



UNIVERSIDADE ESTADUAL PAULISTA  
“JÚLIO DE MESQUITA FILHO”  
Campus de Ilha Solteira

**Bruno Rodrigues de Oliveira**

**Reconhecimento de Contração Prematura Ventricular  
utilizando separação cega de fontes e comitê de máquinas  
bayesianas**

Ilha Solteira

2020

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**Bruno Rodrigues de Oliveira**

**Reconhecimento de Contração Prematura Ventricular  
utilizando separação cega de fontes e comitê de máquinas  
bayesianas**

Texto submetido ao Programa de Pós Graduação  
em Engenharia Elétrica - UNESP - Campus de Ilha  
Solteira, Como requisito para obtenção do título de  
Doutor em Engenharia Elétrica.

Área de Conhecimento: Automação.

**Orientador:** Prof. Dr. Jozue Vieira Filho

Ilha Solteira

2020

FICHA CATALOGRÁFICA

Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

O48r Oliveira, Bruno Rodrigues de.  
Reconhecimento de contração prematura ventricular utilizando separação  
cega de fontes e máquinas bayesianas / Bruno Rodrigues de Oliveira. -- Ilha  
Solteira: [s.n.], 2020  
151 f. : il.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia  
de Ilha Solteira. Área de conhecimento: Engenharia Elétrica, Automação, 2020

Orientador: Jozue Vieira Filho  
Inclui bibliografia

1. Contração prematura ventricular. 2. Inteligência artificial. 3. Aprendizado  
de máquina. 4. Naive bayes. 5. Separação cega de fontes.

  
Raiane da Silva Santos

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Reconhecimento de contração ventricular prematura utilizando separação cega de fontes e comitê de máquinas bayesianas

AUTOR: BRUNO RODRIGUES DE OLIVEIRA

ORIENTADOR: JOZUE VIEIRA FILHO

Aprovado como parte das exigências para obtenção do Título de Doutor em ENGENHARIA ELÉTRICA, área: Automação pela Comissão Examinadora:

Prof. Dr. JOZUE VIEIRA FILHO

Coordenadoria Executiva / Câmpus Experimental de São João da Boa Vista - UNESP

Prof. Dr. EDSON DONIZETE DE CARVALHO

Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira - UNESP

Prof. Dr. FRANCISCO VILLARREAL ALVARADO

Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira - UNESP

Prof. Dr. ROBERTO KAWAKAMI HARROP GALVÃO

Departamento de Sistemas e Controle / Instituto Tecnológico de Aeronáutica - ITA

Prof. Dr. PAULO EDUARDO TEODORO

Departamento de Agronomia / Universidade Federal de Mato Grosso do Sul - UFMS

Ilha Solteira, 17 de setembro de 2020

# Agradecimentos

Agradeço a Deus pelo fôlego de vida. Aos meus pais Mauro e Fátima, pela dedicação e incentivo. Ao meu irmão Lucas, pela ajuda na gramática. A minha esposa Camila e filhos Felipe e Maria Helena, por emprestarem o tempo que lhes devo e pelos momentos de alegria.

Agradeço especialmente ao meu orientador Prof. Dr. Jozue Vieira Filho pela oportunidade e paciência e ao Prof. Dr. Marco Aparecido Queiroz Duarte, que apesar de não constar oficialmente como orientador, vem desde a graduação orientando meus passos na aplicação da matemática para a solução de problemas da vida real.

Agradeço a todos meus colegas de estudo, principalmente ao Prof. Dr. Caio Cesar Enside de Abreu e ao quase doutor Fabrício Ely Gossler, os quais ajudaram a esclarecer muitas dúvidas, e com os quais tive a oportunidade de publicar alguns trabalhos.

Agradeço também a todos os servidores e professores da UNESP de Ilha Solteira, em especial aqueles do Programa de Pós-Graduação em Engenharia Elétrica que nunca mediram esforços para me auxiliarem naquilo que fosse necessário.

*“Não abandone a sabedoria, e ela o protegerá;  
ame-a, e ela cuidarã de vocẽ. (Bíblia Sagrada, Provérbios 4:6)*

# Resumo

As Contrações Prematuras Ventriculares (CPV) estão entre as arritmias que podem desencadear doenças cardíacas fatais. Seu diagnóstico é mais acurado quando utilizado monitoramento ambulatorial de 24 ou 48 horas. Estes exames geram muitos dados para serem analisados pelo especialista, logo, ferramentas computacionais são imprescindíveis para auxiliarem o diagnóstico. Neste trabalho é proposta uma nova abordagem para reconhecimento das CPVs, por meio da análise de segmentos dos sinais de Eletrocardiograma (ECG), baseada na utilização do método de separação cega de fontes AMUSE para extração dos atributos característicos, seguido da Análise Discriminante Linear para redução da dimensionalidade, e o algoritmo Naive Bayes para indução das máquinas de aprendizado, as quais são ponderadas em um comitê de máquinas cujas importâncias são obtidas pelo método Análise Hierárquica de Processos (AHP) empregando uma função de conversão adequada, também proposta. Vários experimentos são implementados a fim de verificar a performance das abordagens propostas, incluindo: dados balanceados e desbalanceados, performances individuais por registros ECG, variações dos parâmetros das máquinas de aprendizado, ambiente ruidoso e diversas composições de comitês. No melhor dos casos, obteve-se medidas de acurácia, sensibilidade e especificidade nos valores de 99,57%, 98,64% e 99,65%, respectivamente. A pesquisa apresentada traz contribuições relevantes para a área de aprendizado de máquina de sinais biomédicos, ao apresentar uma metodologia simples, de baixo custo computacional e eficiente para o reconhecimento de arritmias CPV.

**Palavras-chaves:** Contrações Prematuras Ventriculares; Reconhecimento de arritmias; Aprendizado de máquina; Separação cega de fontes; Análise Hierárquica de Processos.

# Abstract

Premature Ventricular Contractions (PVC) are among the arrhythmias that can trigger fatal heart disease. Their diagnose are more accurate when 24- or 48-hour ambulatory monitoring are used. These exams generate a lot of data to be analyzed by the specialist, therefore computational tools are essential to helping the diagnosis. This work proposes a new approach for PVC recognition, analyzing ECG signal segments, using the AMUSE blind source separation method for attributes extraction, Linear Discriminant Analysis for dimensionality reduction and the Naive Bayes algorithm to induce learning machines which are weighted in a machine committee whose importances are obtained by the Analytic Hierarchy Process method through an appropriate conversion function. Several experiments are implemented, such as: balanced and unbalanced data, individual ECG record performances, variations in learning machine parameters, noisy environment, various committee compositions; in order to verify the performance of the proposed approaches. In the best of the cases, accuracy, sensitivity and specificity measures were 99,57%, 98,64% e 99,65%, respectively. The presented survey brings out relevant contributions to the biomedical signal machine learning area, presenting a simple, low computational and efficient methodology for PVC arrhythmias recognition.

**Keywords:** Premature Ventricular Contractions; Arrhythmia recognition; Machine learning; Blind source separation; Analytic Hierarchy Process.

# Lista de ilustrações

Figura 1 – Fisiologia do coração. . . . .	21
Figura 2 – Representação de um ECG típico da derivação II e os potenciais de ação de várias células do miocárdio. . . . .	24
Figura 3 – Pausa compensatória. Aqui PVC significa <i>Premature Ventricular Contraction</i> . . . . .	29
Figura 4 – Exemplos de CPV (rótulo V) distintas do registro 207 da base de dados MIT/BIH - <i>Arrhythmia Database</i> . . . . .	31
Figura 5 – Exemplos de CPV seguido de ondas T distintas do registro 233 da base de dados MIT/BIH - <i>Arrhythmia Database</i> . . . . .	31
Figura 6 – Exemplos de diferenças (a) suave e (b) abrupta. . . . .	37
Figura 7 – Pontos verdes e púrpura representando os padrões do ambiente e as fronteiras de decisão obtidas por uma Regressão Logística. (a) Instâncias fornecidas a uma máquina induzida previamente; (b) principal fronteira de decisão com probabilidade 0,5; e (c) várias fronteiras de decisão com distintos valores de probabilidades, representadas de acordo com as barras de cores. . . . .	41
Figura 8 – Ilustração da curva de aprendizagem de três máquinas distintas $h_1$ , $h_2$ e $h_3$ , do tipo Regressão Logística, <i>Naives Bayes</i> com função de distribuição Gaussiana e $k$ -Vizinhos mais próximos, com $k = 3$ , respectivamente. . . . .	42
Figura 9 – Comitê de máquinas do tipo Média de Ensemble, considerando o espaço de entrada com rótulos desconhecidos. . . . .	45
Figura 10 – Exemplo de aplicação de um comitê de máquinas do tipo média de ensemble. (a) Instâncias para validação dos algoritmos de aprendizado; (b), (c) e (d) são as superfícies de decisão obtidas por um modelo Naive Bayes, Regressão Logística e Média de Ensemble utilizando as probabilidades de classificação com voto majoritário, respectivamente. . . . .	48
Figura 11 – Exemplo da técnica de validação cruzada, com $K = 4$ , utilizando 100 exemplos gerados aleatoriamente, pertencentes a duas classes e a 10 indivíduos. . . . .	51
Figura 12 – Exemplo da técnica de validação cruzada estratificada, com $K = 4$ , utilizando 100 exemplos gerados aleatoriamente, pertencentes a duas classes e a 10 indivíduos. . . . .	53
Figura 13 – Exemplo da técnica de validação cruzada agrupada, com $K = 4$ , utilizando 100 exemplos gerados aleatoriamente, pertencentes a duas classes e a 10 indivíduos. . . . .	54

Figura 14 – Modelo genérico da Análise Hierárquica de Processos, com $N$ critérios e $M$ alternativas. . . . .	74
Figura 15 – Diagrama de blocos da proposta principal. . . . .	87
Figura 16 – Segmentação do sinal de ECG. Segmentos de retas em vermelho e azul destacam o primeiro ciclo cardíaco e o eixo de tempo, respectivamente. O retângulo pontilhado verde representa o janelamento referente a um ciclo cardíaco. . . . .	89
Figura 17 – Diagrama da composição de uma fonte observável $\mathbf{x}(p)$ resultante da sobreposição de fontes secundárias e ruído. . . . .	92
Figura 18 – (a) Sinal $\mathbf{s}_1(p)$ , (b) singularidade $\mathbf{s}_2(p)$ ; e (c) sinal observado $\mathbf{x}(p)$ . . . . .	93
Figura 19 – Representação das componentes estimadas pelo método AMUSE, (a) e (b), considerando o sinal $\mathbf{x}(p)$ , equação (130) e uma versão defasada. . . . .	94
Figura 20 – Representação das componentes estimadas pelo método AMUSE, (a), (b), (c) e (d), considerando o sinal $\mathbf{x}(p)$ , equação (130), e três versões defasadas. . . . .	95
Figura 21 – Exemplos de batimentos cardíacos do tipo (a) Normal e (b) CPV, ambos do registro 207 da base de dados MIT/BIH-ARDB. . . . .	96
Figura 22 – Componentes estimadas pelo método AMUSE para o batimento cardíaco Normal: (a) e (c); e para CPV: (b) e (d). . . . .	97
Figura 23 – Componentes estimadas pelo método AMUSE para o batimento cardíaco Normal: (a), (c), (e) e (g); e para CPV: (b), (d), (f) e (h). . . . .	98
Figura 24 – Modelo hierárquico para o método AHP, com o objetivo de obter a importância de cada máquina no comitê. . . . .	101
Figura 25 – Performance para diferentes cortes $\alpha$ de probabilidade. . . . .	106
Figura 26 – Boxplot para cada medida de performance exibida na Tabela 18 para todas as 22 subconjuntos. . . . .	119
Figura 27 – Resultados do Experimento IV, considerando ruído adicionado sobre o conjunto de treinamento e de teste. . . . .	121
Figura 28 – Resultados do Experimento IV, considerando ruído adicionado sobre o conjunto de teste apenas. . . . .	122
Figura 29 – Exemplos de batimentos cardíacos (a) Normal e (b) CPV do registro 203 da base de dados MIT/BIH-ARDB. . . . .	145
Figura 30 – Componentes estimadas pelo método AMUSE para o batimento cardíaco (a) e (b) Normal e para (c) e (d) CPV. . . . .	146
Figura 31 – Componentes estimadas pelo método AMUSE para o batimento cardíaco (a), (c), (e) e (g) Normal e para (b), (d), (f) e (h) CPV. . . . .	147

# Lista de tabelas

Tabela 1 – Características eletrofisiológicas para diferentes células cardíacas. A duração se refere à do Potencial de Ação. . . . .	22
Tabela 2 – Discriminação das derivações bipolares. . . . .	25
Tabela 3 – Discriminação das derivações unipolares aumentadas. . . . .	25
Tabela 4 – Classes de batimentos cardíacos de acordo com AAMI (1987) e Goldberger et al. (2000). . . . .	26
Tabela 5 – Exemplo de tabela de contingência. . . . .	56
Tabela 6 – Escala Fundamental de Saaty. . . . .	75
Tabela 7 – Resumo das metodologias pesquisadas para reconhecimento de CPV. . . . .	85
Tabela 8 – Discriminação das bases de dados utilizados nos experimentos. . . . .	104
Tabela 9 – Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN), para distintos cortes $\alpha$ , considerando um comitê de máquinas com voto AHP. . . . .	107
Tabela 10 – Resultados do Experimento I, considerando duas componentes. Valores em negrito são os maiores da coluna. . . . .	108
Tabela 11 – Ranque das máquinas do Experimento I utilizando o método AHP com $\kappa = 500$ na função (137). . . . .	110
Tabela 12 – Resultados do Experimento I, considerando duas componentes com dados balanceados. Valores em negrito são os maiores da coluna. . . . .	111
Tabela 13 – Ranque das máquinas cujas performances estão discriminadas na Tabela 12, considerando dados balanceados utilizando o método AHP com $\kappa = 500$ na função (137). . . . .	112
Tabela 14 – Resultados do Experimento II, considerando três componentes do AMUSE. Valores em negrito são os maiores da coluna. . . . .	113
Tabela 15 – Ranque das máquinas do Experimento II utilizando o método AHP com $\kappa = 500$ na função (137). . . . .	114
Tabela 16 – Resultados do Experimento II, considerando quatro componentes. Valores em negrito são os maiores da coluna. . . . .	115
Tabela 17 – Ranque das máquinas do Experimento II utilizando o método AHP com $\kappa = 500$ na função (137). . . . .	115
Tabela 18 – Resultados do Experimento III, considerando duas componentes do AMUSE, voto AHP com $\alpha = 0,7$ e validação cruzada com 22 subconjuntos. Valores em negrito são os maiores da coluna e $\sigma$ é o desvio padrão. . . . .	117

Tabela 19 – Registros utilizados em cada subconjunto, que geraram os resultados exibidos na Tabela 18. . . . .	118
Tabela 20 – Valores médios das SNRs das duas componentes consideradas, para os conjuntos de treinamento e teste. . . . .	120
Tabela 21 – Resultados do Experimento IV, considerando 269 amostras e duas componentes do AMUSE, e a inserção de algum desvio na detecção das ondas R. Valores em negrito são os maiores da coluna. . . . .	123
Tabela 22 – Ranque das máquinas do Experimento IV utilizando o método AHP com $\kappa = 500$ na função (137). . . . .	123
Tabela 23 – Resultados do Experimento V, considerando o método AdaBoost induzido com duas máquinas . . . . .	124
Tabela 24 – Resultados do Experimento V, considerando o método AdaBoost induzido com três máquinas. . . . .	125
Tabela 25 – Resultados do Experimento VI, considerando duas componentes do AMUSE, e as classes CPV e outras. Valores em negrito são os maiores da coluna. . .	126
Tabela 26 – Ranque das máquinas do Experimento VI utilizando o método AHP com $\kappa = 500$ na função (137). . . . .	127
Tabela 27 – Resultados do Experimento VII, considerando duas componentes sem a técnica ADL. Valores em negrito são os maiores da coluna. . . . .	128
Tabela 28 – Resultados do Experimento VII, considerando duas componentes, utilizando máquinas do tipo KNN. Valores em negrito são os maiores da coluna. . . .	129
Tabela 29 – Resultados do Experimento VII, considerando duas componentes, sem o esquema de normalização. Valores em negrito são os maiores da coluna. . . .	130
Tabela 30 – Comparação das abordagens propostas com aquelas do estado da arte. A coluna “conjuntos”, diz respeito aqueles apresentados na Tabela 8. . . . .	132
Tabela 31 – Quantidade de batimentos cardíacos para cada classe de arritmia. . . . .	144
Tabela 32 – Resultados do Experimento VII, considerando 267 atributos e quatro componentes do AMUSE e o método AdaBoost induzido com quatro máquinas . . . . .	148
Tabela 33 – Resultados considerando as bases DS3 e DS4 para treinamento e teste, respectivamente. Valores em negrito são os maiores da coluna. . . . .	149
Tabela 34 – Resultados considerando as bases DS5 e DS6 para treinamento e teste, respectivamente. Valores em negrito são os maiores da coluna. . . . .	150
Tabela 35 – Resultados considerando as bases DS7 e DS8 para treinamento e teste, respectivamente. Valores em negrito são os maiores da coluna. . . . .	150
Tabela 36 – Resultados considerando a base DS9 para treinamento e teste. Valores em negrito são os maiores da coluna. . . . .	151

# Lista de abreviaturas e siglas

AAMI	<i>Association for the Advancement of Medical Instrumentation</i>
ADL	Análise Discriminante Linear
AHP	Análise Hierárquica de Processos
bpm	batimentos-por-minuto
CPV	Contração Prematura Ventricular
ECG	Eletrocardiograma
EMG	Eletromiograma
FN	Falso Negativo
FP	Falso Positivo
IAM	Infarto Agudo do Miocárdio
KNN	K Nearest Neighbors
MIT/BIH-ARDB	MIT/BIH Arrhythmia Database
ms	milisegundos
mV	milivolts
RMS	Root Mean Square
s	segundos
SNR	Signal-to-Noise-Ratio (Relação sinal-ruído)
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

# Lista de símbolos

$\exp$	Função exponencial cuja base é o número de Euler.
$\mathbb{R}_{>0}$	Números reais positivos.
$\cdot^T$	Transposição de matriz ou vetor
$\text{tra}(\cdot)$	Traço de uma matriz
$\forall x$	Qualquer que seja $x$

# Sumário

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>16</b>
<b>I</b>	<b>REVISÃO BIBLIOGRÁFICA E DISCUSSÕES</b>	<b>19</b>
<b>2</b>	<b>PROBLEMÁTICA . . . . .</b>	<b>20</b>
<b>2.1</b>	<b>Eletrofisiologia Cardíaca . . . . .</b>	<b>20</b>
<b>2.2</b>	<b>Eletrocardiograma . . . . .</b>	<b>23</b>
<b>2.3</b>	<b>Ritmicidade Cardíaca . . . . .</b>	<b>26</b>
<b>3</b>	<b>FUNDAMENTOS . . . . .</b>	<b>33</b>
<b>3.1</b>	<b>Processamento digital de sinais e Álgebra Linear . . . . .</b>	<b>33</b>
<b>3.2</b>	<b>Classificação/Reconhecimento de Padrões . . . . .</b>	<b>38</b>
<b>3.2.1</b>	<b>Comitê de máquinas . . . . .</b>	<b>43</b>
<b>3.2.2</b>	<b>Validação Cruzada . . . . .</b>	<b>50</b>
<b>3.2.3</b>	<b>Medidas de performance . . . . .</b>	<b>55</b>
<b>3.2.4</b>	<b>Naive Bayes . . . . .</b>	<b>56</b>
<b>3.2.5</b>	<b>Análise Discriminante Linear de Fisher . . . . .</b>	<b>59</b>
<b>3.3</b>	<b>Separação Cega de Fontes . . . . .</b>	<b>61</b>
<b>3.3.1</b>	<b>Um Algoritmo para solução do problema BSS . . . . .</b>	<b>65</b>
<b>3.3.1.1</b>	<b>Identificabilidade e Ortogonalidade . . . . .</b>	<b>67</b>
<b>3.3.1.2</b>	<b>Algoritmo AMUSE . . . . .</b>	<b>71</b>
<b>3.4</b>	<b>Análise Hierárquica de Processos (AHP) . . . . .</b>	<b>73</b>
<b>4</b>	<b>ESTADO DA ARTE . . . . .</b>	<b>79</b>
<b>II</b>	<b>METODOLOGIA PROPOSTA, RESULTADOS E DIS-</b>	
	<b>CUSSÕES</b>	<b>86</b>
<b>5</b>	<b>ABORDAGENS PROPOSTAS . . . . .</b>	<b>87</b>
<b>5.1</b>	<b>Segmentação do sinal de ECG . . . . .</b>	<b>87</b>
<b>5.2</b>	<b>Extração de atributos e normalização . . . . .</b>	<b>89</b>
<b>5.3</b>	<b>Redução da Dimensionalidade . . . . .</b>	<b>99</b>
<b>5.4</b>	<b>Indução das máquinas de aprendizado e obtenção das prioridades pelo método AHP . . . . .</b>	<b>100</b>

<b>6</b>	<b>RESULTADOS E DISCUSSÕES . . . . .</b>	<b>103</b>
<b>6.1</b>	<b>Experimentos . . . . .</b>	<b>103</b>
6.1.1	Experimento I . . . . .	105
6.1.2	Experimento II . . . . .	113
6.1.3	Experimento III . . . . .	116
6.1.4	Experimento IV . . . . .	120
6.1.5	Experimento V . . . . .	124
6.1.6	Experimento VI . . . . .	125
6.1.7	Experimento VII . . . . .	128
<b>6.2</b>	<b>Comparações entre abordagens do estado da arte . . . . .</b>	<b>131</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>135</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS . . . . .</b>	<b>137</b>
<b>A</b>	<b>QUANTIDADE DE BATIMENTOS CARDÍACOS POR CLASSE</b>	<b>143</b>
<b>B</b>	<b>EXPERIMENTOS AUXILIARES DA ABORDAGEM PROPOSTA</b>	<b>145</b>
<b>C</b>	<b>EXPERIMENTO COM QUATRO COMPONENTES DO AMUSE</b>	<b>148</b>
<b>D</b>	<b>EXPERIMENTOS ADICIONAIS PARA COMPARAÇÃO COM OUTRAS ABORDAGENS . . . . .</b>	<b>149</b>

# 1 Introdução

O coração é um dos órgãos mais importantes do corpo humano. Ele é responsável por bombear sangue oxigenado vindo dos pulmões para todos os órgãos periféricos e também cabe a ele receber o sangue rico em dióxido de carbono proveniente do sistema circulatório e enviá-lo para os pulmões para oxigená-lo. Embora as primeiras pesquisas sobre o funcionamento do coração remontem a 1885, quando Willem Einthoven desenvolveu o galvanômetro de corda possibilitando assim o estudo eletrofisiológico do coração, ainda hoje pesquisas relacionadas são imprescindíveis, pois as doenças cardíacas estão entre as principais causas de óbitos no mundo. No Brasil a taxa de mortalidade devida a infarto agudo do miocárdio (IAM) é de 183,3 por 100 mil habitantes, estando entre uma das maiores do mundo (SANTOS et al., 2018).

A Contração Prematura Ventricular (CPV) é uma das arritmias que podem desencadear um IAM quando não diagnosticada e devidamente tratada. Sua ocorrência elevada em um indivíduo pode indicar: problemas hemodinâmicos ou mecanismos de reentrada do impulso elétrico provenientes das áreas infartadas. No pior caso, pode desencadear uma fibrilação ventricular ou morte súbita (LATCHAMSETTY; BOGUN, 2015).

Uma das dificuldades no diagnóstico de CPV é que sua prevalência nem sempre pode ser observada no Eletrocardiograma (ECG) simples de 12 derivações, sendo necessários exames do tipo monitoramento ambulatorial de 24 ou 48 horas, a fim de determinar a ocorrência destas arritmias. Tais exames geram uma quantidade enorme de dados para análise do especialista, por isso, ferramentas computacionais são imprescindíveis no reconhecimento da ocorrência destas arritmias.

Para resolver este problema, nesta pesquisa é proposta uma nova abordagem para reconhecimento de CPV utilizando um método de separação cega de fontes para extração de atributos, uma técnica de redução de dimensionalidade, e um comitê de máquinas de aprendizado baseadas em conceitos estatísticos tal como o Teorema de Bayes, as quais são ponderadas pela sua importância, que é determinada pelo método de Análise Hierárquica de Processos (AHP) por meio de uma função de conversão das diferenças de performance.

A abordagem proposta é principalmente fundamentada no fato de que os sinais captados pelo eletrocardiograma podem ser modelados como a sobreposição de distintos sinais fontes provenientes de regiões diferentes do músculo cardíaco. Encarando o problema desse modo, a técnica de separação cega de fontes torna-se adequada para a finalidade de obter um conjunto de atributos que sejam bons discriminadores dos sinais propagados pelo miocárdio, os quais refletem as anomalias eletrocardiográficas decorrentes de um coração doente. Mas para tornar possível o emprego das técnicas de separação de fontes de um sinal unidimensional, como o caso do sinal de ECG, é proposto um esquema que transforma o vetor (sinal) em uma matriz,

utilizando o sinal original e uma versão defasada desse.

Com o objetivo de obter um sistema que seja capaz de aprender os padrões descritos pelos atributos obtidos de cada batimento cardíaco, é proposta a implementação de máquinas de aprendizado baseadas na regra de probabilidade condicional de Bayes. Estas máquinas apresentam maior simplicidade computacional frente a outras abordagens de aprendizado de máquinas como Redes Neurais Artificiais, Árvores de Decisão, Sistemas Imunológicos Artificiais, dentre outras técnicas, pois utilizam apenas o cálculo de probabilidades e a estimação dos parâmetros de média e desvio padrão para cada atributo. Além disso, as máquinas bayesianas não tendem ao superajuste, sendo esta uma boa característica, pois máquinas superajustadas mostram excelentes resultados em ambientes controlados de teste, mas resultados inferiores em aplicações práticas.

Adicionalmente, é proposta a utilização de um comitê de máquinas, ao invés de máquinas isoladas, por meio de uma ponderação dada pela importância que estas têm nas medidas de performance, cujos pesos são obtidos pelo método AHP e uma função de conversão adequada. Esta abordagem é necessária, pois cada componente da separação de fontes representa um vetor de atributos distinto do mesmo padrão, assim, as predições individuais das máquinas devem ser combinadas a fim de produzir uma única predição.

Na validação das abordagens propostas, uma série de sete experimentos são realizados, considerando: variações na quantidade de componentes empregadas pelo método de separação cega de fontes; dados desbalanceados e balanceados por meio da supressão de amostras excedentes; simulação de ambiente ruidoso com ruído branco adicionado aos sinais de ECG e desvios na detecção das ondas R desses sinais; e modificações nos conjuntos de dados de treinamento e teste.

O restante do texto está assim organizado: no Capítulo 2 são apresentados os principais conceitos relacionados a eletrofisiologia cardíaca e discutido como algumas das principais arritmias se formam e são observadas nos sinais de ECG; no Capítulo 3 os fundamentos matemáticos utilizados para o desenvolvimento das abordagens propostas são enunciados e algumas discussões sobre suas aplicações são realizadas, tais como: conceitos gerais de processamento digital de sinais baseados na álgebra linear; separação cega de fontes; métodos e conceitos de aprendizado de máquina e análise hierárquica de processos. Algumas das metodologias mais recentes para reconhecimento de CPVs são apresentadas e discutidas no Capítulo 4, com o intuito de localizar o leitor no panorama das pesquisas relacionadas. No Capítulo 5 as abordagens propostas são anunciadas e todo o referencial teórico anteriormente apresentado é conectado a fim de produzir uma camada de conhecimentos específicos para a problemática de reconhecimento de CPV; no Capítulo 6, sete experimentos são implementados buscando obter medidas de performance das abordagens propostas, verificando sua validade e eficiência na resolução do problema de identificação de arritmias do tipo CPV; e, por fim, no Capítulo 7, as conclusões e considerações

finais são apresentadas juntamente com algumas ideias para trabalhos futuros.

# Parte I

Revisão bibliográfica e discussões

## 2 Problemática

No capítulo anterior foi brevemente mencionado o objetivo primeiro desta pesquisa, que é o reconhecimento de arritmias do tipo Contração Prematura Ventricular (CPV). Antes de definir propriamente o que caracteriza este tipo de arritmia, é necessário compreender a estrutura e funcionamento do coração, resumidamente, porque há minúcias envolvidas nos padrões arrítmicos que só podem ser entendidas portando tais saberes.

### 2.1 Eletrofisiologia Cardíaca

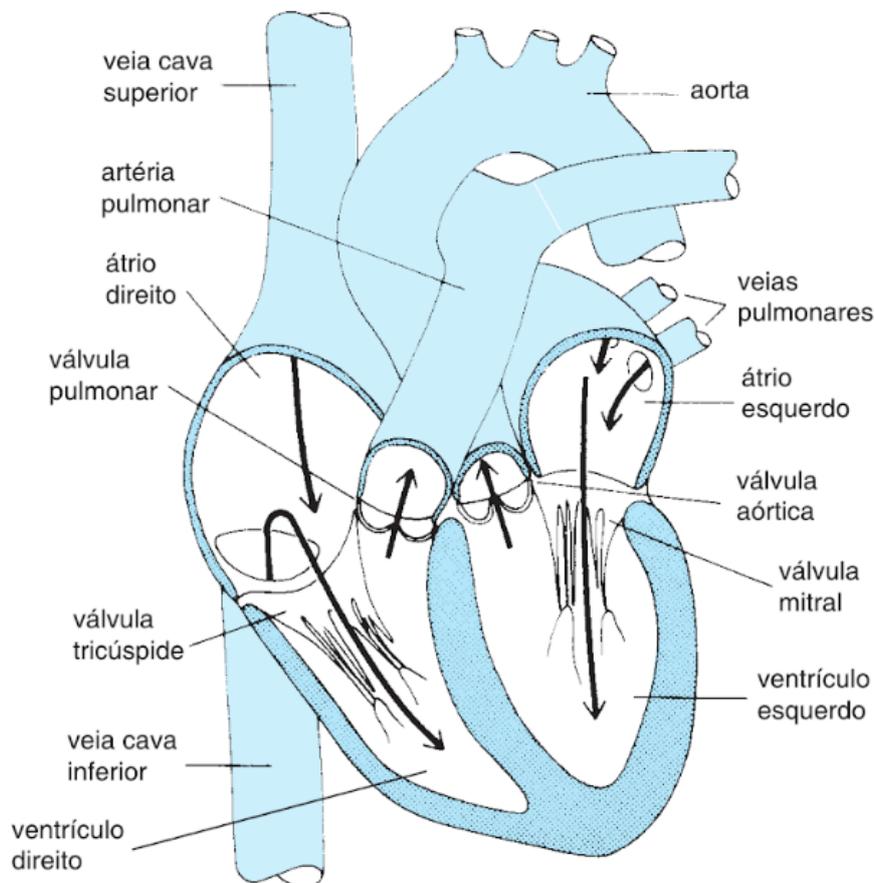
No coração há quatro câmaras, sendo dois átrios, na parte superior, e dois ventrículos, na inferior. O átrio direito recebe sangue via veias cava superior e inferior, e o esquerdo via veias pulmonares, e então bombeiam este sangue para os ventrículos direito e esquerdo, que por sua vez bombeia o sangue para a artéria pulmonar e aórtica, respectivamente, conforme ilustrado na Figura 1. A ação de bombeamento mais potente é executada pela contração dos ventrículos, enquanto os átrios são tidos como uma bomba fraca que apenas os alimenta, pois cerca de 80% do volume de sangue proveniente do sistema circulatório apenas passa pelos átrios, e os restantes 20% são bombeados (GUYTON; HALL, 2006).

Entre os átrios e os ventrículos e, estes e as artérias, existem válvulas<sup>1</sup> que somente se abrem e fecham quando há diferença de pressão. Estas válvulas somente permitem a passagem de sangue nos sentidos ilustrados na Figura 1 (MOHRMAN, 2011), que se dá do local de maior pressão para o de menor.

---

<sup>1</sup> Alguns autores, dentre eles Guyton e Hall (2006, p. 103), utilizam a expressão valva no lugar de válvula.

Figura 1 – Fisiologia do coração.



Fonte: Mohrman (2011, p. 9).

Funcionalmente, o coração pode ser separado em direito e esquerdo, sendo o primeiro responsável por bombear o sangue proveniente do sistema circulatório para os pulmões, e o segundo responsável por receber o sangue advindo dos pulmões e distribuí-lo aos órgãos sistêmicos, incluindo o próprio coração.

A atividade contrátil da bomba cardíaca depende de impulsos elétricos que normalmente se iniciam no nódulo SA (ou nodo SA), localizado na porção superior do átrio direito.

Em condições normais, um primeiro potencial de ação surge espontaneamente nas células localizadas no nódulo SA, considerado o marca-passo cardíaco, e se propaga para as células adjacentes através de sinapses, de uma maneira síncrona, possibilitando uma contração precisa das células isoladas do miocárdio (MOHRMAN, 2011). O impulso elétrico gerado atinge o Nódulo Atrioventricular (Nódulo AV ou nodo AV), onde as células de condução mais lentas retardam o impulso em cerca de 130 ms (milissegundos), possibilitando que os átrios encerrem suas contrações antes de iniciar as contrações ventriculares (GUYTON; HALL, 2006). Em seguida, o impulso elétrico chega as fibras de Purkinge, que é um sistema de células de condução rápida, formado pelo feixe de His e seus ramos esquerdo a direito, que por sua vez conduz o

impulso elétrico à todas as células ventriculares (MOHRMAN, 2011).

Na Tabela 1 estão elencados os potenciais para as células cardíacas, e discriminados os valores para cada tipo de célula, além da duração do potencial de ação.

Tabela 1 – Características eletrofisiológicas para diferentes células cardíacas. A duração se refere à do Potencial de Ação.

	Condução Rápida			Condução Lenta	
	Purkinge	Músculo Atrial	Músculo Ventricular	Nódulo SA	Nódulo AV
Potencial de repouso	-95 a -90 mV	-90 a -80 mV	-90 a -80 mV	-60 a -50 mV	-70 a -60 mV
Potencial de limiar	-70 a -60 mV			-40 a -30 mV	
Potencial de Ação	120 mV	110 a 120 mV	110 a 120 mV	60 a 70 mV	70 a 80 mV
Duração	300 a 500 ms	100 a 300 ms	200 a 300 ms	100 a 300 ms	100 a 300 ms

Fonte: Adaptado de Wood, Froelicher e Motzer (2005, p. 27).

Na Tabela 1, os valores do potencial de ação são as amplitudes que a voltagem atinge antes de entrar na fase 1, que é o início da repolarização. Notavelmente estes valores são maiores para as células de condução rápida. Já a duração do Potencial de Ação para cada célula especializada é discriminada na última linha da Tabela 1. Ela é maior para as células que formam as fibras de Purkinge, para que o impulso elétrico se propague por todas as células do músculo ventricular.

Em resumo, de acordo com Guyton e Hall (2006, p. 120), o impulso elétrico produzido no Nódulo SA chega as paredes externas dos átrios entre 30 e 90 ms e ao Nódulo AV após cerca de 30 ms, sofre um retardo em torno de 100 ms e alcança as paredes dos ventrículos mais próximas ao feixe de His e as mais distantes em aproximadamente 170 ms e 190 ms, respectivamente. Atinge, após cerca de 200 ms a 210 ms, a superfície pericárdica, que é a membrana que envolve o músculo cardíaco.

Além das células do Nódulo SA as fibras do Nódulo AV e do sistema de Purkinge também tem capacidade excitatória, logo recebem a denominação de marca-passo latente (COSTANZO, 2014). Elas emitem impulsos elétricos rítmicos, numa frequência aproximada de 40 bpm a 60 bpm (batimentos-por-minuto) e de 15 bpm a 40 bpm, respectivamente, enquanto que a do Nódulo SA é normalmente em torno de 70 bpm a 80 bpm (GUYTON; HALL, 2006). Como as células do Nódulo SA se despolarizam mais rapidamente, na fase 4 do potencial de ação, e pelo fato de seu potencial de ação ter duração menor, em condições normais estas células assumem o marca-passo cardíaco, pois seu potencial de ação está apto para ser disparado antes das demais células (COSTANZO, 2014). Porém, se a descarga elétrica do Nódulo SA não for suficiente para suplantam a descarga elétrica das demais células, ou se houver um bloqueio na condução

do impulso elétrico no Nódulo AV impedindo o início do potencial de ação das demais células devido ao impulso elétrico proveniente do Nódulo SA, então outras células cardíacas podem assumir a função de marca-passo. Quando isso ocorre diz-se que se trata de um **marca-passo ectópico** (GUYTON; HALL, 2006). Como ficará mais evidente nas próximas seções, o padrão eletrocardiográfico para o qual a presente pesquisa é direcionada se enquadra nessa categoria.

## 2.2 Eletrocardiograma

Parte da corrente elétrica produzida pelas células cardíacas, que é dissipada por meio dos tecidos adjacentes até a superfície corporal, pode ser medida utilizando eletrodos posicionados em locais específicos, os quais obtêm a diferença de potencial entre dois pontos. O sinal resultante, denominado de Eletrocardiograma (ECG), é a soma dos potenciais elétricos produzidos individualmente pelas células do miocárdio (MOHRMAN, 2011; GUYTON; HALL, 2006) e reflete as variações desses potenciais no tempo.

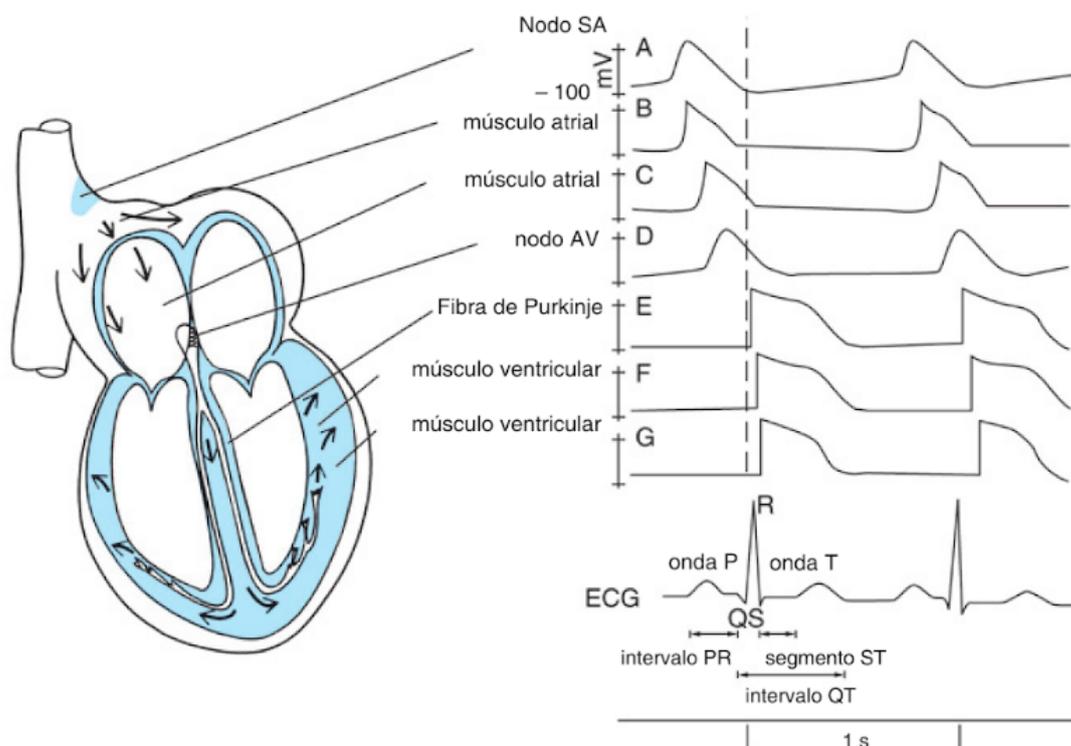
Quando as células despolarizam elas atingem um pico de voltagem antes de entrarem na fase de platô, em cerca de 110 mV a 120 mV (linha 3 da Tabela 1), que é representado no ECG pela onda R, a qual normalmente é aquela deflexão de maior amplitude do sinal de ECG. Ela compõe o complexo QRS<sup>2</sup>, formado também pelas ondas Q e S, e é uma onda com mudança abrupta de amplitude. Por outro lado, a onda P reflete a despolarização atrial, cuja amplitude máxima do potencial de ação fica em torno de 60 mV a 80 mV (linha 3 da Tabela 1), por isso essa onda tem menor amplitude no ECG. Ela também é uma onda mais suave devido a velocidade de propagação do impulso elétrico pelos átrios ser inferior àquela dos ventrículos. Do mesmo modo, a onda T tem amplitude menor, porque representa a repolarização ventricular, que reflete a volta ao repouso do potencial de ação da membrana.

Visto que apenas parte do potencial elétrico dissipado chega a superfície corporal, então os potenciais de ação exibidos na Tabela 1 são drasticamente reduzidos. Para eletrodos posicionados nos dois braços, ou em um braço e uma perna, o potencial medido está em torno de 1 mV a 1,5 mV para o complexo QRS, 0,1 a 0,3 mV para a onda P e 0,2 mV a 0,3 mV para a onda T (GUYTON; HALL, 2006).

Na Figura 2 está ilustrado, no lado inferior direito, um ECG típico da derivação II, com dois complexos QRS, representando um ciclo cardíaco completo. Na parte superior direita, está ilustrado o potencial de ação de diferentes células do músculo cardíaco, considerando o tempo de propagação a partir do impulso elétrico do marca-passo sinusal. Na parte esquerda, destaca-se o sentido de propagação do impulso elétrico associado ao potencial de ação de cada tipo de célula do miocárdio.

<sup>2</sup> Em condições anômalas ou em certas derivações o complexo QRS não é necessariamente composto das ondas Q, R e S, como por exemplo no Bloqueio do ramo esquerdo, com ausência da onda Q (PASTORE et al., 2009).

Figura 2 – Representação de um ECG típico da derivação II e os potenciais de ação de várias células do miocárdio.



Fonte: Mohrman (2011, p. 30).

Nota-se na Figura 2 que existem dois potenciais de ação para células do músculo atrial, sinais B e C. Eles refletem a diferença no tempo, resultante do deslocamento do impulso elétrico do Nódulo SA (sinal A) até um certo local no átrio direito. Também é interessante notar que o potencial de ação no sistema de Purkinje ocorre abruptamente, porque as células ventriculares se despolarizam quase que simultaneamente e são em grande número, o que explica a forma de onda peculiar do complexo QRS (MOHRMAN, 2011). Sua duração varia de 60 ms e 100 ms, em condições normais. É importante destacar que o sinal de ECG não é somente a soma dos potencial de ação de cada célula individualmente, porque o sinal dissipado é contaminado no caminho que percorre até os eletrodos, além de sofrer interferências devido às contrações musculares, por exemplo.

O intervalo discriminado no ECG como PR ou PQ, que varia de 120 ms a 200 ms, se inicia juntamente com a onda P e termina no início do complexo QRS (MOHRMAN, 2011). Ele é determinado pelo tempo que o potencial elétrico leva para atingir o Nódulo AV partindo do Nódulo SA. Após a ocorrência do complexo QRS, no sinal de ECG aparece o segmento ST, que se inicia na onda S indo até o final da onda T. Nota-se que tanto no intervalo PR como no segmento ST existem porções do sinal de ECG sem registro de voltagem significativa, as quais estão localizadas no final do intervalo e início do segmento (MOHRMAN, 2011). Estes padrões

são explicados porque as células respectivas estão na fase de platô, e portanto o potencial elétrico dissipado até a superfície não é detectado pelos eletrodos. Por último, o intervalo QT quase que corresponde ao tempo de despolarização ventricular do músculo cardíaco, durando em torno de 380 ms para uma frequência cardíaca de 60 bpm (MOHRMAN, 2011).

Uma das derivações mais convencionais empregadas no registro do ECG é a bipolar dos membros, que utiliza eletrodos posicionais nos braços e pernas (punhos e tornozelos) (MOHRMAN, 2011). De acordo com Guyton e Hall (2006, p. 129-130), estas derivações seguem o padrão discriminado na Tabela 2.

Tabela 2 – Discriminação das derivações bipolares.

<b>Derivação</b>	<b>Sigla</b>	<b>Terminal Positivo</b>	<b>Terminal Negativo</b>
I	DI	braço esquerdo	braço direito
II	DII	perna esquerdo	braço direito
III	DIII	perna esquerda	braço esquerdo

Fonte: Adaptado de Guyton e Hall (2006, p. 127-128).

Estas derivações formam o que é denominado de Triângulo de Einthoven<sup>3</sup>, onde cada derivação mede a diferença de potencial entre os vértices deste triângulo (MOHRMAN, 2011). O sinal de ECG tem aspectos distintos para cada uma destas derivações, porque o eletrocardiograma “enxerga” a atividade elétrica dissipada de forma distinta dependendo do posicionamento dos eletrodos.

Outros conjuntos de derivações comuns são as unipolares aumentadas, que se distinguem das bipolares porque nestas dois membros ficam conectadas ao terminal negativo e somente um ao positivo. A Tabela 3 discrimina estas derivações. As derivações unipolares e bipolares têm certa similaridade no registro eletrocardiográfico, exceto pela aVR, para a qual a polaridade das deflexões é sempre invertida.

Tabela 3 – Discriminação das derivações unipolares aumentadas.

<b>Sigla</b>	<b>Terminal Positivo</b>	<b>Terminal Negativo</b>
aVR	braço direito	perna esquerda e braço esquerdo
aVL	braço esquerdo	perna esquerda e braço direito
aVF	perna esquerda	braço esquerdo e direito

Fonte: Adaptado de Guyton e Hall (2006, p. 129-130).

Além destas derivações, há ainda outro conjunto denominado de torácicas ou precordiais, cujos seis terminais positivos são colocados no tórax, próximos ao coração, e os terminais

<sup>3</sup> Willem Einthoven inventou o primeiro eletrocardiograma realizando posteriormente estudos sobre a eletrofisiologia cardíaca. (MOHRMAN, 2011)

negativos nos membros. Estas derivações recebem as siglas  $V_1$ ,  $V_2$ ,  $V_3$ ,  $V_4$ ,  $V_5$  e  $V_6$ . Para um ECG normal, os complexos QRS vão apresentando maior porção positiva a medida que os índices das siglas aumentam, porque, para as derivações  $V_1$  e  $V_2$  os eletrodos estão dispostos mais próximos aos átrios, enquanto que para os demais estão próximos dos ventrículos (GUYTON; HALL, 2006).

## 2.3 Ritmicidade Cardíaca

O ritmo cardíaco é determinante no funcionamento do coração e conseqüentemente do sistema cardiovascular. As arritmias, ou seja, ritmos anormais, são principalmente causadas por: (I) ritmo não-normal do marca-passo sinusal, (II) bloqueios do impulso elétrico e (III) marca-passo ectópico (GUYTON; HALL, 2006).

Embora o objetivo primordial do presente trabalho seja o reconhecimento de CPV, é necessário compreender as demais arritmias e padrões normais para que seja possível fazer a distinção entre estas e as CPV. A AAMI (*Association for the Advancement of Medical Instrumentation*) (AAMI, 1987) estabeleceu um padrão para classificação dos diferentes batimentos cardíacos, que é amplamente utilizado nas pesquisas envolvendo eletrocardiograma e reconhecimento de arritmias. Neste trabalho serão consideradas estas classes de acordo com a Tabela 4, onde o rótulo é aquele utilizado pelas bases de dados disponibilizadas por Goldberger et al. (2000).

Tabela 4 – Classes de batimentos cardíacos de acordo com AAMI (1987) e Goldberger et al. (2000).

Classe AAMI	Batimento Cardíaco (Rótulo)
Normal	Normal (N)
	Bloqueio do ramo esquerdo (L)
	Bloqueio do ramo direito (R)
	Escape atrial (e)
	Escape atrioventricular (j)
Supraventricular Ectópico	Contração prematura atrial (A)
	Contração prematura atrial aberrante (a)
	Contração atrioventricular prematura (J)
	Contração prematura supraventricular (S)
Ventricular Ectópico	Contração prematura ventricular (V)
	Escape ventricular (E)
Fusão	Fusão de normal e ventricular (F)
Desconhecido	Ritmado (/)
	Fusão ritmado e normal (f)
	Não classificado (Q)

Fonte: AAMI (1987).

**Bloqueio de ramo:** o sistema de Purkinge é responsável por conduzir o impulso elétrico

quase que simultaneamente para ambos ventrículos via ramos direito e esquerdo do feixe de His. Quando estes ramos são bloqueados, então o músculo ventricular assume a responsabilidade pela condução do impulso elétrico, o que retarda sua propagação em cerca de um terço (GUYTON; HALL, 2006). Se o bloqueio é apenas em um dos ramos, então ocorre que o ventrículo relativo ficará polarizado cerca de 10 ms após a despolarização do outro. O efeito do bloqueio de ramo total no ECG é o alargamento do complexo QRS, cuja duração é alterada para mais que 120 ms, enquanto no bloqueio parcial essa duração permanece no intervalo de 100 a 120 ms (LOSCALZO, 2013). Dependendo da derivação considerada e do tipo do bloqueio também se observa uma onda S mais proeminente, podendo ocorrer a inversão da onda T em relação ao eixo isoeletrico (HESS, 2002).

**Escape atrial:** anteriormente foi mencionado que o Nódulo SA assume a função de marca-passo devido a sua frequência de emissão de impulsos elétricos rítmicos ser maior que a frequência das demais células cardíacas. No entanto, pode ocorrer que algum foco no átrio sobrepuje esta frequência, assumindo o marca-passo cardíaco (HAMPTON, 2014), quando alguma doença do Nódulo SA<sup>4</sup> prejudica seu potencial de ação, ou também devido à drogas ou a algum bloqueio do impulso elétrico (WOODS; FROELICHER; MOTZER, 2005), tendo como resultado o escape do batimento para outro lugar do átrio diferente do Nódulo SA.

**Escape atrioventricular:** semelhantemente ao que ocorre no escape atrial, o Nódulo AV também pode assumir a função de marca-passo cardíaco, quando há bradicardia sinusal, algum bloqueio do impulso elétrico ou mesmo a parada sinusal. Neste último caso a frequência ventricular pode ser mais rápida que a atrial (WOODS; FROELICHER; MOTZER, 2005). As alterações no ECG dependem do que ocasionou o escape atrioventricular, mas geralmente a falha do Nódulo SA é compensada pelo Nódulo AV assumindo a função de marca-passo.

Quando alguma fibra cardíaca se contrai antes do tempo normal de contração, ditado pelo Nódulo SA, então se tem uma contração prematura<sup>5</sup>, que pode ser classificadas em atrial, atrioventricular, supraventricular ou ventricular.

**Contração Prematura Atrial:** geralmente as arritmias do tipo contração prematura atrial ocorrem isoladamente, apresentando algum padrão quando coexistem com ritmos adjacentes (CATALANO, 2002). O efeito desta arritmia no sinal de ECG é uma antecipação temporal do início de suas ondas características e leves mudanças no segmento PR. Além disso, ocorre uma pausa compensatória, devido ao Nódulo SA assumir novamente a função de marca-passo (GUYTON; HALL, 2006). Neste caso, como os átrios se contraem prematuramente então o débito cardíaco sistólico ficará prejudicado porque aqueles 20% de sangue que normalmente os átrios injetariam nos ventrículos ainda não estão nas câmaras atriais (GUYTON; HALL, 2006).

<sup>4</sup> A Síndrome da Doença Sinusal se caracteriza pela bradicardia acentuada, pausas ou períodos de parada alternada sinusal (WOODS; FROELICHER; MOTZER, 2005).

<sup>5</sup> Outras denominações para a contração prematura é extrassistolia, batimento prematuro ou ectópico (GUYTON; HALL, 2006).

Alguns indivíduos podem experimentar palpitações no peito se as contrações prematuras atriais são frequentes (CATALANO, 2002). O resultado desta arritmia no ECG é um complexo QRS mais estreito que o normal (STONE; HUMPHRIES, 2013) e uma onda P deformada (GERTSCH, 2009). Por outro lado, quando ocorre uma contração prematura atrial aberrante, o complexo QRS é mais largo (STONE; HUMPHRIES, 2013), sendo esta arritmia devida a uma precocidade maior da contração prematura que impede a propagação do impulso elétrico gerado nos átrios, para os ventrículos, pelo sistema de condução especializado (EPSTEIN, 2018).

**Contração Prematura Atrioventricular:** se o batimento prematuro ocorre no Nódulo AV, esta arritmia é denominada de Contração Prematura Atrioventricular, cujo registro eletrocardiográfico é distinto dependendo da localização das células juncionais que despolarizam (NETTER, 2015). Quando ocorre no alto do Nódulo AV, as ondas P aparecem invertidas nas derivações DII e DIII e o intervalo PR é mais curto, porque a despolarização atrial percorre o sentido retrógrado, do Nódulo AV para o Nódulo SA. Quando são as células centrais do Nódulo AV que assumem o marca-passo cardíaco, então o efeito no registro do ECG é a supressão das ondas P, que são sobrepostas pelos complexos QRS, alterando ligeiramente sua forma de onda. Por fim, se a despolarização ocorre nas células inferiores, aquelas mais próximas ao feixe de His, então os ventrículos se despolarizam antes que os átrios e, portanto, o complexo QRS precede a onda P, que também aparece invertida nas derivações DII e DIII (NETTER, 2015).

**Contração Prematura Ventricular (CPV):** se o foco ectópico estiver localizado abaixo do Nódulo AV, então está caracterizada uma CPV ou Extrassístole Ventricular. Os ventrículos podem assumir a função de marca-passo pelos mesmos motivos que os átrios e também devido a reentrada, que ocorre quando um impulso elétrico gerado nos ventrículos volta a eles, seja devido a uma diminuição do período refratário de suas células ou a um alongamento da fibra ventricular que faz o impulso percorrer um caminho maior, e por conseguinte encontra as células fora do período refratário, ou ainda se a velocidade de condução do impulso elétrico diminuir (GUYTON; HALL, 2006).

As CPV nem sempre estão associadas a doenças cardíacas, podendo ser causadas pela falta de sono, irritabilidade emocional, cigarro, café e certos medicamentos (GUYTON; HALL, 2006). Sua ocorrência é considerada benigna se o paciente não apresentar nenhuma doença cardíaca estrutural (KENNEDY et al., 1985 apud AHN, 2013). Por outro lado, se sua taxa de ocorrência é alta, podem ocorrer problemas hemodinâmicos (GARCIA; MILLER, 2004), sendo ainda um indicativo da existência de algum mecanismo de reentrada decorrente de áreas infartadas ou isquêmicas, agravando o risco de fibrilação ventricular letal (GUYTON; HALL, 2006) ou mesmo morte súbita (FRED, 2009).

A prevalência das CPV foi estimada em alguns estudos realizados por Latchamsetty e Bogun (2015, p. 2-10). No estudo com 301 homens de meia-idade, em 62% foi diagnosticado alguma arritmia ventricular, incluindo a CPV. Para aqueles indivíduos que apresentaram alto risco de

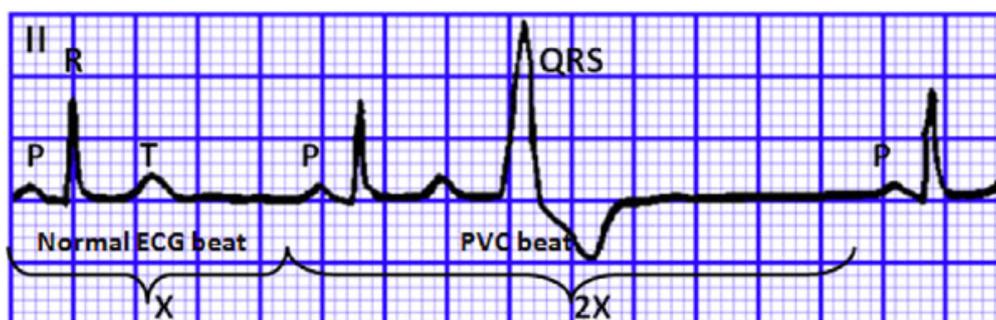
doença coronariana as CPV ocorreram com maior frequência. Em outro estudo realizado com mais de 122 mil homens das forças armadas, a maioria jovem e saudável, a ocorrência de CPV foi menor que 1%. Estima-se que a prevalência de CPV se dê entre 20% a 30% da população em geral (FRED, 2009). Quando é considerado o monitoramento ambulatorial (Holter) de 24 ou 48 horas, este percentual sobe para 50% da população (LATCHAMSETTY; BOGUN, 2015). Enquanto que para um ECG simples de 12 derivações, a ocorrência é reduzida para apenas 1% a 4% da população em geral.

Algumas das características eletrocardiográficas das CPV são (SANTOS, 2015; GARCIA; MILLER, 2004; GUYTON; HALL, 2006):

- Ausência da onda P precedendo o complexo QRS;
- Complexo QRS alargado (maior que 0,12 s), podendo apresentar aparência bizarra e com voltagem elevada;
- Complexo QRS com aspecto de Bloqueio de Ramo Direito ou Esquerdo, quando a extrasístole é proveniente do ventrículo direito ou esquerdo respectivamente;
- Polaridades da onda T e do segmento ST opostas a do complexo QRS;
- Complexo QRS prematuro.

Além destas características, quando ocorre um batimento CPV isoladamente, no ECG é observada uma pausa compensatória, que significa que o intervalo entre dois ciclos cardíacos normais, em torno de um batimento CPV, é duas vezes o comprimento de um ciclo cardíaco normal (LIU et al., 2015). A Figura 3 ilustra este evento, onde X representa o tamanho do ciclo cardíaco e a pausa compensatória ocorre após o terceiro batimento que é uma CPV.

Figura 3 – Pausa compensatória. Aqui PVC significa *Premature Ventricular Contraction*.



Fonte: Liu et al. (2015).

**Contração Prematura Supraventricular:** se o foco ectópico não é nos ventrículos, mas ocorre nos átrios ou Nódulo AV, então diz-se que este é uma Contração Prematura Supraven-

tricular. No ECG esta arritmia é manifestada tendo uma pausa não-compensatória, ou seja, o intervalo RR para este batimento e o anterior é menor que o normal devido a despolarização prematura do Nódulo SA. As ondas P podem ser encobertas pelas ondas T precedentes. Já o intervalo PR pode ser mais prolongado, dependendo da refratariedade do Nódulo AV. Em relação ao complexo QRS, este pode ser mais alargado ou estar ausente dependendo do período refratário da fibra ventricular (WOODS; FROELICHER; MOTZER, 2005).

**Escape Ventricular:** quando o foco ectópico é a nível ventricular, isto é, abaixo do Nódulo AV incluindo aí o feixe de His e a periferia do Sistema de Purkinge, trata-se então de um Escape Ventricular ou ritmo Idioventricular. Análogo ao escape atrial, o escape ventricular também pode ocorrer devido a alguma anormalidade na geração dos potenciais de ação do Nódulo SA ou Nódulo AV, quando então os ventrículos assumem a função de marca-passo, com uma frequência cardíaca de 20 a 40 bpm (AEHLERT, 2018; MAIA, 2018). Outra causa é um bloqueio total do impulso elétrico no Nódulo AV. Em ambas as circunstâncias a morfologia do complexo QRS se apresenta mais alargada que o normal, podendo ter a forma de onda de um bloqueio de ramo esquerdo ou direito, dependendo de qual ventrículo assume a ritmicidade. No entanto, mesmo na presença de um bloqueio total do Nódulo AV o complexo QRS pode apresentar uma morfologia normal, se o foco do escape estiver no início do tronco do feixe de His, uma vez que, neste caso, o percurso da despolarização segue o mesmo caminho da onda de despolarização proveniente do Nódulo SA (MAIA, 2018).

O objetivo primordial desta pesquisa é o reconhecimento das arritmias classificadas como CPV, analisando os sinais de ECG. Sua identificação entre os demais tipos de arritmias é essencial porque ajuda na determinação do correto tratamento, que pode envolver a utilização de medicamentos ou até implantes de desfibriladores caso as CPV sejam seguidas de Taquicardias Ventriculares (AHN, 2013).

Para atingir esta finalidade, é imprescindível a utilização de ferramentas matemáticas e computacionais, porque conforme já destacado anteriormente a prevalência das CPV é mais acuradamente determinada quando os sinais de ECG são obtidos via monitoramento ambulatorial utilizando Holter de 24 ou 48 horas, o que resulta em uma quantidade de informação muito extensa, dificultando ou até mesmo impossibilitando a análise de todo o volume de dados pelo especialista (cardiologista).

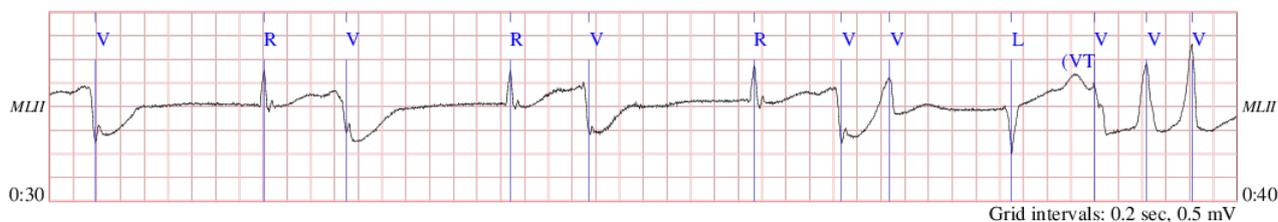
Um dos principais obstáculos para obtenção de tais ferramentas reside no fato de que, embora as CPV possuam algumas características eletrocardiográficas próprias, muitas delas são coincidentes, em certas circunstâncias, com outros tipos de batimentos cardíacos<sup>6</sup>, como por exemplo, os bloqueios de ramo direito e esquerdo e o escape ventricular. Além disso, não existe um padrão recorrente destas morfologias, já que os sinais são estocásticos e ruidosos, logo as CPV podem ser distintas até para o mesmo paciente.

---

<sup>6</sup> Entenda-se também complexos QRS.

A Figura 4 ilustra a ocorrência de 3 morfologias diferentes de CPV em apenas 10 s do sinal de ECG do paciente (registro) 207 da base de dados MIT/BIH - *Arrhythmia Database* (MIT/BIH-ARDB) (GOLDBERGER et al., 2000).

Figura 4 – Exemplos de CPV (rótulo V) distintas do registro 207 da base de dados MIT/BIH - *Arrhythmia Database*.

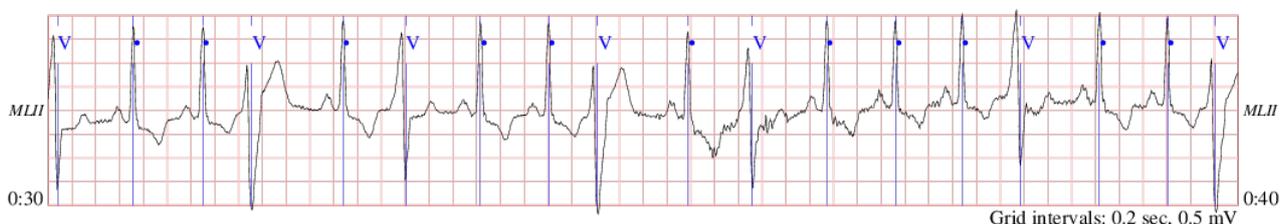


Fonte: Goldberger et al. (2000).

Nota-se da Figura 4 que os quatro primeiros batimentos CPV, referenciados pela letra V, têm morfologia semelhante entre si, assim como os dois últimos, exceto pela diferença de voltagem. Entretanto, o quinto batimento CPV apresenta uma morfologia completamente dispare dos demais. Destaca-se ainda que, neste exemplo existe a ocorrência de Taquicardia Ventricular, representada pelo símbolo “(VT)”, que é uma arritmia fatal caso não haja nenhuma intervenção. Antes desse evento, ocorrem duas CPV sucessivas, indicando a ausência de atividade atrial. Neste instante o ritmo cardíaco é de Bigeminia Ventricular, que é caracterizado por batimentos sinusais seguidos de CPV, geralmente ocasionado por algum mecanismo de reentrada do impulso elétrico ou despolarização tardia (LEWIS, 2010).

Outro exemplo de disparidade entre os padrões das CPV está exibido na Figura 5, obtida do registro 233 da base de dados MIT/BIH-ARDB.

Figura 5 – Exemplos de CPV seguido de ondas T distintas do registro 233 da base de dados MIT/BIH - *Arrhythmia Database*.



Fonte: Goldberger et al. (2000).

Observa-se na Figura 5 que o primeiro e o terceiro batimentos CPV apresentam ondas T com amplitude semelhante às ondas P, enquanto o segundo, quarto e sétimo apresentam ondas T com amplitudes muito maiores. Além disso, o quinto batimento apresenta uma anomalia

no início da onda T, provavelmente devido a algum padrão ruidoso. Este trecho está marcado como ritmo cardíaco normal.

Em linhas gerais, é notável que as morfologias podem ser distintas de um paciente para outro ou até para o mesmo paciente, variando também conforme o ritmo cardíaco. Nota-se que os padrões exibidos no registro apresentado na Figura 4 não são encontrados na Figura 5. Além disso, se observa pelas marcações verticais, que os complexos QRS de ambos registros estão com polaridades invertidas, ou seja, a onda R está abaixo da linha isoeletrica, exceto para os quatro últimos complexos QRS do registro 207 na Figura 4, e ainda, as ondas T do segundo registro são proeminentes enquanto que as do primeiro sequer aparecem.

Enfim, tais considerações evidenciam que a tarefa de reconhecer padrões de arritmias CPV é complexa e sutil, dada a inexistência de padrões fixos. Sendo assim, é necessário implementar um algoritmo que seja capaz de estabelecer alguma similaridade latente entre as CPV, além de ferramentas matemáticas apropriadas que sejam robustas às interferências que prejudiquem a interpretação destas similaridades.

## 3 Fundamentos

### 3.1 Processamento digital de sinais e Álgebra Linear

A área de processamento digital de sinais emprega muitas técnicas baseadas em transformadas, sendo as mais comuns a de Laplace, Z, Fourier, Hilbert, Cosseno e Wavelet. No estudo de filtros digitais invariavelmente as abordagens e/ou análises são baseadas na transformada de Fourier e no ferramental que a acompanha. No entanto, neste trabalho, pretende-se encarar os conceitos de filtragem e processamento de sinais do ponto de vista da álgebra matricial e nos conceitos de transformações lineares.

**Definição 1** *Um espaço vetorial real  $V$  é um conjunto não vazio munido das operações de adição e multiplicação por escalar, de modo que, dados  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$  e  $\alpha, \beta \in \mathbb{R}$ , as seguintes propriedades são satisfeitas: i)  $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$  e  $(\alpha\beta)\mathbf{u} = \alpha(\beta\mathbf{u})$ ; ii)  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ ; iii)  $\exists \mathbf{0} \in V$  tal que  $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$ ; iv)  $\exists -\mathbf{u} \in V, \forall \mathbf{u}$ , tal que  $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$ ; v)  $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$  e  $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ ; vi)  $1\mathbf{u} = \mathbf{u}$  (LIMA, 2006).*

**Definição 2** *Dados dois espaços vetoriais reais  $V$  e  $W$ , uma transformação linear  $T : V \rightarrow W$  é um mapeamento dos elementos de  $V$  em elementos de  $W$ , que satisfaz as propriedades: a)  $\forall \mathbf{v}_1, \mathbf{v}_2 \in V, T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$  e b)  $\forall k \in \mathbb{R}$  e  $\forall \mathbf{v} \in V, T(k\mathbf{v}) = kT(\mathbf{v})$ .*

Um matriz  $\mathbf{A} \in \mathbb{R}^{M \times N}$  é uma transformação linear  $T_{\mathbf{A}} : \mathbb{R}^N \rightarrow \mathbb{R}^M$ , tal que, para um vetor/matriz  $\mathbf{v} \in \mathbb{R}^{N \times K}, K > 0, T_{\mathbf{A}}(\mathbf{v}) = \mathbf{A}\mathbf{v}$  (MEYER, 2000). Assim, dada uma matriz  $\mathbf{X}$  pode-se obter uma matriz de transformação  $\mathbf{A}$  que quando aplicada a  $\mathbf{X}$  produz  $\mathbf{Y}$ , isto é:

$$\begin{aligned} \mathbf{Y} &= T_{\mathbf{A}}(\mathbf{X}) \\ &= \mathbf{A}\mathbf{X}. \end{aligned}$$

No contexto de processamento de sinais  $\mathbf{X}$  é a entrada,  $\mathbf{Y}$  a saída correspondente e  $\mathbf{A}$  é um sistema/filtro que transforma  $\mathbf{X}$  em  $\mathbf{Y}$ . A recuperação do sinal original  $\mathbf{X}$  é obtida simplesmente calculando  $\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}$ , quando  $\mathbf{A}$  é invertível.

Neste trabalho os sinais estudados são digitais, ou seja, eles passaram por um processo de conversão analógico para digital, que amostrou o sinal contínuo para uma versão discreta. Geralmente o período de amostragem é considerado por segundos. O inverso do período de amostragem,  $1/T_s$ , denominado de taxa de amostragem, é considerado em hertz (Hz), ciclos por segundo, que é a unidade de medida empregada quando se fala em frequência de sinais. Sobre este tema, diz-se que um sinal  $\mathbf{x} \in \mathbb{R}^{N \times 1}$  é de tom puro se ele é formado por uma única

componente frequencial de  $\alpha$  Hz, isto é;

$$\mathbf{x}(p) = A \cos(2\pi\alpha p + \varphi). \quad (1)$$

onde  $A$  é a amplitude do sinal,  $\varphi$  é o deslocamento de fase e  $\omega = 2\pi\alpha$  a frequência angular (WEEKS, 2012). No entanto, os sinais de tom puro são exceções nas aplicações de análise e processamento de sinais, sendo que são comumente estudados aqueles sinais que correspondem a uma sobreposição de tons puros, harmônicos, ruídos e distorções, dentre outros. Aliás, este consiste um dos fundamentos do estudo das séries de Fourier, que propõe que um sinal  $\mathbf{x}(p)$  com período  $2L$  pode ser escrito como a sobreposições de senoides (FIGUEREDO, 2005, p. 18), desde que  $\mathbf{x}(p)$  seja absolutamente integrável:

$$\mathbf{x}(p) \approx a_0 + \sum_{k=1}^{\infty} [a_k \cos(k\pi p/L) + b_k \sin(k\pi p/L)], \quad (2)$$

onde os  $a_k$  e  $b_k$  são os coeficientes de Fourier a determinar.

Uma operação corriqueira em processamento de sinais é o atraso unitário, que resulta em uma defasagem do sinal original, cuja matriz é dada por

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad (3)$$

cujos elementos da diagonal abaixo da principal são iguais a 1 e todos os demais elementos iguais a 0.

Outro tipo de matriz muito importante, em processamentos de sinais, são as matrizes ortogonais  $\mathbf{U} \in \mathbb{R}^{N \times N}$  cujas colunas ou linhas formam uma base ortonormal de  $\mathbb{R}^N$ , o que implica que se  $\mathbf{u}_j$  representa a  $j$ -ésima coluna da matriz ortogonal  $\mathbf{U}$ , então (MEYER, 2000):

$$\mathbf{u}_i^T \mathbf{u}_j = \begin{cases} 1 & , \text{ se } i = j \\ 0 & , \text{ se } i \neq j \end{cases} \Leftrightarrow \mathbf{U}^T \mathbf{U} = \mathbf{I}. \quad (4)$$

Alguns exemplos de matrizes ortogonais incluem a matriz identidade  $\mathbf{I}$ , as matrizes de rotação e reflexão e as matrizes de permutação. Para matrizes de ordem 2, por exemplo:

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}. \quad (5)$$

Aplicando a propriedade de ortogonalidade para  $\mathbf{U}$ , resulta que

$$1 = u_{11}^2 + u_{21}^2 \quad (6)$$

$$1 = u_{12}^2 + u_{22}^2 \quad (7)$$

$$0 = u_{11}u_{12} + u_{21}u_{22} \quad (8)$$

cujas soluções são

$$u_{21} = \pm \sqrt{1 - u_{11}^2} \quad (9)$$

$$u_{12} = \pm \sqrt{1 - u_{11}^2} \quad (10)$$

$$u_{22} = \begin{cases} u_{11} & , \text{ se } \operatorname{sgn}(u_{21}) \neq \operatorname{sgn}(u_{12}) \\ -u_{11} & , \text{ se } \operatorname{sgn}(u_{21}) = \operatorname{sgn}(u_{12}) \end{cases} \quad (11)$$

onde  $|u_{11}| < 1$  e  $\operatorname{sgn}(\cdot)$  retorna o sinal do número. Logo, as possibilidades para uma matriz ortogonal  $\mathbf{U}_{2 \times 2}$  são:

$$\mathbf{U} = \begin{bmatrix} u_{11} & \sqrt{1 - u_{11}^2} \\ \sqrt{1 - u_{11}^2} & -u_{11} \end{bmatrix} \text{ ou} \quad (12)$$

$$\mathbf{U} = \begin{bmatrix} u_{11} & -\sqrt{1 - u_{11}^2} \\ -\sqrt{1 - u_{11}^2} & -u_{11} \end{bmatrix} \text{ ou} \quad (13)$$

$$\mathbf{U} = \begin{bmatrix} u_{11} & -\sqrt{1 - u_{11}^2} \\ \sqrt{1 - u_{11}^2} & u_{11} \end{bmatrix} \text{ ou} \quad (14)$$

$$\mathbf{U} = \begin{bmatrix} u_{11} & \sqrt{1 - u_{11}^2} \\ -\sqrt{1 - u_{11}^2} & u_{11} \end{bmatrix}. \quad (15)$$

Fazendo  $u_{11} = 1$  e utilizando o modelo de matriz ortogonal (15), o resultado é a matriz identidade, e pelo modelo (12) fazendo  $u_{11} = 0$ , se obtém uma matriz de permutação.

Tomando o modelo (12) de matriz ortogonal, sem perda de generalidade, basta escolher um valor  $u_{11} < 1$  para obter uma matriz  $\mathbf{U}$  ortogonal. Seja  $u_{11} = 0,70$ , por exemplo. Um sinal de entrada  $\mathbf{x} = [\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2), \mathbf{x}(3)]^\top$  é transformado por  $\mathbf{U}$  em duas etapas, já que a quantidade de elementos do vetor de entrada precisa ser igual ao número de elementos da coluna de  $\mathbf{U}$ :

$$\mathbf{y}_1 = \begin{bmatrix} 0,70 & 0,71 \\ 0,71 & -0,70 \end{bmatrix} \begin{bmatrix} \mathbf{x}(0) \\ \mathbf{x}(1) \end{bmatrix} = \begin{bmatrix} 0,70\mathbf{x}(0) + 0,71\mathbf{x}(1) \\ 0,71\mathbf{x}(0) - 0,70\mathbf{x}(1) \end{bmatrix} \quad (16)$$

$$\mathbf{y}_2 = \begin{bmatrix} 0,70 & 0,71 \\ 0,71 & -0,70 \end{bmatrix} \begin{bmatrix} \mathbf{x}(2) \\ \mathbf{x}(3) \end{bmatrix} = \begin{bmatrix} 0,70\mathbf{x}(2) + 0,71\mathbf{x}(3) \\ 0,71\mathbf{x}(2) - 0,70\mathbf{x}(3) \end{bmatrix}, \quad (17)$$

logo,  $\mathbf{y} = [\mathbf{y}_1^\top, \mathbf{y}_2^\top]^\top$  consiste das somas e diferenças das componentes, consideradas aos pares, de  $\mathbf{x}$ . Para uma transformação direta, as duas etapas podem ser agregadas, escrevendo  $\mathbf{U}$  da forma:

$$\mathbf{U} = \begin{bmatrix} 0,70 & 0,71 & 0 & 0 \\ 0,71 & -0,70 & 0 & 0 \\ 0 & 0 & 0,70 & 0,71 \\ 0 & 0 & 0,71 & -0,70 \end{bmatrix}, \quad (18)$$

sendo que a transformação agora é aplicada para todo o sinal. Se mais coeficientes são implementados, mais zeros são acrescentados na matriz  $\mathbf{U}$ . Neste caso, se a quantidade de elementos do vetor de entrada for menor que a quantidade de coeficientes dos filtros, então são acrescentadas amostras nulas neste vetor até que a quantidade de elementos seja igual ao tamanho do filtro, processo denominado de *zero-padding*.

Dentre os vários métodos de decomposição matricial em Álgebra Linear, a Decomposição em Valores Singulares (DVS) utiliza-se de matrizes ortogonais para a fatoração matricial, e é apoiada nos conceitos de autovalores e autovetores.

**Definição 3** Para uma matriz  $\mathbf{A} \in \mathbb{R}^{N \times N}$ , os vetores  $\mathbf{x} \in \mathbb{R}^{N \times 1}$  e escalares  $\lambda$  que satisfazem a igualdade

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \quad (19)$$

são denominados de autovetores e autovalores, respectivamente. O conjunto  $\Lambda(\mathbf{A})$  dos distintos autovalores é denominado de espectro de  $\mathbf{A}$ , enquanto que o conjunto dos autovetores  $N(\mathbf{A} - \lambda\mathbf{I})$  é conhecido como autoespaço (MEYER, 2000).

Uma propriedade importante da decomposição espectral (ou em autovalores e autovetores), é que a transformação  $\mathbf{A}$  não altera a direção do vetor transformado. Isto implica que

$$\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{x}\|_2, \quad (20)$$

onde  $\|\mathbf{u}\|_2 = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$  é a norma euclidiana ou comprimento do vetor, e  $\langle \mathbf{u}, \mathbf{u} \rangle = u(0)u(0) + \dots + u(N-1)u(N-1)$  é o produto interno (LIMA, 2006).

**Definição 4** A DVS de uma matriz  $\mathbf{A} \in \mathbb{R}^{M \times N}$  é uma decomposição em um produto de três matrizes, duas ortogonais,  $\mathbf{L} \in \mathbb{R}^{N \times N}$  e  $\mathbf{R} \in \mathbb{R}^{M \times M}$  e uma diagonal,  $\mathbf{\Psi} \in \mathbb{R}^{M \times N}$  (ANTON, 2010):

$$\mathbf{A} = \mathbf{L}\mathbf{\Psi}\mathbf{R}^\top \quad (21)$$

onde  $\mathbf{\Psi} = \text{diag}(\psi_1, \psi_2, \dots, \psi_r, 0, \dots, 0)$ , sendo  $r$  o posto<sup>1</sup> de  $\mathbf{A}$ ;  $\mathbf{L} = [l_1 \dots l_N]$ ,  $l_i \in \mathbb{R}^N$  o  $i$ -ésimo autovetor associado aos autovalores de  $\mathbf{A}\mathbf{A}^\top$  e  $\mathbf{R} = [r_1 \dots r_M]$ ,  $r_i \in \mathbb{R}^M$  o  $i$ -ésimo autovetor associado aos autovalores de  $\mathbf{A}^\top\mathbf{A}$ . Os valores singulares  $\psi_i$  de  $\mathbf{A}$  e os autovalores  $\lambda_1, \dots, \lambda_r$  não nulos e únicos de  $\mathbf{A}\mathbf{A}^\top$  e  $\mathbf{A}^\top\mathbf{A}$ , satisfazem

$$\psi_i = \sqrt{\lambda_i}, \quad i = 1, \dots, r \quad (22)$$

sendo que  $\psi_1 \geq \psi_2 \geq \dots \geq \psi_r \geq 0$ .

<sup>1</sup> O posto de uma matriz é a quantidade de linhas não nulas, quando esta está escrita na forma escalonada por linhas.

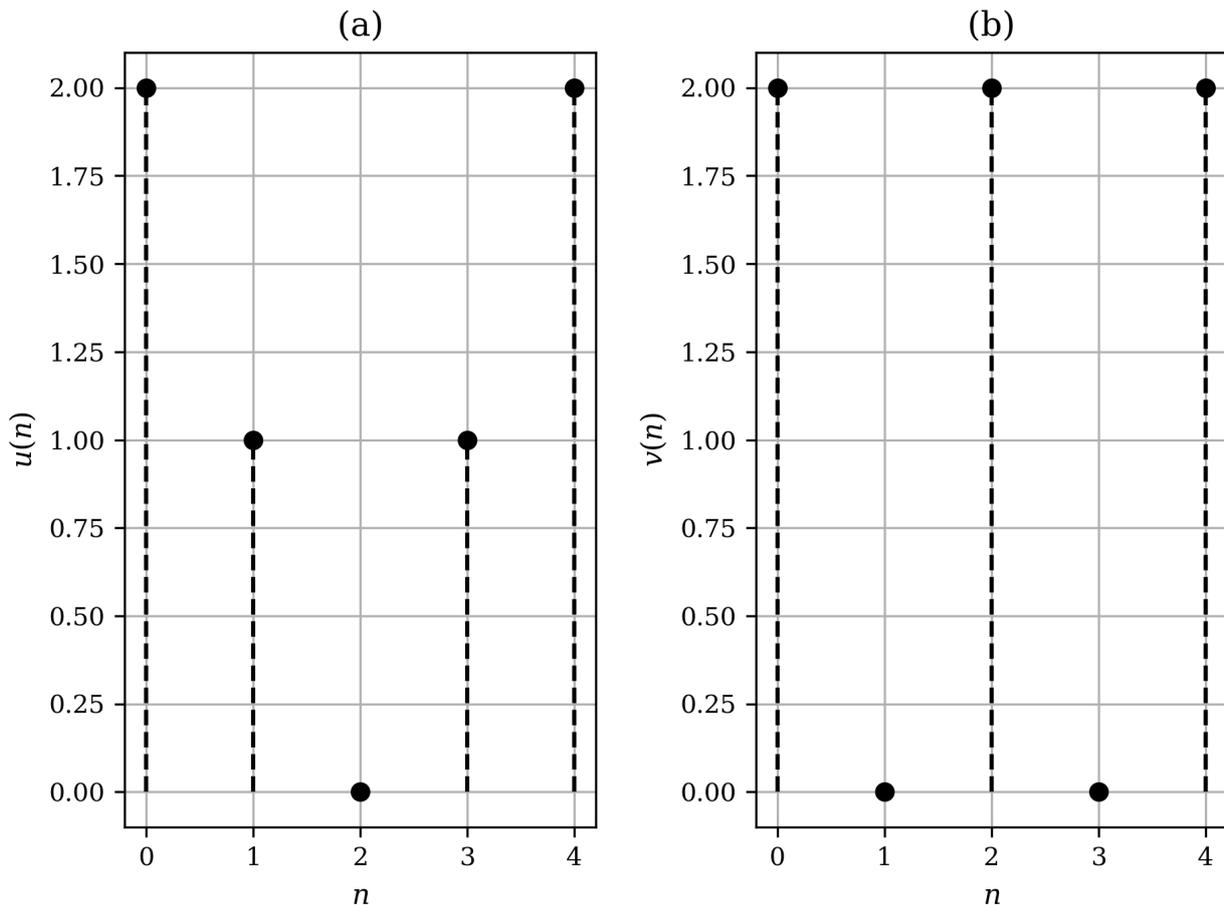
Dados dois vetores  $\mathbf{u}$  e  $\mathbf{v}$  amostrados com período  $T_{s_u}$  e  $T_{s_v}$ , respectivamente, as diferenças absolutas

$$|u(n) - u(n+1)| = \Delta_{\mathbf{u}} \quad (23)$$

$$|v(n) - v(n+1)| = \Delta_{\mathbf{v}}. \quad (24)$$

para todo  $n \in [a, b]$ , podem ser comparadas. Se  $\Delta_{\mathbf{u}}T_{s_u} < \Delta_{\mathbf{v}}T_{s_v}$ , então a diferença entre os elementos de  $\mathbf{u}$ , em um gráfico Cartesiano  $n \times u(n)$ , é mais suave do que a diferença dos elementos de  $\mathbf{v}$ , num intervalo  $[a, b]$ . Caso contrário, as mudanças de um elemento para outro são mais abruptas. Na Figura 6 estão ilustrados estes dois vetores no intervalo  $[0, 4]$ , com uma diferença suave na Figura 6 (a) com  $\Delta_{\mathbf{u}}T_{s_u}/4 = 5$  (considerando a média das diferenças) e outra mais abrupta na Figura 6 (b) com  $\Delta_{\mathbf{v}}T_{s_v}/4 = 10$ . Estes sinais  $\mathbf{u}$  e  $\mathbf{v}$  são denominados de baixa e alta frequência, respectivamente, no contexto de processamento de sinais.

Figura 6 – Exemplos de diferenças (a) suave e (b) abrupta.



Fonte: Autoria própria.

Comparando os pontos nas Figuras 6 (a) e (b), não é difícil constatar que as magnitudes dos elementos do vetor  $\mathbf{u}$ , em valor absoluto, aumentam/diminuem a metade das magnitudes

dos elementos de  $\mathbf{v}$ . Por exemplo, a diferença entre os elementos  $|u(1) - u(0)| = 1$ , enquanto que  $|v(1) - v(0)| = 2$ .

Voltando ao exemplo anterior, onde  $\mathbf{x}$  foi transformada em  $\mathbf{y}$  por meio de  $\mathbf{U}$ , equação (18), e substituindo os elementos de  $\mathbf{x}$  pelos primeiros quatro elementos de  $\mathbf{u}$ , isto é,  $\mathbf{x} = [2, 1, 0, 1]^\top$ , e aplicando a transformação  $\mathbf{U}$ , resulta em  $\mathbf{y} = [2,11; 0,72; 0,71; -0,70]^\top$ <sup>2</sup>. Como já observado, o sinal transformado  $\mathbf{y}$  consiste da soma e diferença entre os elementos de  $\mathbf{x}$ . Considerando esta distinção, pode-se separar  $\mathbf{y}$ , assim:

$$\mathbf{y}_s = [2,11; 0,71] \quad (25)$$

$$\mathbf{y}_d = [0,72; -0,70], \quad (26)$$

onde  $\mathbf{y}_s$  e  $\mathbf{y}_d$  representam a soma e a diferença, respectivamente, os quais dizem respeito a um sinal de aproximação e outro de detalhes, em mais baixa resolução (menos amostras), relativo ao sinal original. Isto porquê, as médias das amostras de  $\mathbf{x}$ , duas a duas, resultam em

$$\mathbf{x}_s = \left[ \frac{x(0) + x(1)}{2}; \frac{x(2) + x(3)}{2} \right] = [1,5; 0,5],$$

e a razão  $\mathbf{x}_s/\mathbf{y}_s \approx [0,71; 0,70]^\top$ . Logo, o sinal  $\mathbf{y}_s$  consiste de uma aproximação de  $\mathbf{x}$ , e a diferença é dada por

$$\mathbf{x}_d = [x(0) - x(1); x(2) - x(3)] = [1, -1],$$

e a razão  $\mathbf{x}_d/\mathbf{y}_d = [0,72; -0,70]^\top$ . Este tipo de transformação ilustra o espírito da Transforma Wavelet (WEEKS, 2012; NIEVERGELT, 1999). Aliás, os valores da matriz  $\mathbf{U}$  são aproximações do filtro empregado na wavelet de Haar.

Além disso, uma característica importante a ser observada, é que o sinal diferença, mede, para cada  $n$ , a variação do sinal, semelhantemente a aplicação da derivada do cálculo diferencial e integral, mas aqui para um vetor discreto. Enquanto o sinal soma, altera a amplitude do sinal apenas, mantendo, em geral, a forma de onda do sinal original. Associando esta observação àquela dos sinais com mudanças suaves e abruptas, pode-se afirmar que o sinal soma é suave e o sinal diferença é abrupto.

## 3.2 Classificação/Reconhecimento de Padrões

O conceito de máquina é bastante abrangente e foge ao escopo deste trabalho uma definição mais formal. Assim, por máquina entende-se qualquer sistema capaz de processar uma entrada produzindo uma saída.

Na teoria de aprendizado de máquina um **ambiente** é definido como o espaço de entrada  $\mathbb{X}$  a partir do qual são retiradas amostras de dados que são fornecidas as máquinas para que estas

<sup>2</sup> Aqui foi utilizado ";" no lugar de "," para separar os elementos do vetor a fim de evitar confusão.

sejam devidamente ajustadas ao ambiente, ou seja, para que elas “aprendam”. O processo de ajuste é denominado **indução** ou **treinamento**, e consiste, em linhas gerais, da apresentação de inúmeras amostras de dados ao algoritmo de aprendizagem implementado pela máquina, utilizando um conjunto de regras previamente definidas. Após a etapa de indução, a máquina é capaz de realizar previsões sobre dados estranhos ao espaço de treinamento, ou seja, dados desconhecidos (HAYKIN, 2003; KUBAT, 2015; DIETTERICH, 2000; LAMPROPOULOS; TSIHRINTZIS, 2015; BELL, 2014).

Se um esquema de aprendizado supervisionado é utilizado, as entradas fornecidas devem estar associadas a algum rótulo, que representa uma classe, o qual possibilita agrupar estas entradas em relação as suas características (KUBAT, 2015). O espaço dos rótulos é designado por  $\mathbb{Y}$ . No contexto do reconhecimento de arritmias cardíacas, uma entrada, pode ser constituída de informações estatísticas das ondas do sinal de ECG, por exemplo. A cada entrada pode estar associado um rótulo do tipo *batimento arritmico* ou *batimento normal*. Na etapa de indução estas estatísticas são calculadas a partir de vários segmentos dos sinais de ECG. A máquina ajustada ao ambiente, que modela um coração normal e arritmico, pode então prever uma nova entrada, executando assim a função de diagnosticadora, realizada na prática por um cardiologista.

Dado um vetor de entrada  $\mathbf{x}_k = [x_{1k} \ x_{2k} \ \cdots \ x_{Mk}]$ , retirado do ambiente, que consiste de uma observação do vetor aleatório  $\mathbf{X}$ , e um rótulo associado  $y_k$ , realização da variável aleatória  $Y$ , o par  $(\mathbf{x}_k, y_k)$  fornece uma instância individual para a máquina, que a utiliza para alterar seus próprios parâmetros ajustando-se ao ambiente a medida que mais instâncias são fornecidas, obtendo ao final uma hipótese  $\hat{h}(\mathbf{x}, \beta) = \hat{y}$  que aproxima a função verdadeira  $h(\mathbf{x}) = y$ , a qual modela o ambiente perfeitamente, onde  $\beta$  é um vetor de parâmetros particulares da máquina, e o rótulo  $\hat{y}$  é uma estimativa do valor real  $y$ .

Assim, quanto mais exemplos do ambiente são submetidos à máquina melhor esta se aproxima, com uma certa limitação, do ambiente a ser modelado. Tais máquinas recebem o nome de aproximadores universais (HAYKIN, 2003).

As componentes do vetor de entrada  $\mathbf{x}_k$  são também denominadas **atributos** e especificam alguma informação particular do ambiente. Comumente não é possível extrair todas as características do ambiente, por isso, alguns atributos significativos devem ser escolhidos a fim de representá-lo com uma precisão estipulada (KUBAT, 2015).

Para executar a estimação da função  $h$ , neste trabalho é considerado um tipo de aprendizado supervisionado. Este procedimento é realizado em duas etapas: 1) indução ou treinamento, onde as amostras do ambiente são utilizadas no ajuste de uma máquina obtendo um modelo matemático do ambiente; e 2) validação, onde o modelo de máquina obtido é validado, por meio de medidas de performance.

Na segunda etapa, quando é observado que a máquina não apresenta bom desempenho, a

primeira etapa pode ser reexecutada, mas considerando algumas modificações nas configurações dos conjuntos de dados ou mesmo nos algoritmos internos à máquina. Neste caso, uma terceira etapa pode ser utilizada, onde um novo conjunto de dados, não utilizado previamente, é considerado, a fim de evitar qualquer super-ajuste do modelo aos dados utilizados na indução da máquina. Os conjuntos de dados empregados em cada etapa recebem as denominações: **conjunto de treinamento**, de **validação** e de **teste**, respectivamente, sendo que não existe concordância sobre os dois últimos nomes, e por vezes eles aparecem trocados na literatura especializada (KUBAT, 2015; HAYKIN, 2003; LAMPROPOULOS; TSIHRINTZIS, 2015; BELL, 2014; CAMASTRA; VINCIARELLI, 2015).

Portanto, na etapa de indução, uma amostra do ambiente contendo  $K$  observações (instâncias) é fornecida por meio de um conjunto  $T = \{(\mathbf{x}_k, y_k)\}_{k=1}^K$ , para que a máquina seja ajustada com o intuito de aprender um mecanismo que possibilite agrupar os padrões de entrada  $\mathbf{x}_k$  nas classes rotuladas como  $y_k$ . A maneira específica por meio da qual uma máquina aprende é denominada de **algoritmo de aprendizagem**. Após a máquina ter aprendido sobre o ambiente, ela será capaz de prever, com uma certa precisão, o rótulo para um padrão desconhecido que não foi submetido a ela na etapa de indução (KUBAT, 2015).

Em problemas de classificação ou reconhecimento de padrões, do tipo binário, cada ponto  $\mathbf{x}_k$  no espaço  $K$ -dimensional representa um padrão cujo rótulo é designado por  $y_k \in \{0, 1\}^3$ , onde 1 é a classe positiva e 0, a classe negativa. Deste modo, o ajuste da máquina consiste em obter uma fronteira de decisão capaz de separar os pontos em lados opostos ou grupos distintos, onde cada lado/grupo corresponde a uma classe específica. Esta fronteira é a hipótese  $\hat{h}$  e ela determina superfícies de decisão, que são denominadas de **hiperplanos** quando a dimensão do vetor de atributos é maior que 4.

Na Figura 7 (a) está ilustrado um ambiente desconhecido representado pelos pontos de cores verde e púrpura, designados pelos valores categóricos 1 e 0, respectivamente. Nota-se que os pontos  $(x_{1k}, x_{2k}), k = 1, 2, \dots, K$  se aglomeram em diferentes regiões do espaço de duas dimensões porque eles representam padrões distintos, logo, seus atributos  $x_{1k}$  e  $x_{2k}$ , também apresentam valores díspares. Na Figura 7 (b) estão ilustradas as superfícies de decisão obtidas a partir da fronteira de decisão, linha branca, gerada por uma Regressão Logística (JAMES et al., 2017), dada pela equação

$$y = -0,10 - 1,08\mathbf{x}_1 - 3,21\mathbf{x}_2. \quad (27)$$

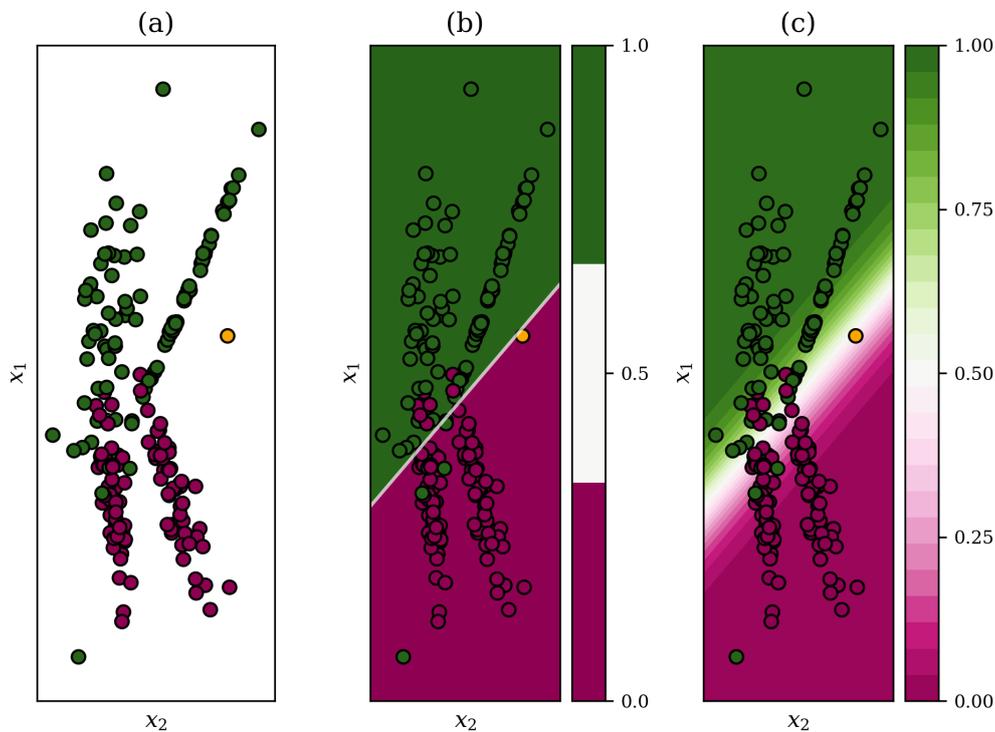
A barra de cores ao lado da Figura 7 (b) descreve as probabilidades das superfícies e da fronteira de decisão, onde a cor branca designa probabilidade de 0,5, ou seja, qualquer padrão situado sobre essa fronteira tem a mesma probabilidade de pertencer a ambas classes, logo, a decisão sobre a qual classe pertence dependerá do ponto de corte escolhido; a cor púrpura

<sup>3</sup> Qualquer outro conjunto binário também pode ser utilizado.

indica que o padrão pertence a classe 0 enquanto que a cor verde indica que ele pertence a classe 1.

Na Figura 7 (c) estão ilustradas mais fronteiras de decisão de acordo com as probabilidades de separação das classes. Observa-se que, quanto mais distante da fronteira de decisão principal, representada pela probabilidade 0,5, maior é a probabilidade do padrão pertencer a umas das classes, e se um ponto está sobre a fronteira principal então nada se pode afirmar sobre sua classe.

Figura 7 – Pontos verdes e púrpura representando os padrões do ambiente e as fronteiras de decisão obtidas por uma Regressão Logística. (a) Instâncias fornecidas a uma máquina induzida previamente; (b) principal fronteira de decisão com probabilidade 0,5; e (c) várias fronteiras de decisão com distintos valores de probabilidades, representadas de acordo com as barras de cores.



Fonte: Autoria própria.

Na Figura 7 está também representado um ponto alaranjado, cujas coordenadas são  $x_1 = 2,62$  e  $x_2 = 0,83$ . Utilizando a equação (27) de regressão o valor obtido é de aproximadamente  $y = -0,26$ . Este valor aplicado a função logística retorna os valores de probabilidades de 0,44 e 0,56 de pertencer a classe 1 e 0, respectivamente. Logo, o padrão é classificado como pertencente a classe 0, embora ele esteja bem próximo a fronteira de decisão, conforme se observa na Figura 7 (b). Nota-se ainda que alguns pontos não poderão ser corretamente classificados utilizando apenas dois atributos, como o padrão verde no canto esquerdo inferior na Figura 7.

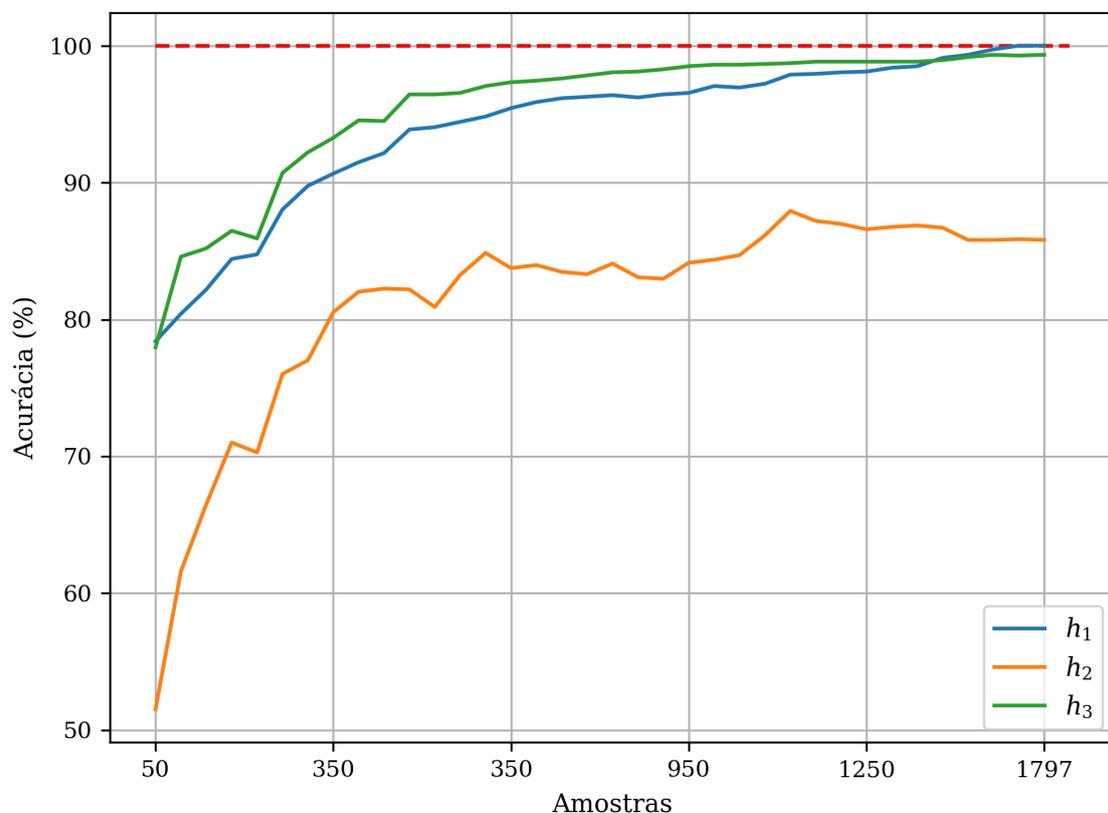
Como dito anteriormente, a estimação da função  $h$  é realizada com uma certa precisão, logo existe uma taxa de erro associada a esta estimação. Esta precisão pode ser ampliada

fornecendo a máquina mais amostras do ambiente, ou seja, aumentando seu conhecimento deste, ou refinando os atributos fornecidos como entrada, ou ainda, alterando configurações internas da máquina, tais como o algoritmo de aprendizagem utilizado, valores iniciais, dentre outros.

É esperado que diferentes máquinas gerem distintas taxas de acurácia, caso seu aprendizado tenha ocorrido utilizando amostras de dados diferentes, ou caso seus algoritmos de aprendizado converjam para soluções distintas. Isso ocorre porque o espaço de hipóteses  $\mathcal{H}$ , de onde os algoritmos de aprendizagem retiram as hipóteses  $\hat{h}$ , não é limitado, logo, para um mesmo ambiente várias máquinas podem ser induzidas (DIETTERICH, 2000).

Na Figura 8 estão ilustradas três curvas de aprendizagem de três máquinas diferentes, cujas hipóteses são  $h_1$ ,  $h_2$  e  $h_3$ . As curvas foram obtidas computando a acurácia de cada máquina utilizando diferentes quantidades de instâncias para indução, mas a mesma quantidade para validação. Estes resultados foram obtidos utilizando máquinas do tipo Regressão Logística ( $h_1$ ), *Naives* Bayes ( $h_2$ ) com função de distribuição Gaussiana e  $k$ -Vizinhos mais próximos ( $h_3$ ), com  $k = 3$ .

Figura 8 – Ilustração da curva de aprendizagem de três máquinas distintas  $h_1$ ,  $h_2$  e  $h_3$ , do tipo Regressão Logística, *Naives* Bayes com função de distribuição Gaussiana e  $k$ -Vizinhos mais próximos, com  $k = 3$ , respectivamente.



Fonte: Autoria própria.

Observa-se, na Figura 8, que a máquina com hipótese  $h_1$  atinge a acurácia máxima (linha vermelha tracejada) quando o número de amostras está por volta de 1600, enquanto que a máquina com hipótese  $h_3$  perde performance, pois com menos de 1300 amostras, aproximadamente, ela apresenta valores melhores que a máquina com hipótese  $h_1$ . Por outro lado, a máquina com hipótese  $h_2$  obteve resultados bem inferiores, independentemente da quantidade de amostras fornecidas.

O exemplo ilustrado na Figura 8 elucidada que a máquina com hipótese  $h_1$  está mais próxima da máquina ideal  $h$  no espaço de hipóteses. Esta proximidade deve ser aqui entendida como o menor erro, ou diferença entre as hipóteses, quando é conhecida uma expressão analítica dos modelos.

Naturalmente o objetivo na solução de problemas de identificação de padrões é encontrar no espaço de hipóteses aquelas que geram a maior acurácia utilizando a menor quantidade de instâncias, porque nem sempre é viável o acesso a grandes amostras de dados de uma população (KUBAT, 2015).

### 3.2.1 Comitê de máquinas

Uma técnica que visa driblar a dificuldade na procura de um classificador mais ajustado para um certo ambiente é denominada de **comitê de máquinas** ou métodos de ensemble<sup>4</sup> (DIETTERICH, 2000; HAYKIN, 2003; POLIKAR, 2006). Por meio dela, as várias máquinas induzidas para um mesmo ambiente são combinadas com o intuito de melhorar a precisão de classificação. Esta técnica é inspirada no princípio de dividir para conquistar, considerando que máquinas mais singelas podem ser agregadas para a solução de problemas difíceis, ao contrário de obter uma única máquina com alta complexidade, para solução do mesmo problema, pois esta, sozinha, demanda maior carga computacional que a utilização de várias máquinas simples (HAYKIN, 2003, p. 385).

Esta solução é análoga a resolução de um problema qualquer de engenharia, onde diversos especialistas em suas áreas agregam seus conhecimentos específicos buscando resolver um problema complexo, o qual não seria razoavelmente solúvel se os engenheiros atuassem individualmente (KUBAT, 2015). Além disso, observa-se nos resultados da aplicação de métodos de aprendizado de máquina que distintas máquinas possuem distintos conjuntos de resultados para os exemplos corretamente/incorrectamente classificados, sugerindo então que, sobre o mesmo ambiente, cada máquina aprende de modo diverso (KITTLER et al., 1998).

<sup>4</sup> Aqui a palavra ensemble é utilizada para designar, especificamente, um conjunto de máquinas de aprendizado. Outro significado, conforme dicionário Michaelis, é “Conjunto numeroso de sistemas preparados similarmente, usado para descrever uma situação do ponto de vista estatístico. A probabilidade de ocorrência de uma característica específica é determinada pela fração de elementos do ensemble que a apresentam” (WEISZFLOG, 2015).

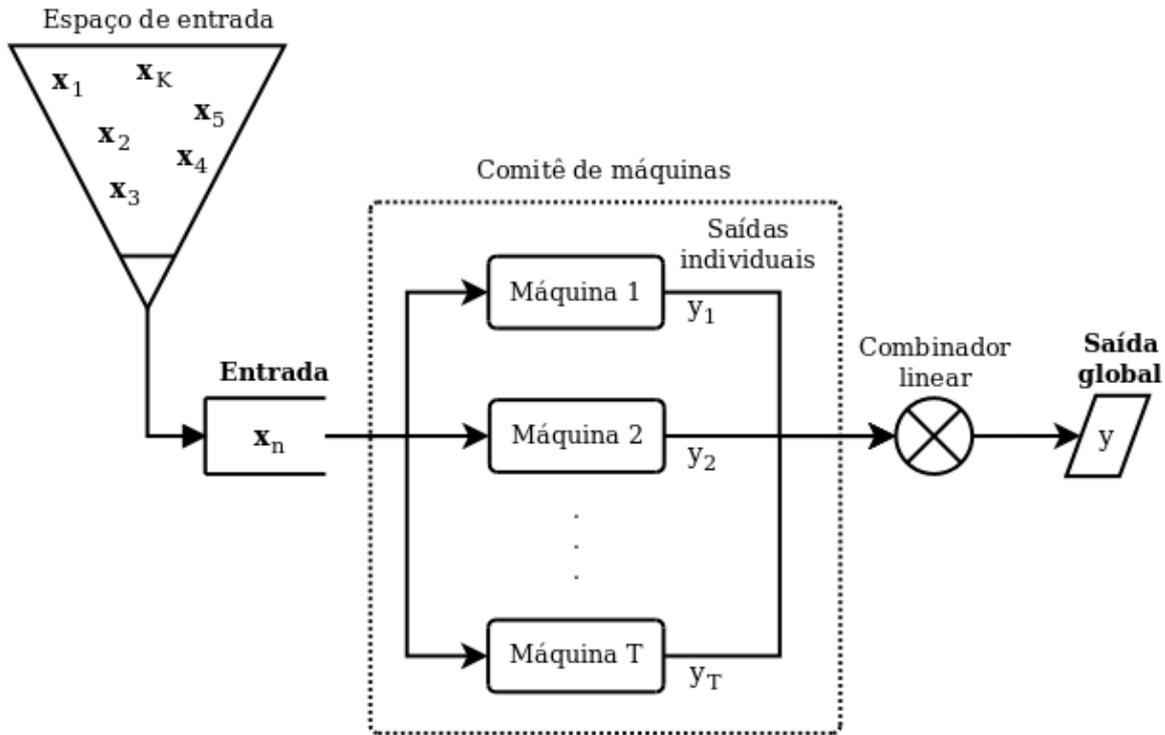
Quando o comitê de máquinas utiliza os mesmos modelos e algoritmos de indução, distinguindo apenas na sua inicialização e algumas configurações, denomina-se **comitê de máquinas homogêneo**. Caso contrário, será um comitê de máquinas não-homogêneo (KUBAT, 2015, p. 184).

Dentre as várias técnicas de comitê de máquinas, neste trabalho são descritas três delas:

1. **Média de Ensemble:** consiste em combinar linearmente a saída de várias máquinas induzidas a partir do mesmo conjunto de dados, mas variando suas condições de inicialização, para produzir uma única saída global (HAYKIN, 2003, p. 387, 391). Esta abordagem é uma variação da técnica denominada *bagging* que é um acrônimo para *bootstrap aggregation*, e consiste na reamostragem aleatória de um mesmo conjunto de dados, onde cada amostra induz uma máquina distinta (KUBAT, 2015, p. 173-174);
2. **Reforço:** é caracterizada pela agregação de máquinas individualmente induzidas com amostras de dados possuindo diferentes distribuições (HAYKIN, 2003, p. 386), de modo que haja uma complementação em relação a estes conjuntos de dados, sendo que aqueles subconjuntos com exemplos mais difíceis são relegados a outras máquinas (KUBAT, 2015, p. 176). Esta técnica também é denominada de *boosting* (KUBAT, 2015, p. 176), (HAYKIN, 2003, p. 391);
3. **Subespaços de Atributos:** tem como peculiaridade a seleção de subconjuntos de atributos ao invés de subconjuntos de exemplos de dados (HO, 1998, p. 832).

Na Figura 9 está representada a abordagem Média de Ensemble.

Figura 9 – Comitê de máquinas do tipo Média de Ensemble, considerando o espaço de entrada com rótulos desconhecidos.



Fonte: Autoria própria, adaptado de Haykin (2003, p. 387).

Observa-se na Figura 9 que a saída (rótulo) com maior frequência será predominante na escolha da saída global. Isso significa que as máquinas escolhidas para comporem o comitê, do tipo Média de Ensemble, de tal modo que elas não falhem, simultaneamente, na classificação dos mesmos exemplos (KUBAT, 2015, p. 175). Assim, existirá uma garantia de robustez, pois quando uma máquina falhar existe alguma chance das outras acertarem a predição.

Sobre o subsistema “Combinador Linear”, ele pode ser implementado de várias maneiras, a saber: (I) voto majoritário/média aritmética simples ou ponderada, que podem ser executadas diretamente sobre os rótulos ou sobre suas probabilidades (PEDREGOSA et al., 2011); (II) uma outra máquina de aprendizado (KUBAT, 2015); ou ainda (III) alguma outra técnica customizada.

A técnica voto majoritário é uma das mais empregadas nos comitês de máquinas e ela pode ser implementada nas versões rígida e suave. Seja  $\chi_q(\mathbf{x}) = y_q$  uma função que retorna o rótulo  $y_q$  predito pela  $q$ -ésima máquina referente a instância desconhecida  $\mathbf{x}$ . Para um problema binário, o número de predições das classes 0 e 1 é dado pela função característica (28) para  $Q$  máquinas induzidas:

$$v_c(\mathbf{x}) = \sum_{q=1}^Q \omega_q \begin{cases} 1, & \text{se } \chi_q(\mathbf{x}) = c \\ 0, & \text{caso contrário} \end{cases}, \quad c = 0, 1, \quad (28)$$

onde  $\omega_q$  é o peso associado à máquina  $q$ , sendo que  $\sum_{q=1}^Q \omega_q = 1$ . Desta maneira, utilizando a

função (28) a saída predita pelo comitê de máquinas para o voto majoritário rígido é dada por:

$$\hat{y} = \arg \max_c \{v_c(\mathbf{x})\}. \quad (29)$$

Por outro lado, para o voto majoritário suave, a predição do comitê de máquinas é simplesmente fornecida por:

$$\hat{y} = \arg \max_c \left\{ \frac{1}{Q} \sum_{q=1}^Q \omega_q p_q(c|\mathbf{x}) \right\}, \quad (30)$$

onde  $p_q(c|\mathbf{x})$  representa a probabilidade, fornecida pela  $q$ -ésima máquina, da instância  $\mathbf{x}$  ser reconhecida como da classe  $c$ .

Adicionalmente, Dietterich (2000, p. 2-4) fornece três razões para explicar porque um comitê de máquinas fornece resultados superiores, são elas: 1) estatística; 2) computacional; e 3) representativa.

A razão estatística leva em conta que os algoritmos de aprendizado de máquina têm por princípio a procura de uma determinada hipótese  $h$  que explique o relacionamento das instâncias de entrada com as saídas, em um espaço de hipóteses  $\mathcal{H}$ . Quando as instâncias fornecidas a máquina não são suficientes para obter uma hipótese próxima a ideal, então o algoritmo de aprendizado pode convergir para hipóteses distintas sendo que cada uma delas fornece acurácia similar. No entanto, estes resultados não significam exatamente que elas classificam na mesma classe um mesmo exemplo, mas que distintas máquinas erram e acertam aproximadamente a mesma quantidade de vezes. É neste sentido que, ao implementar um comitê de máquinas, atenua-se as predições errôneas, exceto quando a maior parte das máquinas falham.

Com algumas semelhanças à razão estatística, a computacional está associada a maneira pela qual os algoritmos de aprendizado de máquina procuram as soluções (funções) que melhor se ajustam aos dados fornecidos. Visto que o espaço de hipóteses  $\mathcal{H}$  é amplo, os algoritmos tendem a soluções locais dos problemas, ou seja, é comum que os algoritmos convirjam para mínimos (máximos) locais das funções a serem otimizadas ao invés dos pontos extremos globais<sup>5</sup>. Assim, dependendo de condições iniciais de inicialização, que mais frequentemente são aleatórias, os mesmos algoritmos podem convergir para soluções distintas, sem garantia que seja a solução ótima do espaço de hipóteses  $\mathcal{H}$ . Logo, a agregação de várias máquinas supera os resultados de uma máquina solitária, porque algumas delas podem convergir para pontos extremos locais que estão mais próximos dos globais do que outras máquinas, melhorando então as predições.

A última razão é a representativa. Muitos algoritmos de aprendizado de máquina possuem uma limitação implícita em relação ao espaço de hipóteses tangível, logo eles somente têm a capacidade de estimar funções dentro de um subconjunto do espaço de hipóteses real  $\mathcal{H}$ . Este

<sup>5</sup> Este comportamento é bastante comum em algoritmos baseados na técnica de gradiente descendente, tais como retro propagação utilizado em Redes Neurais Artificiais (HAYKIN, 2003) e em algoritmos do tipo árvore de decisão devido as regras de divisão dos nós (DIETTERICH, 2000).

é o caso daqueles algoritmos que somente podem separar classes linearmente separáveis<sup>6</sup>, tal como o Perceptron de camada única (HAYKIN, 2003). Desta maneira, ao combinar diferentes máquinas, que utilizam distintos subconjuntos do espaço de hipóteses real para suas buscas, melhora-se a capacidade de predição global.

Polikar (2006, p. 22-24) expõe os cinco principais motivos para a utilização de um comitê de máquinas, sendo que dois deles, o estatístico e a técnica dividir para conquistar já foram explanadas. Os demais motivos dizem respeito ao volume e a fonte dos dados.

Em conjuntos de dados muito volumosos o treinamento de uma única máquina é geralmente impraticável, por isso uma estratégia para abordar estes problemas é o particionamento dos dados em volumes menores, sobre os quais máquinas individuais podem ser induzidas. No problema oposto, com pequeno volume de dados, um comitê de máquinas também é uma boa estratégia, mas neste caso ele atua amenizando a falta de padrões para treinamento, porque possibilita a tomada de subconjuntos de dados com alguma sobreposição, sobre os quais a indução de várias máquinas é realizada.

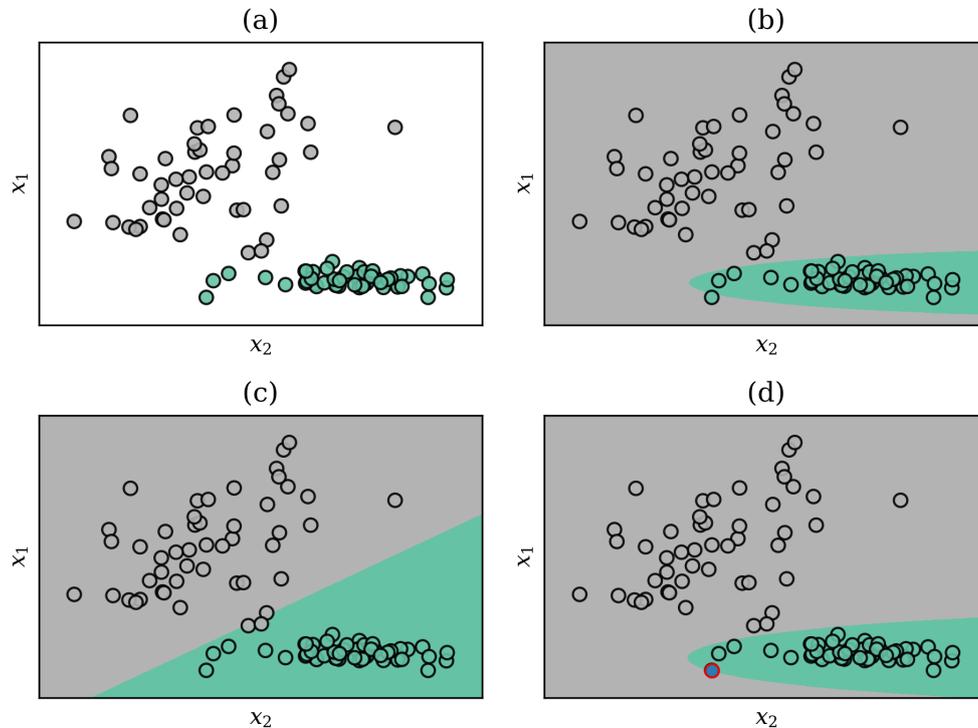
O último motivo diz respeito aos dados provenientes de diferentes fontes, cenário este que pode inviabilizar a utilização de uma única máquina porque os atributos extraídos de cada dado deve variar entre as fontes. Um exemplo fornecido por Polikar (2006, p. 23) é o diagnóstico neurológico que geralmente utiliza vários tipos de exames médicos, tal como imagens de ressonância magnética, eletroencefalograma, testes de sangue, dentre outros.

Para exemplificar a atuação de um Comitê de Máquinas do tipo Média de Ensemble, foram consideradas 100 instâncias de dados aleatórias ilustradas na Figura 10 (a), correspondendo a duas classes, verde e cinza. Selecionou-se duas máquinas de aprendizado distintas: *Naive Bayes* e Regressão Logística, cujas superfícies de decisão obtidas estão exibidas nas Figuras 10 (b) e (c), respectivamente.

---

<sup>6</sup> Um problema linearmente separável é resolvido com apenas um hiperplano separando os dados (HAYKIN, 2003).

Figura 10 – Exemplo de aplicação de um comitê de máquinas do tipo média de ensemble. (a) Instâncias para validação dos algoritmos de aprendizado; (b), (c) e (d) são as superfícies de decisão obtidas por um modelo Naive Bayes, Regressão Logística e Média de Ensemble utilizando as probabilidades de classificação com voto majoritário, respectivamente.



Fonte: Autoria própria.

Na Figura 10 (d), está ilustrada a fronteira de decisão da Média de Ensemble, utilizando a técnica de voto majoritário, calculando a média aritmética das probabilidades retornadas (PEDREGOSA et al., 2011). As acurácias obtidas pelas máquinas *Naive Bayes*, Regressão Logística e Média de Ensemble foram 0,99, 0,98 e 1,00, respectivamente. Portanto, houve uma melhora na performance de predição em torno de 0,01, devido a inclusão do exemplo, destacado em vermelho na Figura 10 (c), na borda da fronteira de decisão, o que não ocorreu na fronteira gerada pela máquina *Naive Bayes* ilustrada na Figura 10 (b).

Voltando a análise do exemplo na Figura 9, observa-se que todas as máquinas são induzidas utilizando o mesmo conjunto de dados, logo distinguem-se em relação aos parâmetros de inicialização ou ainda em relação ao paradigma <sup>7</sup> de indução. Ao contrário desta abordagem, na técnica de Reforço cada máquina é induzida com um conjunto de dados com distribuição distinta. No entanto, após a etapa de indução, as técnicas de Média de Ensemble e Reforço trabalham da mesma forma.

Uma das técnicas de comitê de máquinas mais utilizadas é denominada de AdaBoost, e

<sup>7</sup> Um paradigma de aprendizado (indução) é uma técnica/algoritmo específica que certo tipo de máquina utiliza para aprender sobre o ambiente. Uma mesma máquina, pode utilizar diferentes paradigmas.

utiliza um Reforço por Subamostragem, onde os subconjuntos de dados para indução são obtidos de acordo com uma regra que envolve uma medida de probabilidade destes subconjuntos, de tal forma que exista um maior esforço na classificação dos exemplos mais difíceis (KUBAT, 2015; HAYKIN, 2003). O algoritmo empregado pelo AdaBoost, pode ser resumido nos seguintes passos (KUBAT, 2015, p. 180-182) e (HAYKIN, 2003, p. 396):

1. Selecione  $N$  amostras do ambiente, isto é,  $T = \{(\mathbf{x}_n, y_n)\}_{n=1}^N$ ;
2. Calcule a distribuição  $p_n(\mathbf{x}_n) = 1/N$  para todo  $\mathbf{x}_n \in T$ ;
3. Faça  $r = 1, 2, \dots, R$ :
  - 3.1. E induza uma máquina  $\mathcal{M}_r$  utilizando um subconjunto  $T_r$  de  $T$  para uma dada probabilidade  $p_r$ , obtendo uma hipótese  $h_r$ ;
  - 3.2. Calcule o erro de predição  $\epsilon_r = \sum_{n=1}^N p_r(\mathbf{x}_n) e_r(\mathbf{x}_n)$  avaliando a hipótese  $h_r$  para cada  $\mathbf{x}_n \in T$ , onde

$$e_r(\mathbf{x}_j) = \begin{cases} 1, & \text{se o exemplo } \mathbf{x}_n \text{ é classificado erroneamente} \\ 0, & \text{caso contrário} \end{cases} ;$$

- 3.3. Calcule o peso  $\gamma_r = \epsilon_r / (1 - \epsilon_r)$ ;
- 3.4. Atualize a distribuição fazendo  $p_{r+1}(\mathbf{x}_n) = p_r(\mathbf{x}_n) \gamma_r$ , normalizando-a, de tal modo que  $\sum_{n=1}^N p_{r+1}(\mathbf{x}_n) = 1$ ;
4. Utilize as  $R$  máquinas induzidas  $\mathcal{M} = \{\mathcal{M}_1, \dots, \mathcal{M}_R\}$  para predição de novos padrões. Para o caso da classificação binária, os pesos das hipóteses são somados resultando nos escalares  $\Gamma_1$  e  $\Gamma_0$ , para as classes rotuladas como 1 e 0, respectivamente. Assim, a classe predita é o resultado de

$$\arg \max_{s \in \{0,1\}} \Gamma_s.$$

Pormenorizadamente, o algoritmo AdaBoost inicia todos os padrões do conjunto de dados  $T$  tendo a mesma probabilidade  $1/N$ , logo, qualquer um destes exemplos podem ser selecionados para a indução da primeira máquina na iteração inicial. Entretanto, no final desta iteração as probabilidades são modificadas, conforme passo 3.4), sendo que os exemplos classificados erroneamente têm suas probabilidades acrescidas devido ao peso  $\gamma_r$ . Assim, um novo conjunto de dados  $T_r \in T$  será selecionado, mas desta vez escolhendo os padrões cujas probabilidades são mais altas, que serão aqueles mais difíceis de classificar, uma vez que eles foram mal classificados pela máquina daquela iteração.

O procedimento se repete até que  $R$  máquinas tenham sido induzidas, sendo que cada nova máquina, se concentrará na classificação dos padrões mais difíceis. Note que, diferentemente da

técnica de Reforço por Filtragem, no AdaBoost, os mesmos exemplos podem compor diferentes conjuntos de dados, porque pode ocorrer que existam vários padrões para os quais as máquinas sempre falhem, logo estes terão probabilidades altas e serão selecionados sucessivamente para indução das demais máquinas.

A última técnica de comitê de máquinas, denominada de Subespaços de Atributos, considera a indução de várias máquinas de aprendizado utilizando distintos subpadrões  $\mathbf{a}_i \subset \mathbf{a}$ , onde  $\mathbf{a}$  é o vetor com todos os atributos extraídos do ambiente. A aplicação desta técnica tem fornecido melhorias para máquinas do tipo árvores de decisão, evitando o sobre-ajustamento dos dados e aumentando a acurácia (HO, 1998). Seu foco é obter máquinas que sejam especialistas em certas características do domínio do problema.

Pode-se extrair das abordagens de comitê de máquinas mencionados que a “pedra angular” destas é o conceito de **diversidade**, pois, se várias máquinas induzidas de maneiras distintas previrem os mesmos resultados a aplicação do comitê é inútil. Assim, visto que é improvável a indução de uma máquina que obtenha performance plena de generalização, para um comitê de máquinas devem ser selecionadas aquelas máquinas que geram erros de predição para exemplos diferentes, garantindo assim que a falha de uma será suprimida pelo acerto das outras.

### 3.2.2 Validação Cruzada

Um problema comum ao induzir uma máquina de aprendizado, independentemente da utilização de um comitê, é denominado de super-ajuste aos dados de treinamento, que ocorre quando uma máquina apresenta excelente performance para estes dados mas não possui a capacidade de generalizar o modelo aprendido para dados estranhos (BISHOP, 2006; AWAD; KHANNA, 2015; JAMES et al., 2017).

Para driblar este problema, na prática o conjunto de dados do ambiente é separado em conjuntos disjuntos de treinamento e de validação, forçando a máquina a prever padrões que não foram apresentados a ela (JAMES et al., 2017).

Em certas circunstâncias, quando o conjunto de validação é iteradamente exposto à máquina, pode-se utilizar um outro conjunto, denominado de teste, para verificar o superajuste (BISHOP, 2006, p. 32). Neste caso, o conjunto de validação é utilizado para ajustar certas configurações fixas das máquinas ou seus hiper-parâmetros, e por isso pode também ocorrer superajuste em relação a este conjunto de dados. Sendo assim, o conjunto de teste, que em momento algum foi submetido à máquina, deve ser utilizado.

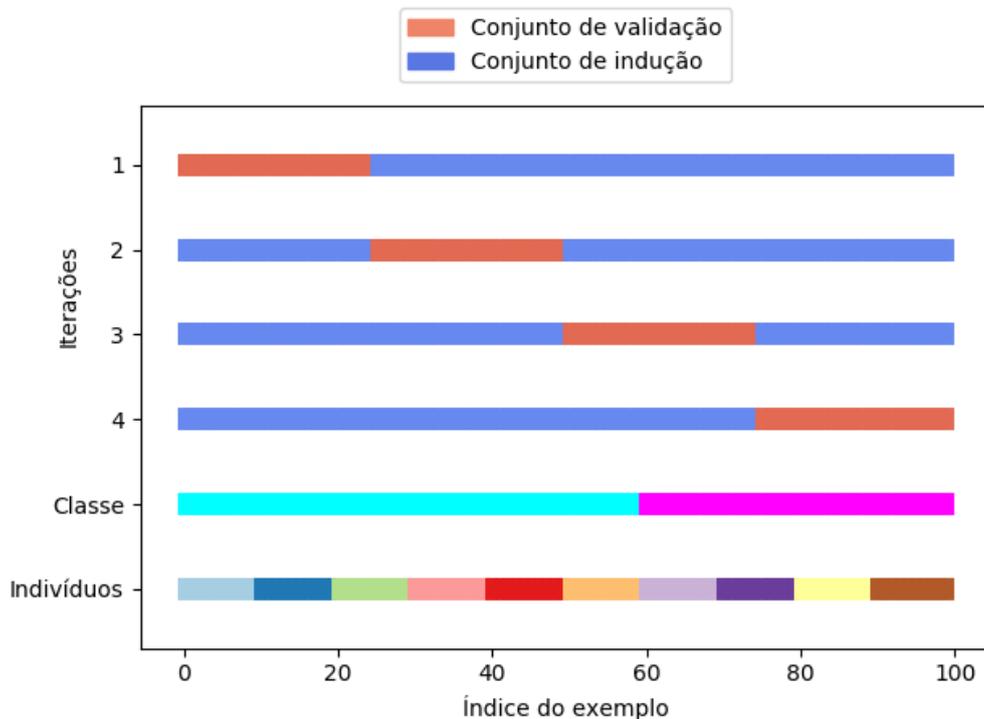
Entretanto, esta abordagem não evita que o super-ajuste aconteça, mas apenas evidencia sua ocorrência, ao gerar resultados inferiores para o conjunto de validação ou teste em relação ao conjunto de treinamento. Como o super-ajuste está associado a variância do modelo e esta aos dados de treinamento, a utilização de outros conjuntos de dados para indução das máquinas de aprendizado pode reduzir a variância.

Uma técnica concebida para driblar este problema é denominada de Validação Cruzada (*cross-validation*) com  $K$  dobras (*folds*). Nesta técnica, o conjunto de dados  $T$ , de tamanho  $N$ , simbolicamente  $\#T = N$ , é particionado em  $K$  subconjuntos disjuntos aleatórios de mesmo tamanho<sup>8</sup>, isto é,  $T_1, T_2, \dots, T_K \subset T$  com  $T_i \cap T_j = \emptyset$  para todo  $i, j = 1, 2, \dots, K$ , com  $i \neq j$ , sendo que  $\#T_i = N/K$  para todo  $i$ .

Na  $p$ -ésima iteração, o subconjunto  $T_p$  é selecionado para validação e os demais  $T_m$ , sendo  $m = 1, 2, \dots, K$ , com  $p \neq m$ , são utilizados para indução (treinamento) (AWAD; KHANNA, 2015). Este procedimento é executado de modo sequencial, ou seja, na primeira iteração  $T_1$  corresponde aos primeiros  $N/K$  elementos da amostra, enquanto  $T_2 \cup T_3 \cup \dots \cup T_K$ , corresponde aos  $(K-1)N/K$  elementos; na segunda iteração,  $T_2$  possui  $N/K$  elementos, enquanto os demais subconjuntos,  $T_1 \cup T_3 \cup \dots \cup T_K$ , possuem os  $(K-1)N/K$  elementos, e assim sucessivamente.

Na Figura 11 está ilustrado o *modus operandi* da técnica de validação cruzada, exemplificada com 4 dobras, representadas pelos quatro primeiros gráficos de cores azul e marrom claro designando os subconjuntos de indução e validação, respectivamente, enquanto que os dois últimos gráficos mostram a distribuição dos 100 exemplos considerados, nas classes e indivíduos respectivos.

Figura 11 – Exemplo da técnica de validação cruzada, com  $K = 4$ , utilizando 100 exemplos gerados aleatoriamente, pertencentes a duas classes e a 10 indivíduos.



Fonte: Autoria própria a partir do *script* disponibilizado por (PEDREGOSA et al., 2011).

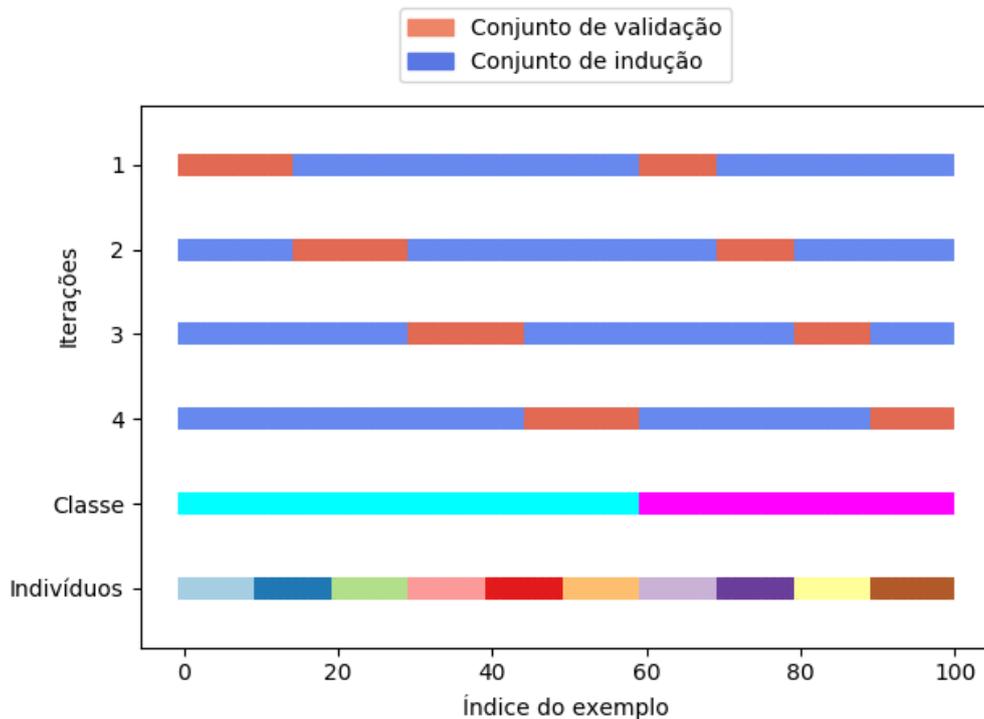
<sup>8</sup> O tamanho retrata a quantidade de exemplos do conjunto.

É importante ainda ressaltar que na prática, o valor de  $K$  na técnica de validação cruzada tem sido escolhido entre 5 e 10, sendo o segundo valor mais comum, porque existem razões empíricas que sugerem estes valores. Neste caso, 10% dos dados são utilizados para validação e 90% para indução das máquinas de aprendizado (LANTZ, 2015).

Dois inconvenientes podem ocorrer com a validação cruzada. O primeiro acontece quando os dados estão desbalanceados, isto é, quando existem mais exemplos de uma classe do que da outra. Na Figura 11, nota-se que existem mais exemplos para a classe azul que para a classe rosa. Neste cenário, as dobras selecionadas pela validação cruzada podem muito bem conter apenas exemplos de uma mesma classe, ocorrência esta que acarretaria na indução de uma máquina de aprendizado fraca, conforme se observa na quarta iteração exibida na Figura 11, para a qual a validação é realizada apenas considerando exemplos da classe rosa. Para evitar este problema, deve-se implementar uma abordagem de validação cruzada **estratificada** (PEDREGOSA et al., 2011), na qual cada subconjunto de dados  $T_p$  é particionado em outros  $s$  subconjuntos disjuntos, ou seja,  $T_{p1}, T_{p2}, \dots, T_{ps} \subset T_p$  sendo que  $T_{pi} \cap T_{pj} = \emptyset$ , para todo  $i, j = 1, 2, \dots, s$ , com  $i \neq j$ , de tal modo que a frequência da classe  $c_r$ ,  $r = 1, 2, \dots, s$ , representada por  $\natural c_r$ , seja aproximadamente a mesma em cada partição  $T_{pu}$ ,  $u = 1, 2, \dots, s$ . Visto que os dados estão desbalanceados, então deve acontecer que  $\natural c_r \neq \natural c_q$ ,  $r, q = 1, 2, \dots, s$ , com  $r \neq q$ . Por isso, são construídas  $s$  partições para os conjuntos de validação e treinamento, diferentemente da abordagem sem estratificação, isto é, na  $p$ -ésima iteração os subconjuntos  $T_p$  e  $T_1 \cup T_2 \cup \dots \cup T_{p-1} \cup T_{p-1} \cup \dots \cup T_K$ , são subdivididos em  $s$  partições.

Na Figura 12 está ilustrado o mesmo exemplo da Figura 11, mas empregando a abordagem de validação cruzada estratificada. Observa-se que na quarta iteração, diferentemente do que se notou anteriormente, a validação é realizada sobre exemplos de ambas classes. Além disso, como neste caso  $s = 2$  (número de classes), então em cada iteração os subconjuntos de validação e indução são subdivididos em dois outros subconjuntos, conforme já explicado.

Figura 12 – Exemplo da técnica de validação cruzada estratificada, com  $K = 4$ , utilizando 100 exemplos gerados aleatoriamente, pertencentes a duas classes e a 10 indivíduos.



Fonte: Autoria própria a partir do *script* disponibilizado por (PEDREGOSA et al., 2011).

É interessante também notar na Figura 12 que as partições dos subconjuntos são proporcionais a quantidade de exemplos que cada classe engloba. Em outros termos, para a classe azul os subconjuntos de validação e indução são maiores que para a classe rosa, que representam, 60% e 40% dos dados, respectivamente.

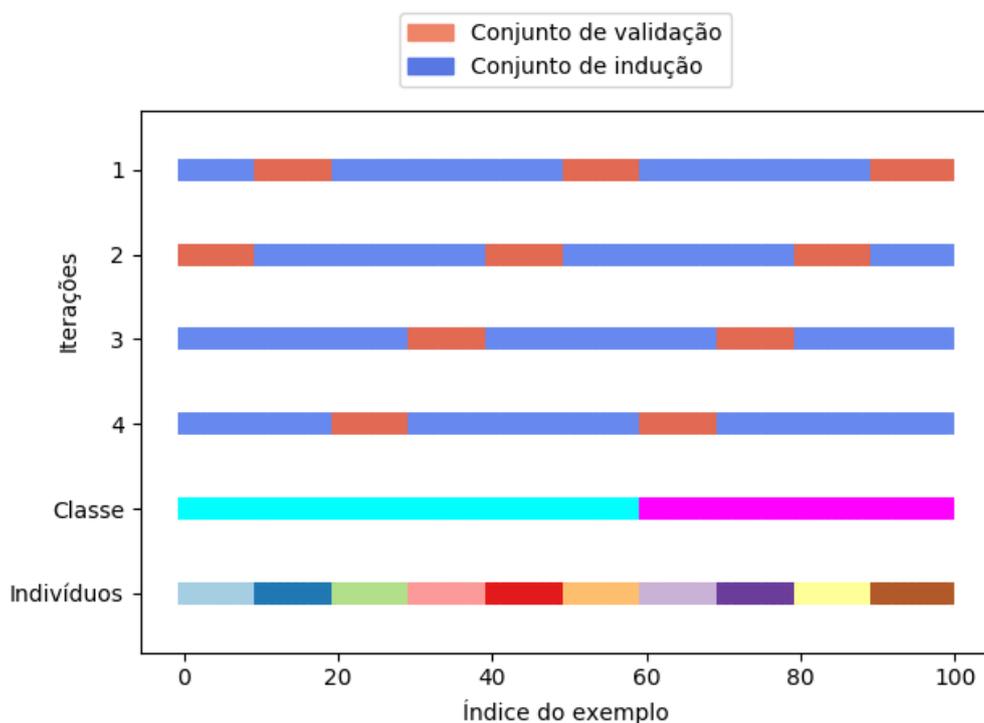
O segundo inconveniente em relação a técnica de validação cruzada simples, diz respeito aos agrupamentos individuais existentes no conjunto de dados. É comum e necessário, em banco de dados da área médica principalmente, que os dados disponibilizados sejam provenientes de vários pacientes, a fim de evitar análises enviesadas<sup>9</sup> (CHAZAL; O'DWYER; REILLY, 2004). Uma prática que tem sido evitada na comunidade de aprendizado de máquina é a utilização de dados intra-pacientes para indução e também validação, pois as máquinas tendem ao superajuste em relação aos pacientes (GARCIA et al., 2017). Em outras palavras, se uma máquina é induzida e validada com dados do mesmo paciente, ela tende a apresentar performance mais elevada. No entanto, ao expor esta máquina a conjuntos de dados desconhecidos ela falha, porque na realidade ela apenas aprendeu os padrões peculiares de um indivíduo. Por isso, é necessária a implementação da **validação cruzada agrupada**, garantindo que os dados do mesmo paciente não sejam utilizados em um dobra separada para indução e em outra para

<sup>9</sup> Neste caso, se refere ao viés estatístico.

validação (PEDREGOSA et al., 2011).

Para clarificar a validação cruzada agrupada, na Figura 13 está ilustrado o mesmo exemplo exibido nas Figuras 11 e 12, mas as dobras são implementadas tendo os indivíduos como foco. Nota-se portanto, que o conjunto de dados é particionado de tal maneira que todos os dados individuais participem da etapa de validação ao menos uma vez. O cerne desta técnica consiste em certificar que, em uma iteração particular, os dados de um mesmo indivíduo não ocorram simultaneamente nos subconjuntos de indução e validação.

Figura 13 – Exemplo da técnica de validação cruzada agrupada, com  $K = 4$ , utilizando 100 exemplos gerados aleatoriamente, pertencentes a duas classes e a 10 indivíduos.



Fonte: Autoria própria a partir do *script* disponibilizado por (PEDREGOSA et al., 2011).

Embora as técnicas de validação cruzada assim como o comitê de máquinas empreguem vários modelos de máquinas de aprendizado para obter um único resultado de predição, estas técnicas são bastantes distintas. A validação cruzada, em suas várias facetas, é utilizada principalmente para seleção de um modelo, ou ainda para obtenção dos hiper-parâmetros incluídos nestes, onde uma busca é realizada no espaço de hipóteses tentando encontrar os melhores parâmetros em termos de performance (PEDREGOSA et al., 2011).

Isto significa que esta técnica retorna qual o melhor modelo obtido ao considerar a variação dos conjuntos de treinamento e validação. Por outro lado, em um comitê de máquinas a saída de cada máquina é avaliada por um subsistema de agregação o qual decide sobre o valor predito.

Na validação cruzada, cada máquina modelada gera seus próprios resultados de performance, sobre os quais é calculado o resultado médio.

### 3.2.3 Medidas de performance

Para avaliar a performance de uma máquina de aprendizado várias medidas e abordagens têm sido propostas. Nesta seção as mais utilizadas serão elencadas de acordo com Bishop (2006, p. 601-615), Zaki e Junior (2014, p. 601-610) e Kubat (2015, p. 213-230).

Após a indução de uma máquina de aprendizado, sua avaliação pode ser realizada sobre os conjuntos de treinamento, validação e teste.

Para problemas de classificação ou reconhecimento de padrões, a performance pode ser calculada com base na quantidade de verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN), sendo que o total de exemplos considerados é dado por

$$D = VP + FP + VN + FN. \quad (31)$$

A **taxa de erro** é dada pela razão entre o total de falsas predições pelo total de exemplos, isto é,

$$E_r = \frac{FP + FN}{D}, \quad (32)$$

e a medida oposta, denominada de **acurácia**, mede a frequência de acertos sobre o total de exemplo, ou seja,

$$A_{cc} = \frac{VP + VN}{D}. \quad (33)$$

Outra medida comumente utilizada é denominada de **preditividade**, e ela mede a frequência das classificações de uma mesma classe, logo, duas medidas de preditividade são definidas, uma para a classe positiva e outra para a negativa, no caso de problemas de classificação binários:

$$P^+ = \frac{VP}{VP + FP} \text{ e} \quad (34)$$

$$P^- = \frac{VN}{VN + FN}. \quad (35)$$

Também há a medida de cobrimento (*Recall*), que retorna a probabilidade de um exemplo ser classificado corretamente. Na área médica, principalmente, esta medida recebe as denominações de **sensibilidade** (considerando os verdadeiros positivos) e **especificidade** (considerando os verdadeiros negativos), definidas conforme equações a seguir, respectivamente:

$$S_e = \frac{VP}{VP + FN} \quad (36)$$

$$S_p = \frac{VN}{VN + FP}. \quad (37)$$

Destaca-se que a distinção entre a preditividade positiva e a sensibilidade é que a primeira medida retorna a frequência de exemplos positivos sobre todos os exemplos positivos, incluindo aqueles que a máquina prediz serem mas não são, enquanto que a segunda medida retorna a frequência de exemplos positivos dentre todos os exemplos positivos do conjunto. Explicação análoga é aplicada na comparação das medidas de preditividade negativa e especificidade, mas para os exemplos negativos.

Também utiliza-se de uma combinação das medidas que são opositivas, resultando na medida de escore  $F(\varsigma)$ , onde  $\varsigma \in [0, \infty)$  é um parâmetro de ponderação para as medidas de performance envolvidas. Esta medida é expressa para cada classe como:

$$F(\varsigma)^+ = \frac{(\varsigma^2 + 1)P^+S_e}{\varsigma^2P^+ + S_e}, \tag{38}$$

$$F(\varsigma)^- = \frac{(\varsigma^2 + 1)P^-S_p}{\varsigma^2P^- + S_p}. \tag{39}$$

Outro modo para analisar simultaneamente a performance de uma máquina de aprendizado é por meio de tabelas de contingência, a qual é construída dispondo nas colunas os valores verdadeiros dos exemplos e nas linhas os valores preditos destes. A Tabela 5 ilustra um modelo de tabela de contingência para problemas de classificação binária.

Tabela 5 – Exemplo de tabela de contingência.

		Exemplos Verdadeiros		
		Positivo	Negativo	Total
Exemplos Preditos	Positivo	VP	FP	FN + VN
	Negativo	FN	VN	VP + FP
	Total	VP + FN	FP + VN	VP + FP + FN + VN

Observa-se na Tabela 5 que a coluna dos exemplos positivos é utilizada para calcular a sensibilidade enquanto a coluna dos exemplos negativos a especificidade. Já a linha dos exemplos positivos é utilizada para calcular a preditividade positiva e a linha dos exemplos negativos a preditividade negativa. Desta maneira, as tabelas de contingência fornecem um panorama da performance de classificação.

### 3.2.4 Naive Bayes

O algoritmo *Naive Bayes*, denominado aqui de Máquina Bayesiana Ingênua, é essencialmente baseado no cálculo de probabilidades e no Teorema de Bayes.

Dado um conjunto de dados  $T = \{(\mathbf{x}_k, y_k)\}_{k=1}^N$ , onde  $y_k \in \{0, 1\}$  e  $\mathbf{x}_k \in \mathbb{R}^M$ , supondo que este seja balanceado e haja  $N_1$  e  $N_0$  instâncias das classes positivas e negativas, respectivamente, ou seja,  $N_0 = N_1$  e  $N_0 + N_1 = N$ , então a probabilidade de uma instância  $\mathbf{x}_k$ , selecionada

aleatoriamente, ser da classe 0 ou 1 é de 50%. No entanto, esta informação sozinha não permite classificar um exemplo desconhecido. Entretanto, levando em consideração que as instâncias  $\mathbf{x}_k$  possuem  $M$  atributos, rotulados como  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_M$ , o conhecimento das probabilidades desses, auxilia no reconhecimento da classe de uma instância. Por exemplo, os atributos  $\mathcal{A}_m$ ,  $m = 1, 2, \dots, M$ , geralmente ocorrem em intervalos de valores restritos. Assim, as faixas de valores podem ser associadas a uma determinada classe. Logo, conhecendo-se os valores dos atributos e utilizando a probabilidade conjunta, mais informação é acrescida reduzindo a incerteza na determinação da probabilidade.

Para exemplificar, toma-se  $M = 1$ . Seja a frequência do atributo  $\mathcal{A}_1$ , para a classe positiva (1) e negativa (0), designadas por  $N_{\mathcal{A}_1|0}$  e  $N_{\mathcal{A}_1|1}$ , tal que  $N_{\mathcal{A}_1|0} \neq N_{\mathcal{A}_1|1}, \forall k$ . Consequentemente, a probabilidade de uma instância  $\mathbf{x}_k$  ser positiva ou negativa, dado o valor do atributo, é fornecida por

$$p(y_k|\mathcal{A}_1) = \frac{N_{\mathcal{A}_1|y_k}}{N_{\mathcal{A}_1}}.$$

Assim, considerando a informação adicional dada pelo atributo  $\mathcal{A}_1$ , uma instância  $\mathbf{x}_k$  pode ser atribuída a classe 0 ou 1, tomando o maior valor da probabilidade, que é obtido pela fórmula de Bayes (KUBAT, 2015, p. 24)

$$p(y_i|\mathbf{x}) = \frac{p_{\mathbf{x}}(\mathbf{x}|y_i)p(y_i)}{p(\mathbf{x})}, \quad (40)$$

onde  $p_{\mathbf{x}}(y_i|\mathbf{x})$  é a probabilidade condicional da instância  $\mathbf{x}$  pertencer a classe  $y_i$ , dados os atributos do vetor  $\mathbf{x}$ . Visto que o numerador é independente da classe, então é escolhida aquela classe que maximiza o numerador. No entanto, a verossimilhança  $p(\mathbf{x}|y_i)$  é extremamente difícil de computar devido a dimensão de  $\mathbf{x}$ , quando muitos atributos são considerados, pois o numerador da equação (40) é igual a probabilidade conjunta  $p(y_i \cap x_1 \cap x_2 \cap \dots \cap x_M)$  que é computada, utilizando a regra da cadeia, como

$$\begin{aligned} p(y_i \cap x_1 \cap \dots \cap x_M) &= p(x_1 \cap x_2 \cap \dots \cap x_M \cap y_i) \\ &= p(x_1|x_2 \cap \dots \cap x_M \cap y_i)p(x_2 \cap \dots \cap x_M \cap y_i) \\ &= p(x_1|x_2 \cap \dots \cap x_M \cap y_i)p(x_2|x_3 \cap \dots \cap x_M \cap y_i)p(x_3 \cap \dots \cap x_M \cap y_i) \\ &= \dots \\ &= p(x_1|x_2 \cap \dots \cap x_M \cap y_i)p(x_2 \cap \dots \cap x_M \cap y_i) \\ &= p(x_1|x_2 \cap \dots \cap x_M \cap y_i)p(x_2|x_3 \cap \dots \cap x_M \cap y_i) \dots p(x_{M-1}|x_M \cap y_i)p(x_M|y_i)p(y_i). \end{aligned}$$

A fim de driblar esta dificuldade, faz-se a suposição de independência estatística mútua dos atributos em relação as classes, o que implica na representação da verossimilhança como um produtório (KUBAT, 2015, p. 24):

$$p(\mathbf{x}|y_i) = \prod_{m=1}^M p(x_m|y_i) \quad (41)$$

e conseqüentemente na alteração da probabilidade condicional

$$p(y_i|\mathbf{x}) \propto p(y_i) \prod_{m=1}^M p(x_m|y_i). \quad (42)$$

onde o denominador  $p(\mathbf{x})$  da equação (40) foi suprimido pois não depende da classe.

Portanto, a classe predita é dada por

$$\hat{y}_i = \arg \max_i p(y_i|\mathbf{x}). \quad (43)$$

Ou então, fixa-se um corte de probabilidade  $\alpha$ , relativo a borda da superfície de decisão, resultando na classe predita

$$\hat{y} = \begin{cases} 0, & \text{se } p(y_i|\mathbf{x}) > \alpha \\ 1, & \text{caso contrário} \end{cases}, \quad (44)$$

para o caso de classificação binária.

Embora essa suposição de independência possa não ocorrer na maioria dos problemas reais<sup>10</sup>, sua violação não impede a aplicação da regra de Bayes dada na equação (42) (MCCALLUM; NIGAM, 1998). Nos casos onde a dependência é muito forte, em termos de correlação, deve-se então considerar alguma técnica de pré-processamento para diminuir ou excluir esta dependência (KUBAT, 2015, p. 25).

O motivo deste algoritmo funcionar bem em situações reais é que a classificação em si é uma estimativa do sinal do modelo da máquina de aprendizado e não um valor exato, como ocorre nos problemas de regressão. Em outros termos, a comparação dos valores das probabilidades resultantes possibilitam escolher um lado do hiperplano, não importando realmente o valor exato destas probabilidades.

As probabilidades condicionais anteriores podem ser calculadas considerando uma distribuição de dados específicos, sendo que as distribuições mais comumente empregadas são a Gaussiana, a Multinomial e a Bernoulli. Neste trabalho será empregado uma máquina de aprendizado com distribuição Gaussiana devido às características das instâncias utilizadas para indução. Neste caso a verossimilhança é calculada por meio da função Gaussiana:

$$p(x_m|y_i) = \frac{1}{\sqrt{2\pi\sigma_{m,y_i}^2}} \exp \left[ -\frac{(x_m - \mu_{m,y_i})^2}{2\sigma_{m,y_i}^2} \right], \quad (45)$$

onde os parâmetros de desvio padrão  $\sigma_{m,y_i}^2$  e média  $\mu_{m,y_i}$ , relativos a classe  $y_i$ , são estimados para cada atributo  $m = 1, 2, \dots, M$  para ajustar o modelo, utilizando, para tanto, os dados de treinamento (PEDREGOSA et al., 2011).

As máquinas de aprendizado Bayesianas Ingênuas possuem algumas vantagens e desvantagens em relação aos demais algoritmos de aprendizado de máquinas, as quais são listadas

<sup>10</sup> Por isso o nome do algoritmo é *Naive*, que significa ingênuo.

a seguir (PEDREGOSA et al., 2011; KUBAT, 2015; MCCALLUM; NIGAM, 1998; BISHOP, 2006):

**Vantagens** : (I) velocidade e simplicidade: tanto na etapa de indução como na validação, pois não depende de esquemas de otimização, mas apenas do cálculo de probabilidades; (II) boa performance, mesmo quando a suposição de independência não é satisfeita; (III) altamente escalável, pois sua estrutura possibilita esquemas de aprendizado *on-line* ou mesmo aprendizado em lote, sem a necessidade de retreinamento considerando todos os dados novamente; (IV) é adequada para classificação binária ou multi-classe; (V) pode ser empregada tanto na classificação de dados contínuos quanto discretos; (VI) não é sensível a escala e irrelevância dos atributos.

**Desvantagens** : (I) se a dependência dos atributos for muito forte a máquina pode apresentar um desempenho pobre; (II) devido a suposição de independência, não é adequada para considerar interações entre os atributos; (III) dados escassos podem induzir máquinas de baixa performance; (IV) se um certo atributo não ocorre no conjunto de treinamento este terá probabilidade zero, acarretando em problemas de predição.

### 3.2.5 Análise Discriminante Linear de Fisher

Analogamente ao método *Naive Bayes*, a Análise Discriminante Linear (ADL) também é empregada para classificação de padrões. Entretanto, devido a sua característica implícita de diminuição da dimensionalidade do espaço de atributos, ela será empregada na pesquisa ora apresentada com esta finalidade, antes da aplicação da classificação pelo método *Naive Bayes*.

No contexto da classificação linear, dado um conjunto de dados  $T = \{(\mathbf{x}_k, y_k)\}_{k=1}^K$ , onde  $y_k \in \{0, 1\}$  e  $\mathbf{x}_k \in \mathbb{R}^M$ , pode-se delimitar as superfícies de decisões tomando o seguinte esquema (WEBB, 2002):

$$\mathbf{w}^\top \mathbf{x}_k + w_0 \begin{cases} \geq 0, & \mathbf{x}_k \in 0 \\ < 0, & \mathbf{x}_k \in 1 \end{cases}, k = 1, 2, \dots, K, \quad (46)$$

onde  $\mathbf{w} = [w_0, w_1, w_{M-1}]$  é um vetor de pesos,  $w_0$  um limiar, e  $\mathbf{x}_k \in 1$  significa que a instância  $\mathbf{x}_k$  é rotulada na classe 1.

Em outros termos, a equação (46) pode ser encarada como a projeção do vetor de atributos de um espaço  $K$ -dimensional para um espaço unidimensional, pois

$$z_k = \mathbf{w}^\top \mathbf{x}_k + w_0, k = 1, 2, \dots, K. \quad (47)$$

Neste caso, o objetivo é encontrar um vetor  $\mathbf{w}$  que torne a separação entre as classes 0 e 1 a maior possível. Para tanto, utiliza-se de uma medida das distâncias entre as médias dos exemplos de dados apresentados. As médias vetoriais das instâncias de cada classe  $i \in \{0, 1\}$  e

suas projeções são dadas, respectivamente, por:

$$\mathbf{m}_i = \frac{1}{K_i} \sum_{\mathbf{x}_k \in i} \mathbf{x}_k \quad (48)$$

$$\tilde{\mathbf{m}}_i = \frac{1}{K_i} \sum_{z_k \in i} z_k = \frac{1}{K_i} \sum_{\mathbf{x}_k \in i} \mathbf{w}^\top \mathbf{x}_k = \mathbf{w}^\top \mathbf{m}_i, \quad (49)$$

onde  $K_i$  representa a quantidade de exemplos da classe  $i$ , e  $\mathbf{m}_i, \tilde{\mathbf{m}}_i \in \mathbb{R}^M$ .

Portanto, uma função objetivo pode incluir a distância entre as projeções como uma medida de separabilidade das classes, do seguinte modo, para o caso binário:  $J(\mathbf{w}) = |\tilde{\mathbf{m}}_0 - \tilde{\mathbf{m}}_1| = |\mathbf{w}^\top (\mathbf{m}_0 - \mathbf{m}_1)|$ . Entretanto, essa função não leva em consideração a dispersão dos dados em torno dos centroides  $\mathbf{m}_i$ , logo, pode resultar em uma medida de separação imprecisa. Considerando tal dispersão, toma-se uma medida equivalente a variância, assim definida:

$$s_i^2 = \sum_{z_k \in i} (z_k - \tilde{\mathbf{m}}_i)^2 \quad (50)$$

$$= \sum_{\mathbf{x}_k \in i} (\mathbf{w}^\top \mathbf{x}_k - \mathbf{w}^\top \mathbf{m}_i)^2 \quad (51)$$

$$= \sum_{\mathbf{x}_k \in i} \mathbf{w}^\top (\mathbf{x}_k - \mathbf{m}_i) (\mathbf{x}_k - \mathbf{m}_i)^\top \mathbf{w} \quad (52)$$

$$= \mathbf{w}^\top \mathbf{S}_i \mathbf{w}. \quad (53)$$

onde  $\mathbf{S}_i$  é a matriz de dispersão.

De maneira similar, obtêm-se a matriz de dispersão entre-classes  $\mathbf{S}_B$ :

$$(\tilde{\mathbf{m}}_0 - \tilde{\mathbf{m}}_1)^2 = (\mathbf{w}^\top \mathbf{m}_0 - \mathbf{w}^\top \mathbf{m}_1)^2 \quad (54)$$

$$= \mathbf{w}^\top (\mathbf{m}_0 - \mathbf{m}_1) (\mathbf{m}_0 - \mathbf{m}_1)^\top \mathbf{w} \quad (55)$$

$$= \mathbf{w}^\top \mathbf{S}_B \mathbf{w}, \quad (56)$$

onde  $\mathbf{S}_W = s_0^2 + s_1^2$  é a matriz de dispersão inter-classes. Por fim, a função objetivo que retorna o melhor vetor-peso  $\mathbf{w}$  é dada por (MIKA et al., 1999):

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_W \mathbf{w}}. \quad (57)$$

A maximização de  $J(\mathbf{w})$  resulta em um vetor  $\mathbf{w}$  que minimiza a variância inter-classes e maximiza a distância entre as médias projetadas (centroides) (MIKA et al., 1999).

Esta otimização é obtida tomando o gradiente da função (57) igual a zero:

$$\frac{dJ(\mathbf{w})}{d\mathbf{w}} = 0 \quad (58)$$

$$\frac{(2\mathbf{S}_B \mathbf{w}) \mathbf{w}^\top \mathbf{S}_W \mathbf{w} - (2\mathbf{S}_W \mathbf{w}) \mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{(\mathbf{w}^\top \mathbf{S}_W \mathbf{w})^2} = 0 \quad (59)$$

$$\mathbf{S}_B \mathbf{w} - \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w} (\mathbf{S}_W \mathbf{w})}{\mathbf{w}^\top \mathbf{S}_W \mathbf{w}} = 0 \quad (60)$$

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} \quad (61)$$

com  $\lambda = \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_W \mathbf{w}}$ , transformando um problema de otimização em um de autovalores e autovetores (SUGIYAMA, 2007).

Considerando que  $\mathbf{S}_W$  tenha posto completo, logo possui uma inversa  $\mathbf{S}_W^{-1}$ , então a equação (61) toma a forma:

$$\mathbf{S}_W^{-1} \mathbf{S}_B \mathbf{w} = \lambda \mathbf{w}. \quad (62)$$

No entanto, a matriz extra-classes  $\mathbf{S}_B$  tem a característica de projetar qualquer vetor  $\mathbf{v} \in \mathbb{R}^M$  na mesma direção que  $\mathbf{m}_0 - \mathbf{m}_1$ , pois

$$\mathbf{S}_B \mathbf{v} = (\mathbf{m}_0 - \mathbf{m}_1) (\mathbf{m}_0 - \mathbf{m}_1)^\top \mathbf{v} \quad (63)$$

$$= (\mathbf{m}_0 - \mathbf{m}_1) \omega \quad (64)$$

onde  $\omega = (\mathbf{m}_0 - \mathbf{m}_1)^\top \mathbf{v}$  é um escalar.

Desse modo, a equação (62) é modificada para retornar diretamente o autovetor  $\mathbf{w}$  que projeta os dados em um espaço unidimensional de tal modo que os centroides tenham a distância máxima enquanto que a dispersão inter-classe é minimizada, assim (ZHU, 2006):

$$\mathbf{w} \propto \mathbf{S}_W^{-1} (\mathbf{m}_0 - \mathbf{m}_1). \quad (65)$$

Portanto, o vetor  $\mathbf{w}$  obtido acima é utilizado para projetar um padrão  $\mathbf{x}_k$  de alta dimensão em um espaço unidimensional.

Pelo que foi discutido nos parágrafos anteriores, ficou claro que para induzir uma máquina de aprendizado, é necessário, além de um algoritmo de aprendizagem, também um conjunto de atributos que são utilizados para discriminar o ambiente a ser modelado. Na seção seguinte, será apresentado um método que pode ser empregado para a extração de atributos do ambiente.

### 3.3 Separação Cega de Fontes

Em problemas de separação cega de fontes (*Blind Source Separation* - BSS) o objetivo é estimar as fontes sem possuir informações de como estas foram misturadas e como são suas formas de onda ou conteúdo espectral, mas apenas impondo algumas suposições no intuito de possibilitar a modelagem matemática do problema.

Ao menos quatro variáveis são consideradas em problemas BSS: 1) os sinais das fontes; 2) o sistema de mistura composto pelos sensores que captam os sinais; 3) um ruído aditivo nos sensores; e 4) os sinais observados. As três primeiras variáveis são latentes ao problema, enquanto a última é aquela a que se tem acesso pleno (CICHOCKI; AMARI, 2003).

Os modelos BSS geralmente estão subdivididos em dois tipos básicos, sendo que em um deles a mistura é convolutiva e no outro instantânea. No caso instantâneo, que será abordado neste trabalho, assume-se que os sinais das fontes são captados pelos sensores assim que são

emitidos e todos ao mesmo tempo, não havendo atrasos, mesmo que as fontes em princípio estejam a distâncias diferentes dos sensores.

Um sistema BSS é formado por dois subsistemas principais: sistema de mistura e sistema de separação. O sistema de mistura geralmente é desconhecido, ou tem-se pouca informação sobre como as fontes foram sobrepostas (misturadas). Os problemas BSS procuram encontrar um filtro que seja capaz de estimar as fontes desconhecidas, ou seja, eles atuam no subsistema correspondente a separação das fontes.

Matematicamente cada fonte é representada por um vetor indexado no tempo, isto é,  $\mathbf{s}_i(p) = [s_i(1) \ s_i(2) \ \cdots \ s_i(P)]$ , sendo que as  $i$  fontes são representadas em uma matriz

$$\mathbf{S}_0(p) = \begin{bmatrix} \mathbf{s}_1(p) \\ \mathbf{s}_2(p) \\ \vdots \\ \mathbf{s}_N(p) \end{bmatrix} \quad (66)$$

$$= \begin{bmatrix} s_1(1) & s_1(2) & \cdots & s_1(P) \\ s_2(1) & s_2(2) & \cdots & s_2(P) \\ \vdots & \vdots & \ddots & \vdots \\ s_N(1) & s_N(2) & \cdots & s_N(P) \end{bmatrix}. \quad (67)$$

Analogamente, os sinais observados  $\mathbf{x}_i(p)$  também são representados em forma matricial, compondo a, assim denominada, matriz de observação:

$$\mathbf{X}(p) = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(P) \\ x_2(1) & x_2(2) & \cdots & x_2(P) \\ \vdots & \vdots & \ddots & \vdots \\ x_M(1) & x_M(2) & \cdots & x_M(P) \end{bmatrix}. \quad (68)$$

O processo ruidoso é representado por uma matriz  $\mathbf{V}(p) = (v_i(p))_{M \times P}$ , e a matriz de mistura por  $\mathbf{A} = (a_{ij})_{M \times N}$ , que é a única variável não indexada temporalmente, pois está sendo considerado um sistema de mistura instantâneo.

Com base nos fundamentos apresentados em Cichocki e Amari (2003), Tong et al. (1990), Tong et al. (1991), Amari, Cichocki e Yang (1996) e Hyvärinen et al. (2001) as definições a seguir são sintetizadas.

**Definição 5** *Um modelo BSS é uma dupla  $(\mathbf{A}_0, \mathbf{S}_0)$ , tal que*

$$\mathbf{X}(p) = \mathbf{A}_0 \mathbf{S}_0(p) + \mathbf{V}(p) \quad (69)$$

$$\mathbf{x}_m(p) = \sum_{n=1}^N a_{mn} s_n(p) + \mathbf{v}_m(p), m = 1, 2, \dots, M \quad (70)$$

e as hipóteses são válidas:

- (S1)  $\mathbf{A}_0 \in \mathbb{R}^{M \times N}$  tem posto completo;
- (S2) Os sinais em  $\mathbf{S}_0(p) \in \mathbb{R}^{N \times P}$  e  $\mathbf{V}(p) \in \mathbb{R}^{M \times P}$  são estacionários<sup>11</sup> com média nula;
- (S3)  $\mathbf{V}(p)$  é um processo ruidoso branco com matriz de covariância  $\mathbf{R}_V = \sigma^2 \mathbf{I}$ ;
- (S4)  $\mathbf{S}_0(p)$  e  $\mathbf{V}(p)$  são mutuamente estatisticamente independentes;
- (S5) Os sinais em  $\mathbf{S}_0(p)$  são descorrelacionados;
- (S6)  $M \leq N$ , ou seja, o número de sensores  $N$  é maior ou igual ao número de fontes  $M$ .

As variáveis envolvidas na equação (69) recebem as seguintes denominações:  $\mathbf{X}(p) \in \mathbb{R}^{M \times P}$  é a matriz de observação;  $\mathbf{S}_0(p)$  é a matriz das fontes;  $\mathbf{V}(p)$  a matriz de ruído e  $\mathbf{A}_0$  a matriz de mistura, ou matriz de parâmetros ou canal de comunicação.

A matriz de covariância  $\mathbf{R}_V$ , acima mencionada, em relação a uma matriz  $\mathbf{X} \in \mathbb{R}^{M \times P}$ , é assim definida:

$$\mathbf{R}_X = \begin{bmatrix} r_1^2 & r_{12} & r_{13} & \cdots & r_{1P} \\ r_{21} & r_2^2 & r_{23} & \cdots & r_{2P} \\ r_{31} & r_{32} & r_3^2 & \cdots & r_{3P} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & r_P^2 \end{bmatrix}, \quad (71)$$

onde  $r_j^2 = 1/M \sum_{i=1}^M (\mathbf{X}_{i,j} - \bar{\mathbf{X}}_j)^2$  e  $r_{jk} = 1/M \sum_{i=1}^M (\mathbf{X}_{i,j} - \bar{\mathbf{X}}_j)(\mathbf{X}_{i,k} - \bar{\mathbf{X}}_k)^\top$  são as covariâncias, e  $\bar{\mathbf{X}}_j = 1/M \sum_{i=1}^M \mathbf{X}_{i,j}$  é o vetor médio.

**Definição 6** Dizer que uma dupla  $(\mathbf{A}, \mathbf{S}) \in \mathbb{BSS}$ , significa que essa dupla satisfaz a equação (69) e que as suposições (S1) a (S5) são válidas.

Os termos  $\mathbf{A}_0$  e  $\mathbf{S}_0(p)$ , na Definição (5), representam os valores reais do sistema de mistura e das fontes, respectivamente. Quaisquer outras matrizes  $\mathbf{A}$  e  $\mathbf{S}$ , tal que a dupla formada por elas pertença ao modelo BSS, conforme definição anterior, representam uma estimação dos valores reais. Teoricamente, uma matriz  $\mathbf{A}_0$  pode ser qualquer matriz não-singular<sup>12</sup>. No entanto, em aplicações práticas, estimar inadequadamente  $\mathbf{A}_0$  ou sua pseudo-inversa, pode por consequência acarretar numa estimação precária dos sinais fontes, no sentido que estes não sejam inteligíveis. Por isso, as matrizes  $\mathbf{A}_0$  e uma estimação  $\mathbf{A}$  devem estar relacionadas por força de algum critério que garanta a inteligibilidade dos sinais fontes estimados.

<sup>11</sup> Sinais estacionário são aqueles cujo aspecto frequencial não muda com o tempo.

<sup>12</sup> Uma matriz é singular se seu determinante é nulo. Consequentemente, uma matriz singular não possui uma inversa.

Em grande parte das aplicações, desde que a forma de onda de uma determinada fonte seja preservada, pois a informação mais relevante do sinal está presente nela e nem tanto na magnitude (TONG et al., 1991), multiplicar essa fonte por uma constante não afeta sua estimação, tampouco a ordem em que tais fontes são estimadas é realmente importante. Portanto, para uma matriz de permutação  $\mathbf{P}$  (composta de zeros e uns) a solução  $\mathbf{S} = \mathbf{P}\mathbf{S}_0$  é aceitável (BUC-CIGROSSI, 1997), assim como é também a solução  $\mathbf{S} = \mathbf{\Lambda}\mathbf{S}_0$ , sendo  $\mathbf{\Lambda}$  uma matriz diagonal e não singular.

Conforme apresentado por Tong et al. (1990) e Tong et al. (1991), segue a definição de uma relação entre a matriz de mistura e sua estimação, e também entre as fontes reais e as estimadas, relativa ao conceito de preservação de forma de onda, acima mencionado.

**Definição 7** *Uma relação  $\mathfrak{R}$  que preserva a forma de onda dos sinais fontes, simbolizada por  $(\mathbf{A}_0, \mathbf{S}_0)\mathfrak{R}(\mathbf{A}, \mathbf{S})$ , é assim descrita:*

$$\mathbf{A} = \mathbf{A}_0\mathbf{\Lambda}^{-1}\mathbf{P}^T \quad (72)$$

$$\mathbf{S}(p) = \mathbf{P}\mathbf{\Lambda}\mathbf{S}_0(p) \quad (73)$$

onde  $\mathbf{P}$  é uma matriz de permutação e  $\mathbf{\Lambda}$  uma matriz diagonal e não-singular. O par  $(\mathbf{A}, \mathbf{S})$  é uma estimação de  $(\mathbf{A}_0, \mathbf{S}_0)$ .

A equação (73) foi discutida no parágrafo anterior e seu entendimento é claro. Mas a equação (72) não é tão evidente. Para obtê-la basta lembrar que pelo modelo BSS, considerando por simplicidade o caso sem ruído, tem-se

$$\mathbf{S}_0(p) = \mathbf{A}_0^{-1}\mathbf{X}(p) \quad (74)$$

e para uma determinada estimação

$$\mathbf{X}'(p) = \mathbf{A}\mathbf{S}'(p). \quad (75)$$

Dada a equação (73), substituindo  $\mathbf{S}_0(p)$  na equação (74), obtêm-se

$$\mathbf{S}'(p) = \mathbf{P}\mathbf{\Lambda}\mathbf{A}_0^{-1}\mathbf{X}(p). \quad (76)$$

Fazendo  $\mathbf{X}'(p) = \mathbf{X}(p)$ , já que pretende-se encontrar a matriz de parâmetros e as fontes que produzam o mesmo espaço de observação, independente da ordem destas e da amplitude, e substituindo (75) em (76) resulta que

$$\mathbf{S}'(p) = \mathbf{P}\mathbf{\Lambda}\mathbf{A}_0^{-1}\mathbf{A}'\mathbf{S}'(p), \quad (77)$$

que simplificando fornece

$$\mathbf{1} = \mathbf{P}\mathbf{\Lambda}(\mathbf{A}_0)^{-1}\mathbf{A}'. \quad (78)$$

Multiplicando pela esquerda, ambos os lados da equação acima sucessivamente por  $\mathbf{P}^T$ ,  $\mathbf{\Lambda}^{-1}$  e  $\mathbf{A}_0$  tem-se por fim a equação (72).

**Definição 8** Resolver um modelo BSS é encontrar um par  $(\mathbf{A}, \mathbf{S})$ , tal que seja válida uma relação da forma

$$(\mathbf{A}_0, \mathbf{S}_0)\mathfrak{R}(\mathbf{A}, \mathbf{S}).$$

Em outras palavras, resolver um problema BSS é encontrar uma matriz que informa como o meio de comunicação misturou as fontes, e como são as formas de onda destas fontes não importando sua magnitude, tampouco a ordem em que são estimadas.

Por outra perspectiva, muitos algoritmos que encontram esta solução não fornecem diretamente a matriz  $\mathbf{A}'$ , mas sua pseudo-inversa  $\mathbf{W}$ , denominada de matriz de separação, a qual é utilizada para estimar as fontes multiplicando esta matriz pelo vetor de observação.

### 3.3.1 Um Algoritmo para solução do problema BSS

AMUSE (*Algorithm for Multiple Unknown Signal Extraction*) é um algoritmo utilizado para solução do problema BSS. Ele considera as estatísticas de segunda ordem das fontes (matriz de covariância) e sua estrutural temporal, pois assume as fontes descorrelacionadas no tempo, para estimar a matriz de mistura, convertendo o problema de identificação das fontes em um de diagonalização simultânea utilizando a DVS (CICHOCKI; AMARI, 2003).

Domingues e Iezzi (1982) fornecem alguns conceitos da Álgebra necessários para a compreensão do algoritmo AMUSE, que são a seguir enunciados, sobre um conjunto  $E \neq \emptyset$ .

**Definição 9** Uma relação  $\mathfrak{R}$  sobre um conjunto  $E$  é uma **relação de equivalência** sobre esse conjunto se, e somente se,  $\mathfrak{R}$  é reflexiva, simétrica e transitiva, ou seja:

1.  $\forall x, x \in E \Rightarrow x\mathfrak{R}x$
2.  $\forall x \text{ e } \forall y, x\mathfrak{R}y \Rightarrow y\mathfrak{R}x$
3.  $\forall x, \forall y \text{ e } \forall z, \text{ se } x\mathfrak{R}y \text{ e } y\mathfrak{R}z \Rightarrow x\mathfrak{R}z$

**Definição 10** Dado um elemento  $a \in E$ , uma **classe de equivalência** sobre  $a$ , módulo  $\mathfrak{R}$ , é o subconjunto

$$\bar{a} = \{x \in E \mid x\mathfrak{R}a\} \quad (79)$$

**Definição 11** Uma partição  $\mathcal{F}$  de  $E$  é uma família de subconjuntos não vazios onde:

1. Se  $A, B \in \mathcal{F}$ , então ou  $A = B$  ou  $A$  e  $B$  são disjuntos.
2.  $\bigcup_{A \in \mathcal{F}} A = E$ .

Dado um problema BSS, Tong et al. (1990) e Tong et al. (1991) introduzem o conceito de espaço de identificação, sendo este um conjunto com todas as matrizes de mistura  $\mathbf{A}$  e fontes  $\mathbf{S}$  que produzem as mesmas observações  $\mathbf{X}$  produzidas pelos valores reais  $\mathbf{A}_0$  e  $\mathbf{S}_0$  quando ocorre o mesmo processo ruidoso  $\mathbf{V}$ . Em outras palavras, esse conjunto contém todas as soluções potenciais dados os vetores de observação e o processo ruidoso.

**Definição 12** *Um Espaço de Identificação  $\mathbb{I}_{(\mathbf{x}, \mathbf{v})}$  é um conjunto das duplas  $(\mathbf{A}, \mathbf{S}) \in \mathbb{BSS}$ , isto é:*

$$\mathbb{I}_{(\mathbf{x}, \mathbf{v})} = \{(\mathbf{A}, \mathbf{S}) \in \mathbb{BSS}\}. \quad (80)$$

No espaço  $\mathbb{I}_{(\mathbf{x}, \mathbf{v})}$  todo  $\mathbf{A}$  e todo  $\mathbf{S}$  são candidatos legítimos a solução do problema BSS. Para entender esta afirmação basta considerar uma matriz  $\mathbf{N}$  não-singular. Do modelo  $\mathbb{BSS}$  para as variáveis reais tem-se que

$$\begin{aligned} \mathbf{X}(p) &= \mathbf{A}_0 \mathbf{S}_0(p) + \mathbf{V}(p) \\ &= \mathbf{A}_0 \mathbf{N} \mathbf{N}^{-1} \mathbf{S}_0(p) + \mathbf{V}(p) \\ &= \mathbf{A} \mathbf{S}(p) + \mathbf{V}(p), \end{aligned}$$

o que implica que qualquer matriz  $\mathbf{A} = \mathbf{A}_0 \mathbf{N}$  e  $\mathbf{S} = \mathbf{N}^{-1} \mathbf{S}_0$  com mesmo posto de  $\mathbf{A}_0$  e  $\mathbf{S}_0$ , respectivamente, são candidatos a solução (TONG et al., 1990). Isto significa que uma dupla  $(\mathbf{A}, \mathbf{S})$  pertence a  $\mathbb{I}_{(\mathbf{x}, \mathbf{v})}$  se, e somente se,  $(\mathbf{A}, \mathbf{S})$  satisfaz a Definição (5) e vale a condição  $\mathbf{A} \mathbf{S}(p) \cong \mathbf{A}_0 \mathbf{S}_0(p)$  (TONG et al., 1991).

Em termos do espaço de identificação, a Definição (10) pode ser reformulada:

**Definição 13** *Dado  $(\mathbf{A}', \mathbf{S}') \in \mathbb{I}_{(\mathbf{x}, \mathbf{v})}$ , uma classe de equivalência sobre  $(\mathbf{A}', \mathbf{S}')$ , módulo  $\mathfrak{R}$ , é o subconjunto*

$$\overline{(\mathbf{A}', \mathbf{S}')} = \{(\mathbf{A}, \mathbf{S}) \in \mathbb{I}_{(\mathbf{x}, \mathbf{v})} \mid (\mathbf{A}, \mathbf{S}) \mathfrak{R} (\mathbf{A}', \mathbf{S}')\} \quad (81)$$

*Cada  $(\mathbf{A}, \mathbf{S}) \in \overline{(\mathbf{A}', \mathbf{S}')}$  é dito um membro dessa classe de equivalência.*

Esta definição é extremamente importante, pois ela estabelece o principal objetivo do algoritmo AMUSE para solução do problema BSS: identificar as classes de equivalência contendo  $(\mathbf{A}', \mathbf{S}')$  no espaço de identificação  $\mathbb{I}_{(\mathbf{x}, \mathbf{v})}$  (TONG et al., 1990). Este caminho rumo a solução é necessário porque a identificação do valor real  $(\mathbf{A}_0, \mathbf{S}_0)$  é impossível diretamente, devido a indeterminação já evidenciada anteriormente, onde se constatou que qualquer matriz  $\mathbf{A}$  com mesmo posto de  $\mathbf{A}_0$  é candidata a solução.

A proposição a seguir estabelece que a relação  $\mathfrak{R}$  que preserva as formas de onda dos sinais fontes goza das propriedades de reflexividade, simetria e transitividade.

**Proposição 1** *A relação  $\mathfrak{R}$  da definição (7) é uma relação de equivalência sobre  $\mathbb{I}_{(\mathbf{x}, \mathbf{v})}$  (TONG et al., 1991).*

Assim, dadas duas estimativas  $(\mathbf{A}', \mathbf{S}')$  e  $(\mathbf{A}, \mathbf{S})$  de  $(\mathbf{A}_0, \mathbf{S}_0)$ , podem-se fazer as seguintes observações: a) Os pares soluções do problema BSS se relacionam consigo mesmos (propriedade reflexiva), o que implica neste caso que  $\mathbf{P} = \mathbf{\Lambda} = \mathbf{I}$ ; b) se um par se relaciona com o outro, este segundo também se relaciona com o primeiro (propriedade simétrica); e c) como os pares ordenados  $(\mathbf{A}', \mathbf{S}')$  e  $(\mathbf{A}, \mathbf{S})$  se relacionam com  $(\mathbf{A}_0, \mathbf{S}_0)$  então eles se relacionam consigo mesmos (propriedade transitiva).

A proposição abaixo destaca as propriedades estatísticas e algébricas dos membros  $(\mathbf{A}, \mathbf{S})$  das classes de equivalência. Ela evidencia que se um membro de uma classe de equivalência satisfaz estas propriedades algébricas e estatísticas, então qualquer membro desta mesma classe também satisfaz este mesmo conjunto de propriedades (TONG et al., 1990).

**Proposição 2** *Se  $(\mathbf{A}_0, \mathbf{S}_0) \mathfrak{R}(\mathbf{A}', \mathbf{S}')$  então:*

1. *Os elementos de  $\mathbf{S}'$  são mutuamente independentes se, e somente se, os elementos de  $\mathbf{S}_0$  também são;*
2. *A matriz de covariância  $\mathbf{R}_{\mathbf{S}'} = E\{\mathbf{S}'(p)\mathbf{S}'(p - \tau)^\top\}$  é diagonal se, e somente se,  $\mathbf{R}_{\mathbf{S}_0} = E\{\mathbf{S}_0(p)\mathbf{S}_0(t - \tau)^\top\}$  é diagonal, onde  $E\{\cdot\}$  é o operador esperança matemática ou média<sup>13</sup>, onde  $\tau > 0$ .*
3. *As colunas de  $\mathbf{A}'$  são ortogonais se, e somente se, as colunas de  $\mathbf{A}_0$  são ortogonais.*

Em outras palavras, os resultados da Proposição 2 significam que, se  $\mathbf{S}'(p)$  é uma estimação de  $\mathbf{S}(p)$  então esses vetores têm as mesmas propriedades estatísticas, e se  $\mathbf{A}'$  é uma estimação de  $\mathbf{A}$  então essas matrizes compartilham as mesmas propriedades algébricas.

**Definição 14** *Uma classe de equivalência  $\overline{(\mathbf{A}', \mathbf{S}')}$  é ortogonal,  $\overline{(\mathbf{A}', \mathbf{S}')}^\perp$ , se as matrizes de mistura dos membros dessa classe de equivalência têm suas colunas ortogonais.*

### 3.3.1.1 Identificabilidade e Ortogonalidade

Viu-se até aqui que  $\mathfrak{R}$  induz uma classe de equivalência dada a dupla  $(\mathbf{A}, \mathbf{S}) \in \mathbb{I}_{(\mathbf{x}, \mathbf{v})}$ . Tong et al. (1990) questionam sobre quantas seriam as classes de equivalência nesse espaço de identificação que satisfazem as propriedades (S1) a (S5) na Definição 5, dados o vetor de observação e o processo ruidoso. Para responder essa questão utiliza-se a definição a seguir.

**Definição 15**  *$(\mathbf{A}_0, \mathbf{S}_0)$  é identificável, com respeito ao modelo BSS, se para algum  $(\mathbf{A}, \mathbf{S}) \in \mathbb{I}_{(\mathbf{x}, \mathbf{v})}$ ,  $(\mathbf{A}_0, \mathbf{S}_0) \mathfrak{R}(\mathbf{A}, \mathbf{S})$ .*

<sup>13</sup> Para um vetor aleatório  $\mathbf{x} \in \mathbb{R}^{K \times 1}$  a média é dada por  $\hat{m}_{\mathbf{x}} = 1/K \sum_{k=1}^K \mathbf{x}_k$  que é uma aproximação da esperança matemática, ou primeiro momento,  $m_{\mathbf{x}} = E\{\mathbf{x}\} = \int \mathbf{x} p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}$  onde  $p_{\mathbf{x}}(\mathbf{x})$  é a função densidade de probabilidade (HYVÄRINEN et al., 2001).

Em termos gerais, a noção de identificabilidade envolve uma dupla  $(\mathbf{A}_0, \mathbf{S}_0) \in \mathbb{BSS}$ , onde esta é identificável se as formas de onda dos sinais fontes são preservadas e há uma relação de equivalência entre essa dupla e a dupla  $(\mathbf{A}, \mathbf{S})$ , solução do problema BSS.

Em outros termos, tem-se uma explicação dada por Buccigrossi (1997), onde o autor enfatiza que o conjunto de propriedades de  $\mathbf{A}_0$  e  $\mathbf{S}_0$  mitigam as duplas  $(\mathbf{A}, \mathbf{S}) \in \mathbb{I}_{(\mathbf{X}, \mathbf{V})}$  que são soluções aceitáveis. Então, deve-se procurar um conjunto de propriedades que, além de reduzir as soluções ainda preserve as formas de onda das fontes. Estas propriedades, quando existirem, farão  $(\mathbf{A}_0, \mathbf{S}_0)$  ser identificável.

O resultado apresentado no Teorema 1 estabelece uma restrição às soluções do problema BSS, a medida que, dada uma propriedade da matriz de covariância dos vetores de observação, as soluções se apresentam em um espaço ortogonal.

**Teorema 1** *Dada  $\mathbf{R}_\mathbf{X}$  (matriz de covariância do vetor de observação), se  $(\mathbf{A}_0, \mathbf{S}_0)$  é identificável, então os  $m$  maiores valores singulares de  $\mathbf{R}_\mathbf{X}$  são distintos e  $(\mathbf{A}_0, \mathbf{S}_0)$  pertence a uma classe de equivalência ortogonal.*

O Teorema 2 fornece um resultado que permite chegar à solução do problema BSS, pois conforme já mencionado esta é encontrada quando for possível identificar uma classe de equivalência contendo  $(\mathbf{A}_0, \mathbf{S}_0)$ , e pelo Teorema 1 quando existir uma certa quantidade de valores singulares da matriz de covariância distintos, então existirá uma única classe de equivalência.

**Teorema 2** *Se os  $m$  maiores valores singulares de  $\mathbf{R}_\mathbf{X}$  são distintos, existe uma, e somente uma, classe de equivalência ortogonal no  $\mathbb{I}_{(\mathbf{X}, \mathbf{V})}$  que satisfaz as suposições (S1) a (S5).*

Em resumo ambos os teoremas permitem a conclusão que se  $(\mathbf{A}_0, \mathbf{S}_0)$  é identificável então  $(\mathbf{A}_0, \mathbf{S}_0) \in \overline{(\mathbf{A}, \mathbf{S})}^\perp$  e os valores singulares de  $\mathbf{R}_\mathbf{X}$  são distintos, daí então existe uma, e somente uma, classe de equivalência ortogonal (TONG et al., 1990).

Como não se pode inferir sobre a ortogonalidade das colunas de  $\mathbf{A}_0$ , visto ser latente no problema BSS, deve-se buscar uma transformação no espaço de observação que seja possível ortogonalizá-las, onde se garanta que  $(\mathbf{A}, \mathbf{S})$  pertença a uma classe de equivalência ortogonal. Uma transformação desse tipo reduz a complexidade do algoritmo para solução do problema BSS (TONG et al., 1991), e ela é obtida a partir da manipulação da matriz de covariância e da DVS.

A matriz de covariância do vetor de observação pode ser decomposta da seguinte forma,

considerando as suposições (S1) a (S5) da Definição 5:

$$\mathbf{R}_X = E \{ \mathbf{X}(p) \mathbf{X}(p)^\top \} \quad (82)$$

$$= E \{ (\mathbf{A}_0 \mathbf{S}_0(p) + \mathbf{V}(p)) (\mathbf{S}_0(p)^\top \mathbf{A}_0^\top + \mathbf{V}^\top) \} \quad (83)$$

$$= \mathbf{A}_0 E \{ \mathbf{S}_0(p) \mathbf{S}_0(p)^\top \} \mathbf{A}_0^\top + E \{ \mathbf{V}(p) \mathbf{V}(p)^\top \} \quad (84)$$

$$= \mathbf{A}_0 \mathbf{R}_{S_0} \mathbf{A}_0^\top + \sigma^2 \mathbf{I}. \quad (85)$$

Se as fontes forem consideradas com potência unitária, pela hipótese (S5) na Definição 5,  $\mathbf{R}_{S_0} = \mathbf{I}$ . Logo, a matriz de mistura pode ser determinada computando  $\mathbf{R}_X$  se o ruído puder ser subtraído do sistema, pois

$$\mathbf{R}_X = \mathbf{A}_0 \mathbf{A}_0^\top + \sigma^2 \mathbf{I}. \quad (86)$$

Calculando a DVS da matriz  $\mathbf{A}_0$  e sua transposta, obtém-se:

$$\mathbf{A}_0 = \mathbf{L} \Psi \mathbf{R}^\top \quad (87)$$

$$\mathbf{A}_0^\top = \mathbf{R} \Psi^\top \mathbf{L}^\top. \quad (88)$$

Substituindo as equações (87) e (88) no primeiro termo do lado direito da equação (85) tem-se

$$\mathbf{A}_0 \mathbf{A}_0^\top = \mathbf{L} \Psi \mathbf{R}^\top \mathbf{R} \Psi^\top \mathbf{L}^\top \quad (89)$$

$$= \mathbf{L} \Psi^2 \mathbf{L}^\top, \quad (90)$$

sendo  $\Psi^2 = \text{diag}(\psi_1^2, \dots, \psi_m^2)$ . Considerando  $\Phi = \text{diag}(\psi_1^2 + \sigma^2, \dots, \psi_m^2 + \sigma^2, \sigma^2, \dots, \sigma^2)$ , tem-se a seguinte decomposição em valores singulares da matriz de autocovariância (ou somente covariância) do vetor de observação

$$\mathbf{R}_X = \mathbf{L} \Psi^2 \mathbf{L}^\top + \sigma^2 \mathbf{I} \quad (91)$$

$$= \mathbf{L} \Phi \mathbf{L}^\top. \quad (92)$$

Define-se então a transformação de ortogonalização como

$$\mathcal{T} = \Psi^{-1} \mathbf{L}^\top. \quad (93)$$

Para verificar que esta transformação de fato ortogonaliza, considera-se a DVS da matriz  $\mathbf{A}_0$ , isto é:  $\mathbf{A}_0 = \mathbf{L} \Psi \mathbf{R}^\top$ . Multiplicando a matriz de transformação  $\mathcal{T}$  por  $\mathbf{A}_0$  a direita, se obtém:

$$\mathcal{T} \mathbf{A}_0 = \Psi^{-1} \mathbf{L}^\top \mathbf{L} \Psi \mathbf{R}^\top \quad (94)$$

$$= \mathbf{R}^\top, \quad (95)$$

logo os vetores coluna de  $\mathcal{T} \mathbf{A}_0$  são ortogonais, pela Definição 4 da matriz  $\mathbf{R}$ .

Aplicando essa transformação a cada elemento do modelo BSS, tem-se um modelo ortogonal equivalente, explicitado pelo símbolo  $\mathbb{BSS}_O$ , com equações a seguir definidas.

**Definição 16** Um modelo BSS ortogonal,  $\mathbb{BSS}_O$ , consiste de duplas  $(\mathbf{B}_0, \mathbf{S}_0)$  e  $(\mathbf{Y}, \mathbf{N})$  nos moldes do problema BSS, sendo que suas variáveis se relacionam da seguinte maneira, dada a transformação ortogonal da equação (93):

$$\mathbf{Y}(p) = \mathcal{T}\mathbf{X}(p) \quad (96)$$

$$\mathbf{B}_0 = \mathcal{T}\mathbf{A}_0 \quad (97)$$

$$\mathbf{N}(p) = \mathcal{T}\mathbf{V}(p). \quad (98)$$

Como pode-se notar, a transformação definida na equação (93), sobre o espaço de observação, induz um mapeamento do espaço de identificação definido pelo modelo BSS no espaço de identificação definido pelo modelo  $\mathbb{BSS}_O$ :

$$\Omega : \mathbb{I}_{(\mathbf{X}, \mathbf{V})} \rightarrow \mathbb{I}_{(\mathbf{Y}, \mathbf{N})} \quad (99)$$

$$(\mathbf{A}, \mathbf{S}) \mapsto (\mathcal{T}\mathbf{A}, \mathbf{S}) \quad (100)$$

Dado então esse mapeamento, o problema da identificação de  $(\mathbf{A}_0, \mathbf{S}_0)$  no espaço de observação de  $\mathbf{X}$  torna-se um problema de identificação da dupla  $(\mathbf{B}_0, \mathbf{S}_0)$  no espaço de observação de  $\mathbf{Y}$ , sendo que  $\mathbf{B}_0$  tem a característica de ter suas colunas ortogonais, pelo que ficou evidenciado na equação (94).

É claro que tal mapeamento não pode alterar a relação de equivalência induzida por  $\mathfrak{R}$  sobre  $\mathbb{I}_{(\mathbf{Y}, \mathbf{N})}$ , no sentido que  $\mathbf{B}_0$  e  $\mathbf{S}_0(p)$  ainda devem estar relacionados conforme explicitado na Definição (7), e portanto as imagens em  $\mathbb{I}_{(\mathbf{X}, \mathbf{V})}$  devem ser equivalentes em  $\mathbb{I}_{(\mathbf{Y}, \mathbf{N})}$  (TONG et al., 1991). A Proposição 3 fornece esta garantia ao mapeamento  $\Omega$ .

**Proposição 3** O mapeamento  $\Omega$  preserva a relação de equivalência  $\mathfrak{R}$ , ou seja, se

$$(\mathbf{A}_0, \mathbf{S}_0) \mathfrak{R} (\mathbf{A}, \mathbf{S})$$

então,

$$\Omega((\mathbf{A}_0, \mathbf{S}_0)) \mathfrak{R} \Omega((\mathbf{A}, \mathbf{S})).$$

Em outras palavras este resultado significa que uma solução do problema BSS no espaço de observação de  $\mathbf{X}$  terá uma solução correspondente no espaço de observação ortogonal de  $\mathbf{Y}$ . O Teorema 3 a seguir certifica esta afirmação, mostrando que para uma solução do modelo ortogonal existe outra para o modelo real, onde a matriz de mistura ortogonal é solução quando multiplicada pela pseudo-inversa da transformação definida na equação (93).

**Teorema 3** Se  $(\mathbf{B}, \mathbf{S}) \mathfrak{R} (\mathbf{B}_0, \mathbf{S}_0)$ , então  $(\mathcal{T}^\dagger \mathbf{B}, \mathbf{S}) \mathfrak{R} (\mathbf{A}_0, \mathbf{S}_0)$ , onde  $\mathcal{T}^\dagger$  é a pseudo-inversa de  $\mathcal{T}$ .

A Definição 15 estabeleceu que  $(\mathbf{A}_0, \mathbf{S}_0)$  é identificável quando existir outro par ordenado com o qual vale  $\mathfrak{R}$ . No entanto, não se pode certificar que  $(\mathbf{A}_0, \mathbf{S}_0)$  seja identificável em relação ao modelo BSS, pelo que foi mencionado no parágrafo sobre o espaço de identificação  $\mathbb{I}_{(\mathbf{x}, \mathbf{v})}$ . Por outro lado, a identificabilidade pode ser garantida para outros modelos que sejam extensões do modelo BSS, conforme Definição 17 e Teorema 4.

**Definição 17** Um modelo  $\mathbb{BSS}_1$  é um modelo BSS que satisfaz as seguintes condições:

1. as fontes em  $\mathbf{S}$  estão descorrelacionadas;
2. existe  $\tau > 0$  tal que, para  $i \neq j$

$$\frac{E\{\mathbf{S}_i(p)\mathbf{S}_i(p-\tau)\}}{E\{\mathbf{S}_i^2\}} \neq \frac{E\{\mathbf{S}_j(p)\mathbf{S}_j(t-\tau)\}}{E\{\mathbf{S}_j^2\}}. \quad (101)$$

**Teorema 4**  $(\mathbf{A}_0, \mathbf{S}_0)$  é identificável em relação ao modelo  $\mathbb{BSS}_1$ .

Portanto, quando as fontes são consideradas descorrelacionadas e com autocorrelações diferentes para determinado atraso temporal  $\tau > 0$ , de acordo com a Definição 17, e pelo Teorema 4, o par  $(\mathbf{A}_0, \mathbf{S}_0)$  é identificável em relação ao modelo mencionado, logo, este é uma solução do problema BSS.

### 3.3.1.2 Algoritmo AMUSE

Como foi mencionado anteriormente, para estimar  $(\mathbf{A}_0, \mathbf{S}_0)$  primeiro deve ser encontrado um mapeamento de  $\mathbb{I}_{(\mathbf{x}, \mathbf{v})}$  em  $\mathbb{I}_{(\mathbf{y}, \mathbf{N})}$  que é dado por uma transformação ortogonal que depende da matriz de covariância de  $\mathbf{X}$ .

Assim, o primeiro passo a fim de obter um algoritmo para separação das fontes é calcular  $\mathbf{R}_\mathbf{X}$  e sua DVS, ou seja,

$$\mathbf{R}_\mathbf{X} = \mathbf{L}\mathbf{\Phi}\mathbf{L}^\top,$$

onde  $\mathbf{\Phi} = \mathbf{\Psi}^2 + \sigma^2\mathbf{I}$ , conforme equação (91), que é uma matriz diagonal da forma:

$$\mathbf{\Phi} = \begin{bmatrix} \phi_1^2 & 0 & \dots & 0 \\ 0 & \phi_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \phi_n^2 \end{bmatrix}, \quad (102)$$

onde  $\phi_i^2 = \psi_i^2 + \sigma^2$ , para  $1 \leq i \leq n$ , sendo  $n$  a quantidade de valores singulares mais significantes na decomposição e  $\sigma^2$  a variância do ruído. A matriz diagonal  $\mathbf{\Phi}$  é então obtida resolvendo  $\phi_i = \sqrt{\psi_i^2 + \sigma^2}$ .

Após este passo, o espaço de observação de  $\mathbf{X}$  então é transformado, por  $\mathcal{T}$ , em um espaço de observação ortogonal  $\mathbf{Y}$ , conforme equação (96), e sua matriz de covariância é calculada para algum parâmetro de atraso  $\tau > 0$ , pois pelo Teorema 4, já que as fontes foram decorrelacionadas pela transformação  $\mathcal{T}$ , pode-se afirmar que  $(\mathbf{A}_0, \mathbf{S}_0)$  é identificável em relação ao modelo BSS<sub>1</sub>. Isso implica, pelos Teoremas 1 e 2, que os  $n$  valores singulares, na decomposição anterior, serão distintos e existirá uma, e somente uma classe de equivalência ortogonal sobre  $\mathbb{I}_{(\mathbf{X}, \mathbf{V})}$ . Portanto, a matriz de covariância de  $\mathbf{Y}$ , considerando um atraso  $\tau$ , é dada por:

$$\mathbf{R}_{\mathbf{Y}}(\tau) = \mathbb{E} \{ \mathbf{Y}(p) \mathbf{Y}(p - \tau)^\top \}. \quad (103)$$

Segundo Tong et al. (1990), visto que o ruído considerado no sistema é branco, essa matriz admite a seguinte decomposição em autovalores:

$$\mathbf{R}_{\mathbf{Y}}(\tau) = \mathbf{B}_0 \mathbb{E} \{ \mathbf{Y}(p) \mathbf{Y}(p - \tau)^\top \} \mathbf{B}_0^\top \quad (104)$$

$$= \mathbf{B}_0 \text{diag}(\gamma_1(\tau), \dots, \gamma_n(\tau)) \mathbf{B}_0^\top, \quad (105)$$

onde  $\gamma_i(\tau) = \mathbb{E} \{ \mathbf{s}_i(p) \mathbf{s}_i(p - \tau)^\top \}$ , para  $i = 1, \dots, n$ . No entanto, na prática  $\mathbf{R}_{\mathbf{X}}$  e  $\mathbf{R}_{\mathbf{Y}}(\tau)$  não podem ser simétricas já que são estimadas, o que implica que a solução da equação (104) não é realizável Tong et al. (1990). Por isto, toma-se uma decomposição simétrica:

$$\mathbf{R}_{\mathbf{Y}_{sym}} = \frac{\mathbf{R}_{\mathbf{Y}}(\tau) + \mathbf{R}_{\mathbf{Y}}(\tau)^\top}{2}. \quad (106)$$

Agora é possível encontrar uma solução para o problema BSS ortogonal, já que a decomposição em autovalores de  $\mathbf{R}_{\mathbf{Y}_{sym}}$  fornece uma estimação  $\mathbf{B}$  de  $\mathbf{B}_0$ . Já a matriz das fontes  $\mathbf{S}$  é obtida resolvendo

$$\mathbf{S}(p) = \mathbf{B}^\top \mathbf{Y}(p) \quad (107)$$

$$= \mathbf{B}^{-1} \mathbf{Y}(p). \quad (108)$$

Pelo Teorema 3 esta dupla  $(\mathbf{B}, \mathbf{S})$  solução do problema ortogonal, também é solução do problema real, se  $\mathbf{B}$  for multiplicado pela pseudo-inversa da transformação de ortogonalização. Esta dupla também é membro da classe de equivalência ortogonal  $\overline{(\mathbf{B}, \mathbf{S})}$ , e portanto  $(\mathbf{B}, \mathbf{S}) \in \mathbb{I}_{(\mathbf{Y}, \mathbf{N})}$ .

Por conseguinte, os autovetores obtidos da decomposição de  $\mathbf{R}_{\mathbf{Y}_{sym}}$  formam as colunas da matriz  $\mathbf{B}$  que é utilizada para obter uma estimação tanto da matriz de mistura como das fontes, como anteriormente observado. Logo,

$$\mathbf{A}_0 = \mathcal{T}^\dagger \mathbf{B} \quad (109)$$

$$\mathbf{S}_0(p) = \mathbf{B}^\top \mathbf{Y}(p). \quad (110)$$

Em situações práticas as igualdades acima nunca se verificarão já que calculam-se estimativas dos valores reais. Por isso, as equações (111) e (112) a seguir representam melhor o que

ocorre sem situações práticas, considerando que  $\hat{\mathbf{A}}$  e  $\hat{\mathbf{S}}(p)$  são as estimativas de  $\mathbf{A}_0$  e  $\mathbf{S}_0(p)$ , e portanto  $(\hat{\mathbf{A}}, \hat{\mathbf{S}})\mathfrak{R}(\mathbf{A}_0, \mathbf{S}_0)$ .

$$\hat{\mathbf{A}} = \mathcal{T}^\dagger \mathbf{B} \quad (111)$$

$$\hat{\mathbf{S}}(p) = \mathbf{B}^\top \mathbf{Y}(p). \quad (112)$$

Por fim, resume-se o algoritmo AMUSE nas etapas abaixo:

1. Estimar a matriz de covariância  $\mathbf{R}_{\mathbf{X}}$ , conforme equação (82);
2. Computar a DVS de  $\mathbf{R}_{\mathbf{X}}$ , de acordo com a equação (91);
3. Estimar a quantidade  $M$  de fontes pelo número de valores singulares mais significantes e a variância do ruído  $\sigma^2$  pelo número daqueles menos significantes;
4. Obter a transformação de ortogonalização  $\mathcal{T}$ , dada pela equação (93);
5. Obter o espaço de observação ortogonal  $\mathbf{Y}$  utilizando a transformação  $\mathcal{T}$ ;
6. Calcular a DVS da equação (106) para um certo  $\tau$  que resulte em distintos autovalores;
7. Construir a automatriz  $\mathbf{B}$  com os autovetores obtidos na etapa anterior;
8. Estimar enfim a matriz de mistura e as fontes, por meio das equações (111) e (112), respectivamente.

Até o momento foi abordado o modo como uma máquina de aprendizado pode ser induzida, a partir de um conjunto de vetores de atributos, a fim de modelar o ambiente. Como mencionado anteriormente, o método AMUSE estima vários sinais a partir de uma mistura de fontes latentes. Cada uma dessas é tida como um vetor de atributos que explica, a seu modo, características distintas do ambiente. Deste modo, é necessário implementar mais de uma máquina de aprendizado. Com o objetivo de escolher qual a melhor máquina e qual têm mais peso na decisão sobre a classe de cada padrão analisado, na próxima seção é apresentado um método para tomada de decisão, que auxiliará nessa tarefa.

### 3.4 Análise Hierárquica de Processos (AHP)

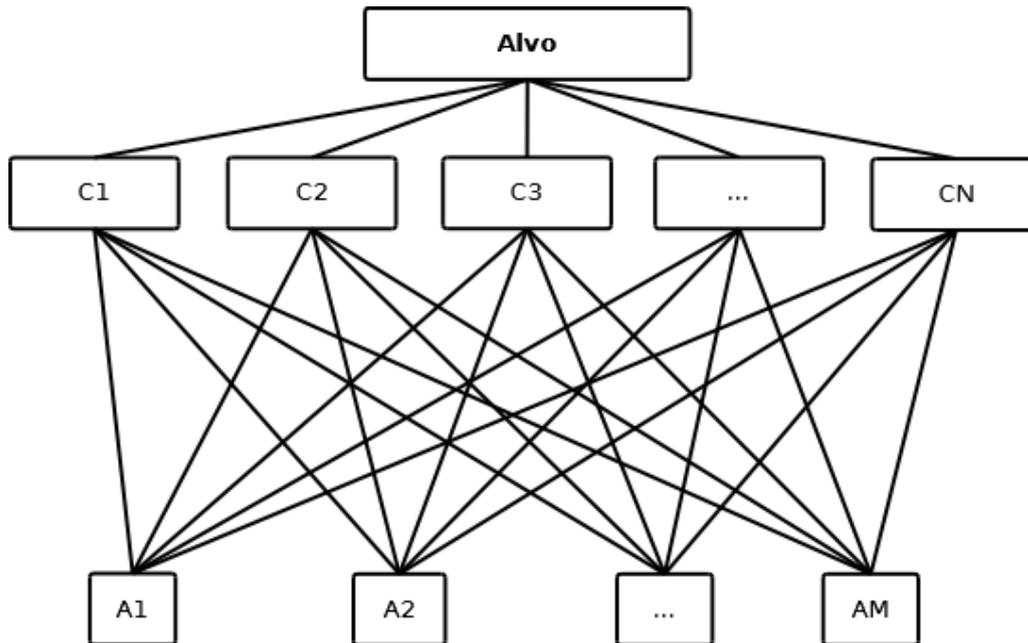
Os processos de tomada de decisão permeiam todas as áreas dos saberes. Várias metodologias têm sido desenvolvidas com o objetivo de auxiliar o tomador de decisão fornecendo a este ferramentas que agreguem mais robustez a tarefa decisória. A Análise Hierárquica de Processos (AHP) é uma destas metodologias. Ela se destaca das demais porque além de ser composta por um ferramental matemático que considera os aspectos tangíveis do processo de tomada de

decisão, que inclui medições, estatísticas, etc, ela permite adicionar aqueles aspectos que são intangíveis, tais como a percepção do tomador de decisão sobre determinados critérios, sua experiência em situações similares, etc (SAATY; SAGIR, 2009).

O método AHP foi proposto por Thomas L. Saaty (SAATY, 1987). Esta metodologia tem sido empregada exaustivamente no campo de Pesquisa Operacional, mas também em outras áreas tais como gestão de recursos humanos, educação, manufatura, política e governo (VAIDYA; KUMAR, 2006), forças armadas, psicologia (SAATY, 2008), judiciário (OLIVEIRA; OLIVEIRA; DUARTE, 2016) e computação (aprendizado de máquina) (JAYA; TAMILSELVI, 2014; KHANMOHAMMADI; REZAEIAHARI, 2014; KOU; WU, 2014).

Para resolver um problema de tomada de decisão utilizando o método AHP, o primeiro passo é a modelagem deste problema numa estrutura hierárquica, cujo topo consiste do alvo da tomada de decisão, logo abaixo listam-se os critérios e subcritérios que balizam as escolhas, e na base da hierarquia encontram-se as alternativas disponíveis para escolha. A Figura 14 ilustra um modelo hierárquico genérico.

Figura 14 – Modelo genérico da Análise Hierárquica de Processos, com  $N$  critérios e  $M$  alternativas.



Fonte: Autoria própria.

Definem-se como  $\mathcal{C}$  e  $\mathcal{A}$  os conjuntos dos critérios e alternativas, respetivamente. Também existe o conjunto dos subcritérios, mas neste texto somente os critérios serão considerados, sem perda de generalidade. Os elementos do conjunto de critérios  $\mathcal{C}$  são basilares para as comparações das alternativas no conjunto  $\mathcal{A}$ . Seja  $\triangleright_{\mathcal{C}}$  uma relação binária sobre o conjunto  $\mathcal{C}$  que significa “maior preferência que”, e  $\nabla_{\mathcal{C}}$  a relação que indica “indiferença a”, ambas para um elemento  $C \in \mathcal{C}$  em relação a alguma alternativa. Neste sentido,  $A_i \triangleright_{\mathcal{C}} A_j$ , significa que

$A_i$  tem maior preferência que  $A_j$ , isto é,  $A_i$  domina  $A_j$ , em relação ao critério  $C$ . Dadas duas alternativas quaisquer  $A_i, A_j \in \mathcal{A}$ , há três resultados possíveis para as comparações binárias em relação a algum critério  $C \in \mathcal{C}$ : 1)  $A_i \triangleright_C A_j$ ; ou 2)  $A_j \triangleright_C A_i$ ; ou 3)  $A_i \nabla_C A_j$  (SAATY, 2008).

As entidades (critérios, subcritérios e alternativas) em cada nível da hierarquia são comparadas uma com a outra, comparações paritárias, em relação a uma entidade em um nível menor, conforme especificado pelas relações binárias definidas anteriormente. Assim sendo, a partir da Figura 14, o tomador de decisão faz uma pergunta do tipo: “a alternativa A1 é preferível a alternativa A2 em relação ao critério C1?”. A resposta a esta pergunta é um julgamento verbal que deve ser convertido para um julgamento numérico a partir de uma escala pré-definida. Nesta escala existe uma associação entre a preferência, dada de modo verbal, e sua intensidade respectiva, dada de modo numérico. Geralmente a Escala Fundamental de Saaty é utilizada. Ela está discriminada na Tabela 6.

Tabela 6 – Escala Fundamental de Saaty.

Preferência	Intensidade	Explicação
Igual	1	As entidades analisadas contribuem do mesmo modo para o alvo
Leve	2	A experiência e o julgamento são levemente a favor de uma entidade sobre outra
Moderada	3	Idem ao anterior
Mais que moderada	4	A experiência e o julgamento são fortemente a favor de uma entidade sobre outra
Forte	5	Idem ao anterior
Mais que forte	6	Uma entidade é muito favorável a outra e sua dominância é percebida na prática
Muito forte	7	Idem ao anterior
Muito, muito forte	8	A evidência de uma entidade sobre outra é altamente confirmada
Extrema	9	Idem ao anterior

Fonte: Adaptado de Saaty e Vargas (2012, p. 5-7).

O método AHP a partir destes julgamentos possibilita ranquear as alternativas de modo que o tomador de decisão possa escolher, por ordem de preferência, qual a melhor alternativa para o problema em questão. Além disso, ele disponibiliza um modo de verificar a consistência dos julgamentos (SAATY; VARGAS, 2012).

Ao final de cada julgamento o resultado obtido é uma matriz de comparações paritárias

$\mathbf{A} \in \mathbb{R}_{>0}^{n \times n}$ , com  $n \geq 2$ :

$$\mathbf{A}_k^{(h)} = \begin{bmatrix} 1 & a_{12}^{(h)} & \cdots & a_{1n}^{(h)} \\ 1/a_{12}^{(h)} & 1 & \cdots & a_{2n}^{(h)} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n}^{(h)} & 1/a_{2n}^{(h)} & \cdots & 1 \end{bmatrix}, \quad (113)$$

onde  $a_{ij}^{(h)} = w_i^{(h)}/w_j^{(h)}$  representa a intensidade do julgamento da alternativa<sup>14</sup>  $A_i$  sobre a alternativa  $A_j$  em relação a outra entidade no nível anterior;  $w_i^{(h)} > 0, \forall i$ ;  $h$  representa o nível da hierarquia;  $n$  é a quantidade de alternativas no nível  $h$ ; e  $k = 1, 2, 3, \dots$  é utilizado aqui somente para identificar as matrizes.

O relacionamento entre as matrizes de comparações paritárias e a escala de Saaty resulta em uma escala derivada que fornece um mapeamento entre dois sistemas numéricos relacionais (SAATY; VARGAS, 2012). Isso implica que as intensidades dos julgamentos obtidas após a tradução da escala verbal para a escala de Saaty são convertidas para outros valores, os quais permitirão que a dominância relativa das alternativas seja aferida.

Voltando a atenção para os elementos destas matrizes, viu-se que eles são escritos na forma  $a_{ij} = w_i/w_j$ <sup>15</sup>. Saaty e Vargas (2012, p. 4) afirmam que as comparações que realizamos ou são relativas ou absolutas, de acordo com estudos realizadas pela psicologia. Assim sendo, visto que os valores  $a_{ij}$  são os níveis de intensidade obtidos a partir da escala fundamental ao traduzir os julgamentos, estes valores são então do tipo absoluto, pois permitem ranquear as alternativas em relação a algum critério (SAATY, 2008, p. 5). Tais valores explicam o quanto uma alternativa é mais preferível do que outra. Por outro lado, estes valores absolutos são expressos como a razão entre os valores relativos  $w_i$  e  $w_j$ .

No entanto, não se tem acesso aos elementos  $w_i$  e  $w_j$  ao final dos julgamentos, porque apenas as intensidades  $a_{ij} = w_i/w_j$  das matrizes de comparações paritárias estão disponíveis. Naturalmente estes elementos compõe um vetor de pesos  $\mathbf{w} = [w_1 \ w_2 \ \cdots \ w_n]^T \in \mathbb{R}^{n \times 1}$ , denominado vetor de prioridades. Uma das propriedades que este vetor deve possuir está relacionada a retenção da invariância sobre a estrutura hierárquica. Isso significa que, exceto por uma constante multiplicativa  $c$ , este vetor não deve sofrer qualquer alteração na sua orientação quando multiplicado por uma matriz de comparações paritárias  $\mathbf{A}$  que provém de uma estrutura hierárquica, ou seja,  $\mathbf{A}\mathbf{w} = c\mathbf{w}$  (SAATY, 2003, p. 86), mas apenas a sua magnitude pode ser modificada (MEYER, 2000, p. 490). De outro modo, pensando na matriz de comparações paritárias como uma transformação linear, e tendo em conta que ela representa a estrutura hierárquica concebida, ela somente poderá ampliar ou reduzir o vetor de prioridades, mantendo a orientação e o sinal inalterados, já que  $n > 0$ .

<sup>14</sup> Embora tenha sido mencionado alternativa, a mesma explicação é estendida para qualquer entidade.

<sup>15</sup> Note que o índice  $(h)$  foi omitido, porque o raciocínio é o mesmo independente do nível da hierarquia que se considera.

A fim de obter os valores dos pesos relativos, Saaty (2008, p. 261) sugere a multiplicação da matriz de comparações paritárias, considerando as razões  $w_i/w_j$ , pelo vetor  $\mathbf{w}$ , para algum nível de hierarquia  $h$  que será omitido a partir deste ponto por conveniência de notação. Percebe-se que este vetor representa uma escala, porque possui os valores relativos dos julgamentos. Esta abordagem resulta na igualdade:

$$\begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \dots & \frac{w_2}{w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & \frac{w_n}{w_n} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} w_1 + w_1 + \dots + w_1 \\ w_2 + w_2 + \dots + w_2 \\ \vdots \\ w_n + w_n + \dots + w_n \end{bmatrix}, \quad (114)$$

$$\mathbf{A}\mathbf{w} = n \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}, \quad (115)$$

$$\mathbf{A}\mathbf{w} = n\mathbf{w}. \quad (116)$$

Com o intuito de resolver a equação (116) recorre-se ao Teorema de Perron, o qual estabelece que: para uma dada matriz  $\mathbf{A}$  existe um autovetor  $\mathbf{w} > \mathbf{0}$  tal que a equação (116) é satisfeita, sendo que  $\lambda = n > 0$ <sup>16</sup> é um autovalor associado (SAATY, 2003), (MEYER, 2000, p. 666-667). Além disso, o autovetor  $\mathbf{w}$  é o único que satisfaz a equação (116) sob a restrição adicional de norma unitária  $\|\mathbf{w}\| = 1$ , exceto para múltiplos positivos de  $\mathbf{w}$ .

Sobre os autovalores, dado que o traço<sup>17</sup> das matrizes de comparações paritárias são iguais a  $n$ , porque os elementos da diagonal principal são unitários,  $\text{tra}(\mathbf{A}) = n = \lambda$  e, de acordo com Bronson e Costa (2007, p. 232), como a soma dos autovalores de uma matriz é igual ao seu traço, então, tomando estes dois resultados, e considerando o que foi dito anteriormente sobre  $\lambda$  ser um autovalor de  $\mathbf{A}$ , conclui-se que todos os autovalores de  $\mathbf{A}$  são nulos, exceto um, o qual é  $\lambda$ .

Portanto, este autovalor está associado ao autovetor a direta de  $\mathbf{w}$  que é solução do sistema de equações homogêneas. Visto que o Teorema de Perron, anteriormente mencionado, fixa a norma do vetor de prioridades para a unidade, então este vetor deve ser normalizado, fazendo:

$$\mathbf{w}' = \frac{\mathbf{w}}{\|\mathbf{w}\|}. \quad (117)$$

Como os elementos de  $\mathbf{w}$  representam valores relativos de julgamento, então esta normalização acarreta em uma conversão para uma escala absoluta possibilitando que os valores da

<sup>16</sup> A variável  $n$  foi substituída por  $\lambda$  apenas para evitar confusão posterior com a ordem da matriz.

<sup>17</sup> O traço de uma matriz é a soma dos elementos de sua diagonal principal.

matriz de comparações paritárias possam ser recuperados, mas não na mesma escala (SAATY, 2008).

Até o momento foram explicadas três etapas da Análise Hierárquica de Processos: 1) construção da estrutura hierárquica; 2) obtenção das matrizes de comparações paritárias; e 3) cálculo dos vetores de prioridades locais. Estes vetores, portanto, explicam somente a preferência local de uma determinada alternativa sem levar em consideração as preferências das demais.

Para cada nível  $h$  da hierarquia haverá tantas matrizes de comparações paritárias quanto são as entidades envolvidas nos níveis superiores, além de um vetor de prioridades associado a cada uma delas, conforme constatado anteriormente. Para agregar os vetores de prioridades de cada nível, os quais informam sobre as prioridades locais dos critérios, subcritérios ou alternativas, é necessária a obtenção de um outro vetor de prioridades global que combina linearmente as prioridades dos demais vetores. Assim, se existem  $l$  critérios e  $p$  alternativas, então existirão  $l$  vetores de prioridade no nível  $h = 2$  e no nível  $h = 1$ , porque os critérios somente são comparados entre si em relação a um alvo. A expressão resultante para o vetor de prioridades global é dada por:

$$\mathbf{v} = [v_1 \ v_2 \ \cdots \ v_p] \text{ com} \quad (118)$$

$$v_k = \sum_{l=1}^o \mathbf{s}_l^{(h-1)} \mathbf{t}_{kl}^{(h)}, \quad k = 1, \cdots, p, \quad (119)$$

onde  $\mathbf{s}_l^{(h-1)}$  reproduz o peso do  $l$ -ésimo critério/subcritério obtido dos julgamentos feitos em relação ao alvo,  $\mathbf{t}_{kl}^{(h)}$  é o vetor de prioridade local obtido ao final dos julgamentos da alternativa  $A_k$  em relação as demais, tendo em mente o  $l$ -ésimo critério/subcritério, sendo que  $o$  e  $p$  dizem respeito a quantidade de critérios/subcritérios e alternativas, respectivamente.

Por ora, foram apresentados alguns métodos que possibilitarão a obtenção de soluções para o problema de reconhecimento de CPV. No próximo capítulo é apresentado um panorama dos métodos já propostos para solução deste problema, com o intuito de elencar as direções que as propostas têm tomado, os resultados atingidos, as desvantagens e vantagens de cada proposição, a fim de orientar a concepção de uma nova metodologia.

## 4 Estado da Arte

Várias pesquisas têm sido propostas com o objetivo de reconhecer/classificar as arritmias do tipo CPV a partir dos sinais de ECG. Elas utilizam diferentes abordagens e métodos para extração de atributos. A seguir, algumas das mais recentes pesquisas sobre este tema estão elencadas, no intuito de vislumbrar o que já tem sido feito para resolução deste problema de classificação para então poder balizar a presente pesquisa rumo a obtenção de novas metodologias que melhorem os indicadores de performance já obtidos.

No trabalho proposto por Liu et al. (2015), na etapa de pré-processamento, é utilizada a transformada wavelet com uma técnica de limiar auto-adaptativo e a função wavelet db6, com a finalidade de remover ruído dos sinais de ECG. A escolha dessa wavelet, segundo os autores, se justifica por ela apresentar maior similaridade com o complexo QRS. Eles também propõe um método para detecção dos picos das ondas R, ou complexos QRS, utilizando a banda de frequência no banco de filtros wavelets com maior concentração de energia. Com este método obtiveram 98,74% de acurácia na detecção.

Para obtenção dos atributos, eles propõem o cálculo do expoente de Lyapunov, que é uma medida da taxa média de separação de duas trajetórias no espaço de fase trata-se da taxa de crescimento médio, em relação a distância inicial:

$$\lambda_i = \frac{1}{t_i - t_{i-1}} \ln \frac{D'(t_i)}{D(t_{i-1})}, i = 1, 2, \dots, N - (m - 1)\tau,$$

onde  $t_i$  é o  $i$ -ésimo índice temporal,  $D'(t_i) = |x(t_i) - x_0(t_i)| > \epsilon$  e  $D(t_i) = |x(t_i) - x_1(t_i)| < \epsilon$  sendo  $\epsilon > 0$ ,  $x(t_i)$  é uma série temporal defasada pelo parâmetro  $\tau$  e  $x_k(t)$  a  $k$ -ésima posição nesta série.

As curvas obtidas pelo cálculo deste expoente, para cada ponto do complexo QRS, apresentam comportamentos distintos para batimentos CPV, Normal e Contração Atrial Prematura. Para as CPV a curva resultante é muito mais suave, enquanto para os demais batimentos ela apresenta várias oscilações, sendo que, para batimentos normais tais oscilações são mais proeminentes. Por isso, eles propõe calcular a derivada destas curvas com a intenção de capturar tais oscilações.

Neste caminho, dois atributos foram calculados a partir da derivada do expoente de Lyapunov, dada pela equação

$$f'_i = \frac{\lambda_{i+1} - \lambda_i}{\Delta}.$$

O primeiro atributo foi obtido pela soma dos pontos da derivada e o segundo pela quantidade de valores absolutos da derivada maiores que o limiar 0,001. Esta escolha é utilizada porque a curva da derivada para os batimentos CPV apresenta algumas flutuações apenas nos instantes

iniciais resultando em uma amplitude inferior, enquanto que para os demais batimentos as flutuações se prolongam e têm amplitudes mais acentuadas.

Ambos atributos são utilizados para treinamento e teste de uma Rede Neural com Quantização Vetorial por Aprendizagem, que é baseada na classificação pelo princípio do vizinho mais próximo e utiliza uma aprendizagem competitiva. Esta rede é formada por três camadas: uma de entrada, uma camada de Kohonen com 50 neurônios e uma camada de saída com 2 neurônios, representando as classes CPV ou não CPV. Esta última inclui batimentos normais e Contração Atrial Prematura. Os atributos são calculados para cada batimento cardíaco. Foram utilizados 20 registros da base de dados MIT/BIH-ARDB. Os resultados obtidos foram: 90,26% de sensibilidade, 92,31% de preditividade positiva e 98,90% de acurácia.

No trabalho de Ataollah e Ali (2010) foi proposta uma metodologia para classificação de CPV baseada em três estágios: no primeiro, a Transformada Wavelet Estacionária foi utilizada para remoção de ruídos devido a atividade muscular e ao mal contato dos eletrodos com a pele e para suavização do sinal resultante foi aplicado um filtro Savitsky-Golay<sup>1</sup>; no segundo, 10 atributos morfológicos foram calculados para cada ciclo cardíaco, sendo as amplitudes e posições dos picos (valores máximos) das ondas P, R e T e vales (valores mínimos) das ondas S e Q, e um atributo temporal dado pela equação

$$IR_i = \frac{T_i - T_{i-1}}{T_{i+1} - T_i}, \quad (120)$$

onde  $T_i$  representa a localização temporal do pico da  $i$ -ésima onda R.

No terceiro estágio, os autores utilizam uma Rede Neural Perceptron de múltiplas camadas. Estas redes neurais foram estudadas para diferentes números de camadas e algoritmos de aprendizagem. Quatro arquiteturas de redes foram propostas, distinguindo-se apenas na quantidade de camadas ocultas e de neurônios nestas camadas. Duas redes neurais, NET1 e NET2, foram configuradas com uma camada oculta, tendo 35 e 45 neurônios em cada uma delas, respectivamente. As demais redes neurais foram projetadas com duas camadas ocultas. Para a rede neural NET3 foram considerados 30 e 20 neurônios na primeira e segunda camadas, respectivamente, e para a NET4, 30 e 15 neurônios. Para todas as arquiteturas propostas, as camadas de entrada e saída foram fixadas com 11 e 3 neurônios, respectivamente, sendo que estes três neurônios correspondem as classes Normal, CPV e Outras. O algoritmo Backpropagation foi aquele com a melhor taxa de convergência enquanto que, com o método Levenberg-Marquardt, eles obtiveram a melhor acurácia. Os autores Ataollah e Ali (2010) compararam duas funções de ativação na camada de saída, sigmoide e linear, e concluíram que a segunda fornece melhores resultados. Os autores obtiveram uma acurácia de 95,40% utilizando 7 registros da base de dados MIT/BIH-ARDB.

<sup>1</sup> Filtro digital de ajuste polinomial que utiliza o método dos mínimos quadrados.

Outro trabalho que também utiliza redes neurais é aquele apresentado por Omer, Laurent e Gregory (2006). Os autores propõem a utilização da transformada wavelet diádica para gerar o vetor de atributos para treinamento/teste da rede neural. São utilizados seis conjuntos de vetores de atributos espectrais, obtidos a partir de cinco níveis de decomposição da transformada wavelet, e um atributo temporal, obtido pela razão normalizada entre os intervalos RR do sinal original, calculada pela equação (120).

Segundo os autores, a utilização de atributos espectrais e temporais garantem uma precisão melhor na tarefa de classificação, visto que somente as morfologias dos complexos QRS dos sinais de ECG não são suficientes para distinguir as arritmias, já que algumas apresentam similaridades. Assim, no método proposto pelos autores, para cada nível de decomposição da transformada wavelet, os coeficientes wavelets retornados são utilizados juntamente com a razão  $IR_i$ , para algum  $i$ , como entrada para uma rede neural. Assim, são projetadas redes neurais com 23, 27, 33, 44 e 65 neurônios na camada de entrada. Na única camada oculta existente, a mesma quantidade de neurônios é utilizada. A camada de saída é composta por três neurônios, correspondentes as classes Normal, CPV e Outros.

Na etapa de treinamento são utilizados 540 batimentos de 18 registros da base de dados MIT/BIH-ARDB. Estes registros foram escolhidos de modo que possuíssem exemplos de várias arritmias que ocorrem com maior frequência. Para a etapa de teste, 93.281 batimentos foram selecionados de 40 registros. Segundo os autores, apenas 2% dos dados foram utilizados simultaneamente nas etapas de treinamento e teste. O melhor resultado obtido foi com a utilização dos coeficientes wavelets na quarta escala juntamente com a informação temporal retida na série  $IR_i$ . As medidas de performance obtidas foram 95,20% de acurácia, 85,20% de sensibilidade e 92,40% de preditividade positiva. No entanto, os autores também apresentam medidas de performance maiores que estas, mas para uma base de dados muito reduzida.

Na pesquisa proposta por Li et al. (2014), na etapa de pré-processamento os autores primeiramente utilizam um filtro média móvel para remover mudança de linha de base e um filtro FIR passa-baixa 12-taps<sup>2</sup> para remoção de ruído de alta frequência. Diferentemente dos demais trabalhos citados, eles optaram por construir um mecanismo de aprendizagem próprio, baseado em correspondência de modelos<sup>3</sup>. Para atingir este objetivo, dois modelos foram propostos, nomeados como  $T1$  e  $T2$ , relativos ao intervalo entre batimentos normais e ao intervalo entre as amplitudes das formas de ondas normais, respectivamente.

Para construir estes modelos os autores tomaram cinco minutos de registro de ECG. Para o modelo  $T1$ , a distância entre as posições das ondas R foram calculadas, obtendo um sinal denominado de Variabilidade da Frequência Cardíaca. Os pontos deste foram ordenados em ordem crescente, e divididos em cinco subconjuntos com o mesmo tempo de duração. O subconjunto com maior quantidade de pontos foi utilizado para designar o intervalo batimento-

<sup>2</sup> A expressão taps designa o número de coeficientes do filtro.

<sup>3</sup> O termo empregado na literatura é *Template matching*.

por-batimento normal. Analogamente foi construído o modelo T2, mas neste caso as amplitudes dos complexos QRS foram consideradas.

Para o reconhecimento dos batimentos CPV, os autores propuseram a utilização das equações (121) e (122) que retornam o coeficiente de correlação normalizado (Li et al., 2014):

$$x_k = \frac{\sum_{n=0}^{L-1} [b_k(n) - \bar{b}_k] [T1(n) - \bar{T}1]}{\sqrt{\sum_{n=0}^{L-1} [b_k(n) - \bar{b}_k]^2 [T1(n) - \bar{T}1]^2}} e \quad (121)$$

$$y_k = \frac{\sum_{n=0}^{M-1} [c_k(n) - \bar{c}_k] [T2(n) - \bar{T}2]}{\sqrt{\sum_{n=0}^{M-1} [c_k(n) - \bar{c}_k]^2 [T2(n) - \bar{T}2]^2}}, \quad (122)$$

onde  $b_k(n)$  é o  $k$ -ésimo complexo QRS,  $L$  é o comprimento do modelo  $T1$ ,  $\bar{b}_k$  e  $\bar{T}1$  são os valores médios destas sequências,  $c_k(n)$  representa o episódio entre os batimentos  $k$  e  $k + 1$  e  $M$  é o comprimento do modelo  $T2$ .

Assim, o coeficiente de correlação, conforme equações (121) e (122), de um batimento CPV deve apresentar um valor baixo, já que os modelos  $T1$  e  $T2$  representam eventos não-CPV do ECG. Para considerar simultaneamente os coeficientes de correlação para ambos modelos, os autores propuseram a utilização da função normalizada

$$z_k = f(x_k, y_k) = \frac{e^{x_k^r} + e^{y_k^r}}{2e},$$

onde  $r$  é a taxa de declive da curva e  $e$  é a base de Euler.

Para reconhecer um batimento como não-CPV, os autores compararam os valores de  $z_k$  com um limiar  $z_{thre}$ , obtido experimentalmente, cujo valor ótimo foi estabelecido em 0,55 para  $r = 4$ .

Além da base de dados MIT/BIH-ARDB os autores também experimentaram com a base INCART, no entanto, os resultados foram inferiores. O melhor resultado foi 98,2% de acurácia, 93,1% sensibilidade e 98,5% de especificidade, 81,4% e 99,5% de preditividade positiva e negativa, respectivamente. Assim como em Omer, Laurent e Gregory (2006) os autores Li et al. (2014) obtiveram resultados muito melhores utilizando uma base de dados com apenas seis registros.

Zarei et al. (2016) também propuseram um algoritmo de aprendizagem própria que toma por parâmetros as direções dos vetores fornecidos pela Análise de Componentes Principais. A metodologia proposta é dividida em dois estágios denominados processamento e detecção de anormalidade. No primeiro estágio três módulos foram implementados, a saber: detecção de batimento cardíaco, segmentação e normalização. No segundo estágio, os autores empregaram a Análise de Componentes Principais e uma estratégia de substituição de um segmento normal por um anormal, analisando em seguida a variação sofrida pelos vetores.

Em resumo, para cada segmento de dois minutos do sinal de ECG, os autores Zarei et al. (2016) obtiveram sub-segmentos  $\mathbf{s}_i = [s_1, \dots, s_{150}]$ , consistindo de 50 pontos antes e 100 pontos

após a localização das ondas R. Eles então selecionaram  $k$  destes sub-segmentos do tipo Normal e criaram uma matriz  $\mathbf{M} = [\mathbf{s}_1 \cdots \mathbf{s}_k]^\top \in \mathbb{R}^{150 \times k}$ , calculando subseqüentemente seu vetor principal. Em seguida, o último sub-segmento da matriz  $\mathbf{M}$  é substituído por um segmento Normal e outro sub-segmento CPV, resultando em novas matrizes  $\mathbf{M}_1$  e  $\mathbf{M}_2$ , e novamente os vetores principais foram calculados. Os autores constataram que quando se compararam os vetores  $\mathbf{v}_M$ ,  $\mathbf{v}_{M_1}$  e  $\mathbf{v}_{M_2}$  associados as matrizes  $\mathbf{M}$ ,  $\mathbf{M}_1$  e  $\mathbf{M}_2$ , respectivamente, notava-se um desvio significativo na direção destes na comparação entre  $\mathbf{v}_M$  e  $\mathbf{v}_{M_2}$ , enquanto que na comparação entre  $\mathbf{v}_M$  e  $\mathbf{v}_{M_1}$ , o desvio era insignificante.

No estágio de detecção de anormalidade, os ângulos entre as direções principais são calculados pela fórmula

$$\theta = \arccos \left( \frac{\mathbf{v}_M \cdot \mathbf{v}_{M_2}}{|\mathbf{v}_M| |\mathbf{v}_{M_2}|} \right).$$

Primeiramente, na etapa de treinamento, os valores destes ângulos foram armazenados no vetor  $\theta_O$ . Na etapa de detecção dois limiares adaptativos foram calculados

$$\begin{aligned} T_1 &= \bar{\theta}_u r_1, \\ T_2 &= \bar{\theta}_q r_2, \end{aligned}$$

onde  $\theta_u$  e  $\theta_q$  armazenam os ângulos para as direções principais e para os batimentos não-CPV, respectivamente. Os valores iniciais destes ângulos foram obtidos do vetor  $\theta_O$ . Assim, quando um novo ângulo  $\theta_x$  foi apresentado ao algoritmo, eles verificaram se este era superior ao limiar  $T_2$ , ficando caracterizado como batimento CPV e, caso contrário, como um batimento Normal, e para este último o vetor  $\theta_q$  foi atualizado com o novo valor. Para a obtenção dos parâmetros  $k$ ,  $r_1$  e  $r_2$  foram realizados experimentos com nove registros da base de dados MIT/BIH-ARDB, sendo fixados como  $k = 10$ ,  $r_1 = 3$  e  $r_2 = 9$ . Os resultados gerais de performance foram 98,77% de acurácia, 96,12% de sensibilidade e 86,48% de preditividade positiva.

No trabalho proposto por Shen et al. (2011) os autores implementaram três diferentes metodologias para extração de seis atributos, as quais foram divididas em temporais, frequenciais e morfológicas. O único atributo temporal consistiu da distância entre duas ondas R sucessivas. Os atributos morfológicos também foram baseados nestas ondas, como os resultados da altura, largura e inclinação destas, sendo que a inclinação diz respeito a onda estar abaixo ou acima do eixo isoeletrico. Já os atributos frequenciais foram obtidos a partir da aplicação da Transformada de Fourier a um segmento do sinal com 512 pontos centrado na onda R, a saber: amplitude máxima dos coeficientes da transformada e a frequência correspondente da densidade espectral de potência, obtida pelo método de Welch.

Na fase de pré-processamento os autores Shen et al. (2011) propuseram a aplicação de dois filtros passa-baixa com frequência de corte de 3 dB sobre 35 Hz e 4 Hz, cujas equações de diferenças são  $y(n) = y(n-1) + x(n) - x(n-6)$  e  $y(n) = y(n-1) + x(n) - x(n-60)$ , respectivamente, onde  $y(n)$  é a saída e  $x(n)$  a entrada.

Na etapa de treinamento/teste do classificador os atributos obtidos foram fornecidos para uma Máquina de Suporte Vetorial, cuja função núcleo utilizada foi uma de base radial, dada por  $K(x_i, x_j) = e^{\left[-\frac{(x_i - x_j)^2}{2\sigma^2}\right]}$ . Os autores apenas apresentaram a medida de acurácia no valor de 97,00%, utilizando nove registros da base de dados MIT/BIH-ARDB. Eles ainda ressaltaram que sua metodologia é adequada para aplicações em tempo-real.

No trabalho proposto por Ebrahimpour et al. (2013) é implementado uma estrutura de comitê de máquinas onde cada máquina é uma Rede Neural Perceptron com uma cama oculta. Cada uma dessas redes gera uma determinada saída as quais são ponderadas por outra rede do tipo *gating*, com uma camada de saída do tipo softmax. A combinação ponderada das saídas das redes são então utilizada no algoritmo *backpropagation* a fim de determinar o peso de cada uma delas no comitê de máquinas. Os autores também implementam um mecanismo de filtragem do espaço de entrada para alimentar cada uma das máquinas.

Na Tabela 7 estão resumidas as principais características das metodologias apresentadas e os resultados obtidos por algumas delas, relativamente àquelas que serão utilizadas para comparação com as abordagens propostas.

Tabela 7 – Resumo das metodologias pesquisadas para reconhecimento de CPV.

Referência	Pré-processamento	Etapas		Resultados		
		Atributos	Classificação	$A_{cc}$	$P_+$	$S_e$
Liu et al. (2015)	Transformada Wavelet limiar auto-adaptativo	A partir da derivada do expoente de Lyapunov	Rede neural com Quantização Vetorial por Aprendizagem	98,90%	92,31%	90,26%
Ataollah e Ali (2010)	Transformada Wavelet estacionária e filtro Savitsky-Golay	Amplitudes e posições das ondas do ECG e distância RR normalizada	Rede neural Perceptron com múltiplas camadas	95,40%	-	-
Omer, Laurent e Gregory (2006)	Não consta	Coefficientes wavelets e distância RR normalizada	Rede neural Perceptron com múltiplas camadas	95,20%	92,40%	85,20%
Li et al. (2014)	Filtro Média Móvel e FIR de 12-taps	Coefficientes de correlação dos modelos	Correspondência de modelos	98,20%	90,45%	95,80%
Zarei et al. (2016)	Normalização	Ângulo entre vetores principais	Varição das direções principais	98,77%	86,48%	96,12%
Shen et al. (2011)	Dois filtros passa-baixa	Distância entre ondas R, sua altura, largura e inclinação. Frequência máxima de Fourier e densidade espectral correspondente	Máquina de Suporte Vetorial	97,00%	-	-

Fonte: Autoria própria.

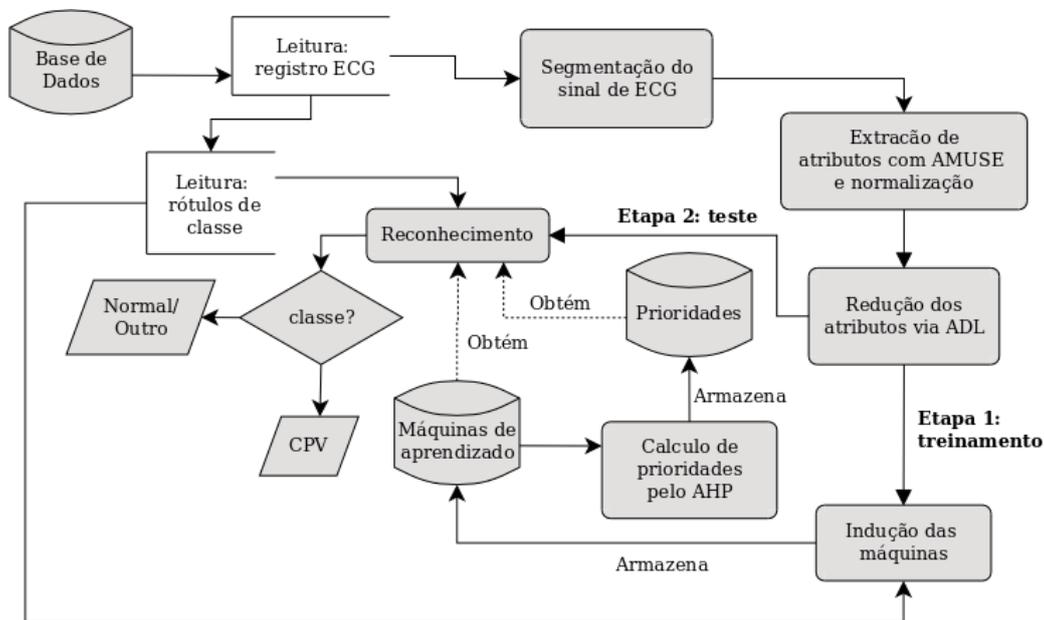
## Parte II

Metodologia proposta, resultados e discussões

## 5 Abordagens Propostas

Neste capítulo são descritas as abordagens propostas para solução do problema de reconhecimento de CPV. São vários os blocos de procedimentos necessários para atingir esta finalidade, e dentro de cada escopo uma abordagem específica precisa ser implementada. Na Figura 15 está ilustrado um panorama da abordagem proposta, onde cada um dos retângulos com cantos arredondados representa um bloco de procedimentos, os quais são pormenorizadamente explicados a seguir.

Figura 15 – Diagrama de blocos da proposta principal.



Fonte: Autoria própria.

### 5.1 Segmentação do sinal de ECG

Após a leitura de um sinal de ECG a partir de alguma base de dados é necessária a implementação de alguma segmentação/janelamento deste sinal porque sua extensão é demasiado longa para processamento por completo. Por exemplo, a base de dados MIT/BIH-ARDB foi amostrada pelo eletrocardiógrafo a 360 Hz, e seus sinais possuem em média 30 minutos de duração, logo, cada sinal tem cerca de 648.000 amostras cada um.

A segmentação implementada aqui consiste primeiramente da localização das ondas fiduciais do sinal de ECG, a fim de que cada segmento contenha as mesmas deflexões. A deflexão de maior significância nos sinais de ECG é geralmente a onda R, e é sobre esta que a maioria dos métodos de segmentação se baseiam. Para atingir este fim é necessário a concepção de um

algoritmo para detecção das ondas R. Entretanto, neste trabalho tal algoritmo será suprimido, porque serão utilizadas as informações disponíveis na própria base de dados sobre a localização de tais ondas.

Assim, após obter estas localizações a partir da base de dados, o sinal de ECG inteiro é truncado multiplicando-o por uma janela retangular de tamanho  $J^s \Delta^s$ , onde  $J^s$  é o comprimento (em segundos) da janela e  $\Delta^s$  é a taxa de amostragem, em Hz, dos sinais de ECG. Esta janela é posicionada de tal modo que seu fim ocorra pelo menos 0,38 segundos após a localização da onda R e seu início pelo menos 0,2 segundos antes, de acordo com a duração média dos segmentos PR e QT, que totalizam cerca de 0,58 segundos, conforme mencionado na Seção 2.2. Logo,

$$J^s = J_e^s + J_d^s \geq 0,58, \quad (123)$$

$$J_e^s \geq 0,2 \text{ e} \quad (124)$$

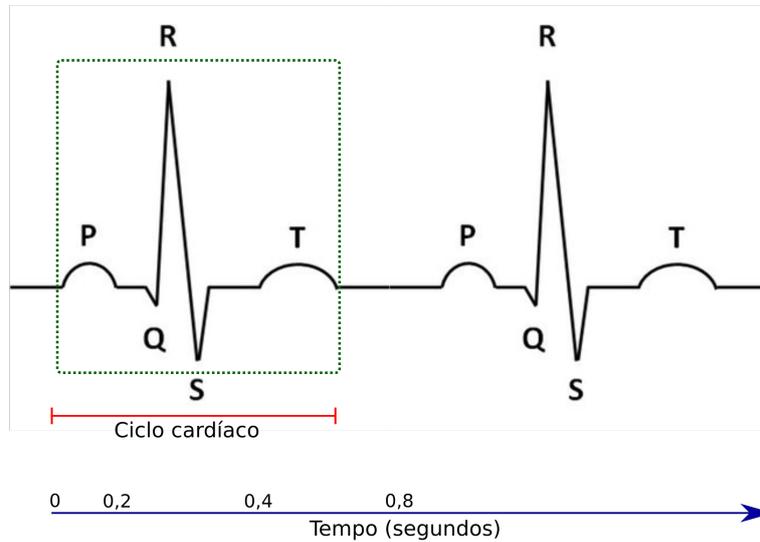
$$J_d^s \geq 0,38, \quad (125)$$

sendo  $J_e^s$  e  $J_d^s$  os segmentos da metade esquerda e direita da janela de comprimento  $J^s$ .

Se o comprimento  $J^s$  é inferior a 0,75 segundos, que corresponde à duração média de um ciclo cardíaco (GUYTON; HALL, 2006, p. 124), partes das ondas P ou T poderão não ser consideradas. Portanto, esse comprimento de janela será utilizado nos experimentos a seguir, e consequentemente  $J_e^s = 0,37$  e  $J_d^s = 0,38$ .

Na Figura 16 está ilustrado o bloco de segmentação de um ciclo cardíaco, representado pelo retângulo pontilhado verde. Este procedimento é realizado para cada onda R, sucessivamente, sendo que na implementação computacional, cada um destes segmentos é armazenado para posterior execução dos demais blocos.

Figura 16 – Segmentação do sinal de ECG. Segmentos de retas em vermelho e azul destacam o primeiro ciclo cardíaco e o eixo de tempo, respectivamente. O retângulo pontilhado verde representa o janelamento referente a um ciclo cardíaco.



Fonte: Autoria própria.

Portanto, dado um sinal de ECG,  $\mathbf{x}(p)$ , amostrado a  $\Delta^s$  Hz, o  $k$ -ésimo segmento  $\mathbf{x}_k(m)$ , de comprimento  $J^s$ , é obtido assim:

$$\mathbf{x}_k(m) = \mathbf{x}(p), \text{ com } R_k - J_e^s \Delta^s \leq p \leq R_k + J_d^s \Delta^s, \quad (126)$$

onde  $p = 0, 1, \dots, P-1$ ,  $m = 0, 1, \dots, J^s \Delta^s - 1$ ,  $k = 1, 2, \dots, \#R$ , sendo  $R_k < P$  a localização da  $k$ -ésima onda R,  $P$  o comprimento do sinal de ECG e  $\#R$  a quantidade de ondas R presentes no sinal. Deste modo, a segmentação resulta em uma matriz  $\mathbf{X} \in \mathbb{R}^{\#R \times J^s}$ :

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1(0) & \mathbf{x}_1(1) & \cdots & \mathbf{x}_1(J^s - 1) \\ \mathbf{x}_2(0) & \mathbf{x}_2(1) & \cdots & \mathbf{x}_2(J^s - 1) \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{x}_{\#R}(0) & \mathbf{x}_{\#R}(1) & \cdots & \mathbf{x}_{\#R}(J^s - 1) \end{bmatrix}. \quad (127)$$

## 5.2 Extração de atributos e normalização

A extração de atributos é uma tarefa essencial em sistemas de reconhecimentos de padrões, conforme já destacado na Seção 3.2. Neste trabalho propõe-se a utilização do método de separação cega de fontes AMUSE para esta finalidade.

Na Seção 3.3 foi mencionado que o método AMUSE, dada uma matriz constituída de sinais fontes misturados, é capaz de estimar as fontes isoladamente, mas sem garantir a manutenção das amplitudes e a ordem destas fontes, podendo também ocorrer uma defasagem de  $180^\circ$ .

Afim de compreender como o método AMUSE pode ser aplicado para o propósito de reconhecimento de CPV, primeiramente é necessário entender como este método pode ser utilizado para analisar um vetor, já que este deve receber uma matriz como entrada, no entanto, não é viável fornecer toda a matriz de segmentação da equação (127), pois nela estão representados tanto os batimentos cardíacos Normais como os CPV. Mesmo que os tipos de batimentos fossem separados em matrizes distintas, ainda assim, não seria viável utilizar estas matrizes, porque neste caso seria possível apenas extrair os atributos de todos os batimentos representados simultaneamente, logo, uma classificação sobre o sinal pleno seria realizada, mas o que se pretende é reconhecer cada um dos batimentos cardíacos individualmente.

Assim, para resolver o problema da dimensão relativa ao dado de entrada para o método AMUSE, propõe-se criar uma matriz  $\mathbf{Y}_k \in \mathbb{R}^{2 \times J^s - 2}$  a partir de cada segmento  $\mathbf{x}_k(m)$  e uma versão defasada deste, isto é:

$$\mathbf{Y}_k = \begin{bmatrix} \mathbf{x}_k(0) & \mathbf{x}_k(1) & \mathbf{x}_k(2) & \cdots & \mathbf{x}_k(J^s - 2) \\ \mathbf{x}_k(1) & \mathbf{x}_k(2) & \mathbf{x}_k(3) & \cdots & \mathbf{x}_k(J^s - 1) \end{bmatrix}. \quad (128)$$

Explica-se a exigência de inserção de uma versão defasada ao invés da repetição do segmento, resgatando a lembrança do primeiro passo na implementação do método AMUSE, que é o cálculo da matriz de covariância  $\mathbf{R}_Y$ . Se as linhas da matriz  $\mathbf{Y}_k$  forem iguais, ou seja, se não for considerada a segunda linha como uma versão defasada do  $k$ -ésimo segmento do sinal de ECG, isso resultará em uma matriz de covariância nula, pois, conforme a equação (71), a média de cada coluna será igual ao próprio elemento da coluna  $j$ , ou seja,  $\bar{\mathbf{y}}_j = \mathbf{y}_i(j), \forall i$ , logo, os valores  $r_j^2$  e  $r_{jk}$ , equação (71), serão nulos para todo  $j$ .

Outras versões defasadas também podem ser acrescentadas à matriz  $\mathbf{Y}_k$ , sendo que para  $\nu$  versões defasadas o método AMUSE estimará  $\nu + 1$  fontes ou componentes. Nos experimentos realizados no Capítulo 6 serão considerados vários modelos para a matriz  $\mathbf{Y}_k$ .

É importante esclarecer que nesta abordagem estão sendo utilizadas versões do mesmo sinal de ECG, entretanto, o método AMUSE é empregado para a separação de fontes, e a matriz  $\mathbf{Y}_k$  representa apenas uma fonte, mas em duas versões. Recordando que as taxas de amostragens são maiores ou iguais a 360 Hz, é esperado que, na prática, as versões sejam muito próximas em amplitude, já que o conteúdo frequencial dos sinais de ECG não devem ultrapassar 180 Hz, logo, as formas de onda, representadas pelas linhas das matrizes, são semelhantes.

Para resolver esta incompatibilidade, primeiramente é necessário recordar alguns aspectos característicos dos sinais. Pelo que foi apresentado na Seção 3.1, um sinal pode ser representado como a sobreposição de senoides, conforme equação (2). Portanto, se o sinal não for um tom puro, então ele é composto pela sobreposição de outros sinais, logo, é bastante aceitável supor que um sinal natural, tal como o sinal de ECG, seja formado por distintas fontes secundárias de sinais, que juntas são percebidas como uma única fonte principal.

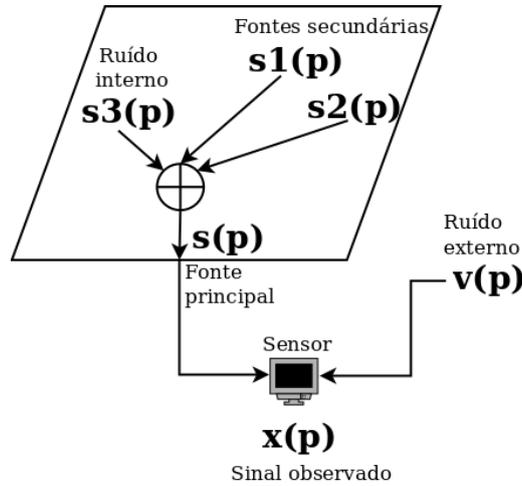
Embora o resultado sobre sobreposição senoidal da equação (2) seja relativo a sinais periódicos, dado um segmento de sinal de ECG, pode-se considerar repetições deste, gerando assim um sinal periódico, cujo período é o próprio comprimento do segmento. Portanto, o conceito de sobreposição pode ser aplicado no caso em estudo, sem violação do teorema sobre séries de Fourier.

O fato que comprova essa afirmação sobre a sobreposição de sinais, foi discutido na Seção 2.2, e particularmente exibido na Figura 2, onde se constata que o sinal de ECG é o resultado da adição dos potenciais de ação disparados pelos cardiomiócitos do músculo atrial e ventricular, nodo SA e AV, e Sistema de Purkinje. Neste domínio, pode-se então considerar cada um dos músculos cardíacos especializados como uma fonte distinta, e os eletrodos do eletrocardiógrafo sendo os sensores que captam a sobreposição destas fontes. Visto que apenas um canal do eletrocardiógrafo está sendo considerado neste trabalho, então somente um sensor e, conseqüentemente, um sinal de ECG são tomados nos experimentos. No entanto, devido a utilização de versões defasadas do sinal de ECG, a modelagem implementada, discriminada pela matriz  $\mathbf{Y}_k$ , refere-se a mais de um sensor. Nesse contexto, o número de linhas da matriz de segmentação  $\mathbf{Y}_k$  diz respeito a quantidade de sensores virtuais considerados no sistema de mistura artificial projetado.

Os dados discriminados na Tabela 1 da Seção 2.1, também auxiliam na constatação de que o sinal de ECG é composto pela sobreposição de fontes. Nota-se que a duração do potencial de ação do Nódulo SA varia entre 100 ms e 300 ms, logo em seguida as células do músculo atrial são disparadas com o mesmo intervalo de duração. Ou seja, mesmo após o impulso elétrico ter atingido o músculo atrial o potencial do Nódulo SA exerce influência sobre o sinal de ECG, já que ele não se dissipa imediatamente. A mesma conclusão pode ser obtida ao analisar a duração do potencial de ação dos demais sistemas de condução.

Na Figura 17 está ilustrado este comportamento de sobreposição dos sinais relativos ao sistema projetado.

Figura 17 – Diagrama da composição de uma fonte observável  $\mathbf{x}(p)$  resultante da sobreposição de fontes secundárias e ruído.



Fonte: Autoria própria.

Observa-se na Figura 17 que são admitidas duas fontes de ruídos, diferentemente do método AMUSE. Esta inserção de uma fonte interna de ruído é necessária devido ao que foi observado sobre a sobreposição dos sinais, isto é, supõe-se que eles sejam compostos pela sobreposição de distintos tons puros, harmônicos e ruídos. Já o ruído externo é recebido pelo sensor juntamente com o sinal  $\mathbf{x}$ .

Logo, o sinal de entrada  $\mathbf{x}(p)$  pode ser escrito como:

$$\mathbf{x}(p) = \sum_k \mathbf{s}_k(p) + \mathbf{v}(p), \quad (129)$$

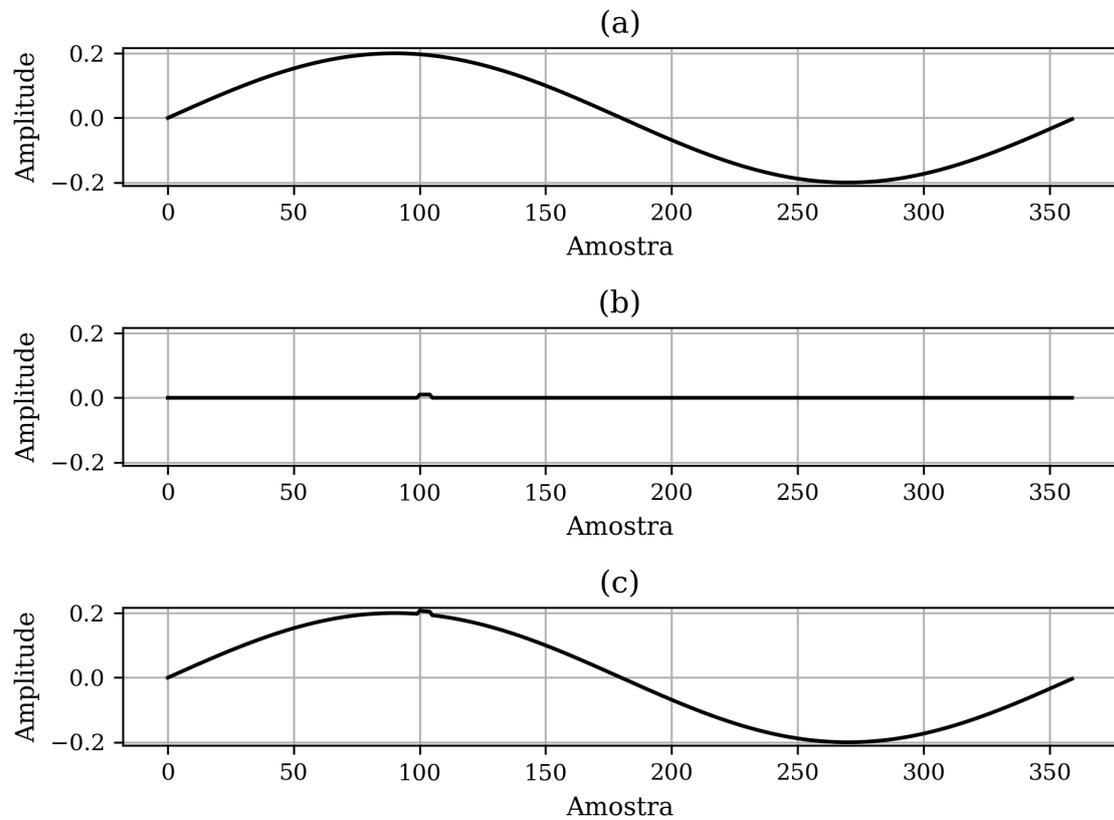
onde  $\mathbf{s}_k$  são as fontes secundárias, incluindo um provável ruído interno, formadoras da fonte principal  $\mathbf{s}$ , e  $\mathbf{v}$  é o ruído externo.

Com o objetivo de compreender como o método AMUSE atua separando as fontes secundárias, formadoras da fonte principal, apresenta-se um experimento onde, a partir de um sinal  $\mathbf{x}(p)$  constrói-se a matriz  $\mathbf{Y}$ , equação (128). O sinal  $\mathbf{x}(p)$  é composto pela sobreposição de dois sinais: uma senoide  $\mathbf{s}_1(p)$  de 1 Hz com duração de 1 segundo, e um pulso  $\mathbf{s}_2(p)$  com duração de 0,0138 segundo com amplitude de 0,001, encarada aqui como uma singularidade:

$$\mathbf{x}(p) = \mathbf{s}_1(p) + \mathbf{s}_2(p) \quad (130)$$

$$= \sin(2\pi t) + \begin{cases} 0,001, & \text{se } 0,27 \leq t \leq 0,29 \\ 0, & \text{caso contrário} \end{cases} \quad (131)$$

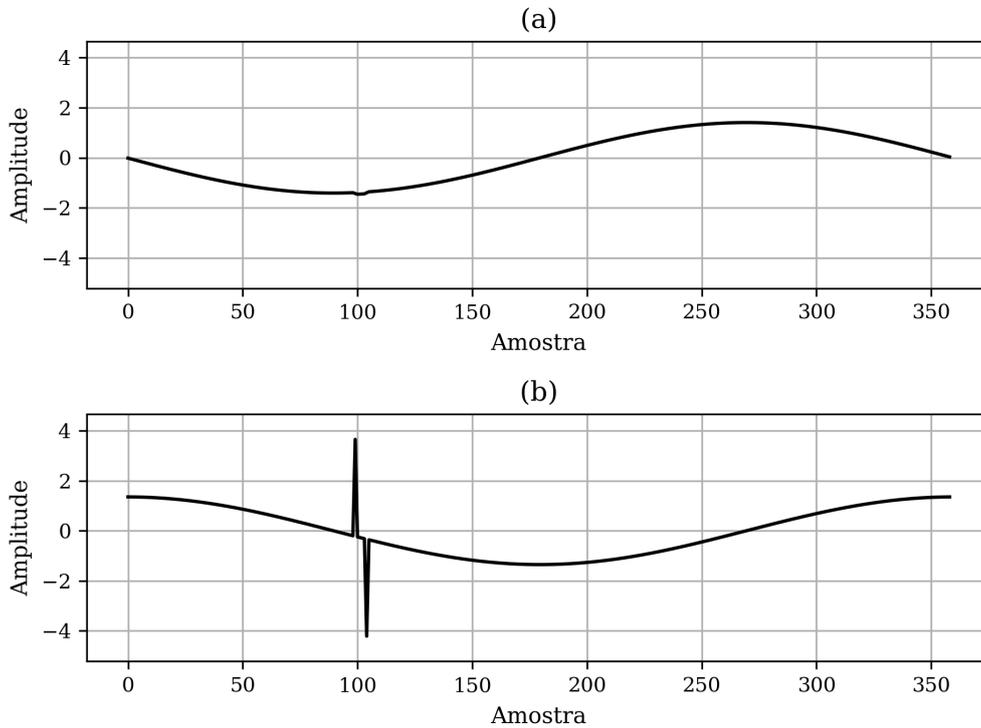
As Figuras 18, 19 e 20 exibem: o sinal  $\mathbf{x}(p)$  e suas componentes e duas e quatro componentes estimadas pelo método AMUSE, respectivamente.

Figura 18 – (a) Sinal  $s_1(p)$ , (b) singularidade  $s_2(p)$ ; e (c) sinal observado  $x(p)$ .

Fonte: Autoria própria.

Na abordagem proposta, o sinal  $x(p)$  é inserido na matriz  $\mathbf{Y}$  juntamente com uma versão defasada sua, sendo então esta matriz submetida ao método AMUSE. Visto que dois sinais compõem a matriz de mistura  $\mathbf{Y}$ , o AMUSE retorna duas componentes, as quais estão exibidas na Figura 19.

Figura 19 – Representação das componentes estimadas pelo método AMUSE, (a) e (b), considerando o sinal  $\mathbf{x}(p)$ , equação (130) e uma versão defasada.



Fonte: Autoria própria.

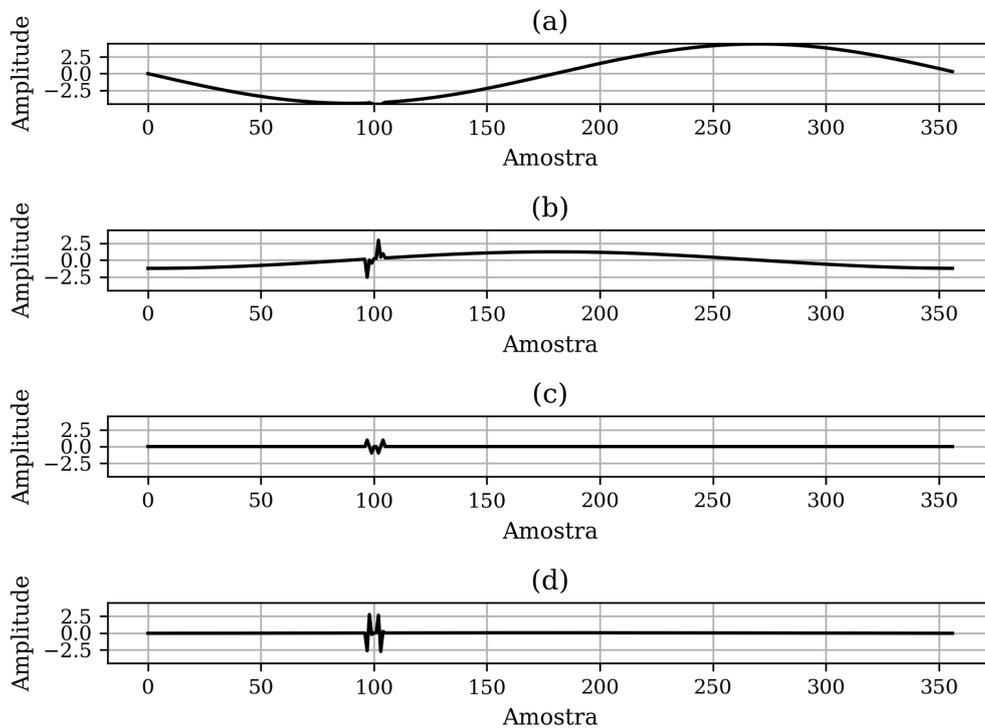
A componente exibida na Figura 19 (a) é semelhante ao sinal observado  $\mathbf{x}(p)$  mas com a amplitude aumentada e a fase alterada em  $180^\circ$ , característica que é esperada na separação de fontes, de acordo com a Definição 7 da Seção 3.3. Por outro lado, na segunda componente estimada, exibida na Figura 19 (b), a ocorrência da singularidade foi ressaltada, embora a singularidade propriamente não foi estimada, mas apenas o entorno de sua localização, qualidade esta que implica em acréscimo de informação.

Ao tomar mais duas componentes defasadas, resultando na matriz:

$$\mathbf{Y}' = \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \mathbf{x}(2) & \mathbf{x}(3) & \cdots & \mathbf{x}(P-4) \\ \mathbf{x}(1) & \mathbf{x}(2) & \mathbf{x}(3) & \mathbf{x}(4) & \cdots & \mathbf{x}(P-3) \\ \mathbf{x}(2) & \mathbf{x}(3) & \mathbf{x}(4) & \mathbf{x}(5) & \cdots & \mathbf{x}(P-2) \\ \mathbf{x}(3) & \mathbf{x}(4) & \mathbf{x}(5) & \mathbf{x}(6) & \cdots & \mathbf{x}(P-1) \end{bmatrix}, \quad (132)$$

o método AMUSE estima quatro componentes, conforme ilustrado pelas suas formas de onda na Figura 20.

Figura 20 – Representação das componentes estimadas pelo método AMUSE, (a), (b), (c) e (d), considerando o sinal  $\mathbf{x}(p)$ , equação (130), e três versões defasadas.



Fonte: Autoria própria.

Observa-se que as duas primeiras componentes estimadas, Figuras 20 (a) e (b) são semelhantes, em relação a forma de onda, às componentes exibidas nas Figuras 19 (a) e (b) que foram estimadas tomando apenas um sinal defasado, exceto pela amplitude das componentes. Diferentemente, nas componentes estimadas exibidas nas Figuras 20 (c) e (d) não houve interferência da componente senoidal  $s_1(p)$ , nem mesmo a componente impulsiva foi estimada, mas apenas sua localização ressaltada, em ambas componentes, Figuras 20 (c) e (d), mas de modo distinto em cada uma delas.

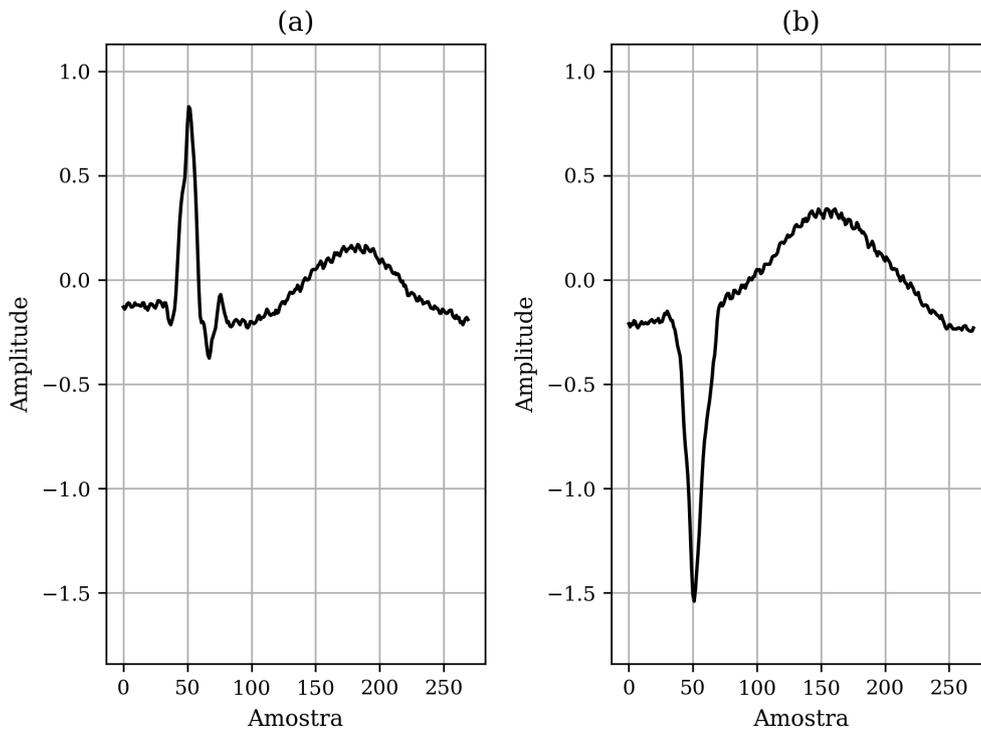
Por meio destas análises verifica-se que o método AMUSE consegue frisar a ocorrência de uma componente pouco significativa no sinal observado, sem estimá-la efetivamente. Além disso, em ambos experimentos o AMUSE estimou uma versão defasada em  $180^\circ$  do sinal observado.

Agora que já se tem uma ideia do comportamento do método AMUSE na estimação de fontes, considerando a modelagem proposta que consiste na criação de uma matriz com o sinal original, menos uma amostra, e uma versão defasada deste, é necessário entender como este comportamento se estende para a análise de sinais de ECG, principalmente na estimação de componentes relacionadas aos batimentos cardíacos Normais e CPV.

As Figuras 21 (a) e (b) ilustram as formas de onda de um exemplo de batimento cardíaco

Normal e outro CPV do registro 207 da base de dados MIT/BIH-ARDB. Em ambos exemplos a onda T é proeminente, e no batimento CPV a onda R tem deflexão negativa.

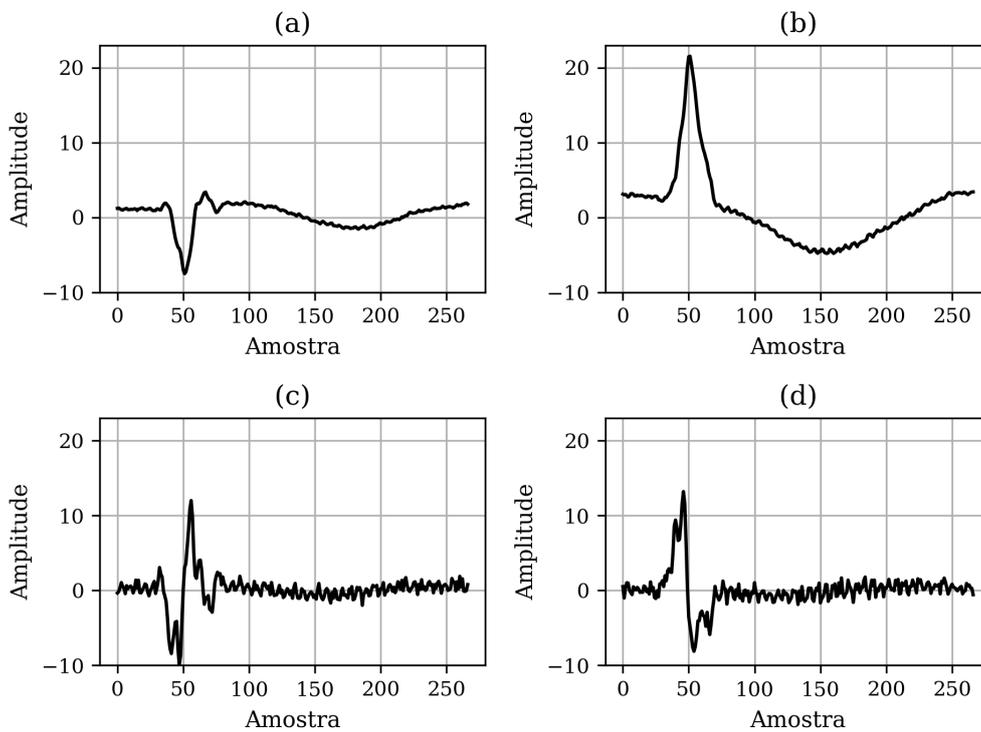
Figura 21 – Exemplos de batimentos cardíacos do tipo (a) Normal e (b) CPV, ambos do registro 207 da base de dados MIT/BIH-ARDB.



Fonte: Autoria própria.

As duas componentes estimadas, para cada tipo de batimento, estão ilustradas na Figura 22. Neste caso foi considerado o sinal original e uma versão defasada, conforme equação (128). Nota-se, que as duas principais componentes, relativas a cada batimento, Figuras 22 (a) e (b), representam versões defasadas em  $180^\circ$  dos batimentos apresentados na Figura 21, com aumento de amplitude. Já as componentes secundárias, Figuras 22 (c) e (d), retratam as variações abruptas do sinal, ou seja, as características de mais alta frequência.

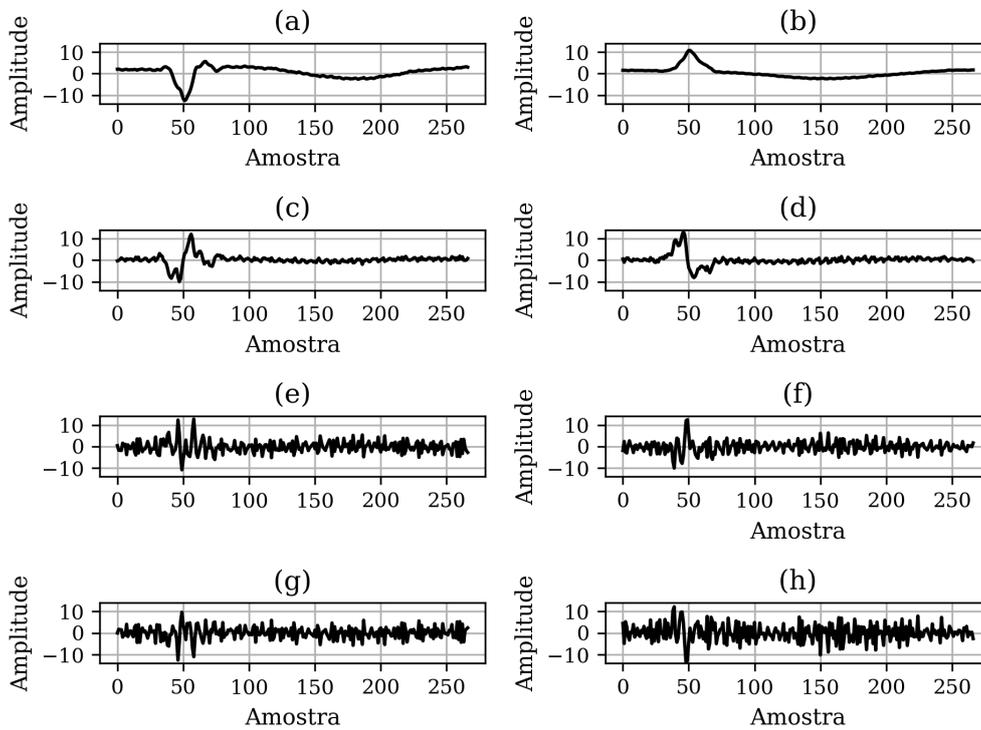
Figura 22 – Componentes estimadas pelo método AMUSE para o batimento cardíaco Normal: (a) e (c); e para CPV: (b) e (d).



Fonte: Autoria própria.

Na Figura 23 estão ilustradas quatro componentes estimadas pelo AMUSE, de cada batimento. Analogamente ao que foi constatado nas análises anteriores, as componentes associadas aos maiores valores singulares, isto é, aqueles que aparecem no topo da Figura 23, representam versões defasadas em  $180^\circ$  dos sinais originais, e a medida que se considera aquelas componentes mais próximas a base da Figura 23, mais as características de alta frequência são ressaltadas.

Figura 23 – Componentes estimadas pelo método AMUSE para o batimento cardíaco Normal: (a), (c), (e) e (g); e para CPV: (b), (d), (f) e (h) .



Fonte: Autoria própria.

No Apêndice B outros batimentos são analisados e suas componentes estimadas, sendo notável que um comportamento semelhante aquele observado acima ocorre, ou seja, as componentes principais, associadas aos maiores valores singulares, são relativas às características de aproximação dos sinais analisados, logo, tais componentes são versões com alguma ambiguidade de amplitude e fase, do sinal original. Por outro lado, as componentes secundárias, associadas aos valores singulares menores, estão relacionadas às mudanças abruptas dos sinais analisados. Este resultado do método AMUSE é análogo aqueles obtidos por bancos de filtros, tal como a Transformada Wavelet Discreta (MALLAT, 2009). A fim de compreender o motivo, a seguir uma discussão mais detalhada é realizada.

Para um sistema de banco de filtros de dois canais a matriz de filtragem implementada é comumente ortogonal. O método AMUSE emprega em sua solução uma transformação de ortogonalização que é aplicada sobre a matriz de mistura latente  $\mathbf{A}_0$  resultando em uma matriz ortogonal  $\mathbf{B}_0^T$ , de acordo com as equações (94) e (97).

Visto que pelo Teorema 3 esta transformação faz com que a solução ortogonal seja também solução do problema real devido a relação de equivalência, a transposta da matriz ortogonal  $\mathbf{B}_0$  é aplicada sobre a matriz de entrada  $\mathbf{Z}$  que contém os sinais observados ortogonalizados,

os quais representam a sobreposição dos vários sinais fontes do sistema, conforme apresentado na equação (108). Esta aplicação resulta na estimação da matriz  $\mathbf{S}_0$ , para a qual cada linha representa a estimativa de uma certa fonte, isto é

$$\hat{\mathbf{S}}(p) = \mathbf{B}^\top \mathbf{Z}(p), \quad (133)$$

onde  $\mathbf{B}\mathfrak{A}\mathbf{B}_0$ , isto é,  $\mathbf{B}$  é uma estimativa de  $\mathbf{B}_0$  no sentido que as fontes estimadas em  $\hat{\mathbf{S}}$  podem estar em ordem distinta e com amplitudes diferentes, conforme Definição (7).

Em outras palavras, o método AMUSE fornece uma matriz ortogonal  $\mathbf{B}^\top$  que separa as fontes numa matriz  $\hat{\mathbf{S}}$ . Logo, a matriz  $\mathbf{B}^\top$  representa uma transformação do tipo filtragem, semelhantemente àquela empregada em bancos de filtros com reconstrução perfeita, e devido sua ortogonalidade sabe-se que ela resultará em sinais com mudanças suaves e abruptas, pelo que foi observado nas seções anteriores. Ou seja, a aplicação desta matriz de filtragem resultará em sinais de mais baixa e mais alta frequências.

Dadas todas as discussões anteriores e o modelo proposto para modificação dos sinais de entrada, que consiste da construção de uma matriz com o sinal original e versões defasadas deste, conclui-se que o método AMUSE pode ser utilizado para extração de atributos dos batimentos cardíacos.

Em seguida a extração dos atributos, propõe-se empregar uma normalização nas componentes estimadas, com o intuito de padronizar, em relação as amplitudes, os dados de entrada para as máquinas de aprendizagem, visto que, as amplitudes das componentes variam significativamente.

Portanto, dada uma componente estimada  $\mathbf{c}(q)$  com  $q = 1, 2, \dots$ , esta é normalizada tomando:

$$\mathbf{c}'(q) = \frac{\mathbf{c}(q)}{\|\mathbf{c}(q)\|_2}, \quad (134)$$

onde  $\|\cdot\|_2$  é a norma Euclidiana.

### 5.3 Redução da Dimensionalidade

O método de extração de atributos proposto na seção anterior retorna um vetor de atributos em um espaço 269-dimensional, quando for considerada apenas uma versão defasada do sinal de ECG, pois os segmentos de ECG são tomados como tendo 0,75 segundos e a taxa de amostragem é de 360Hz, para a base de dados empregada. Isso resulta em segmentos de ECG com 270 amostras, entretanto, é tomada uma amostra a menos para ser factível a construção da matriz de defasagem.

Nota-se nas análises das seções anteriores que alguns atributos adjacentes possuem amplitudes muito próximas. Visto que o algoritmo de classificação utiliza uma função de densidade

gaussiana cujos parâmetros, média e desvio-padrão, são obtidos a partir do conjunto de treinamento, é esperado que as densidades obtidas sejam similares, para atributos adjacentes, acarretando que as probabilidades estimadas para cada atributo também estejam muito próximas. Logo, alguns atributos contribuem com a mesma informação para o reconhecimento do padrão.

A fim de selecionar apenas aqueles atributos mais relevantes para o reconhecimento dos padrões de batimentos cardíacos Normal e CPV, propõe-se o emprego da técnica Análise Discriminante Linear (ADL) de Fisher, apresentada na Seção 3.2.5, para redução da dimensionalidade. No entanto, visto que a classificação pretendida é binária, a aplicação da ADL acarretará na redução para um espaço unidimensional (PEDREGOSA et al., 2011).

Assim, em resumo, após a extração de atributos e normalização, emprega-se a técnica ADL que retornará apenas um simples atributo, que é a combinação linear dos demais, para cada segmento de ECG, o qual será utilizado na etapa ulterior para indução das máquinas de aprendizado, acarretando inerentemente em uma redução no custo computacional para aprendizado e teste das máquinas.

## 5.4 Indução das máquinas de aprendizado e obtenção das prioridades pelo método AHP

Após a etapa de extração de atributos, onde cada vetor de atributos é uma componente fornecida pelo método AMUSE, é necessário induzir máquinas de aprendizado as quais gerarão hiperplanos de decisão para classificação das componentes como pertencentes a um batimento Normal ou CPV.

Pelo esquema proposto, Figura 15, cada batimento cardíaco estará associado a várias componentes do AMUSE, conforme foi mostrado nos experimentos anteriores. Assim, é necessário induzir uma máquina para cada componente (vetor) combinando suas predições posteriormente em um comitê de máquinas, ou, alternativamente, concatenar as componentes a fim de induzir uma máquina solitária.

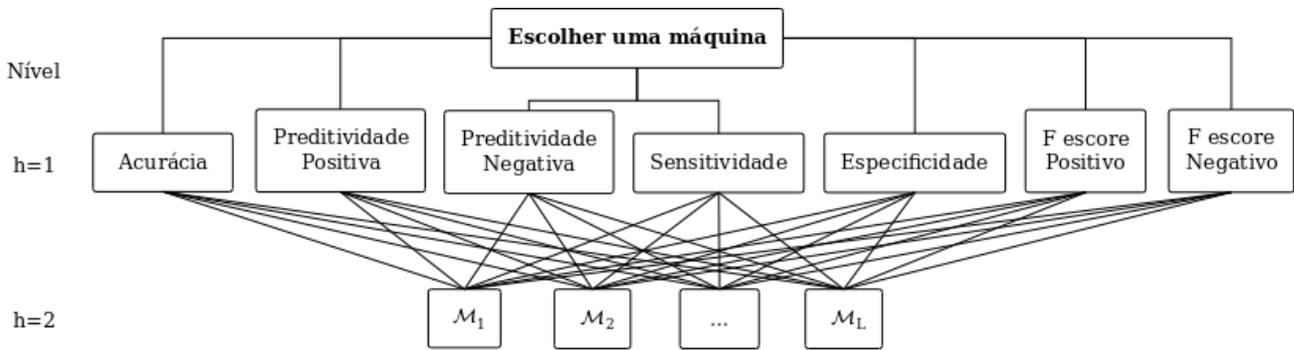
Ambas abordagens podem ser implementadas executando-se um comitê do tipo média de ensemble, sendo que: (I) na primeira abordagem cada máquina especializa-se em particularidades do domínio transformado do problema e (II) na segunda abordagem, cada máquina aprenderá distintas distribuições dos dados. Adicionalmente, a segunda abordagem pode ser empregada utilizando comitês do tipo reforço ou subespaços de atributos, ou apenas uma máquina. É importante destacar que a principal distinção entre as abordagens diz respeito ao domínio em que as máquinas se especializam.

Visto que no comitê de máquinas a saída de cada máquina é combinada linearmente, propõe-se então ponderar essa combinação utilizando o método AHP, a fim de obter os pesos que serão

fixados no combinador linear.

Assim, para um conjunto de dados de treinamento  $\text{Tr} = \{(\mathbf{a}_k, b_k)\}_{k=1}^K$  constituído do valor do atributo  $\mathbf{a}_k$  obtido da aplicação da ADL, cada máquina  $\mathcal{M}_l$ , com  $l = 1, 2, \dots, L$ , é induzida e as medidas de performance Acurácia, Preditividade, Sensitividade,  $F$ -escore e Especificidade são calculadas, comparando os valores preditos por cada máquina  $\hat{b}_l$  ao rótulo verdadeiro  $b_k$  para cada atributo. Após, aplica-se o método AHP para obter as prioridades de importância  $v_l$  de cada máquina  $l$  em relação as demais máquinas, onde  $\mathbf{v}$  é o vetor de prioridades global, conforme equação (118), sendo os valores do vetor  $\mathbf{s}$  fixados em 0,25 garantindo que todas as medidas de performance tenham a mesma prioridade em relação ao Alvo no modelo hierárquico exibido na Figura 24.

Figura 24 – Modelo hierárquico para o método AHP, com o objetivo de obter a importância de cada máquina no comitê.



Fonte: Autoria própria.

Para calcular as prioridades de importância são utilizadas as diferenças  $d_{m, \mathcal{M}_i \mathcal{M}_j}$  entre as medidas de performance  $c_{m, \mathcal{M}_i}$  e  $c_{m, \mathcal{M}_j}$  obtidas pelas máquinas  $\mathcal{M}_i$  e  $\mathcal{M}_j$ ,  $i \neq j$ ;  $i, j = 1, 2, \dots, L$ , e uma nova versão da função de conversão (137) inicialmente proposta em (OLIVEIRA et al., 2019), que converte essas diferenças em valores na escala de Saaty,

$$d_{m, \mathcal{M}_i \mathcal{M}_j} = (c_{m, \mathcal{M}_i} - c_{m, \mathcal{M}_j})\kappa \quad (135)$$

$$\delta_{m, \mathcal{M}_i \mathcal{M}_j} = \begin{cases} 1, & \text{se } |d_{m, \mathcal{M}_i \mathcal{M}_j}| < 1 \\ 9, & \text{se } |d_{m, \mathcal{M}_i \mathcal{M}_j}| > 9 \\ \lceil |d_{m, \mathcal{M}_i \mathcal{M}_j}| \rceil, & \text{caso contrário} \end{cases} \quad (136)$$

$$\Delta_{m, \mathcal{M}_i \mathcal{M}_j} = \begin{cases} \delta_{m, \mathcal{M}_i \mathcal{M}_j}^{-1}, & \text{se } d_{m, \mathcal{M}_i \mathcal{M}_j} < 0 \\ \delta_{m, \mathcal{M}_i \mathcal{M}_j}, & \text{caso contrário} \end{cases}, \quad (137)$$

onde  $m = 1, 2, \dots, D$ ;  $L$  e  $D$  são as quantidades de máquinas e medidas de performance, respectivamente,  $\lceil \cdot \rceil$  é a função teto (maior inteiro), e a constante  $\kappa > 0$  é utilizada para aumentar a sensibilidade da diferença  $|d_{m, \mathcal{M}_i \mathcal{M}_j}|$ .

Para exemplificar a utilização da função (137) considera-se as performances obtidas para uma certa medida  $m$  iguais a  $c_{m\mathcal{M}_i} = 0,9821$  e  $c_{m\mathcal{M}_j} = 0,9732$ . Assim, no modelo hierárquico, Figura 24, ao comparar a máquina  $\mathcal{M}_i$  com a  $\mathcal{M}_j$  a diferença é  $d_{m\mathcal{M}_i\mathcal{M}_j} = 0,0089$ , resultando em  $\delta_{m\mathcal{M}_i\mathcal{M}_j} = 1$ , que é convertido para o mesmo valor na escala de Saaty, significando que as máquinas  $\mathcal{M}_i$  e  $\mathcal{M}_j$  têm a mesma preferência, conforme Tabela 6. Para aumentar a sensibilidade desta diferença, pode-se tomar  $\kappa = 200$ , o que resultará em  $\delta_{m\mathcal{M}_i\mathcal{M}_j} = 2$ , implicando que a máquina  $\mathcal{M}_i$  tem preferência um pouco maior que a máquina  $\mathcal{M}_j$ .

Em seguida, na etapa de treinamento, após a indução de cada máquina  $l$  e a obtenção dos pesos relativos  $v_l$  que serão utilizados no comitê, passa-se para a etapa de teste/validação, onde cada atributo  $\mathbf{a}_{k'}$  do conjunto de dados de teste  $\text{Te} = \{(\mathbf{a}_{k'}, b_{k'})\}_{k'=1}^{K'}$  é fornecido para cada máquina  $l$  resultando nas saídas  $p_{l,k'}$ , as quais são combinadas linearmente considerando as ponderações obtidas pelo AHP, do seguinte modo:

$$\hat{p}(1|\mathbf{a}_{k'}) = \sum_{l=1}^L v_l p_l(1|\mathbf{a}_{k'}), \quad (138)$$

onde  $p_l(1|\mathbf{a}_{k'})$  é a probabilidade do atributo  $\mathbf{a}_{k'}$  pertencer a classe positiva (batimento CPV), ou  $b_{k'} = 1$ , obtida utilizando os modelos de máquinas de aprendizagem Bayesianos Ingênuos, induzidos com os dados do conjunto Tr.

Para classificar o atributo  $\mathbf{a}_{k'}$  faz-se

$$\hat{b}_{k'} = \begin{cases} 1, & \text{se } \hat{p}(1|\mathbf{a}_{k'}) \geq \alpha \\ 0, & \text{se } \hat{p}(1|\mathbf{a}_{k'}) < \alpha \end{cases}, \quad (139)$$

onde  $\hat{b}_{k'}$  é o rótulo predito pelo comitê de máquinas e  $\alpha$  é um corte de probabilidade, que comumente recebe o valor de 0,5.

No Capítulo 6 os resultados e discussões das abordagens propostas são descritos por meio de vários experimentos, a fim de verificar a performance e a robustez destas proposições.

## 6 Resultados e Discussões

Neste capítulo são implementados vários experimentos com o objetivo de verificar a performance das abordagens propostas para o reconhecimento de CPV.

As características extraídas dos registros de ECG são as componentes do método AMUSE, que poderão ser duas, três ou quatro componentes, dependendo da quantidade de sinais de ECG defasados que se considere, cada uma combinada linearmente pelos pesos obtidos pela ADL resultando em um único atributo. Para o reconhecimento das arritmias propriamente, são induzidas máquinas de aprendizado baseadas no algoritmo Naive Bayes, utilizando esse atributo como as características do ambiente que se pretende modelar.

Sendo assim, os experimentos focam em reconhecer as arritmias CPV e os batimentos não-arrítmicos, ou Normais, eis que é necessário que as máquinas de aprendizado induzidas sejam capazes de evitar confusões, gerando falsos alarmes de doenças.

Além disso, neste capítulo também são exploradas as variações nas performances quando os parâmetros livres são modificados, tais como: número de componentes, número de máquinas de aprendizado, tipos de comitês de máquinas, dados balanceados ou não, ambiente ruidoso e desvios nas detecções dos complexos QRS dos sinais de ECG.

### 6.1 Experimentos

As bases de dados utilizadas estão discriminadas na Tabela 8. Elas são obtidas dos registros do banco de dados MIT/BIH-ARDB que é o conjunto de registros de ECG mais comumente utilizado para o propósito de reconhecimento de arritmias. Esse banco de dados contém mais de 109 mil segmentos de determinados tipos de arritmias ou batimentos da classe Normal distribuídos em 48 registros, sendo que dois destes registros pertencem ao mesmo paciente, de acordo com a Tabela 31 no Apêndice A. Nos experimentos implementados os registros 102, 104, 107 e 217 foram descartados de acordo com orientações da AAMI (1987), pois são compostos por ritmos cardíacos anômalos.

Os registros foram coletados durante os anos de 1975 e 1979 nos laboratórios do Hospital Beth Israel de Boston e são formados por meia-hora de registro da atividade elétrica cardíaca a uma taxa de amostragem de 360 Hz e 11 *bits* de resolução, e as anotações foram realizadas por dois ou mais cardiologistas independentes (GOLDBERGER et al., 2000). Os registros são apresentados em dois canais, sendo utilizadas dentre as derivações MLII<sup>1</sup>, V5, V1, V2, V4, mas neste trabalho apenas os registros das duas primeiras derivações foram empregados, pois apenas o primeiro canal foi considerado nas implementações.

<sup>1</sup> Esta derivação é a derivação DII modificada pela inserção de um eletrodo no tórax.

Tabela 8 – Discriminação das bases de dados utilizados nos experimentos.

Rótulo	Registros ECG	Origem	Instâncias	
			CPV	Normal
DS1	101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230	MIT/BIH-ARDB	3.683	38.087
DS2	100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234	MIT/BIH-ARDB	3.219	36.428
DS3	108, 109, 111, 112, 113, 115, 117, 122, 124, 200, 203, 207, 208, 209, 210, 212, 213, 214, 219, 222, 215, 220, 223, 228, 230, 231, 233, 234	MIT/BIH-ARDB	5.040	46.980
DS4	100, 101, 102, 103, 104, 105, 106, 107, 114, 116, 118, 119, 121, 123, 201, 202, 205, 221, 223, 232	MIT/BIH-ARDB	2.400	29.826
DS5	103, 105, 106, 108, 109, 111, 112, 113, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 200, 201, 202, 203, 205, 207, 208, 209, 210, 212, 213, 214, 215, 219, 220, 221, 222, 223, 228, 230, 231, 232, 233, 234	MIT/BIH-ARDB	6.901	70.419
DS6	100, 101, 102, 104, 105, 106, 107	MIT/BIH-ARDB	627	8.391
DS7	100, 102, 104, 105, 106, 107, 118, 119, 200, 201, 203, 205, 208, 212, 213, 214, 215, 217	MIT/BIH-ARDB	4.420	25.124
DS8	111, 115, 116, 119, 221, 230, 231	MIT/BIH-ARDB	953	10.395
DS9	106, 119, 200, 201, 208, 213, 221, 223, 233	MIT/BIH-ARDB	4.899	16.930

Na Tabela 8, exceto as bases DS1 e DS2, que são aquelas utilizadas em todos os experimentos, as demais bases somente são empregadas na comparação com outras metodologias de reconhecimento de CPV, pois cada uma delas emprega um conjunto diferente.

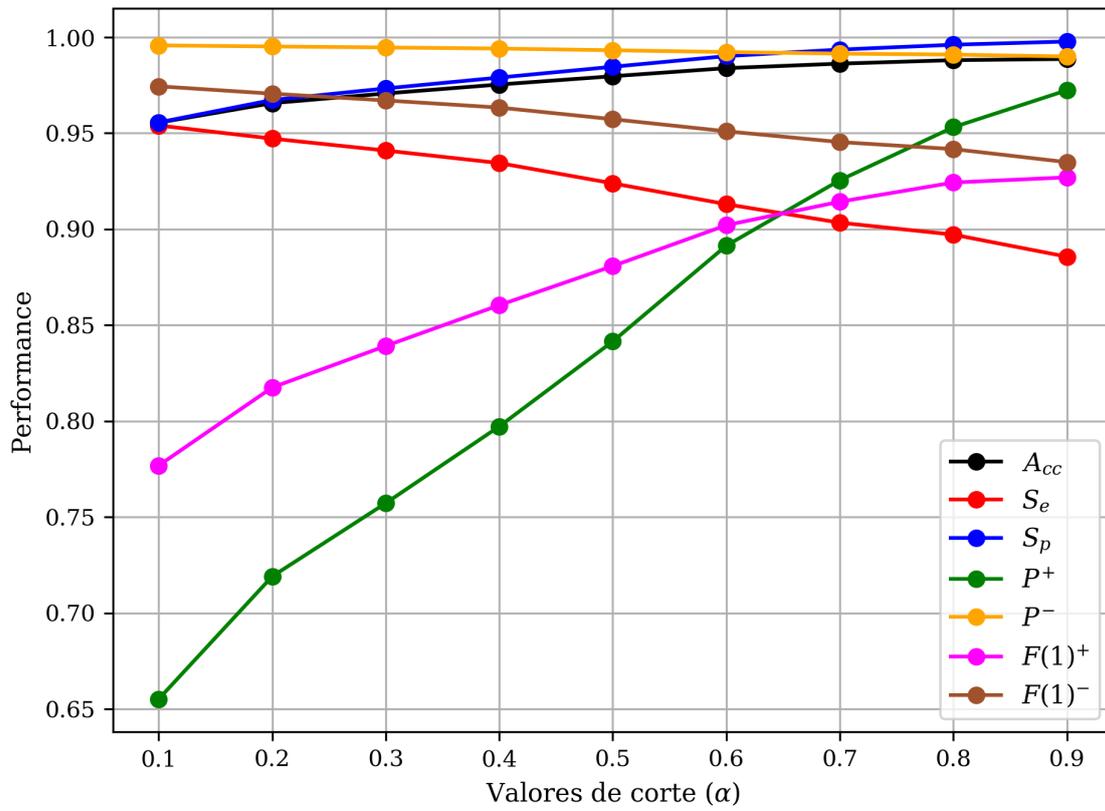
### 6.1.1 Experimento I

Neste experimento são utilizadas a base DS1 para treinamento e DS2 para teste, de acordo com os registros da Tabela 8, e duas componentes latentes da abordagem proposta, considerando o ciclo cardíaco de comprimento padrão, isto é, 0,75 segundos, resultando em 269 atributos para cada componente.

As máquinas  $Clf_0$  e  $Clf_1$  são induzidas utilizando as componentes 1 e 2, após a normalização e redução dos atributos, respectivamente, consistindo em uma abordagem de comitê homogênea, onde cada máquina se especializa em uma componente específica. São utilizadas 41.770 instâncias para treinamento de cada máquina e 39.647 para teste. As máquinas são combinadas por meio das técnicas de voto majoritário rígido e suave, de acordo com as equações (29) e (30), respectivamente, e voto AHP proposto na seção anterior, a fim de melhorar a performance individual.

Na implementação do comitê de máquinas utilizando o voto AHP, por padrão, se  $\hat{p}(1|\mathbf{a}) \geq 0,5$  então o batimento cardíaco é classificado como CPV (classe positiva) e caso  $\hat{p}(1|\mathbf{a}) < 0,5$  o batimento é do tipo Normal, onde  $\hat{p}(1|\mathbf{a})$  é a probabilidade média, ponderada pelo pesos fornecidos pelo método AHP, de uma instância desconhecida  $\mathbf{a}$  pertencer a classe positiva, de acordo com equação (138).

Na Figura 25 são exibidas as curvas de performance tomando diferentes cortes de probabilidade  $\alpha$ . O experimento que resultou nestas curvas fixa o corte acima ou abaixo do valor padrão, ou seja, considera  $\hat{p}(1|\mathbf{a}) > \alpha$  para classificar os exemplos positivos.

Figura 25 – Performance para diferentes cortes  $\alpha$  de probabilidade.

Fonte: Autoria própria.

Da Figura 25 observa-se que a preditividade negativa  $P^-$  recebe pouca influência ao alterar os valores do corte de probabilidade, variando cerca de 0,005 com o aumento de  $\alpha$ . Por outro lado, a preditividade positiva  $P^+$  melhora significativamente, em torno de 0,31. Comportamento análogo é observado para a medida F-escore positivo  $F(1)^+$ , que tem um acréscimo próximo a 0,15, enquanto que o F-escore negativo  $F(1)^-$  tem um decréscimo menor que 0,03 para valores maiores do corte de probabilidade. Outra medida que também decresce com o aumento de  $\alpha$  é a sensibilidade  $S_e$ . Por outro lado, as medidas de especificidade e acurácia melhoram com o aumento de  $\alpha$ , com acréscimos inferiores a 0,05.

Em resumo, estes resultados implicam que a utilização de cortes de probabilidade mais altos é viável pois o decréscimo nas medidas  $F(1)^-$ ,  $S_e$  e  $P^-$  são inferiores aos acréscimos obtidos pelas demais medidas.

De outro ponto de vista, estes resultados refletem a diminuição dos falsos positivos devida ao acréscimo no valor de  $\alpha$ , ao custo de um aumento no número de falsos negativos. Tal fato pode ser verificado na Tabela 9, que consiste dos valores utilizados no cálculo das medidas de performance para distintos cortes  $\alpha$ .

Tabela 9 – Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN), para distintos cortes  $\alpha$ , considerando um comitê de máquinas com voto AHP.

<b>Corte <math>\alpha</math></b>	<b>VP</b>	<b>VN</b>	<b>FP</b>	<b>FN</b>
0,2	3.046	35.250	1.178	173
0,3	3.025	35.455	973	194
0,4	2.991	35.677	751	228
0,5	2.966	35.871	557	253
0,6	2.948	36.070	358	271
0,7	2.918	36.233	195	301

Fonte: Autoria própria.

Observa-se ainda da Tabela 9, que também ocorre a diminuição de VP com conseqüente aumento de VN, a medida que o corte  $\alpha$  aumenta. Este resultado é esperado, pois aquelas instâncias classificadas como FP, quando o corte é baixo, são classificadas como VN quando o valor do corte aumenta.

Os resultados gerais obtidos neste experimento estão exibidos na Tabela 10, para cada máquina individualmente e para os comitês, sendo que para o voto AHP variações do corte de probabilidade foram consideradas, e para o voto majoritário rígido foram fixadas ora a classe positiva (Voto Rígido - P) ora a negativa (Voto Rígido - N), sendo esta fixação necessária pois são duas máquinas apenas, e quando estas discordam, uma classe pré-fixada deve ser atribuída ao resultado, já que nesta abordagem não há pesos associados as máquinas.

Tabela 10 – Resultados do Experimento I, considerando duas componentes. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9800	0,8821	0,9891	0,9214	0,9852	0,8460	0,9930
$Clf_1$	0,9777	0,8708	0,9878	0,9248	0,9824	0,8228	0,9933
Voto majoritário							
Suave	0,9797	0,8808	0,9573	0,9239	0,9846	0,8415	0,9932
Rígido - N	0,9874	0,9214	0,9470	0,9062	<b>0,9946</b>	0,9370	0,9917
Rígido - P	0,9703	0,8371	0,9665	<b>0,9400</b>	0,9730	0,7544	<b>0,9946</b>
Voto AHP							
$\alpha = 0,3$	0,9706	0,8383	<b>0,9664</b>	0,9397	0,9733	0,7566	<b>0,9946</b>
$\alpha = 0,4$	0,9753	0,8594	0,9603	0,9292	0,9794	0,7993	0,9936
$\alpha = 0,5$	0,9796	0,8799	0,9559	0,9214	0,9847	0,8419	0,9930
$\alpha = 0,6$	0,9841	0,9036	0,9526	0,9158	0,9902	0,8917	0,9925
$\alpha = 0,7$	<b>0,9875</b>	<b>0,9217</b>	0,9472	0,9065	<b>0,9946</b>	<b>0,9374</b>	0,9918

Fonte: Autoria própria.

Constata-se dos resultados da Tabela 10 que a máquina  $Clf_1$  obteve uma acurácia  $A_{cc}$  menor que da máquina  $Clf_0$  em cerca de 1%, e também obteve resultado superior em relação as medidas de performance  $F(1)^+$ ,  $F(1)^-$ ,  $S_p$  e  $P^+$ .

Analisando cada exemplo classificado a fim de verificar a diversidade das máquinas, constata-se que ocorreram 898 desacordos entre elas sendo que, para 494 instâncias a máquina  $Clf_0$  esteve correta, enquanto que a máquina  $Clf_1$  predisse a classe correta para 404 instâncias, logo, a máquina  $Clf_1$  erra menos que a máquina  $Clf_0$ .

Por outro lado, em relação aos acordos entre as máquinas houve 3.113 classificações acordadas em relação a classe CPV e 35.636 para a classe Normal, dentre essas, 2.917 e 35.443 para as classes CPV e Normal, respectivamente, estavam corretas. Ou seja, 196 predições da classe positiva e 193 para a classe negativa estavam equivocadas, e mesmo assim as máquinas concordaram com a predição. Tal disparidade entre os valores concordados e o real se deve a performance um pouco superior da máquina  $Clf_0$  na classificação de exemplos positivos, conforme se pode verificar na Tabela 10.

Analisando agora o ensemble (conjunto) de máquinas, para o voto majoritário Suave observa-se que este classificador não conseguiu superar a performance de ambas máquinas, mas apenas da máquina  $Clf_1$  para certas medidas. Já o voto majoritário Rígido - N supera a performance de ambas máquinas para as medidas  $A_{cc}$ ,  $F(1)^+$ ,  $S_p$  e  $P^+$ . Enquanto que o voto Rígido - P, supera em relação as medidas  $S_e$  e  $P^-$  apenas.

Porém, o voto AHP supera os valores das máquinas individuais e dos comitês de voto majoritário para certos valores de corte  $\alpha$ . Tomando  $\alpha = 0,7$ , nota-se que o voto AHP e o voto majoritário Rígido - N obtêm resultados muito próximos, sendo que o voto AHP supera ou empata nas performances.

Para melhor compreender a contribuição das máquinas no comitê de voto AHP, a seguir são analisadas as matrizes de comparações paritárias e os vetores de prioridades obtidos.

As matrizes de comparações paritárias obtidas aplicando a função de conversão (137) para cada medida de performance e fixando  $\kappa = 500^2$  são:

$$\mathbf{M}_{Acc} = \begin{bmatrix} 1 & 2 \\ 1/2 & 1 \end{bmatrix}, \quad (140)$$

$$\mathbf{M}_{F(1)^+} = \begin{bmatrix} 1 & 6 \\ 1/6 & 1 \end{bmatrix}, \quad (141)$$

$$\mathbf{M}_{F(1)^-} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad (142)$$

$$\mathbf{M}_{S_e} = \begin{bmatrix} 1 & 1/2 \\ 2 & 1 \end{bmatrix}, \quad (143)$$

$$\mathbf{M}_{S_p} = \begin{bmatrix} 1 & 2 \\ 1/2 & 1 \end{bmatrix}, \quad (144)$$

$$\mathbf{M}_{P^+} = \begin{bmatrix} 1 & 9 \\ 1/9 & 1 \end{bmatrix} \text{ e} \quad (145)$$

$$\mathbf{M}_{P^-} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad (146)$$

e os respectivos vetores de prioridades locais são:

$$\mathbf{t}_1 = \begin{bmatrix} 0,67 & 0,33 \end{bmatrix}, \quad (147)$$

$$\mathbf{t}_2 = \begin{bmatrix} 0,85 & 0,15 \end{bmatrix}, \quad (148)$$

$$\mathbf{t}_3 = \begin{bmatrix} 0,5 & 0,5 \end{bmatrix}, \quad (149)$$

$$\mathbf{t}_4 = \begin{bmatrix} 0,33 & 0,67 \end{bmatrix}, \quad (150)$$

$$\mathbf{t}_5 = \begin{bmatrix} 0,67 & 0,33 \end{bmatrix}, \quad (151)$$

$$\mathbf{t}_6 = \begin{bmatrix} 0,90 & 0,10 \end{bmatrix} \text{ e} \quad (152)$$

$$\mathbf{t}_7 = \begin{bmatrix} 0,5 & 0,5 \end{bmatrix}, \quad (153)$$

<sup>2</sup> Esse valor foi fixado, pois as máquinas individuais apresentam performances muito próximas.

os quais são utilizados na equação (118), onde  $k = 2$ ,  $h = 2$  e  $s_l \approx 0,1428, \forall l = 1, \dots, 7$  (indicando que todas as medidas de performance têm a mesma importância em relação a escolha da melhor máquina), para obter o vetor de prioridade global

$$\mathbf{v} = [0,63 \ 0,37]. \tag{154}$$

Por meio dos vetores de prioridades locais, equações (147)-(153), fica claro que, analisando o primeiro coeficiente, a máquina  $Clf_0$  é preferível à máquina  $Clf_1$ , exceto para a medida  $S_e$ ; e para as medidas  $F(1)^-$  e  $P^-$  ambas máquinas têm mesma preferência.

A maior preferência ocorre em relação as medidas  $F(1)^+$  e  $P^+$ , sendo que a máquina  $Clf_0$  tem intensidade de importância 6 (mais que forte) e 9 (extrema), na escala Saaty, respectivamente, em relação a máquina  $Clf_1$ , conforme equações (141) e (142).

A preferência da máquina  $Clf_0$  é notada também no vetor de prioridade global (154), onde a importância desta, considerando todas as medidas de performance com mesmo peso, é 0,26 mais importante que a máquina  $Clf_1$ .

Com o objetivo de ranquear as melhores máquinas considerando as performances obtidas, utiliza-se novamente do método AHP juntamente com a função de conversão (137), mas desta vez a prioridade de importância não é utilizada para ponderar alguma máquina, mas tão somente para ranquear as máquinas pelas suas performances. Os resultados do ranqueamento, considerando que todas as medidas têm a mesma importância, são exibidos na Tabela 11.

Tabela 11 – Ranque das máquinas do Experimento I utilizando o método AHP com  $\kappa = 500$  na função (137).

Ordem	Máquina	Ranque (%)
1º	Voto AHP com $\alpha = 0,7$	17,40
2º	Voto Rígido N	17,39
3º	Voto AHP com $\alpha = 0,6$	11,06
4º	$Clf_0$	10,62
5º	$Clf_1$	9,23
6º	Voto Rígido P	8,36
7º	Voto AHP com $\alpha = 0,3$	8,29
8º	Voto Suave	6,42
9º	Voto AHP com $\alpha = 0,5$	5,95
10º	Voto AHP com $\alpha = 0,4$	5,23

Fonte: Autoria própria.

A partir dos resultados ranqueados na Tabela 11 verifica-se que o voto AHP com  $\alpha = 0,7$  e o Voto Rígido - N obtiveram preferências bem próximas e ocupam as primeiras posições. Logo, a proposição de utilização do método AHP na ponderação das máquinas melhora os resultados.

Destaca-se também que, a maior parte dos comitês têm menor preferência que as máquinas solitárias  $Clf_0$  e  $Clf_1$ , logo, estes não contribuem para melhorar o reconhecimento dos padrões.

Visto que os dados, conforme Tabela 8, estão desbalanceados, considera-se agora o mesmo experimento, no entanto, remove-se algumas instância de batimentos da classe Normal, de cada registro, a fim de que haja a mesma quantidade de batimentos que a classe CPV. Desse modo, são utilizados 6.954 batimentos para indução e 5.924 na fase de teste, sendo metade dos batimentos para cada classe. Assim, são considerados menos batimentos CPV que aqueles das bases DS1 e DS2, pois como alguns registros não possuem ambos os batimentos simultaneamente (vide Tabela 31), então estes não foram tomados.

Os resultados obtidos, considerando as mesmas máquinas e comitês implementados anteriormente estão exibidos na Tabela 12.

Tabela 12 – Resultados do Experimento I, considerando duas componentes com dados balanceados. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9789	0,9787	<b>0,9791</b>	0,9689	0,9889	0,9886	0,9695
$Clf_1$	0,9784	0,9781	0,9787	0,9659	0,9909	0,9907	0,9667
Voto majoritário							
Suave	0,9792	0,9790	0,9683	0,9679	0,9905	0,9903	0,9686
Rígido - N	0,9775	0,9771	0,9592	0,9585	<b>0,9966</b>	<b>0,9965</b>	0,9600
Rígido - P	0,9797	0,9797	0,9764	0,9764	0,9831	0,9830	0,9765
Voto AHP							
$\alpha = 0,3$	<b>0,9802</b>	0,9802	0,9788	<b>0,9787</b>	0,9818	0,9817	<b>0,9788</b>
$\alpha = 0,4$	0,9808	<b>0,9806</b>	0,9748	0,9747	0,9868	0,9867	0,9750
$\alpha = 0,5$	0,9792	0,9790	0,9683	0,9679	0,9905	0,9903	0,9686
$\alpha = 0,6$	0,9787	0,9784	0,9631	0,9625	0,9949	0,9948	0,9637
$\alpha = 0,7$	0,9760	0,9755	0,9563	0,9554	<b>0,9966</b>	<b>0,9965</b>	0,9572

Fonte: Autoria própria.

Nota-se da Tabela 12, que a melhor acurácia  $A_{cc}$  e sensibilidade  $S_e$  foi obtida pelo comitê voto AHP com  $\alpha = 0,3$ , enquanto que a melhor especificidade  $S_p$  pelos comitês voto majoritário Rígido - N e Voto AHP com  $\alpha = 0,7$ . Diferentemente dos resultados apresentados na Tabela 10, estes últimos apresentam maior homogeneidade entre os valores de performance devido ao balanceamento dos dados.

Como não houve nenhum comitê absolutamente melhor que outro, recorre-se ao método AHP para ranqueá-los, de acordo com os valores exibidos na Tabela 13.

Tabela 13 – Ranque das máquinas cujas performances estão discriminadas na Tabela 12, considerando dados balanceados utilizando o método AHP com  $\kappa = 500$  na função (137).

Ordem	Máquina	Ranque (%)
1º	Voto AHP com $\alpha = 0,3$	15,54
2º	Voto AHP com $\alpha = 0,4$	11,93
3º	Voto Rígido P	11,65
4º	Clf <sub>0</sub>	9,7
5º	Voto AHP com $\alpha = 0,6$	9,59
6º	Clf <sub>1</sub>	9,25
7º	Voto Rígido N	9,20
8º	Voto AHP com $\alpha = 0,7$	8,19
9º	Voto Suave	7,48
3º	Voto AHP com $\alpha = 0,5$	7,48

Fonte: Autoria própria.

Observa-se, da Tabela 13, que o ranque obtido para dados balanceados é muito distinto do caso desbalanceado, vide Tabela 11. Exceto para o classificador Clf<sub>0</sub>, todos os demais tiveram suas posições alteradas no ranque, sendo que o classificador Clf<sub>1</sub> desceu uma posição ao considerar os dados balanceados. Estes resultados indicam que os comitês são sensíveis a quantidade de instâncias de cada classe. Por exemplo, o voto AHP com  $\alpha = 0,7$  saltou da primeira posição para a oitava, ao tomar os exemplos balanceados. Analisando os resultados gerais de performance, nota-se que esta alteração é devida, principalmente, a perda de performance em relação ao reconhecimento de batimentos da classe Normal.

Comparando o melhor resultado de performance (voto AHP com  $\alpha = 0,3$ ) nas Tabelas 10 e 12, de acordo com o ranque fornecido pelo método AHP, observa-se que há melhora na performance em relação ao reconhecimento de CPV, enquanto que o oposto ocorre no reconhecimento da classe Normal. A preditividade positiva  $P^+$ , por exemplo, aumentou 0,2251 ao tomar os dados balanceados enquanto que a preditividade negativa  $P^-$  diminuiu 0,0158. Já a acurácia  $A_{cc}$  teve um acréscimo de 0,0096 apenas.

Por outro lado, ao comparar diferentes comitês, por exemplo voto AHP com  $\alpha = 0,7$  (dados desbalanceados) e voto AHP com  $\alpha = 0,3$  (dados balanceados), pensando na performance geral do sistema de reconhecimento, pode-se constatar que ocorreu uma piora na acurácia  $A_{cc}$  por volta de 0,007.

Visto que as diferenças observadas são muito pequenas, conclui-se que a utilização de dados balanceados não geram ganhos significativos de performance, logo, o sistema proposto não é muito sensível ao equilíbrio entre instâncias Normais e CPV, uma vez que, conhecendo-se as características do ambiente em relação a diversidade dos dados, pode-se escolher um ou outro comitê de máquinas. Ou seja, se na solução do problema de reconhecimento de CPV

forem utilizados dados desbalanceados, dá-se preferência à utilização do comitê voto AHP com  $\alpha = 0,7$ , e no caso desbalanceado ao comitê voto AHP com  $\alpha = 0,3$ , tendo ao final uma performance geral similar, com uma diferença inferior a 1%.

### 6.1.2 Experimento II

Neste experimento, diferentemente do que foi implementado no Experimento I, duas versões defasadas do segmento de ECG são utilizadas na matriz de entrada para o AMUSE, o que acarreta em três componentes com 268 amostras cada uma. Similarmente ao Experimento I, neste também são considerados diferentes cortes de probabilidade  $\alpha$  para os comitês de voto AHP. Os resultados de performance obtidos estão discriminados na Tabela 14 e as máquinas ranqueadas da melhor para a pior estão exibidas na Tabela 15.

Tabela 14 – Resultados do Experimento II, considerando três componentes do AMUSE. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9651	0,7654	0,9811	0,7018	0,9883	0,8417	0,9740
$Clf_1$	0,9654	0,7663	0,9813	0,6977	0,9891	0,8498	0,9737
$Clf_2$	0,9777	0,8691	<b>0,9878</b>	0,9130	0,9834	0,8293	0,9922
Voto majoritário							
Suave	0,9728	0,8181	0,8517	0,7540	0,9921	0,8943	0,9786
Rígido	0,9694	0,7956	0,8378	0,7335	0,9903	0,8693	0,9768
Voto AHP							
$\alpha = 0,3$	0,9759	0,8608	0,9548	<b>0,9195</b>	0,9808	0,8092	<b>0,9928</b>
$\alpha = 0,4$	<b>0,9786</b>	<b>0,8737</b>	0,9494	0,9102	0,9847	0,8400	0,9920
$\alpha = 0,5$	0,9797	0,8784	0,9441	0,9012	0,9867	0,8568	0,9912
$\alpha = 0,6$	0,9756	0,8444	0,8916	0,8152	<b>0,9898</b>	0,8758	0,9838
$\alpha = 0,7$	0,9741	0,8240	0,8459	0,7453	0,9944	<b>0,9213</b>	0,9779

Fonte: Autoria própria.

Tabela 15 – Ranque das máquinas do Experimento II utilizando o método AHP com  $\kappa = 500$  na função (137).

Ordem	Máquina	Ranque (%)
1º	Clf <sub>2</sub>	15,53
2º	Voto AHP com $\alpha = 0,5$	14,56
3º	Voto AHP com $\alpha = 0,4$	12,98
4º	Voto AHP com $\alpha = 0,3$	12,65
5º	Voto AHP com $\alpha = 0,7$	12,10
6º	Voto AHP com $\alpha = 0,6$	8,34
7º	Voto Suave	8,34
8º	Clf <sub>0</sub>	5,66
9º	Clf <sub>1</sub>	5,19
10º	Voto Rígido	4,64

Fonte: Autoria própria.

Analisando os resultados gerais de performance na Tabela 14 e o ranqueamento na Tabela 15, verifica-se que, de modo diverso ao Experimento I, no Experimento II, a máquina solitária Clf<sub>2</sub> obteve melhor desempenho que todos os demais classificadores, mesmo os comitês. Este desempenho superior é devido a superioridade da medida F-escore negativa  $F(1)^-$ , que é ao menos, 3,3% superior aos comitês. Por outro lado, as demais medidas de performance são inferiores a todos os demais classificadores. No entanto, essa inferioridade é menos significativa.

Ocupando as posições secundárias de performance, estão todos os comitês de voto AHP. Neste experimento, ao utilizar  $\alpha = 0,5$  obteve-se melhor resultado do que com a utilização dos cortes de probabilidade  $\alpha = 0,7$  e  $\alpha = 0,3$ , os quais geraram melhores resultados no Experimento I.

Nas Tabelas 16 e 17 estão exibidos os resultados de performance e os ranque dos classificadores, considerando três versões defasadas dos sinais de ECG.

Tabela 16 – Resultados do Experimento II, considerando quatro componentes. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(\zeta)^+$	$F(\zeta)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9427	0,5690	0,9693	0,4660	0,9848	0,7306	0,9543
$Clf_1$	0,9521	0,6339	<b>0,9743</b>	0,5113	0,9910	0,8338	0,9582
$Clf_2$	0,9318	0,5906	0,9628	0,6058	0,9606	0,5761	0,9650
$Clf_3$	0,9499	0,6941	0,9727	0,6999	0,9720	0,6884	0,9734
Voto majoritário							
Suave	<b>0,9730</b>	0,8089	0,8167	0,7030	0,9969	0,9524	0,9744
Rígido	0,9501	0,5633	0,5590	0,3961	0,9991	0,9748	0,9493
Voto AHP							
$\alpha = 0,3$	0,9615	0,7808	0,9093	<b>0,8437</b>	0,9719	0,7266	<b>0,9860</b>
$\alpha = 0,4$	0,9725	<b>0,8254</b>	0,8830	0,8018	0,9875	0,8504	0,9826
$\alpha = 0,5$	<b>0,9730</b>	0,8116	0,8254	0,7154	0,9958	0,9377	0,9754
$\alpha = 0,6$	0,9558	0,6323	0,6283	0,4682	0,9989	0,9735	0,9551
$\alpha = 0,7$	0,9491	0,5469	0,5409	0,3784	<b>0,9995</b>	<b>0,9862</b>	0,9479

Fonte: Autoria própria.

Tabela 17 – Ranque das máquinas do Experimento II utilizando o método AHP com  $\kappa = 500$  na função (137).

Ordem	Máquina	Ranque (%)
1º	Voto AHP com $\alpha = 0,4$	17,02
2º	Voto AHP com $\alpha = 0,3$	14,03
3º	$\alpha = 0,5$	13,07
4º	Voto Suave	12,27
5º	Voto AHP com $\alpha = 0,7$	7,98
6º	$Clf_3$	5,19
7º	Voto AHP com $\alpha = 0,6$	7,19
8º	$Clf_1$	6,60
9º	Voto Rígido	6,26
10º	$Clf_0$	4,20
11º	$Clf_2$	4,07

Fonte: Autoria própria.

Pelos resultados apresentados nas Tabelas 16 e 17, nota-se que, ao contrário dos resultados das Tabelas 14 e 15, ao tomar três versões defasadas novamente os comitês voto AHP geraram os melhores resultados. Deve-se notar que neste caso, apesar das máquinas terem sido rotuladas

com os mesmos nomes, são induzidas sobre diferentes amostras, pois ao considerar mais versões defasadas, distintas componentes são estimadas pelo método AMUSE, conforme observado no Capítulo 5.

Por fim, pode-se concluir que a utilização de componentes adicionais não melhora a performance geral considerando as instâncias da base de dados testada e a divisão implementada. Além disso, não pôde ser observada uma consistência dos resultados, pois, ora os comitês geraram melhor performance ora uma máquina solitária.

### 6.1.3 Experimento III

Com o objetivo de verificar que as performances obtidas anteriormente não são apenas devidas a escolha dos registros que compõem os dados de indução e teste, nesse terceiro experimento é considerada uma abordagem de validação cruzada agrupada apresentada na Seção 3.2.2, utilizando 22 subconjuntos, garantindo assim que todos os registros participem dos conjuntos de treinamento e teste. Nesse experimento apenas a abordagem voto AHP com  $\alpha = 0,7$  é implementada, pois a intenção não é comparar as abordagens, mas apenas verificar o quão independentes são os resultados de performance dos dados escolhidos. Os resultados obtidos com essa implementação estão discriminados na Tabela 18.

Tabela 18 – Resultados do Experimento III, considerando duas componentes do AMUSE, voto AHP com  $\alpha = 0,7$  e validação cruzada com 22 subconjuntos. Valores em negrito são os maiores da coluna e  $\sigma$  é o desvio padrão.

Subconjunto	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
1	<b>0,9875</b>	0,9217	<b>0,9472</b>	<b>0,9065</b>	0,9946	0,9374	<b>0,9918</b>
2	0,9872	0,9206	0,9457	0,904	0,9946	<b>0,9378</b>	0,9915
3	0,9870	<b>0,9297</b>	0,9447	0,9034	<b>0,9958</b>	0,9575	0,9899
4	0,9758	0,8720	0,9146	0,8541	0,9888	0,8908	0,9844
5	0,9759	0,8739	0,9148	0,8544	0,9891	0,8943	0,9843
6	0,9755	0,8690	0,9110	0,8480	0,9890	0,8910	0,9840
7	0,9772	0,8784	0,9174	0,8585	0,9898	0,8992	0,9850
8	0,9777	0,8814	0,9208	0,8640	0,9898	0,8995	0,9857
9	0,9780	0,8838	0,9238	0,8689	0,9896	0,8992	0,9861
10	0,9779	0,8636	0,9181	0,8579	0,9886	0,8694	0,9874
11	0,9744	0,8559	0,8923	0,8179	0,9904	0,8976	0,9815
12	0,9786	0,8782	0,9286	0,8759	0,9885	0,8804	0,9880
13	0,9779	0,8775	0,9235	0,8678	0,9889	0,8873	0,9867
14	0,9791	0,8857	0,9237	0,8685	0,9904	0,9036	0,9865
15	0,9575	0,7648	0,8507	0,7543	0,9780	0,7756	0,9753
16	0,9579	0,7673	0,8583	0,7655	0,9771	0,7690	0,9766
17	0,9696	0,8241	0,8841	0,8046	0,9856	0,8446	0,9811
18	0,9632	0,8264	0,8572	0,7675	0,9884	0,8950	0,9706
19	0,9671	0,8295	0,8660	0,7789	0,9887	0,8873	0,9750
20	0,9654	0,8146	0,8586	0,7672	0,9872	0,8681	0,9747
21	0,9668	0,8164	0,8674	0,7798	0,9864	0,8565	0,9772
22	0,9644	0,7876	0,8520	0,7555	0,9844	0,8225	0,9768
Média	0,9737	0,8555	0,9009	0,8328	0,9883	0,8801	0,9827
$\pm\sigma$	$\pm 0,0084$	$\pm 0,0453$	$\pm 0,0323$	$\pm 0,0506$	$\pm 0,0043$	$\pm 0,0447$	$\pm 0,0058$

Fonte: Autoria própria.

Pela análise das performances obtidas por cada subconjunto, conforme Tabela 18, constata-se que os melhores resultados são obtidos utilizando as subconjuntos 1, 2 e 3, cujos registros respectivos estão discriminados na Tabela 19.

Tabela 19 – Registros utilizados em cada subconjunto, que geraram os resultados exibidos na Tabela 18.

Subconjunto	Treinamento	Teste
1	101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230	100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234
2	100, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230	101, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234
3	100, 103, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230	101, 106, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234

Fonte: Autoria própria.

Nota-se das Tabelas 18 e 19 que a diferença entre as subconjuntos 1, 2 e 3 consiste apenas das trocas dos registros 100, 101, 103 e 106 da base de treinamento para a base de teste. Especificamente, a subconjunto 1 é apenas uma repetição dos resultados apresentados no Experimento I, Tabela 10, logo, seus resultados são utilizados como parâmetros de comparação com as demais subconjuntos.

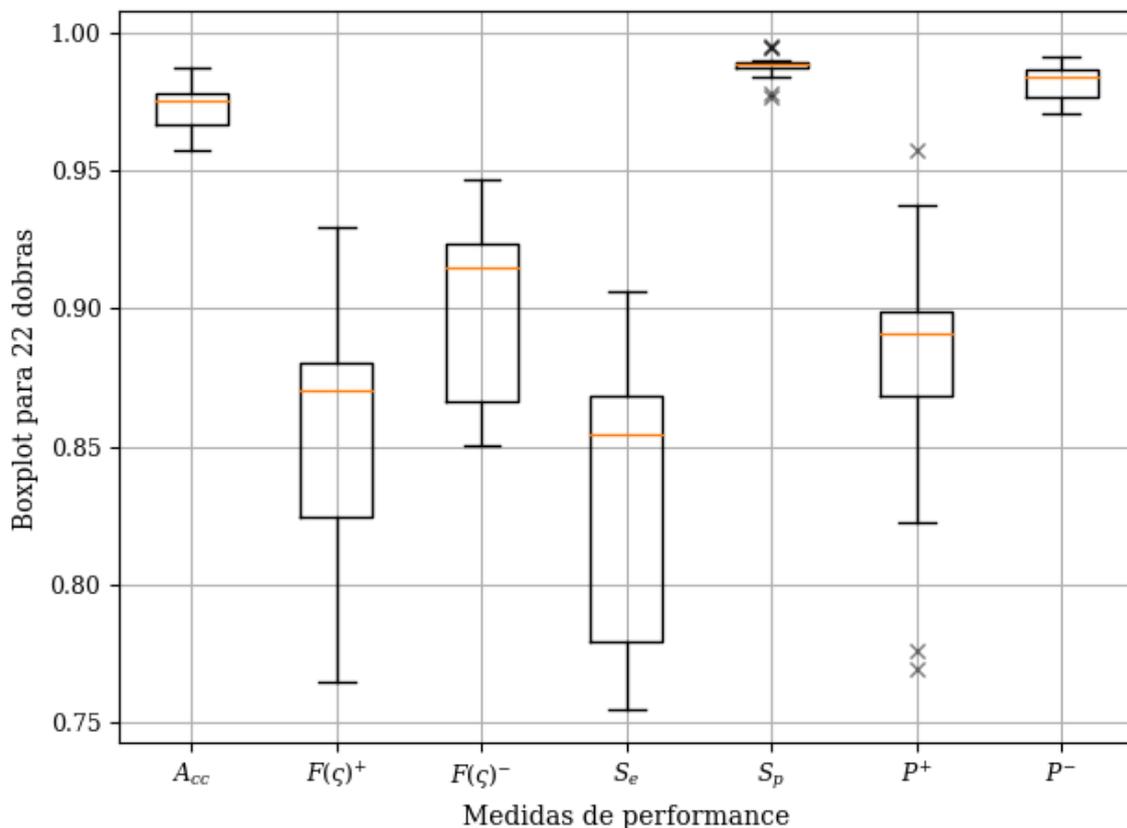
Para as subconjuntos 2 e 3, nota-se que ocorre acréscimo de performance em relação as medidas  $P^+$ ,  $F(1)^+$  e  $S_p$ , respectivamente. No entanto, a significância desse acréscimo é da ordem de  $10^{-3}$  somente. De outro modo, para as demais subconjuntos observa-se uma piora mais significativa de performance, tendo a acurácia  $A_{cc}$  um decréscimo de até 0,03.

Na média, nota-se decréscimos de performance mais significantes para as medidas  $F(1)^+$ ,  $S_e$  e  $P^+$ , consequência das quantidade de instâncias das classes positivas no registros selecionados por algumas subconjuntos. Por exemplo, o pior resultado da preditividade positiva  $P^+$  ocorre na subconjunto 16, que utilizou para indução os registros 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 207, 208, 209, 215, 220, 223 e 230 e para teste 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 221, 222, 228, 231, 232, 233 e 234. Assim, conforme Tabela 31, na subconjunto 16 foram utilizados 321 instâncias na fase de indução a menos que na subconjunto 1, e na fase de testes 321 a mais, resultando na piora da performance, pois as máquinas não tiveram exemplos suficientes para aprender os padrões e além disso, foram expostas a mais exemplos cujos padrões não haviam sido modelados adequadamente.

Por fim, dados os resultados apresentados na Tabela 18 e exibidos na Figura 26, para cada subconjunto, pode-se afirmar que existe sim uma dependência entre a performance de

reconhecimento de arritmias CPV e os registros escolhidos para comporem os conjuntos de treinamento e teste, sendo que esta dependência é menos refletida nas medidas de acurácia  $A_{cc}$ , especificidade  $S_p$  e preditividade negativa  $P^-$ . Nota-se também que a medida de sensibilidade  $S_e$  apresenta o maior desvio em relação à mediana. Tal resultado é devido a variação na quantidade de verdadeiros positivos, já que  $P^-$  varia menos, e ambas fórmulas dessas medidas incluem a quantidade de falsos negativos. Tal resultado é esperado, pois a mudança de registros de um conjunto para outro ocasiona a diminuição/aumento da quantidade de exemplos positivos que são utilizados no treinamento, e visto que essa quantidade é bem inferior aquela dos exemplos de batimento do tipo Normal, então a variabilidade é esperada.

Figura 26 – Boxplot para cada medida de performance exibida na Tabela 18 para todas as 22 subconjuntos.



Fonte: Autoria própria.

Finalmente, embora exista dependência entre a performance e o conjunto de dados escolhido, os resultados são superiores a 96,44% de acurácia para algum comitê de máquina ao menos, independentemente do conjunto de dados selecionado.

### 6.1.4 Experimento IV

Neste quarto experimento é testada a robustez das abordagens propostas por meio da inserção de ruído artificial nos sinais de ECG e da simulação de erros de detecção dos complexos QRS.

Para o primeiro caso, para cada segmento dos sinais de ECG foi adicionado ruído com distribuição contínua uniforme<sup>3</sup>, sendo as demais etapas executadas como no Experimento I.

Na Figura 27 estão ilustrados os resultados da acurácia  $A_{cc}$  para diferentes valores de relação sinal-ruído SNR (*Signal-to-Noise-Ratio*), considerando a média das SNRs das componentes dos segmentos do conjunto de teste. Lembrando que, nesta implementação também o conjunto de treinamento foi contaminado por ruído, conforme especificado na Tabela 20.

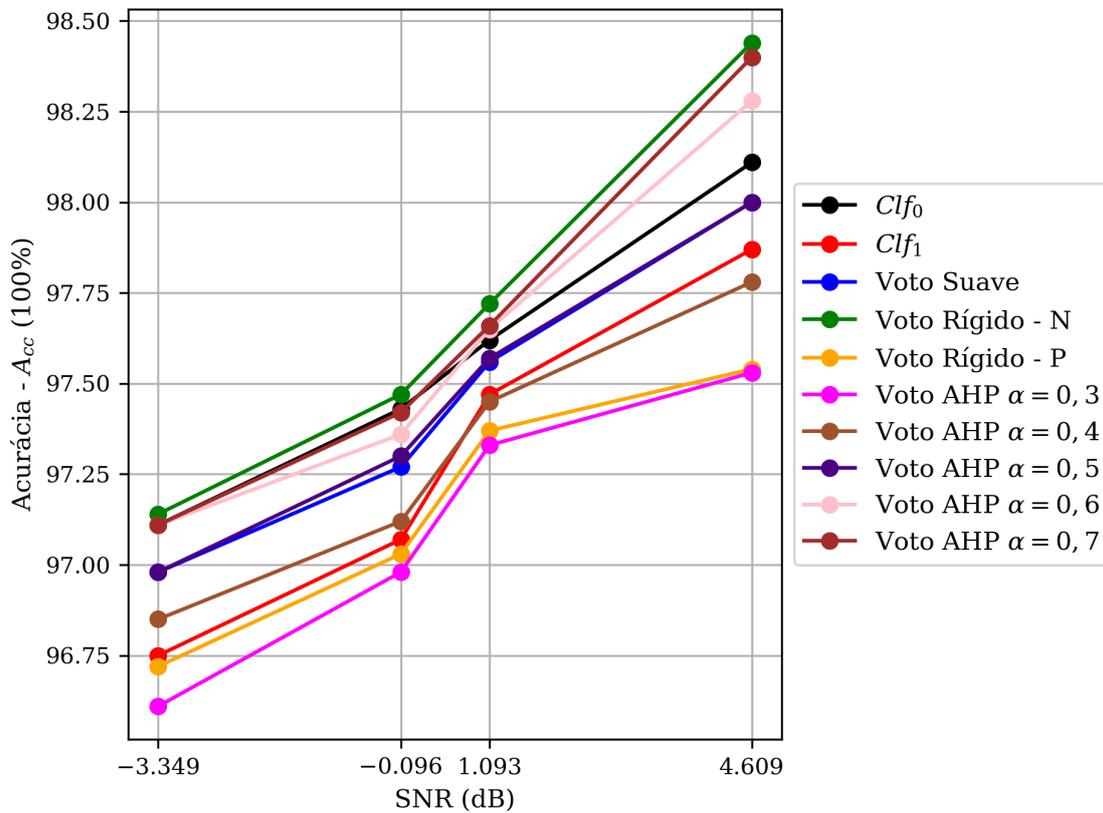
Tabela 20 – Valores médios das SNRs das duas componentes consideradas, para os conjuntos de treinamento e teste.

Implementação	SNR média (dB)	
	Treinamento	Teste
1	-1,4434	-3,3489
2	0,4913	-0,0958
3	2,9936	1,0925
4	9,0125	7,1093
5	12,5346	10,6324

Fonte: Autoria própria.

<sup>3</sup> Números reais aleatórios no intervalo  $[0, 1)$ .

Figura 27 – Resultados do Experimento IV, considerando ruído adicionado sobre o conjunto de treinamento e de teste.



Fonte: Autoria própria.

Nota-se dos resultados exibidos na Figura 27, que as máquinas Voto Rígido - N e Voto AHP com  $\alpha = 0,7$  e com  $\alpha = 0,6$  apresentaram menor sensibilidade ao ruído que as demais máquinas. Fica claro pelos resultados obtidos que o ruído adicionado aos conjuntos de dados acarreta em uma degradação na medida de performance geral em torno de 1,6% no pior caso e 0,30% no melhor, ao comparar com os resultados obtidos no Experimento I, Tabela 10, em relação a máquina Voto Rígido - N, que foi aquela que obteve o melhor resultado de acurácia ao considerar o ruído.

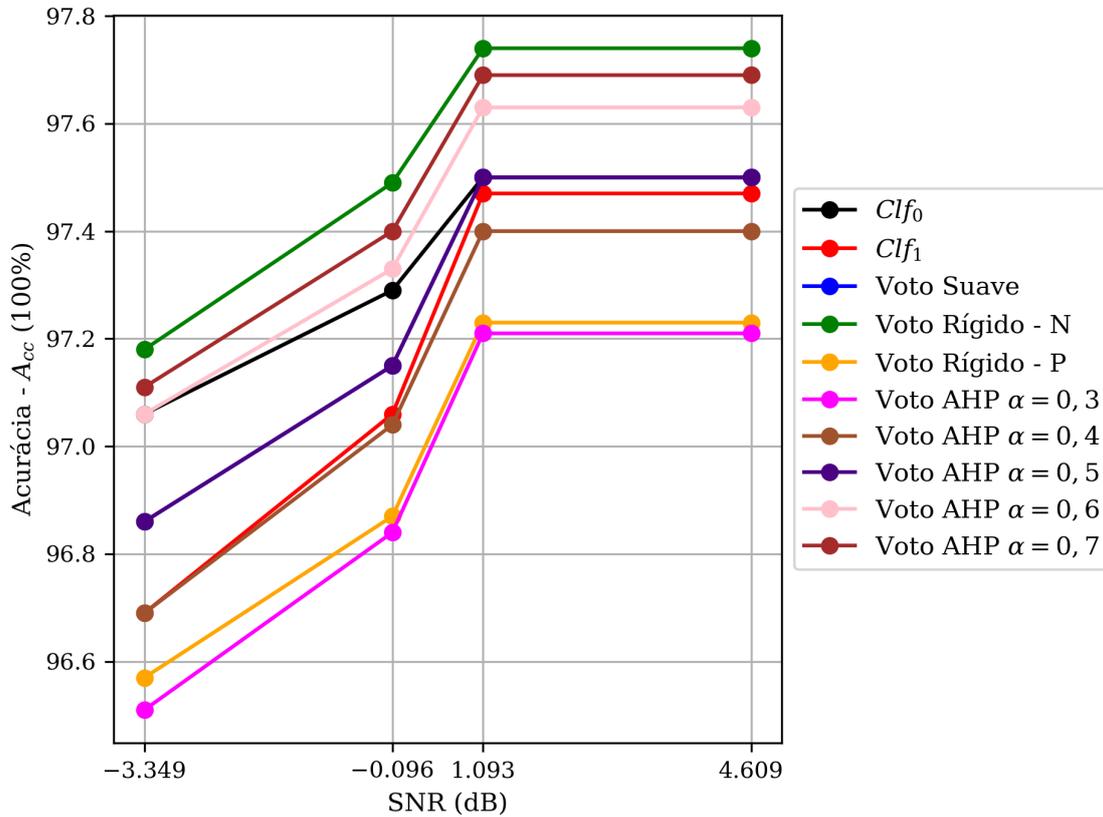
Outro fato notável é que o comitê Voto AHP com  $\alpha = 0,6$  tem um acréscimo de performance superior a máquina solitária  $Clf_0$  a medida que a SNR aumenta.

Visto que a melhor acurácia obtida pela máquina solitária  $Clf_0$  é inferior aquela gerada pelos comitês Voto Rígido - N, Voto AHP com  $\alpha = 0,7$  e  $\alpha = 0,6$ , então a abordagem de ensemble de máquinas é menos sensível ao ruído.

Já os resultados exibidos na Figura 28 são obtidos de uma implementação análoga a anterior. Entretanto, nesse caso apenas ao conjunto de testes é adicionado ruído. Simulação

esta que reflete melhor o que acontece em ambientes reais, onde se tem acesso ao conjunto de treinamento, sendo possível filtrá-lo previamente, enquanto que a aplicação propriamente da técnica é realizada sobre dados que podem ter sido contaminados com ruído.

Figura 28 – Resultados do Experimento IV, considerando ruído adicionado sobre o conjunto de teste apenas.



Fonte: Autoria própria.

Observa-se, comparando as Figuras 27 e 28, que, ao introduzir ruído no conjunto de treinamento, os algoritmos de aprendizado de máquina também aprenderam as modificações inseridas pelos padrões ruidosos, por isso o resultado foi superior àquele onde os dados de treinamento não incluíam ruído. Além disso, nota-se que a melhora da performance se estabilizou após a SNR igual ou superior a 1,093 dB.

De modo geral, as máquinas apresentaram o mesmo comportamento em relação a performance, independente de ter sido considerado ruído no conjunto de treinamento ou não, apenas gerando valores inferiores no segundo caso.

Na Tabela 21 estão exibidos os resultados para as abordagens propostas anteriormente mas tomando um desvio aleatório entre 0 e 10 amostras na localização da onda R, simulando erros

no algoritmo de detecção dos complexos QRS. Na Tabela 22 estão discriminadas as máquinas de aprendizado ranqueadas tomando todas as medidas de performance.

Tabela 21 – Resultados do Experimento IV, considerando 269 amostras e duas componentes do AMUSE, e a inserção de algum desvio na detecção das ondas R. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9710	0,8209	0,9842	0,8186	0,9845	0,8232	0,9840
$Clf_1$	0,9731	0,8379	0,9853	0,8577	0,9833	0,8190	0,9874
Voto majoritário							
Suave	0,9751	0,8473	0,9146	0,8521	0,9859	0,8424	0,9869
Rígido - N	<b>0,9789</b>	0,8603	0,8817	0,7996	<b>0,9948</b>	0,9309	0,9825
Rígido - P	0,9651	0,8033	0,9294	0,9294	0,8767	<b>0,9730</b>	0,9889
Voto AHP							
$\alpha = 0,3$	0,9676	0,8150	0,9313	0,8798	0,9753	0,7590	<b>0,9892</b>
$\alpha = 0,4$	0,9747	<b>0,9130</b>	<b>0,9589</b>	<b>0,9311</b>	0,9820	0,8956	0,9885
$\alpha = 0,5$	0,9747	0,8456	0,9161	0,8546	0,9853	0,8367	0,9871
$\alpha = 0,6$	0,9763	0,8520	0,9081	0,8416	0,9882	0,8627	0,9860
$\alpha = 0,7$	0,9775	0,8545	0,8914	0,8148	0,9918	0,8983	0,9838

Fonte: Autoria própria.

Tabela 22 – Ranque das máquinas do Experimento IV utilizando o método AHP com  $\kappa = 500$  na função (137).

Ordem	Máquina	Ranque (%)
1º	Voto AHP com $\alpha = 0,4$	18,07
2º	Voto Rígido N	14,34
3º	Voto Rígido P	13,34
4º	Voto AHP com $\alpha = 0,7$	9,89
5º	$Clf_1$	9,48
6º	Voto AHP com $\alpha = 0,6$	7,88
7º	$Clf_0$	7,23
8º	Voto AHP com $\alpha = 0,3$	6,77
9º	Voto Suave	6,59
9º	Voto AHP com $\alpha = 0,5$	6,41

Fonte: Autoria própria.

Comparando os resultados das Tabelas 10 e 21, bem como as respectivas tabelas com os ranqueamentos das máquinas, observa-se que a máquina Voto Rígido - N ocupa as segundas

posições em ambos experimentos. Já as máquinas Voto AHP com  $\alpha = 0,7$  e com  $\alpha = 0,6$  que ocupam, respectivamente, a primeira e a terceira posição, perdem performance ao considerar desvios de detecção dos complexos QRS. Assim, pode-se concluir que a máquina Voto Rígido - N é robusta em relação a estas falhas, sendo que a acurácia  $A_{cc}$  teve um decréscimo inferior a 0,009.

Por outro lado, a máquina Voto AHP com  $\alpha = 0,4$  superou todas as demais nesse ambiente com falhas de detecção, tendo obtido acréscimos nas medidas de performance  $F(1)^+$ ,  $S_e$ ,  $S_p$  e  $P^+$ . Mesmo assim, o melhor resultado para este experimento não supera o melhor resultado obtido no Experimento I, fato que é esperado, já que os algoritmos propostos são executados a partir da detecção dos complexos QRS, e falhas nessa detecção fornecem distintos segmentos para processamento e análise.

### 6.1.5 Experimento V

Nos experimentos anteriores foi implementada a metodologia de subespaços de atributos e média de ensemble para o reconhecimento de CPV. Neste experimento será verificada a relevância da abordagem de reforço por filtragem, que consiste em selecionar, para indução e teste, distintos conjuntos do espaço de entrada, por meio do método AdaBoost apresentado na Seção 3.2.1. Intenta-se assim, verificar se a abordagem proposta, que utiliza o método AHP para ponderação das máquinas, é mais eficiente que este método clássico de comitê de máquinas.

Na Tabela 23 estão exibidos os resultados obtidos aplicando o método AdaBoost juntamente com o Naive Bayes para cada componente do AMUSE após aplicada a ADL e também para atributos agregados, resultando em um vetor de 2 atributos (terceira linhas da tabela).

Tabela 23 – Resultados do Experimento V, considerando o método AdaBoost induzido com duas máquinas

Comp.	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Primeira	0,9188	0,0000	0,9577	0,0000	1,0000	0,0000	0,9188
Segunda	0,9564	0,6344	0,9768	0,4660	0,9997	0,9934	0,9549
Todas	0,9188	0,0000	0,9577	0,0000	1,0000	0,0000	0,9188

Fonte: Autoria própria.

Observa-se dos resultados da Tabela 23 que ao tomar a primeira componente do AMUSE o algoritmo AdaBoost obteve máquinas com capacidade apenas de predição de batimentos da classe Normal, pois as medidas de sensibilidade  $S_e$  e preditividade positiva  $P^+$  são iguais a zero, implicando que a quantidade de verdadeiros positivos é nula.

Por outro lado, ao tomar apenas a segunda componente, nota-se uma melhora significativa de performance, já que a preditividade positiva  $P^+$  saltou para cerca de 0,99, embora a sensibilidade  $S_e$  tenha melhorado menos.

Analisando os atributos agregados, terceira linha da Tabela 23, conclui-se que a contribuição da segunda componente não foi significativa para os resultados.

Na comparação entre os resultados obtidos pela abordagem de subespaços de atributos, especificamente a abordagem Voto AHP com  $\alpha = 0,7$ , conforme Experimento I, Tabela 10, embora o melhor resultado do comitê AdaBoost tenha gerado maior especificidade e preditividade positiva em cerca de 0,005 e 0,05, respectivamente, devido ao valor da sensibilidade ser inferior em cerca de 0,44, conclui-se que a abordagem de reforço utilizando AdaBoost não melhora o reconhecimento de CPV.

Assim como no Experimento I, neste também é considerado o reconhecimento com três componentes do AMUSE e conseqüentemente três máquinas induzidas pelo AdaBoost, conforme resultados exibidos na Tabela 24.

Tabela 24 – Resultados do Experimento V, considerando o método AdaBoost induzido com três máquinas.

<b>Comp.</b>	<b><math>A_{cc}</math></b>	<b><math>F(1)^+</math></b>	<b><math>F(1)^-</math></b>	<b><math>S_e</math></b>	<b><math>S_p</math></b>	<b><math>P^+</math></b>	<b><math>P^-</math></b>
Primeira	0,9605	0,7263	0,9787	0,6455	0,9883	0,8302	0,9693
Segunda	0,8890	0,2515	0,9401	0,2296	0,9473	0,2780	0,9330
Terceira	0,9749	0,8278	0,9864	0,7446	0,9952	0,9320	0,9778
Todas	0,9696	0,7829	0,9837	0,6751	0,9956	0,9318	0,9720

Fonte: Autoria própria.

Nota-se da Tabela 24 que a acurácia  $A_{cc}$  piorou apenas para a utilização da segunda componente, na comparação com os resultados exibidos na Tabela 23. A especificidade  $S_e$  e a preditividade positiva  $P^+$  também tiveram reduções para todos os resultados, sendo estas menos significativas ao considerar a terceira componente e as componentes agregadas.

Em geral, assim como nos experimentos utilizando os comitês propostos, ao tomar mais componentes não é garantida uma melhoria na performance.

Comparando o melhor resultado do AdaBoost para três componentes, terceira linha da Tabela 24, e os resultados obtidos utilizando a abordagem de subespaços de atributos, Voto AHP com  $\alpha = 0,4$ , Tabela 14, fica claro que, embora o comitê AdaBoost tenha gerado especificidade  $S_p$  e preditividade positiva  $P^+$  maiores, a sensibilidade está muito aquém dos resultados obtidos ao empregar outras abordagens.

### 6.1.6 Experimento VI

Até o momento, todos os experimentos consideraram o reconhecimento de batimentos cardíacos CPV em oposição aos batimentos Normais. Entretanto, foi anunciado no Capítulo 2 que existem várias arritmias catalogadas pela AAMI. Sendo assim, é preciso verificar se as

metodologias propostas também são capazes de distinguir entre as arritmias CPV e as demais arritmias.

Neste experimento será verificada a performance desse reconhecimento, onde duas classes de batimentos são considerados: CPV e outros, sendo que a segunda classe inclui tanto batimentos Normais como das outras arritmias, sejam ventriculares ou atriais.

Nas Tabelas 25 e 26 estão exibidos os resultados do reconhecimento de batimentos CPV frente ao demais tipos de batimentos, incluindo os da classe Normal, e o ranque das máquinas de aprendizado, respectivamente, tomando o mesmo esquema empregado no Experimento I.

Tabela 25 – Resultados do Experimento VI, considerando duas componentes do AMUSE, e as classes CPV e outras. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9702	0,7753	<b>0,9840</b>	0,8180	0,9804	0,7369	0,9877
$Clf_1$	0,9698	0,7816	0,9838	0,8596	0,9772	0,7167	0,9904
Voto majoritário							
Suave	0,9710	0,7879	0,9187	0,8568	0,9786	0,7292	0,9903
Rígido - N	0,9808	0,8402	0,8843	0,8012	<b>0,9929</b>	0,8832	0,9867
Rígido - P	0,9591	0,7296	0,9304	0,8764	0,9647	0,6249	0,9915
Voto AHP							
$\alpha = 0,3$	0,9596	0,7334	0,9340	<b>0,8826</b>	0,9648	0,6273	<b>0,9919</b>
$\alpha = 0,4$	0,9654	0,7599	0,9269	0,8705	0,9718	0,6742	0,9911
$\alpha = 0,5$	0,9714	0,7903	0,9189	0,8571	0,9791	0,7332	0,9903
$\alpha = 0,6$	0,9766	0,8192	0,9108	0,8437	0,9855	0,7960	0,9895
$\alpha = 0,7$	<b>0,9814</b>	<b>0,8468</b>	0,8949	0,8180	0,9923	<b>0,8777</b>	0,9878

Fonte: Autoria própria.

Tabela 26 – Ranque das máquinas do Experimento VI utilizando o método AHP com  $\kappa = 500$  na função (137).

Ordem	Máquina	Ranque (%)
1º	Voto AHP com $\alpha = 0,7$	18,00
2º	Voto Rígido N	17,38
3º	Voto AHP com $\alpha = 0,6$	10,94
4º	Clf <sub>1</sub>	9,62
5º	Voto AHP com $\alpha = 0,3$	9,26
6º	Clf <sub>0</sub>	8,66
7º	Voto Rígido P	7,00
8º	Voto AHP com $\alpha = 0,5$	6,68
9º	Voto AHP com $\alpha = 0,4$	6,27
10º	Voto Suave	6,19

Fonte: Autoria própria.

Constata-se dos resultados apresentados nas Tabelas 25 e 26, em comparação com aqueles do Experimento I, Tabela 12, em relação a máquina Voto AHP com  $\alpha = 0,7$ , que houve uma melhora na acurácia  $A_{cc}$  em torno de 0,24% apenas. Já a medida de sensibilidade  $S_e$  teve uma piora mais significativa, por volta de 13,74%, enquanto que a especificidade  $S_p$  piorou somente 0,43%. De modo geral, todas as máquinas geraram performances piores ao incluir na classe outras várias arritmias além dos batimentos Normais.

Felizmente, este resultado não implica necessariamente que as abordagens propostas não conseguem distinguir as demais arritmias dos batimentos CPV, já que várias performances ficaram acima dos 90%. O fato que explica a diminuição da performance geral é a escassez de exemplos das outras arritmias, conforme se verifica facilmente pela consulta a Tabela 31 no Apêndice A, onde se nota, por exemplo, que existem 106 exemplos da classe Escape Ventricular (E) e apenas dois da classe Batimento Ectópico supraventricular. Além disso, os exemplos não se distribuem uniformemente para todos os registros da base de dados, mas se concentram em alguns poucos pacientes. Um bom exemplo desta ocorrência é observada na classe Bloqueio de Ramo Esquerdo (L) que possui 8.075 exemplos de batimentos, mas distribuídos em apenas quatro pacientes.

Isto acarreta que vários exemplos são utilizados ou no conjunto de treinamento ou no conjunto de teste, mas não em ambos. Logo, ou os padrões aprendidos não têm candidatos na fase de testes, ou nesta fase o padrão ainda não foi aprendido pela máquina.

## 6.1.7 Experimento VII

No diagrama de blocos do sistema de reconhecimento de CPV proposto, ilustrado pela Figura 15, observa-se quatro esquemas principais de execução: 1) extração de atributos pelo método AMUSE e normalização; 2) redução da dimensão do vetor de atributos utilizando a ADL; e 3) indução das máquinas de aprendizado.

Com o intuito de verificar a contribuição de cada um desses esquemas na performance do sistema de reconhecimento, nesta seção são realizados experimentos substituindo as técnicas empregadas por outras similares ou pela remoção da técnica.

Primeiramente desconsidera-se a técnica ADL, permanecendo os demais esquemas inalterados. Os resultados são apresentados na Tabela 27.

Tabela 27 – Resultados do Experimento VII, considerando duas componentes sem a técnica ADL. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,7812	0,4015	0,8661	0,9037	0,7704	0,2581	0,9891
$Clf_1$	0,9139	0,6392	0,9511	0,9397	0,9116	0,4843	0,9942
Voto majoritário							
Suave	0,9333	0,6914	0,9550	0,9202	0,9345	0,5537	0,9925
Rígido - N	<b>0,9645</b>	<b>0,8011</b>	0,9317	0,8804	<b>0,9719</b>	<b>0,7350</b>	0,9892
Rígido - P	0,7306	0,3673	<b>0,9790</b>	<b>0,9630</b>	0,7100	0,2269	<b>0,9954</b>
Voto AHP							
$\alpha = 0,3$	0,9118	0,6341	0,9673	0,9416	0,9091	0,4780	0,9944
$\alpha = 0,4$	0,9131	0,6375	0,9669	0,9410	0,9107	0,4821	0,9943
$\alpha = 0,5$	0,9143	0,6402	0,9658	0,9391	0,9121	0,4856	0,9941
$\alpha = 0,6$	0,9157	0,6439	0,9655	0,9385	0,9137	0,4900	0,9941
$\alpha = 0,7$	0,9170	0,6472	0,9648	0,9372	0,9153	0,4943	0,9940

Fonte: Autoria própria.

Comparando os resultados das Tabelas 10 e 27 nota-se que para as máquinas individuais  $Clf_0$  e  $Clf_1$  ocorreu uma redução na acurácia  $A_{cc}$  em torno de 19,88% e 6,37%, respectivamente, ao desconsiderar a técnica ADL para atenuação de atributos menos relevantes. Além disso, diferentemente do Experimento I, os comitês de máquinas AHP não geraram os melhores resultados. Levando em conta o melhor resultado obtido no Experimento I (utilizando Voto AHP com  $\alpha = 0,7$ ), observa-se que a piora na acurácia foi de 7,05%. Portanto, conclui-se que os atributos utilizados pela máquina  $Clf_0$ , que são uma versão do sinal original escalonada e defasada, não são bons descritores dos padrões de batimentos cardíacos, no entanto, ao considerar uma

combinação linear destes atributos, por meio da ADL, obtém-se um atributo que é relevante para o reconhecimento dos padrões de batimentos Normal e CPV.

Por outro lado, os atributos utilizados pela máquinas  $Clf_1$  geram melhores resultados, mais ainda assim alguns destes atributos pioram a performance geral, logo, a combinação linear desses atributos, com a conseqüente atenuação daqueles menos relevantes melhora a acurácia em mais de 6%.

O experimento seguinte consiste em substituir as máquinas gaussianas ingênuas por máquinas do tipo  $k$  vizinhos mais próximos (*K Nearest Neighbors* (KNN)), com  $k = 5$ , a fim de verificar a importância que os modelos de máquinas gaussianas ingênuas têm no reconhecimento dos padrões de batimentos cardíacos CPV. Os resultados obtidos estão exibidos na Tabela 28.

Tabela 28 – Resultados do Experimento VII, considerando duas componentes, utilizando máquinas do tipo KNN. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9757	0,8554	0,9867	0,8866	0,9835	0,8263	0,9899
$Clf_1$	0,9822	0,8911	<b>0,9903</b>	0,8959	0,9898	0,8863	0,9908
Voto majoritário							
Suave	<b>0,9856</b>	<b>0,9096</b>	0,9390	0,8925	0,9938	0,9274	0,9905
Rígido - N	0,9852	0,9034	0,9162	0,8546	<b>0,9967</b>	<b>0,9582</b>	0,9873
Rígido - P	0,9727	0,8467	0,9596	0,9279	0,9767	0,7785	0,9935
Voto AHP							
$\alpha = 0,3$	0,9723	0,8453	0,9617	<b>0,9317</b>	0,9759	0,7735	<b>0,9938</b>
$\alpha = 0,4$	0,9812	0,8872	0,9495	0,9105	0,9875	0,8651	0,9921
$\alpha = 0,5$	0,9832	0,8961	0,9402	0,8947	0,9910	0,8975	0,9907
$\alpha = 0,6$	0,9848	0,9035	0,9306	0,8785	0,9942	0,9300	0,9893
$\alpha = 0,7$	0,9844	0,8992	0,9173	0,8565	0,9957	0,9464	0,9874

Fonte: Autoria própria.

Comparando os resultados de performance obtidos ao considerar máquinas do tipo KNN, Tabelas 10 e 28, percebe-se que a maior acurácia  $A_{cc}$  foi gerada pelas máquinas gaussianas ingênuas, mas essa superioridade é de apenas 0,19%. O mesmo ocorre para a sensibilidade e preditividade negativa  $P^-$ , mas nos percentuais de 0,83% e 0,08%, respectivamente. Por outro lado, os valores de especificidade  $S_p$  e preditividade positiva  $P^+$  são maiores ao utilizar máquinas KNN, nos percentuais de 0,21% e 2,08%, respectivamente.

No entanto, essas medidas  $P^+$  superiores, somente são atingidas ao custo de uma diminuição da sensibilidade  $S_e$ . Em outros termos, vê-se que para o Voto Rígido - N a medida  $S_e$  é 5,16% maior utilizando máquinas gaussianas ingênuas que utilizando máquinas KNN. Para o comitê

Voto AHP com  $\alpha = 0,7$  o mesmo ocorre, pois a medida  $P^+$  é superior em torno de 0,9% para o algoritmo KNN, mas a medida  $S_e$  é 5% maior para as máquinas utilizadas na abordagem proposta.

Assim, embora as máquinas KNN tenham gerado um resultado de preditividade positiva  $P^+$  maior ao comparar todos os resultados dos comitês, ao escolher aquele comitê que gerou esse resultado superior nota-se que a performance para outras medidas são inferiores.

Por estes resultados e análises, pode-se concluir que a utilização de máquinas do tipo KNN não é mais vantajosa. Além disso, outra desvantagem das máquinas KNN é o tempo de processamento para reconhecimento dos padrões, pois para o conjunto DS2 ambas as máquinas,  $Clf_0$  e  $Clf_1$ , consomem juntas cerca de 22 segundos utilizando máquinas KNN, enquanto que as máquinas gaussianas ingênuas fazem o mesmo reconhecimento em apenas 0,03 segundos.

O experimento seguinte consiste apenas da remoção da normalização dos atributos, mantendo os demais esquemas sem qualquer alteração. Os resultados de performance obtidos estão descritos na Tabela 29.

Tabela 29 – Resultados do Experimento VII, considerando duas componentes, sem o esquema de normalização. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9759	0,8621	0,9868	0,9264	0,9803	0,8062	0,9934
$Clf_1$	0,9747	0,8560	0,9861	0,9251	0,9791	0,7965	0,9933
Voto majoritário							
Suave	0,9759	0,8620	0,9589	0,9267	0,9803	0,8058	0,9934
Rígido - N	<b>0,9790</b>	0,8754	0,9492	0,9099	<b>0,9851</b>	<b>0,8434</b>	0,9920
Rígido - P	0,9717	0,8438	0,9674	0,9416	0,9744	0,7644	0,9947
Voto AHP							
$\alpha = 0,3$	0,9710	0,8409	<b>0,9678</b>	<b>0,9422</b>	0,9736	0,7592	<b>0,9948</b>
$\alpha = 0,4$	0,9738	0,8533	0,9650	0,9372	0,9771	0,7832	0,9944
$\alpha = 0,5$	0,9758	0,8616	0,9593	0,9273	0,9801	0,8046	0,9935
$\alpha = 0,6$	0,9779	0,8712	0,9546	0,9192	0,9831	0,8279	0,9928
$\alpha = 0,7$	0,9794	<b>0,8775</b>	0,9486	0,9090	0,9856	0,8481	0,9919

Fonte: Autoria própria.

Pelas performances apresentadas nas Tabelas 10 e 29 e considerando os melhores resultados, mas para abordagens distintas, ou seja, considerando o Voto Rígido - N (Tabela 29) e o Voto AHP com  $\alpha = 0,7$  (Tabela 10), observa-se que ocorreu piora na acurácia em 0,85% ao não incorporar o esquema de normalização na abordagem proposta. Analisando as preditividades positiva e negativa,  $P^+$  e  $P^-$ , nota-se, ao tomar os melhores resultados, que ocorreu piora mais

significante para a primeira, em torno de 9,4%, enquanto que para a segunda foi de apenas 0,2%. Também houve redução no valor da especificidade  $S_p$ , mas a sensibilidade  $S_e$  teve um aumento de 0,22% ao desconsiderar a normalização.

Agora, ao comparar a mesma abordagem, por exemplo, Voto Rígido - N, pode-se observar que a acurácia  $A_{cc}$  piora 0,84% sem a normalização, enquanto que as preditividades positiva e negativa e a especificidade  $S_p$  pioraram 9,36%, 0,03% e 0,95%, respectivamente. Por outro lado, a sensibilidade  $S_e$  melhora 0,37%.

Portanto, pode-se concluir que a normalização de fato melhora a performance geral para os dados analisados, sendo que essa melhoria é mais notável para os batimentos cardíacos da classe CPV.

## 6.2 Comparações entre abordagens do estado da arte

Nessa seção são feitas comparações das abordagens propostas com aquelas apresentadas no Capítulo 4, onde foram expostas metodologias que caracterizam o estado da arte. O objetivo é posicionar a contribuição das abordagens propostas nesta pesquisa classificando-as como superior ou inferior em relação as medidas de performance empregadas nos experimentos.

Na Tabela 30 estão discriminados os resultados de performances para todos os conjuntos de dados da Tabela 30. Os resultados obtidos aplicando as metodologias propostas estão exibidos no Apêndice D, pois são específicos para certas bases de dados de acordo com cada implementação dos trabalhos mencionados, conforme discriminado nas últimas colunas da Tabela 30.

Tabela 30 – Comparação das abordagens propostas com aquelas do estado da arte. A coluna “conjuntos”, diz respeito aqueles apresentados na Tabela 8.

Abordagem	$A_{cc}$	Performance			Conjuntos	
		$S_e$	$S_p$	$P^+$	Treino	Teste
Liu et al. (2015)	0,9890	0,9026	-	<b>0,9890</b>	DS3	DS4
Proposto. Voto AHP com $\alpha = 0,7$ (Tabela 33, Apêndice D)	0,9867	0,9058	0,9932	0,9146	DS3	DS4
Zarei et al. (2016)	0,9877	0,9612	-	0,8648	DS1	DS2
Li et al. (2014)	0,9820	0,9310	0,9850	0,8140	DS1	DS2
Oliveira et al. (2019)	0,9840	0,9110	0,9870	0,8570	DS1	DS2
Proposto. Voto AHP com $\alpha = 0,7$ (Tabela 12, Experimento I)	0,9875	0,9065	0,9946	0,9374	DS1	DS2
Proposto. Voto AHP com $\alpha = 0,3$ (Tabela 12, Experimento I)	0,9802	0,9787	0,9818	0,9817	DS1	DS2
Ataollah e Ali (2010)	0,9540	-	-	-	DS5	DS6
Proposto. Voto AHP com $\alpha = 0,7$ (Tabela 34, Apêndice D)	0,9869	0,9761	0,9877	0,8559	DS5	DS6
Omer, Laurent e Gregory (2006)	0,9520	0,8520	-	0,9240	DS7	DS8
Proposto. Voto AHP com $\alpha = 0,7$ (Tabela 35, Apêndice D)	<b>0,9957</b>	<b>0,9864</b>	<b>0,9965</b>	0,9631	DS7	DS8
Shen et al. (2011)	0,9700	-	-	-	DS9	DS9
Proposto. Voto AHP com $\alpha = 0,7$ (Tabela 36, Apêndice D)	0,9878	0,9714	0,9925	0,9740	DS9	DS9

Fonte: Autoria própria.

Nota-se pelos resultados da Tabela 30, em relação às performances para as mesmas bases de dados que, na primeira comparação, relativa aos conjuntos DS3 e DS4, a abordagem proposta superou a abordagem de Liu et al. (2015) apenas na medida de sensibilidade  $S_e$ , tendo performance inferior para as medidas de acurácia  $A_{cc}$  e preditividade positiva  $P^+$ , implicando que a quantidade de falsos negativos é maior para a abordagem proposta. A especificidade  $S_p$  não pode ser comparada pois os autores Liu et al. (2015) não apresentaram a quantidade de verdadeiros negativos.

Na segunda comparação, sobre os conjuntos DS1 e DS2, a abordagem proposta com  $\alpha = 0,7$  supera aquelas do estado da arte em relação as medidas de especificidade  $S_p$  e preditividade positiva  $P^+$ , tendo resultado inferior em relação a acurácia  $A_{cc}$ , mas apenas em 0,02%, e em relação a sensibilidade  $S_e$ , para a qual foi constatada uma diferença de 5,47%. Tais resultados indicam que a abordagem proposta gera menos falsos positivos mas, por outro lado, gera mais falsos negativos. Já para os resultados com  $\alpha = 0,3$ , percebe-se que a acurácia é inferior em 0,75%, mas, ao contrário do caso anterior, agora a sensibilidade é superior em cerca de 1,7% e a preditividade positiva em 11,69%, resultados esses que implicam em uma redução dos falsos negativos, ao tomar um corte de probabilidade mais baixo.

Nas comparações sobre os conjuntos DS5 e DS6, a abordagem proposta superou aquela de

Ataollah e Ali (2010) em relação a acurácia  $A_{cc}$  em torno de 3%.

Na quarta comparação, implementada sobre os conjuntos DS7 e DS8, também a abordagem proposta superou aquela apresentada por Omer, Laurent e Gregory (2006), em 4,37% de acurácia  $A_{cc}$ , em 13,44% de sensibilidade  $S_e$  e em 3,9% de preditividade positiva  $P^+$ . Aliás, a indução/teste das máquinas de aprendizado sobre estes conjuntos gerou os melhores resultados de performance em comparação com todos os demais resultados para as outras bases de dados, exceto para a preditividade positiva  $P^+$ .

Na quinta e última comparação, para o conjunto de dados DS9, a abordagem proposta superou aquela do estado da arte em 1,78%. Neste caso, a mesma base foi utilizada para indução e teste.

Embora nas comparações acima ora a abordagem proposta foi superior ora inferior, obtendo resultados melhores para a maior parte das medidas e comparações, em geral, uma delas gerou maior performance para três das quatro medidas apresentadas na Tabela 30. Também é evidente, nos resultados apresentados, que as abordagens propostas geraram valores altos de sensibilidade  $S_e$  e preditividade positiva  $P^+$ , superiores a 90%, exceto para a implementação nas bases DS5 e DS6 onde obteve-se uma medida  $P^+$  reduzida, implicando em mais ocorrências de falsos positivos.

Outra característica que tem peso na decisão de uma abordagem de reconhecimento de padrões diz respeito às questões computacionais, relacionadas ao tempo de processamento e armazenamento em memória. É popularmente conhecido, no meio da comunidade de aprendizado de máquina, que as máquinas de aprendizado baseadas no algoritmo Naive Bayes são mais rápidas e menos custosas que os demais algoritmos, os quais muitas vezes incluem métodos de otimização, ou ainda, metodologias baseadas em instâncias, o que acarreta em consumo elevado de memória. Além disso, as máquinas Bayesianas têm a capacidade implícita de aprendizado dinâmico, ou seja, elas podem aprender a qualquer momento bastando mostrar ao algoritmo novos exemplos, sem necessidade de reconsideração de todos os dados, mas levando em conta apenas os dados novos, implementação impossível para muitas metodologias de aprendizado.

Devido a sua simplicidade e a inexistência de métodos de otimização, as máquinas Bayesianas somente têm boa performance quando os atributos utilizados em sua indução são adequados para discriminação dos padrões, pois seu algoritmo não faz nada mais que calcular probabilidades. Isso significa que também os atributos obtidos pelo método AMUSE, projetados pela ADL em um espaço de menor dimensão, resultam em eficiência na obtenção da performance pelas máquinas de aprendizado utilizadas. Adicionalmente, a utilização do método AHP gera uma gama de máquinas que podem ser melhores no reconhecimento de uma classe ou outra, característica que pode ser adequada quando exames adicionais são utilizados para confirmação da doença cardíaca.

Outra vantagem da abordagem proposta é que ela não inclui uma etapa de pré-processamento

para eliminação de ruído, como as demais abordagens (vide Tabela 7), o que também reduz a carga computacional. Nos experimentos da Seção 6.1.4 constatou-se que, a ausência deste pré-processamento afeta menos as abordagens baseadas no Voto Rígido - N e Voto AHP com  $\alpha = 0,4$ .

## 7 Considerações finais

Esta pesquisa consistiu da proposição de abordagens para solução do problema de reconhecimento de arritmias do tipo CPV, que são arritmias associadas à morte súbita e ao infarto agudo do miocárdio.

Para tanto, o método AMUSE, utilizado para separação cega de fontes, foi empregado na etapa de extração de atributos, pois constatou-se que os sinais de ECG são compostos pela sobreposição de sinais fontes oriundos de diversas regiões do músculo cardíaco. Por isso, foi proposto um esquema de obtenção de sinais de ECG com atrasos, para construção de uma matriz de defasagem composta por tais sinais e o sinal original subtraído de algumas amostras. Essa matriz assume um formato adequado para decomposição pelo método AMUSE que gera componentes-atributos que são utilizados para indução das máquinas de aprendizado.

Também foi empregada a Análise Discriminante Linear para combinação linear dos atributos e consequente redução da dimensionalidade, pois as amostras das componentes estimadas têm amplitudes muito próximas.

O algoritmo escolhido para indução e validação das máquinas, e consequentemente, do reconhecimento de CPVs, foi o Naive Bayes, devido a sua simplicidade algorítmica e boa performance.

Visto que na abordagem proposta ao menos dois conjuntos de atributos foram gerados, foi preciso o emprego da técnica de comitê de máquinas de aprendizado, utilizando média de ensemble e subespaços de atributos, para agregar as previsões individuais das máquinas em uma previsão global única. Além disso, neste trabalho foi proposta a utilização do método AHP para a ponderação das importâncias de cada máquina no comitê e também uma função de conversão, que converte os valores das diferenças de performance para um escala pré-definida.

Vários experimentos foram implementados a fim de verificar as performances das abordagens propostas. Foram também empregados comitês de máquinas clássicos tais como voto majoritário. Nas comparações gerais, os comitês utilizando o método AHP ora se saíram melhores que as abordagens de voto suave e rígido, ora não.

Outro resultado importante está relacionado as performances gerais obtidas pelos comitês. Seus valores sempre foram superiores àqueles das máquinas individuais, exceto para uma ou outra medida de performance.

Também foi considerada a utilização de bases de dados balanceadas. Nestes casos notou-se uma equalização das medidas de performance, as quais se estabilizaram acima dos 90%, exceto para uma ou outra medida em relação a certos comitês de máquinas.

A melhor performance foi obtida ao considerar a utilização dos conjuntos DS7 e DS8, para indução e validação, respectivamente, consistindo de 99,57% de acurácia, 98,64% de sensibili-

dade e 99,65% de especificidade.

Foi também verificada a robustez das abordagens propostas mediante o acréscimo de ruído gaussiano nos segmentos de ECG e da inserção de desvios, simulando erros de detecção das ondas R. Em ambos casos, as abordagens apresentaram certa robustez dependendo do nível da SNR. Em relação ao ruído, os comitês Voto Rígido - N e Voto AHP com  $\alpha = 0,4$  foram os menos sensíveis as altas SNRs. Já o Voto AHP com  $\alpha = 0,4$  foi a abordagem mais robusta a desvio nas localizações das ondas R.

Nas comparações com outros métodos de reconhecimento de CPVs as abordagens propostas se mostraram superiores a algumas, mas inferiores a outras, para certas medidas de performance. De modo geral, a abordagem Voto AHP com  $\alpha = 0,7$  nas bases de dados DS7 e DS8 superou todas as performances do estado da arte, exceto para a medida de preditividade positiva.

Finalmente, conclui-se que as componentes do AMUSE obtidas da matriz de defasagem proposta combinadas pela Análise Discriminante Linear e o emprego do algoritmo Naive Bayes em comitês de máquinas de aprendizado ponderados pelo Voto AHP, formam um sistema eficaz e de alta performance para o reconhecimento de arritmias do tipo CPV com erros de confusão reduzidos para as demais classes.

Trabalhos futuros incluem a pesquisa sobre diferentes funções de conversão para o método AHP e outras escalas também e investigações mais aprofundadas sobre como o ruído afeta a performance. Além disso, implementação das fases de teste e indução das abordagens propostas em outras bases de dados.

## Referências bibliográficas

- AAMI. *Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms*. Arlington, 1987.
- AEHLERT, B. *ACLS: Suporte Avançado de Vida em Cardiologia*. 5. ed. Rio de Janeiro: Elsevier, 2018.
- AHN, M.-S. Current concepts of premature ventricular contractions. *Journal of Lifestyle Medicine*, v. 3, n. 1, p. 26–33, 2013.
- AMARI, S.; CICHOCKI, A.; YANG, H. H. A new learning algorithm for blind signal separation. MIT Press, p. 757–763, 1996.
- ANTON, H. *Elementary linear algebra*. 10. ed. New York: John Wiley, 2010.
- ATAOLLAH, E.; ALI, K. Detection of premature ventricular contractions using mlp neural networks: A comparative study. *Measurement*, v. 43, p. 103–112, 2010.
- AWAD, M.; KHANNA, R. *Efficient Learning Machines*. New York: Apress Open, 2015.
- BELL, J. *Machine Learning*. Wiley John & Sons, 2014. ISBN 1118889061. Disponível em: <[http://www.ebook.de/de/product/22395855/jason\\_bell\\_machine\\_learning.html](http://www.ebook.de/de/product/22395855/jason_bell_machine_learning.html)>.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006. (Information Science and Statistics).
- BRONSON, R.; COSTA, G. B. *Linear Algebra: An Introduction*. 2. ed. New York: Elsevier, 2007.
- BUCCIGROSSI, R. Breaking up is hard to do: blind signal separation. 1997.
- CAMASTRA, F.; VINCIARELLI, A. *Machine Learning for Audio, Image and Video Analysis*. Springer-Verlag GmbH, 2015. ISBN 1447167341. Disponível em: <[http://www.ebook.de/de/product/24221225/francesco\\_camastra\\_alessandro\\_vinciarelli\\_machine\\_learning\\_for\\_audio\\_image\\_and\\_video\\_analysis.html](http://www.ebook.de/de/product/24221225/francesco_camastra_alessandro_vinciarelli_machine_learning_for_audio_image_and_video_analysis.html)>.
- CATALANO, J. T. *Guide to ECG analysis*. 2. ed. Philadelphia: Lippincott, 2002.
- CHAZAL, P. de; O'DWYER, M.; REILLY, R. B. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, v. 51, n. 7, p. 1196–1206, July 2004. ISSN 0018-9294.
- CICHOCKI, A.; AMARI, S. *Adaptative blind signal e image processing*. New York: Wiley, 2003.
- COSTANZO, L. S. *Fisiologia*. 5. ed. Rio de Janeiro: Elsevier, 2014.
- DIETTERICH, T. G. Ensemble methods in machine learning. *Lecture Notes in Computer Science*, p. 1–15, 2000.

- DOMINGUES, H. H.; IEZZI, G. *Álgebra moderna*. 3. ed. São Paulo: Atual Editora, 1982.
- EBRAHIMPOUR, R.; SADEGHNEJAD, N.; SAJEDIN, A.; MOHAMMADI, N. Electrocardiogram beat classification via coupled boosting by filtering and preloaded mixture of experts. *Neural Comput. and Applic.*, n. 93, p. 1169–1178, 2013.
- EPSTEIN, M. L. *Distúrbios do Ritmo Cardíaco*. 2018. Disponível em: <<http://www.bibliomed.com.br/bibliomed/books/livro8/cap/cap11.htm>>.
- FIGUEREDO, D. G. de. *Análise de Fourier e Equações Diferenciais Parciais*. 4. ed. Rio de Janeiro: IMPA, 2005.
- FRED, K. *ECG Interpretation: From Pathophysiology to Clinical Application*. New York: Springer, 2009.
- GARCIA, G.; MOREIRA, G.; MENOTTI, D.; LUZ, E. Inter-patient ecg heartbeat classification with temporal vcg optimized by pso. *Scientific reports*, v. 7, n. 1, p. 1–11, 2017.
- GARCIA, T. B.; MILLER, G. T. *Arrhythmia Recognition: The Art of Interpretation*. London: Jones and Bartlett Publishers, 2004.
- GERTSCH, M. *The ECG Manual: An Evidence-Based Approach*. New York: Springer, 2009.
- GOLDBERGER, A. L.; AMARAL, L. A. N.; GLASS, L.; HAUSDORFF, J. M.; IVANOV, P. C.; MARK, R. G.; MIETUS, J. E.; MOODY, G. B.; PENG, C.-K.; STANLEY, H. E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, v. 101, n. 23, p. e215–e220, 2000. Circulation Electronic Pages: <http://circ.ahajournals.org/content/101/23/e215.full> PMID:1085218.
- GUYTON, A. C.; HALL, J. E. *Tratado de fisiologia médica*. 11. ed. Rio de Janeiro: Elsevier, 2006.
- HAMPTON, J. R. *ECG Essencial*. 8. ed. Rio de Janeiro: Elsevier, 2014.
- HAYKIN, S. *Redes Neurais: Princípios e prática*. Porto Alegre: Bookman, 2003.
- HESS, M. L. *Doenças cardíacas: primeiros cuidados*. Barueri: Manole, 2002.
- HO, T. K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 8, p. 832–844, 1998.
- HYVÄRINEN, A.; KARHUNEN, J.; ; OJA, E. *Independent Component Analysis*. New York: John Wiley and Sons, Inc., 2001.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An Introduction to Statistical Learning with Applications in R*. New York: Springer, 2017.
- JAYA, Y. B. J.; TAMILSELVI, J. J. Simplified mcdm analytical weighted model for ranking classifiers in financial risk datasets. In: *2014 International Conference on Intelligent Computing Applications*. [S.l.: s.n.], 2014. p. 158–161.

- KENNEDY, H. L.; WHITLOCK, J. A.; SPRAGUE, M. K.; KENNEDY, L. J.; BUCKINGHAM, T. A.; GOLDBERG, R. J. Long-term follow-up of asymptomatic healthy subjects with frequent and complex ventricular ectopy. *New England Journal of Medicine*, v. 312, n. 4, p. 193–197, 1985.
- KHANMOHAMMADI, S.; REZAEIAHARI, M. Ahp based classification algorithm selection for clinical decision support system development. *Procedia Computer Science*, v. 36, p. 328–334, 2014.
- KITTLER, J.; HATEF, M.; DUIN, R. P. W.; MATAS, J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 3, p. 226–239, March 1998. ISSN 0162-8828.
- KOU, G.; WU, W. An analytic hierarchy model for classification algorithms selection in credit risk analysis. *Mathematical Problems in Engineering*, v. 2014, p. 7, 2014.
- KUBAT, M. *An Introduction to Machine Learning*. Springer, 2015. ISBN 978-3-319-20009-5. Disponível em: <[http://www.ebook.de/de/product/24114079/miroslav\\_kubat\\_an\\_introduction\\_to\\_machine\\_learning.html](http://www.ebook.de/de/product/24114079/miroslav_kubat_an_introduction_to_machine_learning.html)>.
- LAMPROPOULOS, A. S.; TSIHRINTZIS, G. A. *Machine Learning Paradigms*. Springer-Verlag, 2015. ISBN 3319191349. Disponível em: <[http://www.ebook.de/de/product/23986573/aristomenis\\_s\\_lampropoulos\\_george\\_a\\_tsihrintzis\\_machine\\_learning\\_paradigms.html](http://www.ebook.de/de/product/23986573/aristomenis_s_lampropoulos_george_a_tsihrintzis_machine_learning_paradigms.html)>.
- LANTZ, B. *Machine Learning with R*. 2. ed. Mumbai: Packt Publishing, 2015. (Community Experience Distilled).
- LATCHAMSETTY, R.; BOGUN, F. Premature ventricular complexes and premature ventricular complex induced cardiomyopathy. *Curr. Probl. Cardiol*, n. 40, p. 379–422, 2015.
- LEWIS, K. M. J. *Multilead ECGs: A practical analysis of Arrhythmias*. 5. ed. Nova York: Delmar Cengage Learning, 2010.
- LI, P.; LIU, C.; WANG, X.; ZHENG, D.; LI, Y.; LIU, C. A low-complexity data-adaptive approach for premature ventricular contraction recognition. *Signal, Image and Video Processing*, n. 8, p. 111–120, 2014.
- LIMA, E. L. *Álgebra Linear*. Rio de Janeiro: IMPA, 2006.
- LIU, X.; DU, H.; WANG, G.; ZHOU, S.; ZHANG, H. Automatic diagnosis of premature ventricular contraction based on lyapunov exponents and lvq neural network. *Computer Methods and Programs in Biomedicine*, v. 122, n. 1, p. 47 – 55, 2015. ISSN 0169-2607. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169260715001777>>.
- LOSCALZO, J. *Medicina cardiovascular de harrison*. 2. ed. Porto Alegre: Artmed e McGraw-Hill, 2013.
- MAIA, I. G. *Ritmos de Escape*. 2018. Disponível em: <<http://www.bibliomed.com.br/bibliomed/books/livro2/cap/cap32.htm>>.

- MALLAT, S. *A Wavelet Tour of Signal Processing: The Sparse Way*. 3. ed. Burlington: Elsevier, 2009.
- MCCALLUM, A.; NIGAM, K. A comparison of event models for naive bayes text classification. In: *AAAI-98 workshop on learning for text categorization*. [S.l.: s.n.], 1998. v. 752, n. 1, p. 41–48.
- MEYER, C. D. *Matrix Analysis and Applied Linear Algebra*. 2. ed. Philadelphia: Siam, 2000.
- MIKA, S.; RATSCH, G.; WESTON, J.; SCHOLKOPF, B.; MULLERS, K.-R. Fisher discriminant analysis with kernels. In: IEEE. *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*. [S.l.], 1999. p. 41–48.
- MOHRMAN, D. E. *Fisiologia cardiovascular*. 6. ed. Porto Alegre: McGraw-Hill, 2011.
- NETTER, F. H. *Coleção Netter de Ilustrações Médicas: Sistema Cardiovascular*. Rio de Janeiro: Saunders Elsevier, 2015.
- NIEVERGELT, Y. *Wavelets Made Easy*. Boston: Birkhäuser, 1999.
- OLIVEIRA, B. R. de; ABREU, C. C. E. de; DUARTE, M. A. Q.; FILHO, J. V. Geometrical features for premature ventricular contraction recognition with analytic hierarchy process based machine learning algorithms selection. *Comp. Methods and Programs in Bio.*, v. 169, p. 59–69, 2019.
- OLIVEIRA, B. R. de; OLIVEIRA, L. R.; DUARTE, M. A. Q. Multicriteria analysis applied at the choice of projects specified by resolution 154/2012 of the national council of justice (in portuguese). *Revista Democracia Digital e Governo Eletrônico*, Revista Democracia Digital e Governo Eletrônico, v. 1, n. 14, p. 121–142, 2016.
- OMER, T. I.; LAURENT, G.; GREGORY, T. A. K. Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. *IEEE Trans. on Biomedical Eng.*, n. 53, p. 2507–2515, 2006.
- PASTORE, C.; PINHO, C.; GERMINIANI, H.; SAMESIMA, N.; MANO, R. Diretrizes da sociedade brasileira de cardiologia sobre análise e emissão de laudos eletrocardiográficos. *Arquivos Brasileiros de Cardiologia*, scielo, v. 93, p. 1–19, 00 2009.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- POLIKAR, R. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, v. 6, n. 3, p. 21–45, Third 2006. ISSN 1531-636X.
- SAATY, R. W. The analytic hierarchy process - what it is and how is is useds. *Mathl Modelling*, v. 9, n. 3, p. 161–176, 1987.

- SAATY, T. L. Decision-making with the ahp: Why is the principal eigenvector necessary. *European Journal of Operational Research*, v. 145, p. 85–91, 2003.
- SAATY, T. L. Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. *Rev. R. Acad. Cien. Serie A. Mat.*, v. 102, n. 2, p. 251–318, 2008.
- SAATY, T. L.; SAGIR, M. Extending the measurement if tangibles to intangibles. *International Journal of Information Technology & Decision Making*, v. 08, n. 01, p. 7–27, 2009.
- SAATY, T. L.; VARGAS, L. G. *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. 2. ed. New York: Springer, 2012.
- SANTOS, J. D.; MEIRA, K. C.; CAMACHO, A. R.; SALVADOR, P. T. C. D. O.; aES, R. M. G.; PIERIN, A. M. G.; oES, F. H. M. D. A. F. T. C. S. v. 23, p. 1621–1634, 2018.
- SANTOS, J. L. F. d. *Eletrocardiograma ao alcance de todos*. 3. ed. São Paulo: Phorte, 2015.
- SHEN, Z.; HU, C.; LI, P.; MENG, M. Q. H. Research on premature ventricular contraction real-time detection based support vector machine. In: *2011 IEEE International Conference on Information and Automation*. [S.l.: s.n.], 2011. p. 864–869.
- STONE, C. K.; HUMPHRIES, R. L. *Current Diagnóstico e Tratamento: Medicina de Emergência*. 7. ed. Porto Alegre: AMGH Editora Ltda, 2013.
- SUGIYAMA, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of machine learning research*, v. 8, n. May, p. 1027–1061, 2007.
- TONG, L.; LIU, R.; SOON, V.; HUANG, Y.-F. Indeterminacy and identifiability of blind identification. *Circuits and Systems, IEEE Transactions on*, v. 38, n. 5, p. 499–509, May 1991. ISSN 0098-4094.
- TONG, L.; SOON, V.; HUANG, Y.; LIU, R. Amuse: a new blind identification algorithm. v. 3, p. 1784–1787, May 1990.
- VAIDYA, O. S.; KUMAR, S. Analytic hierarchy process: An overview of applications. *European Journal of Operational Research*, v. 169, p. 1–29, 2006.
- WEBB, A. *Statistical Pattern Recognition*. England: John Wiley & Sons, Ltd, 2002.
- WEEKS, M. *Processamento Digital de Sinais Utilizando Matlab e Wavelets*. Rio de Janeiro: LTC, 2012.
- WEISZFLOG, W. *Moderno Dicionário da Língua Portuguesa*. Editora Melhoramentos Ltda, 2015. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues>>.
- WOOD, S. L.; FROELICHER, E. S. S.; MOTZER, S. U. *Enfermagem em Cardiologia*. 4. ed. Barueri: Manole, 2005.
- WOODS, S. L.; FROELICHER, E. S. S.; MOTZER, S. U. *Enfermagem em cardiologia*. Barueri: Manole, 2005.

ZAKI, M. J.; JUNIOR, W. M. *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press, 2014. Disponível em: <<http://www.dataminingbook.info/>>.

ZAREI, R.; HE, J.; HUANG, G.; ZHANG, Y. Effective and efficient detection of premature ventricular contractions based on variation of principal directions. *Digital Signal Processing*, v. 50, p. 93–102, 2016.

ZHU, M. *A study of the generalized eigenvalue decomposition in discriminant analysis*. Tese (Doutorado) — The Ohio State University, 2006.

# A Quantidade de batimentos cardíacos por classe

Na Tabela 31 estão discriminadas as quantidades de batimentos cardíacos por tipo de batimento, para a base de dados MIT/BIH-ARDB, onde os símbolos têm os seguintes significados:

**N:** Normal

**L:** Bloqueio de ramo esquerdo

**R:** Bloqueio de ramo direito

**e:** Escape atrial

**j:** Escape Nodal (juncional)

**A:** Batimento atrial prematuro

**a:** Batimento atrial prematuro aberrante

**J:** Batimento prematuro Nodal (juncional)

**S:** Batimento supraventricular prematuro ou Batimento ectópico

**V:** Contração prematura ventricular

**E:** Escape ventricular

**F:** Fusão de batimento ventricular e normal

**f:** Fusão entre batimento ritimado e normal

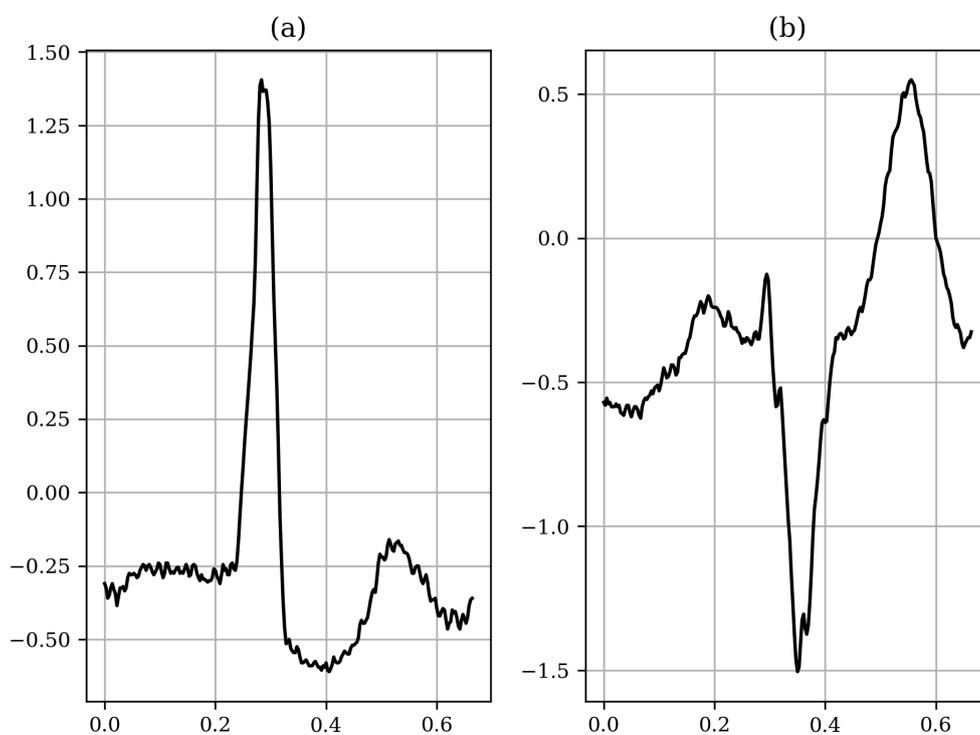
Tabela 31 – Quantidade de batimentos cardíacos para cada classe de arritmia.

Registro	Classes de batimentos cardíacos															Total
	N	L	R	e	j	A	a	J	S	V	E	F	P	f	U	
100	2239	0	0	0	0	33	0	0	0	1	0	0	0	0	0	2273
101	1860	0	0	0	0	3	0	0	0	0	0	0	0	0	2	1865
102	99	0	0	0	0	0	0	0	0	4	0	0	2028	56	0	2187
103	2082	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2084
104	163	0	0	0	0	0	0	0	0	2	0	0	1380	666	18	2229
105	2526	0	0	0	0	0	0	0	0	41	0	0	0	0	5	2572
106	1507	0	0	0	0	0	0	0	0	520	0	0	0	0	0	2027
107	0	0	0	0	0	0	0	0	0	59	0	0	2078	0	0	2137
108	1740	0	0	0	1	4	0	0	0	17	0	2	0	0	0	1763
109	0	2492	0	0	0	0	0	0	0	38	0	2	0	0	0	2532
111	0	2123	0	0	0	0	0	0	0	1	0	0	0	0	0	2124
112	2537	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2539
113	1789	0	0	0	0	0	6	0	0	0	0	0	0	0	0	1795
114	1820	0	0	0	0	10	0	2	0	43	0	4	0	0	0	1879
115	1953	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1953
116	2302	0	0	0	0	1	0	0	0	109	0	0	0	0	0	2412
117	1534	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1535
118	0	0	2166	0	0	96	0	0	0	16	0	0	0	0	0	2278
119	1543	0	0	0	0	0	0	0	0	444	0	0	0	0	0	1987
121	1861	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1863
122	2476	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2476
123	1515	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1518
124	1531	0	0	0	5	2	0	29	0	47	0	5	0	0	0	1619
200	1743	0	0	0	0	30	0	0	0	826	0	2	0	0	0	2601
201	1625	0	0	0	10	30	97	1	0	198	0	2	0	0	0	1963
202	2061	0	0	0	0	36	19	0	0	19	0	1	0	0	0	2136
203	2529	0	0	0	0	0	2	0	0	444	0	1	0	0	4	2980
205	2571	0	0	0	0	3	0	0	0	71	0	11	0	0	0	2656
207	0	1457	86	0	0	107	0	0	0	105	105	0	0	0	0	1755
208	1586	0	0	0	0	0	0	0	2	992	0	373	0	0	2	2955
209	2621	0	0	0	0	383	0	0	0	1	0	0	0	0	0	3005
210	2423	0	0	0	0	0	22	0	0	194	1	10	0	0	0	2650
212	923	0	1825	0	0	0	0	0	0		0	0	0	0	0	2748
213	2641	0	0	0	0	25	3	0	0	220	0	362	0	0	0	3251
214	0	2003	0	0	0	0	0	0	0	256	0	1	0	0	2	2262
215	3195	0	0	0	0	3	0	0	0	164	0	1	0	0	0	3363
217	244	0	0	0	0	0	0	0	0	162	0	0	1542	260	0	2208
219	2082	0	0	0	0	7	0	0	0	64	0	1	0	0	0	2154
220	1954	0	0	0	0	94	0	0	0	0	0	0	0	0	0	2048
221	2031	0	0	0	0	0	0	0	0	396	0	0	0	0	0	2427
222	2062	0	0	0	212	208	0	1	0	0	0	0	0	0	0	2483
223	2029	0	0	16	0	72	1	0	0	473	0	14	0	0	0	2605
228	1688	0	0	0	0	3	0	0	0	362	0	0	0	0	0	2053
230	2255	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2256
231	314	0	1254	0	0	1	0	0	0	2	0	0	0	0	0	1571
232	0	0	397	0	1	1382	0	0	0	0	0	0	0	0	0	1780
233	2230	0	0	0	0	7	0	0	0	831	0	11	0	0	0	3079
234	2700	0	0	0	0	0	0	50	0	3	0	0	0	0	0	2753
<b>Total</b>	<b>76584</b>	<b>8075</b>	<b>5728</b>	<b>16</b>	<b>229</b>	<b>2546</b>	<b>150</b>	<b>83</b>	<b>2</b>	<b>7130</b>	<b>106</b>	<b>803</b>	<b>7028</b>	<b>982</b>	<b>33</b>	<b>109495</b>

## B Experimentos auxiliares da abordagem proposta

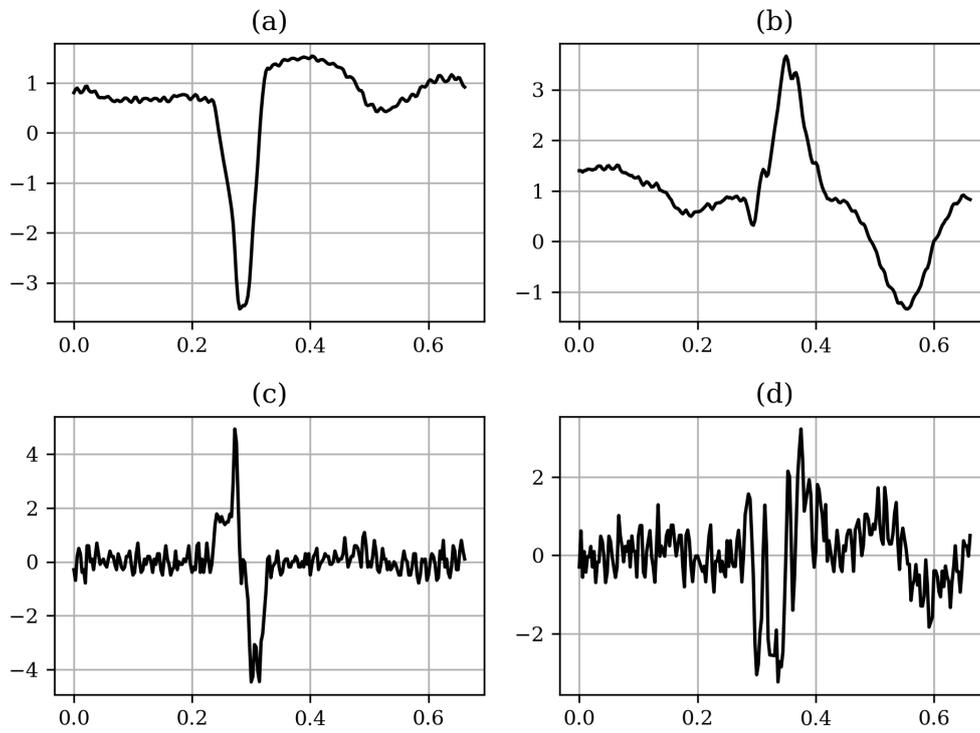
As Figuras 29, 30 e 31, ilustram, respectivamente, dois exemplos de batimentos cardíacos do registro 203 da base de dados MIT/BIH-ARDB, e as componentes estimadas pelo método AMUSE. Essas figuras ajudam a entender o comportamento semelhante ao um banco de filtros, que é realizado pelo método AMUSE, conforme discutido no Capítulo 5.

Figura 29 – Exemplos de batimentos cardíacos (a) Normal e (b) CPV do registro 203 da base de dados MIT/BIH-ARDB.



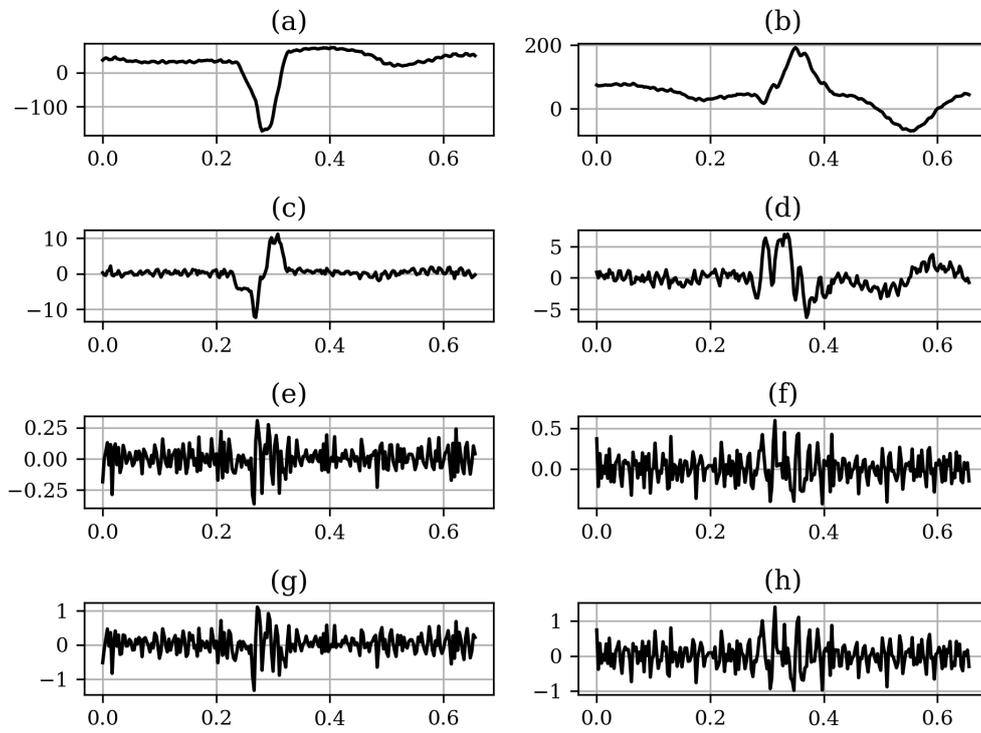
Fonte: Autoria própria.

Figura 30 – Componentes estimadas pelo método AMUSE para o batimento cardíaco (a) e (b) Normal e para (c) e (d) CPV.



Fonte: Autoria própria.

Figura 31 – Componentes estimadas pelo método AMUSE para o batimento cardíaco (a), (c), (e) e (g) Normal e para (b), (d), (f) e (h) CPV.



Fonte: Autoria própria.

# C Experimento com quatro componentes do AMUSE

Tabela 32 – Resultados do Experimento VII, considerando 267 atributos e quatro componentes do AMUSE e o método AdaBoost induzido com quatro máquinas

Comp.	$A_{cc}$	$F(1)^+$	$F(1)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Primeira	<b>0,9188</b>	0,0000	<b>0,9577</b>	0,00000	<b>1,0000</b>	0,0000	0,9188
Segunda	<b>0,9188</b>	0,0000	<b>0,9577</b>	0,0000	<b>1,0000</b>	0,0000	<b>0,9188</b>
Terceira	0,2714	0,1317	0,3724	0,6803	0,2352	<b>0,0729</b>	0,8928
Quarta	0,0940	<b>0,1456</b>	0,0357	<b>0,9509</b>	0,0183	0,0788	0,8080
Todas	<b>0,9188</b>	0,0000	<b>0,9577</b>	0,0000	<b>1,0000</b>	0,0000	<b>0,9188</b>

Fonte: Autoria própria.

## D Experimentos adicionais para comparação com outras abordagens

Os resultados exibidos neste apêndice são necessários para comparação com as outras abordagens para reconhecimento de CPV, eis que cada uma delas implementa suas metodologias sobre bases distintas. Sendo assim, a comparação é mais justa quando realizada sobre os mesmos conjuntos de dados.

Tabela 33 – Resultados considerando as bases DS3 e DS4 para treinamento e teste, respectivamente. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(\zeta)^+$	$F(\zeta)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
Clf <sub>0</sub>	0,9785	0,8568	0,9884	0,8625	0,9879	0,8512	0,9889
Clf <sub>1</sub>	0,9850	0,9019	<b>0,9918</b>	0,9292	0,9894	0,8762	0,9943
Voto majoritário							
Suave	0,9839	0,8942	0,9520	0,9142	0,9895	0,8751	0,9931
Rígido - N	0,9841	0,8890	0,9154	0,8525	<b>0,9947</b>	<b>0,9287</b>	0,9882
Rígido - P	0,9793	0,8713	0,9663	0,9392	0,9826	0,8125	0,9950
Voto AHP							
$\alpha = 0,3$	0,9820	0,8862	0,9672	<b>0,9408</b>	0,9853	0,8375	<b>0,9952</b>
$\alpha = 0,4$	0,9842	0,8984	0,9642	0,9354	0,9882	0,8641	0,9948
$\alpha = 0,5$	0,9857	0,9062	0,9597	0,9275	0,9904	0,8858	0,9941
$\alpha = 0,6$	0,9861	0,9077	0,9530	0,9158	0,9918	0,8997	0,9932
$\alpha = 0,7$	<b>0,9867</b>	<b>0,9102</b>	0,9472	0,9058	0,9932	0,9146	0,9924

Fonte: Autoria própria.

Tabela 34 – Resultados considerando as bases DS5 e DS6 para treinamento e teste, respectivamente. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(\zeta)^+$	$F(\zeta)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
Clf <sub>0</sub>	0,9810	0,8778	0,9897	0,9793	0,9812	0,7953	0,9984
Clf <sub>1</sub>	0,9776	0,8601	0,9878	0,9904	0,9766	0,7601	0,9993
Voto majoritário							
Suave	0,9802	0,8742	0,9957	<b>0,9920</b>	0,9793	0,7814	<b>0,9994</b>
Rígido - N	0,9866	0,9102	0,9879	0,9777	0,9872	0,8514	0,9983
Rígido - P	0,9721	0,8316	<b>0,9957</b>	<b>0,9920</b>	0,9706	0,7158	<b>0,9994</b>
Voto AHP							
$\alpha = 0,3$	0,9709	0,8260	<b>0,9957</b>	<b>0,9920</b>	0,9694	0,7076	<b>0,9994</b>
$\alpha = 0,4$	0,9764	0,8538	<b>0,9957</b>	<b>0,9920</b>	0,9752	0,7494	<b>0,9994</b>
$\alpha = 0,5$	0,9802	0,8733	0,9914	0,9841	0,9799	0,7850	0,9988
$\alpha = 0,6$	0,9839	0,8949	0,9914	0,9841	0,9839	0,8205	0,9988
$\alpha = 0,7$	<b>0,9869</b>	<b>0,9121</b>	0,9870	0,9761	<b>0,9877</b>	<b>0,8559</b>	0,9982

Fonte: Autoria própria.

Tabela 35 – Resultados considerando as bases DS7 e DS8 para treinamento e teste, respectivamente. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(\zeta)^+$	$F(\zeta)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
Clf <sub>0</sub>	0,9901	0,9437	0,9946	0,9853	0,9906	0,9055	0,9986
Clf <sub>1</sub>	0,9914	0,9511	<b>0,9953</b>	<b>1,0000</b>	0,9906	0,9068	<b>1,0000</b>
Voto majoritário							
Suave	0,9922	0,9554	<b>1,0000</b>	<b>1,0000</b>	0,9914	0,9146	<b>1,0000</b>
Rígido - N	0,9955	0,9736	0,9919	0,9853	0,9964	0,9621	0,9987
Rígido - P	0,9860	0,9230	<b>1,0000</b>	<b>1,0000</b>	0,9847	0,8570	<b>1,0000</b>
Voto AHP							
$\alpha = 0,3$	0,9863	0,9243	<b>1,0000</b>	<b>1,0000</b>	0,9850	0,8593	<b>1,0000</b>
$\alpha = 0,4$	0,9894	0,9408	<b>1,0000</b>	<b>1,0000</b>	0,9885	0,8882	<b>1,0000</b>
$\alpha = 0,5$	0,9922	0,9559	<b>1,0000</b>	<b>1,0000</b>	0,9915	0,9155	<b>1,0000</b>
$\alpha = 0,6$	0,9941	0,9660	0,9994	0,9990	0,9937	0,9352	0,9999
$\alpha = 0,7$	<b>0,9957</b>	<b>0,9746</b>	0,9925	0,9864	<b>0,9965</b>	<b>0,9631</b>	0,9987

Fonte: Autoria própria.

Tabela 36 – Resultados considerando a base DS9 para treinamento e teste. Valores em negrito são os maiores da coluna.

Máquina	$A_{cc}$	$F(\zeta)^+$	$F(\zeta)^-$	$S_e$	$S_p$	$P^+$	$P^-$
Individual							
$Clf_0$	0,9785	0,9518	0,9862	0,9443	0,9884	0,9594	0,9839
$Clf_1$	0,9851	0,9671	0,9904	0,9757	0,9878	0,9587	0,9929
Voto majoritário							
Suave	0,9875	0,9723	0,9839	0,9751	0,9911	0,9696	0,9928
Rígido - N	0,9842	0,9639	0,9608	0,9398	<b>0,9970</b>	0,9893	0,9828
Rígido - P	0,9794	0,9553	0,9871	<b>0,9802</b>	0,9792	0,9317	0,9942
Voto AHP							
$\alpha = 0,3$	0,9836	0,9641	<b>0,9865</b>	0,9792	0,9849	<b>0,9939</b>	<b>0,9962</b>
$\alpha = 0,4$	0,9855	0,9681	0,9856	0,9778	0,9878	0,9586	0,9935
$\alpha = 0,5$	0,9861	0,9693	0,9841	0,9755	0,9892	0,9631	0,9929
$\alpha = 0,6$	0,9869	0,9710	0,9831	0,9739	0,9907	0,9681	0,9924
$\alpha = 0,7$	<b>0,9878</b>	<b>0,9727</b>	0,9815	0,9714	0,9925	0,9740	0,9917

Fonte: Autoria própria.