274–281, abr. 2015.
- LI, C.; ZHENG, C.; TAI, C. Detection of ECG characteristic points using wavelet transforms. **IEEE Transactions on Biomedical Engineering**, v. 42, n. 1, p. 21–28, jan. 1995.
- LIU, N.; SUN, M.; WANG, L.; *et al.* A Support Vector Machine Approach for AF Classification from a Short Single-Lead ECG Recording. **Physiological Measurement**, v. 39, n. 6, p. 064004, 2018.
- LLAMEDO, M.; MARTINEZ, J. P. An Automatic Patient-Adapted ECG Heartbeat Classifier Allowing Expert Assistance. **IEEE Transactions on Biomedical Engineering**, v. 59, n. 8, p. 2312–2320, ago. 2012.

- LUZ, E. J. da S.; SCHWARTZ, W. R.; CÁMARA-CHÁVEZ, G.; *et al.* ECG-Based Heartbeat Classification for Arrhythmia Detection: A Survey. **Computer Methods and Programs in Biomedicine**, v. 127, p. 144–164, abr. 2016.
- LYNN, P. A. Recursive Digital Filters for Biological Signals. **Medical & biological engineering**, v. 9, n. 1, p. 37–43, 1 jan. 1971.
- MARCOLINO, M. S.; PALHARES, D. M. F.; BENJAMIN, E. J.; *et al.* Atrial fibrillation: prevalence in a large database of primary care patients in Brazil. **Europace: European Pacing, Arrhythmias, And Cardiac Electrophysiology: Journal Of The Working Groups On Cardiac Pacing, Arrhythmias, And Cardiac Cellular Electrophysiology Of The European Society Of Cardiology**, v. 17, n. 12, p. 1787–1790, dez. 2015.
- MOODY, G. B.; MARK, R. G. The impact of the MIT-BIH Arrhythmia Database. **IEEE Engineering in Medicine and Biology Magazine**, v. 20, n. 3, p. 45–50, maio 2001.
- PAN, J.; TOMPKINS, W. J. A Real-Time QRS Detection Algorithm. **IEEE Transactions on Biomedical Engineering**, v. BME-32, n. 3, p. 230–236, mar. 1985.
- PARK, J.; LEE, S.; JEON, M. Atrial Fibrillation Detection by Heart Rate Variability in Poincare Plot. **BioMedical Engineering OnLine**, v. 8, n. 1, p. 38, 11 dez. 2009.
- RANB, R. G.-B.; NAVOT, A.; TISHBY, N. Margin Based Feature Selection - Theory and Algorithms. In: In International Conference on Machine Learning (ICML, **Anais...ACM Press**, 2004.
- RÓDENAS, J.; GARCÍA, M.; ALCARAZ, R.; *et al.* Wavelet Entropy Automatically Detects Episodes of Atrial Fibrillation from Single-Lead Electrocardiograms. **Entropy**, v. 17, n. 9, p. 6179–6199, 7 set. 2015.
- RUFFY, R. **Cardiac Electrophysiology: from cell to bedside 63 - Atrial Fibrillation**. 2. ed. [s.l.] Zipes and Jalife, 1995.
- SAMENI, R.; SHAMSOLLAHI, M. B.; JUTTEN, C.; *et al.* A Nonlinear Bayesian Filtering Framework for ECG Denoising. **IEEE Transactions on Biomedical Engineering**, v. 54, n. 12, p. 2172–2185, dez. 2007.
- SARKAR, S.; RITSCHER, D.; MEHRA, R. A Detector for a Chronic Implantable Atrial Tachyarrhythmia Monitor. **IEEE Transactions on Biomedical Engineering**, v. 55, n. 3, p. 1219–1224, mar. 2008.
- SAYADI, O.; SHAMSOLLAHI, M. B. Multiadaptive Bionic Wavelet Transform: Application to CG Denoising and Baseline Wandering Reduction. **EURASIP Journal on Advances in Signal Processing**, p. 1–11, jan. 2007.
- SAYADI, O.; SHAMSOLLAHI, M. B. ECG Denoising and Compression Using a Modified Extended Kalman Filter Structure. **IEEE Transactions on Biomedical Engineering**, v. 55, n. 9, p. 2240–2248, set. 2008.

- STATNIKOV, A.; ALIFERIS, C. F.; TSAMARDINOS, I.; *et al.* A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. **Bioinformatics (Oxford, England)**, v. 21, n. 5, p. 631–643, 1 mar. 2005.
- STEWART, S.; HART, C.; HOLE, D.; *et al.* Population prevalence, incidence, and predictors of atrial fibrillation in the Renfrew/Paisley study. **Heart**, v. 86, n. 5, p. 516–521, nov. 2001.
- SUN, Y.; TODOROVIC, S.; GOODISON, S. Local-Learning-Based Feature Selection for High-Dimensional Data Analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 32, n. 9, p. 1610–1626, set. 2010.
- TEIJEIRO, T.; GARCIA, C. A.; CASTRO, D.; *et al.* Arrhythmia Classification from the Abductive Interpretation of Short Single-Lead ECG Records. In: **Anais... In: 2017 COMPUTING IN CARDIOLOGY CONFERENCE**. 14 set. 2017. Disponível em: <<http://www.cinc.org/archives/2017/pdf/166-054.pdf>>. Acesso em: 20 ago. 2018.
- TOMAŠIĆ, I.; TROBEC, R. Electrocardiographic Systems With Reduced Numbers of Leads; Synthesis of the 12-Lead ECG. **IEEE Reviews in Biomedical Engineering**, v. 7, p. 126–142, 2014.
- TURAKHIA, M. P.; SHAFRIN, J.; BOGNAR, K.; *et al.* Estimated prevalence of undiagnosed atrial fibrillation in the United States. **PLoS ONE**, v. 13, n. 4, p. 1–11, 12 abr. 2018.
- XIAN, Y.; O'BRIEN, E. C.; LIANG, L.; *et al.* Association of Preceding Antithrombotic Treatment With Acute Ischemic Stroke Severity and In-Hospital Outcomes Among Patients With Atrial Fibrillation. **JAMA**, v. 317, n. 10, p. 1057–1067, 14 mar. 2017.
- XUE, Q.; HU, Y. H.; TOMPKINS, W. J. Neural-network-based adaptive matched filtering for QRS detection. **IEEE Transactions on Biomedical Engineering**, v. 39, n. 4, p. 317–329, abr. 1992.
- YANG, W.; WANG, K.; ZUO, W. Neighborhood Component Feature Selection for High-Dimensional Data. **Journal of Computers**, v. 7, n. 1, 1 jan. 2012. Disponível em: <<http://ojs.academypublisher.com/index.php/jcp/article/view/5076>>. Acesso em: 5 set. 2017.
- YEH, Y.-C.; WANG, W.-J. QRS complexes detection for ECG signal: The Difference Operation Method. **Computer Methods and Programs in Biomedicine**, v. 91, n. 3, p. 245–254, 1 set. 2008.
- ZABIHI, M.; BAHRAMI RAD, A.; KATSAGGELOS, A. K.; *et al.* Detection of Atrial Fibrillation in ECG Hand-held Devices Using a Random Forest Classifier. In: **Anais... In: 2017 COMPUTING IN CARDIOLOGY CONFERENCE**. 14 set. 2017. Disponível em: <<http://www.cinc.org/archives/2017/pdf/069-336.pdf>>. Acesso em: 20 ago. 2018.