

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E  
INFORMÁTICA INDUSTRIAL

EDRAS REILY PACOLA

**Uso da análise de discriminante linear em conjunto com a transformada  
wavelet discreta no reconhecimento de espículas**

TESE

CURITIBA

2016

**EDRAS REILY PACOLA**

**Uso da análise de discriminante linear em conjunto com a transformada wavelet discreta no reconhecimento de espículas**

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Doutor em Ciências” – Área de Concentração: Engenharia Biomédica.

Orientador: Humberto Remigio Gamba

Coorientador: Miguel Antônio Sovierzoski

**CURITIBA**

**2016**

---

**Dados Internacionais de Catalogação na Publicação**

---

P121u Pacola, Edras Reily  
2016      Uso da análise discriminante linear em conjunto com a transformada wavelet discreta no reconhecimento de espículas / Edras Reily Pacola.-- 2016.  
143 f.: il.; 30 cm

Texto em português, com resumo em inglês.  
Tese (Doutorado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Curitiba, 2016.  
Bibliografia: f. 138-143.

1. Eletroencefalografia. 2. Epilepsia. 3. Processamento de sinais - Técnicas digitais. 4. Detecção de sinais. 5. Wavelets (Matemática). 6. Análise discriminatória. 7. Reconhecimento de padrões. 8. Software - Desenvolvimento. 9. Simulação (Computadores). 10. Engenharia biomédica. 11. Engenharia elétrica - Teses. I. Gamba, Humberto Remígio, orient. II. Sovierzoski, Miguel Antonio, coorient. III. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial. IV. Título.

CDD: Ed. 22 -- 621.3

---

**Biblioteca Central da UTFPR, Câmpus Curitiba**

**Título da Tese Nº. \_\_\_\_\_**

**Uso da Análise de Discriminante Linear em Conjunto com a Transformada Wavelet Discreta no Reconhecimento de Espículas.**

por

**Edras Reily Pacola**

**Orientador:** Prof. Dr. Humberto Remígio Gamba

**Coorientador:** Prof. Dr. Miguel Antonio Sovierzoski

Esta tese foi apresentada como requisito parcial à obtenção do grau de DOUTOR EM CIÊNCIAS – Área de Concentração: Engenharia Biomédica, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI – da Universidade Tecnológica Federal do Paraná – UTFPR, às 08h30m do dia 18 de dezembro de 2015. O trabalho foi aprovado pela Banca Examinadora, composta pelos doutores:

---

Prof. Dr. Humberto Remígio Gamba  
(Presidente – UTFPR)

---

Prof. Dr. Raimes Moraes  
(UFSC)

---

Prof. Dr. Alcimar Soares Barbosa  
(UFU)

---

Prof. Dr. Marcos Vinício Hass Rambo  
(UFPR)

---

Prof. Dr. Sérgio Francisco Pichorim  
(UTFPR)

---

Prof<sup>ª</sup>. Dr. Lúcia Valéria Ramos de Arruda  
(UTFPR)

Visto da Coordenação:

---

Prof. Emilio Carlos Gomes Wille, Dr.  
(Coordenador do CPGEI)

## **AGRADECIMENTOS**

Agradeço primeiramente à minha esposa Verônica Isabela Quandt e às minhas filhas Sabrina e Rafaela pela ajuda e apoio irrestrito dados durante todo o tempo dedicado à realização deste trabalho.

Agradeço ao meu orientador Prof. Dr. Humberto Remigio Gamba, e especialmente ao coorientador Prof. Dr. Miguel Antônio Sovierzoski que me apoiou desde o início, no projeto da tese, ensaios, qualificação, nas publicações e na elaboração da tese.

Agradeço também ao Dr. Paulo B. M. Liberalesso, médico do departamento de neuropediatria do Hospital Pequeno Príncipe.

Muito obrigado!

## RESUMO

PACOLA, Edras R.. Uso da análise de discriminante linear em conjunto com a transformada wavelet discreta no reconhecimento de espículas. 143 f. Tese – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2016.

Pesquisadores têm concentrado esforços, nos últimos 20 anos, aplicando a transformada wavelet no processamento, filtragem, reconhecimento de padrões e na classificação de sinais biomédicos, especificamente em sinais de eletroencefalografia (EEG) contendo eventos característicos da epilepsia, as espículas. Várias famílias de wavelets-mães foram utilizadas, mas sem um consenso sobre qual wavelet-mãe é a mais adequada para essa finalidade. Os sinais utilizados apresentam uma gama muito grande de eventos e não possuem características padronizadas. A literatura relata sinais de EEG amostrados entre 100 a 600 Hz, com espículas variando de 20 a 200 ms. Nesse estudo foram utilizadas 98 wavelets. Os sinais de EEG foram amostrados de 200 a 1 kHz. Um neurologista marcou um conjunto de 494 espículas e um conjunto de 1500 eventos não-espícula. Esse estudo inicia avaliando a quantidade de decomposições wavelets necessárias para a detecção de espículas, seguido pela análise detalhada do uso combinado de wavelets-mães de uma mesma família e entre famílias. Na sequência é analisada a influência de descritores e o uso combinado na detecção de espículas. A análise dos resultados desses estudos indica que é mais adequado utilizar um conjunto de wavelets-mães, com vários níveis de decomposição e com vários descritores, ao invés de utilizar uma única wavelet-mãe ou um descritor específico para a detecção de espículas. A seleção desse conjunto de wavelets, de níveis de decomposição e de descritores permite obter níveis de detecção elevados conforme a carga computacional que se deseje ou a plataforma computacional disponível para a implementação. Como resultado, esse estudo atingiu níveis de desempenho entre 0,9936 a 0,9999, dependendo da carga computacional. Outras contribuições desse estudo referem-se à análise dos métodos de extensão de borda na detecção de espículas; e a análise da taxa de amostragem de sinais de EEG no desempenho do classificador de espículas, ambos com resultados significativos. São também apresentadas como contribuições: uma nova arquitetura de detecção de espículas, fazendo uso da análise de discriminante linear; e a apresentação de um novo descritor, energia centrada, baseado na resposta dos coeficientes das sub-bandas de decomposição da transformada wavelet, capaz de melhorar a discriminação de eventos espícula e não-espícula.

**Palavras-chave:** espícula, wavelet, LDA, EEG, epilepsia

## ABSTRACT

PACOLA, Edras R.. THE LINEAR DISCRIMINANT ANALYSIS USAGE COMBINED WITH THE DISCRETE WAVELET TRANSFORM IN SPIKE DETECTION. 143 f. Tese – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2016.

Researchers have concentrated efforts in the past 20 years, by applying the wavelet transform in processing, filtering, pattern recognition and classification of biomedical signals, in particular signals of electroencephalogram (EEG) containing events characteristic of epilepsy, the spike. Several families of mother-wavelets were used, but there are no consensus about which mother-wavelet is the most adequate for this purpose. The signals used have a wide range of events. The literature reports EEG signals sampled from 100 to 600 Hz with spikes ranging from 20 to 200 ms. In this study we used 98 wavelets. The EEG signals were sampled from 200 Hz up to 1 kHz. A neurologist has scored a set of 494 spikes and a set 1500 non-spike events. This study starts evaluating the amount of wavelet decompositions required for the detection of spikes, followed by detailed analysis of the combined use of mother-wavelets of the same family and among families. Following is analyzed the influence of descriptors and the combined use of them in spike detection. The results of these studies indicate that it is more appropriate to use a set of mother-wavelets, with many levels of decomposition and with various descriptors, instead of using a single mother-wavelet or a specific descriptor for the detection of spikes. The selection of this set of wavelets, decomposition level and descriptors allows to obtain high levels of detection according to the computational load desired or computing platform available for implementation. This study reached performance levels between 0.9936 to 0.9999, depending on the computational load. Other contributions of this study refer to the analysis of the border extension methods for spike detection; and the influences of the EEG signal sampling rate in the classifier performance, each one with significant results. Also shown are: a new spike detection architecture by making use of linear discriminant analysis; and the presentation of a new descriptor, the centred energy, based on the response of the coefficients of decomposition levels of the wavelet transform, able to improve the discrimination of spike and non-spike events.

**Keywords:** spike, wavelet, LDA, EEG, epilepsy

## LISTA DE FIGURAS

FIGURA 1	– Custo do tratamento da epilepsia no Brasil em 2011	16
FIGURA 2	– Logotipo do dia internacional do cérebro e do dia roxo	17
FIGURA 3	– Sistema de montagem de eletrodos 10-20	22
FIGURA 4	– EEG - Descriptor morfológico de evento epileptiforme	24
FIGURA 5	– EEG - Espículas	24
FIGURA 6	– EEG - Ondas agudas	25
FIGURA 7	– EEG - Artefatos - piscada palpebral	25
FIGURA 8	– Teste Diagnóstico - ponto de corte, VP, VN, FP, FN	27
FIGURA 9	– Curva ROC	30
FIGURA 10	– DWT - Decomposição wavelet	34
FIGURA 11	– DWT - Decomposição wavelet multirresolução	35
FIGURA 12	– Algoritmo de coeficientes wavelet	36
FIGURA 13	– Extensão de bordas	45
FIGURA 14	– Diagrama do procedimento de geração de extensão de borda na WT	46
FIGURA 15	– Aviso de atenção do software Matlab®	47
FIGURA 16	– Autovetor e autovalor	48
FIGURA 17	– Projeção LDA unidimensional	51
FIGURA 18	– Histograma de distribuição das espículas da base de sinais.	68
FIGURA 19	– Diagrama em blocos do detector de espículas proposto.	69
FIGURA 20	– Cálculo da energia centrada	73
FIGURA 21	– Energia centrada para janela deslizante	74
FIGURA 22	– Construção da matriz de características	75
FIGURA 23	– Tela do programa proprietário de EEG	76
FIGURA 24	– Tela do programa de seleção e exportação de espículas	77
FIGURA 25	– Espículas selecionadas no programa	77
FIGURA 26	– Exportação de espícula pelo programa	78
FIGURA 27	– Espícula exportada pelo programa	78
FIGURA 28	– Primeira tela do programa	79
FIGURA 29	– Segunda tela do programa	80
FIGURA 30	– Terceira tela do programa	80
FIGURA 31	– Tela de resultados do programa	81
FIGURA 32	– Tela de LDA do programa	82
FIGURA 33	– Extensão de borda SP0 e PPD na espícula	87
FIGURA 34	– Diagrama do classificador para extensão de bordas	88
FIGURA 35	– Índices de desempenho AUC, Acc e FPR para extensão de bordas	91
FIGURA 36	– Diferenças entre índices de desempenho AUC, Acc e FPR	93
FIGURA 37	– Índice de desempenho AAF para extensão de bordas	94
FIGURA 38	– Decomposições wavelet por família de wavelets-mães	105
FIGURA 39	– Combinações de wavelet wavelets-mães da mesma família	111
FIGURA 40	– Distribuição das classes espícula e não espícula após o LDA	112
FIGURA 41	– Wavelets-mães selecionadas, AUC e dimensões	116
FIGURA 42	– Informação discriminatória de 6 descritores	118



FIGURA 43	–	Informação discriminatória de outros 8 descritores .....	119
FIGURA 44	–	Índice de desempenho AUC para descritores e wavelets .....	121
FIGURA 45	–	Tempo de Processamento x Dimensões .....	122
FIGURA 46	–	Tempo de Processamento x AUC .....	123
FIGURA 47	–	Espícula amostrada a 1000 e 200 Hz com DB4 .....	126
FIGURA 48	–	Índice de desempenho AUC de 100 a 1000 Hz .....	128
FIGURA 49	–	Configuração do classificador de eventos epileptiformes .....	134

## LISTA DE TABELAS

TABELA 1	– Sub-bandas de frequências do EEG .....	23
TABELA 2	– Coeficientes Haar .....	36
TABELA 3	– Coeficientes da função escala para DB1 a DB19 .....	38
TABELA 4	– Coeficientes da função escala e wavelet para Bior11 a Bior68 .....	39
TABELA 5	– Coeficientes da função escala Coif1 a Coif5 .....	40
TABELA 6	– Coeficientes da função escala Sym2 a Sym20 .....	41
TABELA 7	– Coeficientes da função escala DMeyer .....	42
TABELA 8	– Coeficientes da função escala da família Legendre .....	43
TABELA 9	– Métodos de extensão de borda .....	44
TABELA 10	– Comparação entre especialistas .....	61
TABELA 11	– Descritores utilizados .....	69
TABELA 12	– Avaliação comparativa de resultados .....	85
TABELA 13	– Número de coeficientes de cada wavelet-mãe .....	89
TABELA 14	– Índices de desempenho para extensão de bordas .....	89
TABELA 15	– Combinações de decomposições wavelet .....	97
TABELA 16	– Grupos de wavelets-mães .....	97
TABELA 17	– Combinação de decomposições para DB1 a DB19 .....	98
TABELA 18	– Combinação de decomposições para DB20 a DB38 .....	99
TABELA 19	– Combinação de decomposições para Bior11 a Bior68 .....	100
TABELA 20	– Combinação de decomposições para RBior11 a RBior68 .....	101
TABELA 21	– Combinação de decomposições para Sym2 a Sym20 .....	102
TABELA 22	– Combinação de decomposições para o grupo 5 de wavelets-mães .....	103
TABELA 23	– Cinco melhores wavelets-mães de cada grupo .....	106
TABELA 24	– Desempenho percentual entre combinações de decomposições .....	107
TABELA 25	– Índice de desempenho AUC para grupos de wavelets do grupo 1 .....	108
TABELA 26	– Índice de desempenho AUC para grupos de wavelets dos grupos 2 e 3 ..	109
TABELA 27	– Índice de desempenho AUC para grupos de wavelets do grupo 4 .....	109
TABELA 28	– Índice de desempenho AUC para grupos de wavelets do grupo 5 .....	110
TABELA 29	– Pontos da curva ROC para DB1-DB5, (D1234-), ApEn .....	113
TABELA 30	– Comparação de energia para decomposições de espículas .....	113
TABELA 31	– Resultado comparativo das wavelets-mães Haar Ort. e RBior31 .....	115
TABELA 32	– Índice de desempenho AUC para descritores individuais .....	120
TABELA 33	– Índice de desempenho AUC para descritores agrupados .....	120
TABELA 34	– Índices de desempenho AUC para grupos de wavelets e descritores ....	121
TABELA 35	– Frequência de amostragem e filtros utilizados .....	125
TABELA 36	– Frequências da WT para sinais amostrados de 100 Hz a 1000 Hz .....	126
TABELA 37	– Níveis de decomposição para cada frequência de amostragem .....	127
TABELA 38	– Resultado de frequências de amostragem para descritores selecionados	127
TABELA 39	– Resultado de frequências de amostragem para descritor ApEn .....	128

## LISTA DE SIGLAS

Acc	<i>Accuracy</i>
ANN	<i>Artificial Neural Network</i>
ANVISA	Agência Nacional de Vigilância Sanitária
ApEn	<i>Approximate Entropy</i>
ASYMH	<i>Antisymmetric-Padding Half-Point</i>
ASYMW	<i>Antisymmetric-Padding Whole-Point</i>
AUC	<i>Area under the ROC Curve</i>
cA	Coeficientes de Aproximação
cD	Coeficientes de Detalhe
CD	<i>Correlation Dimension</i>
CWT	<i>Continuous Wavelet Transform</i>
DBN	Wavelet-Mãe Daubechies $N$
DWT	<i>Discrete Wavelet Transform</i>
EEG	Eletoencefalograma, Eletoencefalógrafo, Eletoencefalografia
ELM	<i>Extreme Learning Machine</i>
FD	<i>Fractal Dimension</i>
FN	Falso Negativo
FP	Falso Positivo
FPR	<i>False Positive Rate</i>
GA	<i>Genetic Algorithm</i>
H	<i>Hurst Exponent</i>
HHT	<i>Hilbert-Huang Transform</i>
HOS	<i>Higher Order Spectra</i>
IBE	<i>International Bureau for Epilepsy</i>
ICA	<i>Independent Component Analysis</i>
ICF	<i>Information Cost Function</i>
ILAE	<i>International League Against Epilepsy</i>
LBWT	<i>Lifting Base Wavelet Transform</i>
LDA	<i>Linear Discriminant Analysis</i>
LLE	<i>Largest Lyapunov Exponent</i>
MLP	<i>Multilayer Perceptron</i>
OMS	Organização Mundial de Saúde
OPF	<i>Optimum Path Forest</i>
PCA	<i>Principal Component Analysis</i>
PCA	<i>Principal Component Analysis</i>
PDF	<i>Probability Density Function</i>
PER	<i>Periodic-Even-Padding</i>
PNN	<i>Probabilistic Neural Network</i>
PPD	<i>Periodic-Padding</i>
ROC	<i>Receiver Operating Characteristics</i>
RP	<i>Recurrence Plot</i>

RQA	<i>Recurrence Quantification Analysis</i>
SampEn	<i>Sample Entropy</i>
SP0	<i>Smooth-Padding of order 0</i>
SP1	<i>Smooth-Padding of order 1</i>
SVM	<i>Support Vector Machine</i>
SYMH	<i>Symmetric-Padding Half-Point</i>
SYMW	<i>Symmetric-Padding Whole-Point</i>
TNR	<i>True Negative Rate</i>
TPR	<i>True Positive Rate</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
WDM	<i>Wavelet Detection Method</i>
WF	<i>Wavelet Frame</i>
WFN	<i>World Federation of Neurology</i>
WP	<i>Wavelet Packet</i>
WPT	<i>Wavelet Packet Transform</i>
WT	<i>Wavelet Transform</i>
ZPD	<i>Zero-Padding</i>

## LISTA DE SÍMBOLOS

$\delta$	Sub-banda do EEG de 0 a 4 Hz
$\theta$	Sub-banda do EEG de 4 a 8 Hz
$\alpha$	Sub-banda do EEG de 8 a 15 Hz
$\beta$	Sub-banda do EEG de 15 a 30 Hz
$\gamma$	Sub-banda do EEG de 30 a 60 Hz
$\psi$	Função Wavelet-Mãe
$\phi$	Função Escala

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>14</b>
1.1 MOTIVAÇÃO .....	15
1.2 VISÃO GERAL DO TRABALHO .....	17
1.3 OBJETIVOS .....	18
1.4 COMENTÁRIOS .....	19
<b>2 FUNDAMENTAÇÃO</b> .....	<b>21</b>
2.1 ELETROENCEFALOGRAMA .....	21
2.1.1 Posicionamento dos Eletrodos .....	21
2.1.2 Montagem dos Eletrodos .....	22
2.1.3 Resposta Elétrica do Eletroencefalograma .....	22
2.2 ESPÍCULAS E ONDAS AGUDAS NO ELETROENCEFALOGRAMA .....	23
2.2.1 Artefatos no Eletroencefalograma .....	24
2.3 TESTE DIAGNÓSTICO .....	26
2.4 CURVA ROC .....	29
2.4.1 Índice de desempenho da curva ROC - AUC .....	29
2.5 WAVELETS .....	30
2.5.1 Transformada Wavelet Contínua .....	31
2.5.2 Transformada Wavelet Discreta .....	33
2.5.3 Famílias de Wavelets-Mães .....	34
2.6 CONVOLUÇÃO E MÉTODOS DE EXTENSÃO DE BORDA .....	43
2.6.1 Extensão de Borda na Transformada Wavelet Discreta .....	45
2.6.2 Extensão de Borda no Matlab® .....	47
2.7 AUTOVETORES E AUTOVALORES .....	48
2.8 ANÁLISE DE DISCRIMINANTE LINEAR .....	49
2.9 COMENTÁRIOS .....	53
<b>3 ESTADO DA ARTE</b> .....	<b>54</b>
3.1 TRANSFORMADA WAVELET APLICADA NO RECONHECIMENTO DE ESPÍCULAS .....	54
3.2 ABORDAGENS DO ESTADO DA ARTE .....	63
3.2.1 Estado da Arte - A Espícula e o Registro do Sinal .....	63
3.2.2 Estado da Arte - Wavelets-Mães na Detecção da Espícula .....	63
3.2.3 Estado da Arte - Extração de Características da Transformada Wavelet .....	64
3.2.4 Estado da Arte - Classificadores .....	64
3.2.5 Estado da Arte - Resultados obtidos na detecção de espículas .....	65
3.3 COMENTÁRIOS .....	65
<b>4 MATERIAIS E FERRAMENTAS</b> .....	<b>67</b>
4.1 BANCO DE SINAIS DE EEG .....	67
4.2 CLASSIFICADOR DE EVENTOS EPILEPTIFORMES .....	68
4.3 DESCRITORES .....	69
4.3.1 Descritor Entropia Aproximada .....	71
4.3.2 Descritor Energia Centrada .....	72

4.4	MATRIZ DE CARACTERÍSTICAS .....	74
4.5	SOFTWARE .....	75
4.5.1	Programa de seleção de espículas .....	76
4.5.2	Programa de processamento de espículas .....	78
4.6	COMENTÁRIOS .....	83
<b>5</b>	<b>MÉTODOS E RESULTADOS .....</b>	<b>84</b>
5.1	AVALIAÇÃO COMPARATIVA DE RESULTADOS .....	84
5.2	AVALIAÇÃO DOS MÉTODOS DE EXTENSÃO DE BORDA .....	86
5.2.1	Comparação de índices de desempenho .....	92
5.3	ANÁLISE DE FAMÍLIAS DE WAVELETS-MÃES E DESCRITORES .....	95
5.3.1	Análise de Decomposições Wavelet .....	96
5.3.2	Análise de Agrupamentos de Wavelets-Mães .....	107
5.3.3	Análise entre Famílias de Wavelets-Mães .....	113
5.3.4	Análise de Descritores .....	117
5.4	ANÁLISE DA TAXA DE AMOSTRAGEM EM SINAIS DE EEG .....	123
5.5	COMENTÁRIOS .....	129
<b>6</b>	<b>CONCLUSÃO E COMENTÁRIOS FINAIS .....</b>	<b>130</b>
6.1	AVALIAÇÃO DOS MÉTODOS DE EXTENSÃO DE BORDA .....	130
6.2	ANÁLISE DE FAMÍLIAS DE WAVELETS-MÃES E DESCRITORES .....	131
6.3	ANÁLISE DA TAXA DE AMOSTRAGEM EM SINAIS DE EEG .....	133
6.4	CLASSIFICADOR DE EVENTOS EPILEPTIFORMES BASEADO EM LDA .....	133
6.5	LIMITAÇÕES DO ESTUDO .....	134
6.6	COMENTÁRIOS FINAIS .....	135
6.7	TRABALHOS FUTUROS .....	136
	<b>REFERÊNCIAS .....</b>	<b>138</b>

## 1 INTRODUÇÃO

A epilepsia é o nome de uma desordem encefálica caracterizada, predominantemente, por interrupções recorrentes e imprevisíveis do funcionamento normal do encéfalo, denominadas de "crises epiléticas". A epilepsia não é uma doença singular, pois engloba um conjunto amplo de doenças ou desordens neurológicas crônicas que se refletem na disfunção cerebral, resultante de diversas causas (FISHER *et al.*, 2005). Por isso, o termo mais correto para caracterizar estas disfunções seja "epilepsias".

A epilepsia é uma das doenças neurológicas mais comuns em todo o mundo. Segundo dados da Organização Pan-Americana de Saúde, 70 milhões de pessoas sofrem de epilepsia no mundo, das quais 6 milhões só nas Américas. Em 2013, o Ministério da Saúde abriu discussão sobre como melhorar o atendimento a epiléticos, e iniciou-se um mapeamento da situação da epilepsia no Brasil no atendimento e diagnóstico da doença nas redes de atenção básica, secundária e terciária (TAMBOURGI, 2012).

A epilepsia atinge homens e mulheres de todas as faixas etárias e grupos étnicos. A maior incidência das epilepsias ocorre nos países em desenvolvimento. Supõe-se que contribuem para essa maior incidência a maior exposição da população desses países às condições que representam riscos de lesões neurológicas como a neurocisticercose, a malária, a meningite, as complicações pré e perinatal, a desnutrição e os traumatismos envolvendo o crânio (POLIMENO, 2008).

De acordo com a ANVISA - Agência Nacional de Vigilância Sanitária, "o tratamento da epilepsia tem muitas variantes e, provavelmente, o custo é uma das mais sérias limitações". Não é todo o paciente que pode custear os vários tipos de tratamentos, especialmente nos países em desenvolvimento" (CALAIS; LIMA, 2011).

A crise epilética é o resultado de um estado de hiperatividade dos neurônios, capaz de gerar descargas elétricas síncronas anormais que se manifestam de diversas formas.

Fatores ambientais como estímulos visuais, sonoros ou olfativos podem desencadear crises epiléticas, além da causa genética. Álcool, privação do sono e estresse estão também



relacionados às crises epilépticas. Alguns tipos de epilepsias benignas que aparecem na infância desaparecem com o tempo (SHNEKER; FOUNTAIN, 2003).

As crises podem ser focais ou parciais, generalizadas, convulsivas, espasmos, tônicas, ausência e outros termos definidos por BLUME *et al.* (2001), com aspectos específicos para cada tipo de crise. Além disso, distúrbios como a epilepsia contribuem com os índices de mortalidade, onde a probabilidade de suicídio do paciente após os seis primeiros meses do diagnóstico é três vezes maior comparado com portadores de outros tipos de doenças (POLIMENO, 2008).

É notório o quanto a epilepsia afeta a vida de quem é acometido por ela e de sua família. Não há cura e o diagnóstico ainda é falho, motivando a pesquisa e procura de soluções nessa área de conhecimento que venham a melhorar a vida de indivíduos diagnosticados como portadores de epilepsia ou de evitar que pessoas erroneamente diagnosticadas sofram com a carga psicológica e financeira no tratamento de algo do qual não são portadores.

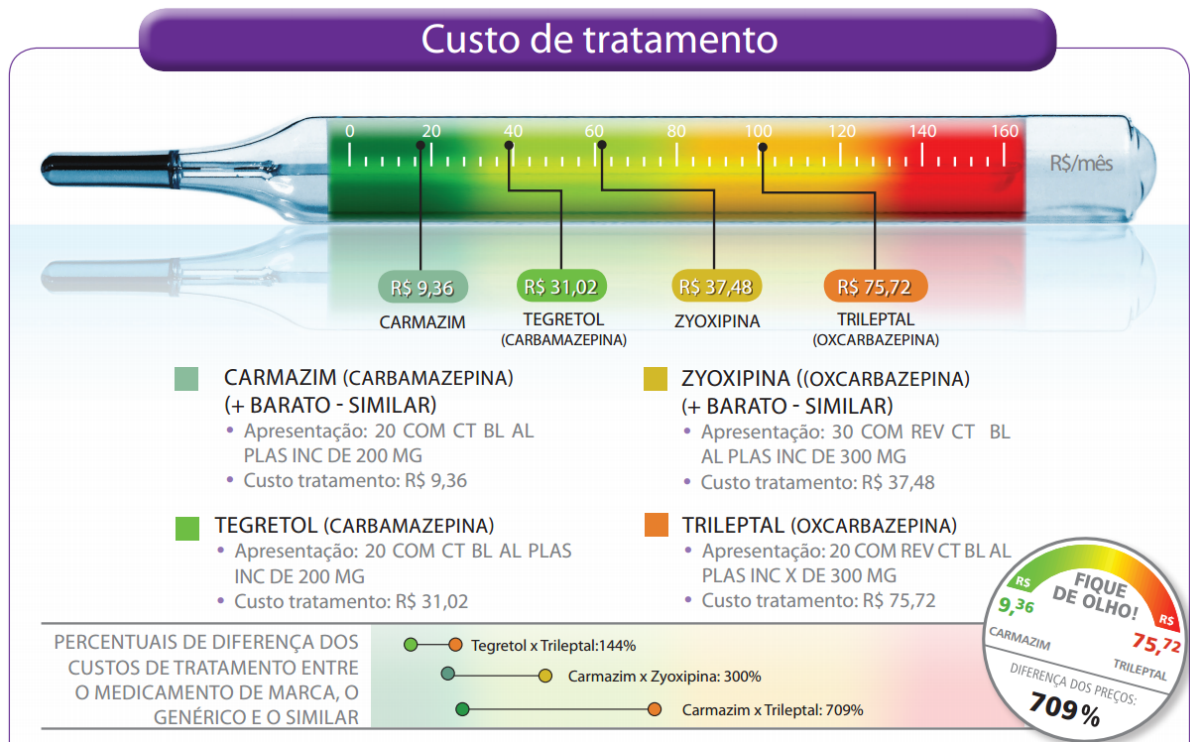
## 1.1 MOTIVAÇÃO

Pela necessidade de parâmetros sobre a epilepsia no Brasil e diretrizes para o diagnóstico, tratamento e acompanhamento dos indivíduos doentes, o Ministério da Saúde publicou a Portaria SAS/MS No 492, de 23 de Setembro de 2010. Essa portaria é um guia que recomenda o uso de carbamazepina em crises focais e fármacos específicos para cada tipo de epilepsia (CALAIS; LIMA, 2011). A figura 1 apresenta o custo para o tratamento da epilepsia com valores do ano de 2011 em material divulgado pela ANVISA.

No serviço ambulatorial brasileiro, a segunda maior causa de atendimento neurológico é a epilepsia. Ainda no Brasil, aproximadamente 20% dos pacientes diagnosticados com epilepsia refratária para possível encaminhamento cirúrgico, **não são portadores da doença** (POLIMENO, 2008).

O procedimento de análise do exame de eletroencefalografia é realizado por um neurologista que identifica (manualmente nos registros dos sinais) os eventos característicos da epilepsia. Este procedimento de análise manual dos registros é cansativo, passível de erro humano por repetição, cansaço mental e visual.

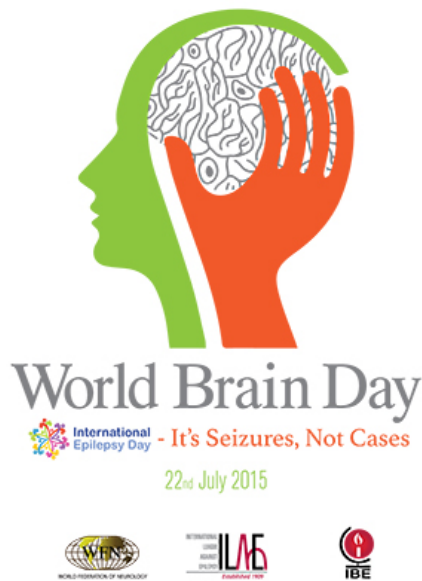
Anualmente, ocorre o dia mundial do cérebro (*World Brain Day*), organizado pela Federação Mundial de Neurologia (*World Federation of Neurology - WFN*), pela Liga Internacional contra a Epilepsia (*International League Against Epilepsy - ILAE*) e o Escritório Internacional para Epilepsia (*International Bureau for Epilepsy - IBE*). No evento de 2015,



**Figura 1:** Custo do tratamento da epilepsia no Brasil em Reais Mensais com valores de 2011.  
**Fonte:** (CALAIS; LIMA, 2011)

a Organização Mundial de Saúde (OMS) aprovou a resolução "Responsabilidade global da epilepsia e da necessidade de ações coordenadas a nível mundial para resolver os problemas na saúde, no social e no ambiente público", atendendo ao apelo de países, incluindo o Brasil. A meta mais importante da campanha é sensibilizar o público quanto ao tema e a formação e treinamento dos clínicos gerais, enfermeiros e paramédicos (WASAY; GRISOLD, 2015).

Ainda sobre sensibilização, o dia 26 de março é tido como o dia do roxo (*purple day*) quando é realizado mundialmente ações de conscientização sobre a epilepsia. O logotipo do dia do roxo e do dia mundial do cérebro podem ser vistos na figura 2 (LEONEL, 2015).



Fonte: WBD2015



Fonte: EpilepsiaBrasil

**Figura 2:** Logotipo do dia internacional do cérebro à esquerda e do dia do roxo à direita.

## 1.2 VISÃO GERAL DO TRABALHO

Este capítulo apresenta o contexto, o objetivo geral e os objetivos específicos do trabalho.

A seguir, no capítulo 2, são explicados os conceitos fundamentais para o entendimento desse trabalho como o exame de eletroencefalografia (seção 2.1) e a espícula (seção 2.2). É explanado o teste-diagnóstico e índices de desempenho na seção 2.3. A transformada wavelet, base desse trabalho, é apresentada na seção 2.5 assim como a análise de discriminante linear (seção 2.8) e conceitos matemáticos nas seções 2.6 e 2.7.

No capítulo 3, são discutidos os trabalhos e estudos existentes na literatura que abordem o tema da detecção de espícula com o uso da transformada wavelet, assim como técnicas de detecção, classificadores, frequência de amostragem do sinal de eletroencefalografia e resultados significativos na área.

No capítulo 4, é apresentada a obtenção dos sinais eletroencefalográficos, o diagrama em blocos do sistema proposto de detecção de espículas (seção 4.2), elemento fundamental na pesquisa, assim como os descritores de extração de características e a geração da matriz de características nas seções 4.3 e 4.4, respectivamente.

No capítulo 5, tem-se os desenvolvimentos, artigos publicados e os resultados obtidos ao longo do trabalho de pesquisa sobre o tema de transformada wavelet e a detecção de

espículas.

E por fim, no capítulo 6, são comentados de modo geral os resultados obtidos no capítulo 5 assim como trabalhos futuros e a conclusão da pesquisa. A seguir os objetivos dessa pesquisa serão apresentados.

### 1.3 OBJETIVOS

O objetivo deste trabalho é investigar o reconhecimento de espículas em sinais de EEG com o uso de transformadas wavelet e delimitar configurações isoladas ou combinadas de descritores e wavelets-mães juntamente com um modelo de classificador de espículas que obtenha índices de desempenho melhores do que os encontrados na literatura.

Objetivos específicos para esse estudo:

#### **1 - Classificador de eventos epileptiformes baseado em LDA**

É proposto um classificador de eventos epileptiformes baseado em transformadas wavelet com redução de dimensionalidade construído sobre a Análise de Discriminante Linear (*Linear Discriminant Analysis - LDA*), usando o critério de Fisher como separador de classes. Esse modelo de classificador não foi encontrado na literatura aplicado a sinais de EEG e se baseia em métodos estatísticos para a separação de classes espícula e não-espícula. Esse item é apresentado em detalhes com seus resultados no item 4.2, 5.1, 5.2, 5.3, e 5.4.

#### **2 - Analisar o Método de Extensão de Borda do Sinal na Transformada Wavelet**

A interpretação de como se comporta o sinal na borda da janela de convolução da transformada wavelet interfere significativamente nos resultados finais. O desenvolvimento sobre esse tópico, seus efeitos e resultados são descritos no item 5.2.

#### **3 - Determinar a melhor wavelet-mãe e o número de decomposições para a detecção de espículas**

A literatura retrata ausência de concordância entre os pesquisadores sobre qual família de wavelets tem melhor desempenho no reconhecimento e/ou caracterização de eventos epileptiformes. Isso se verifica comparando as diversas wavelets utilizadas em estudos na literatura (ERCELEBI; SUBASI, 2006; SADATI; MOHSENI; MAGHSOUDI, 2006; INDIRADEVI *et al.*, 2008; HSU; YU, 2009; MIRZAEI; AYATOLLAHI; VAVADI, 2011; XU; SONG, 2008; RAMANAN; KALPAKAM; SAHAMBI, 2004).

Foram pesquisadas 98 tipos diferentes de wavelets, divididas em oito famílias.

O desenvolvimento do estudo, objetivando determinar famílias mais adequadas para o reconhecimento de espículas, número de decomposições, ensaios e resultados detalhados são descritos no item 5.3.

#### **4 - Analisar descritores de sinal**

Após o processamento do sinal pelas transformadas wavelet, é necessário o pós-processamento que permita a caracterização dos sinais obtidos. Vários descritores são apresentados pela literatura como o valor máximo e mínimo dos valores de um conjunto de dados, assim como a média, mediana, entre outros (GEVA; KEREM, 1998; SADATI; MOHSENI; MAGHSOUDI, 2006; UBEYLI, 2009; WANG *et al.*, 2009; VAVADI; AYATOLLAHI; MIRZAEI, 2010; SONG; LIO, 2011; XU; SONG, 2008; MIRZAEI; AYATOLLAHI; VAVADI, 2011). No item 5.3.4, é verificado o desempenho de um conjunto de 14 descritores na caracterização de espículas.

#### **5 - Analisar a taxa de amostragem de sinais de EEG e sua influência na classificação de espículas**

O exame de EEG é analisado e laudado por neurologista treinado na detecção visual dos eventos epileptiformes. O exame de EEG é obtido a partir do registro eletrográfico dos sinais do encéfalo, feitos através do aparelho de eletroencefalografia. Os exames digitais, feitos por aparelhos comumente encontrados em hospitais, obtém o sinal através da amostragem elétrica do encéfalo a uma frequência de 176,5 a 256 Hz (KALAYCI; OZDAMAR; ERDOL, 1994; KALAYCI; OZDAMAR, 1995; YAMAGUCHI, 2003; SONG; LIO, 2011; ACHARYA *et al.*, 2013; GAJIC *et al.*, 2014; NUNES *et al.*, 2014). Aparelhos de EEG atualmente são capazes de gravar exames com frequência de amostragem superior a 200 Hz, porém, a configuração do aparelho de EEG é feita de modo a facilitar o reconhecimento visual dos eventos epileptiformes pelo neurologista, onde visualmente, não há diferença entre sinais amostrados a 100 Hz ou frequências superiores. A frequência de amostragem está intrinsecamente ligada ao espectro de frequências de um sinal e exames de EEG com amostragem superior a 200 Hz podem apresentar características relevantes para o reconhecimento de eventos epileptiformes. Estudos comparativos e resultados são apresentados na seção 5.4.

### 1.4 COMENTÁRIOS

Foram apresentadas a contextualização do trabalho assim como a sua motivação na área de pesquisa da engenharia biomédica. Na sequência, foram apresentadas a estrutura do trabalho quanto a organização de seu conteúdo e as propostas do trabalho.

O próximo capítulo apresenta a fundamentação teórica necessária para o entendimento do trabalho, como a obtenção do sinal de EEG, a apresentação do espícula, definição de índices de desempenho para análise de classificadores, a transformada wavelet discreta, a análise de discriminante linear, e conceitos como a extensão de borda e o cálculo de autovetores e autovalores.

## 2 FUNDAMENTAÇÃO

Este capítulo aborda a fundamentação teórica para o entendimento do trabalho, apresentando o método de obtenção de sinais e as ferramentas matemáticas aplicadas ao processamento digital de sinais de modo a obter as características de interesse dos eventos epileptiformes. Este capítulo descreve a eletroencefalografia, a espícula, o teste diagnóstico, a transformada wavelet, a correlação de sinais, o cálculo de autovetores e autovalores e a análise de discriminante linear.

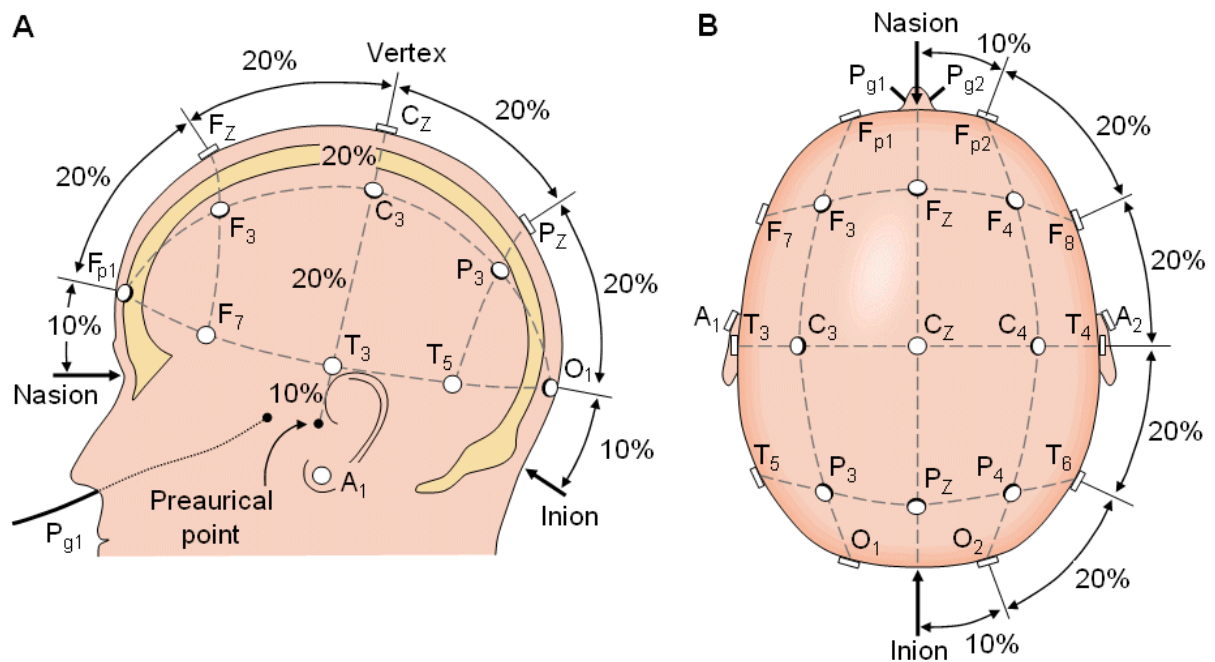
### 2.1 ELETROENCEFALOGRAMA

O eletroencefalograma de superfície (EEG), de acordo com Laget e Salbreux (1967), consiste em detectar e registrar os fenômenos bioelétricos do córtex cerebral, captando esses sinais através de eletrodos posicionados sobre a superfície do couro cabeludo.

Na neurologia, o exame de eletroencefalografia é utilizado como exame auxiliar para uma série de patologias neurológicas, destacando o diagnóstico das epilepsias e os distúrbios do sono. O exame pode ser realizado em ambulatório, com uma duração de 20 a 40 minutos, ou em clínicas especializadas com duração de alguns dias. O exame ambulatorial serve para o acompanhamento médico de rotina e avaliação para prescrição médica. O exame de EEG de longa duração normalmente é realizado em situação pré-operatória para a retirada do tecido encefálico com foco epilético através de procedimento cirúrgico.

#### 2.1.1 POSICIONAMENTO DOS ELETRODOS

O sistema 10-20 é usado internacionalmente como padrão de colocação de eletrodos para a gravação espontânea de EEG. Nesse sistema, são referenciadas 21 posições de colocação de eletrodos (LAGET; SALBREUX, 1967; MALMIVUO; PLONSEY, 1995). A figura 3 apresenta o sistema padronizado internacional 10-20.



**Figura 3:** Sistema Internacional 10-20 de posicionamento de eletrodos para registro de EEG visto pela esquerda (A) e pelo escalpo (B). A=Lóbulo da Orelha, C=Central, Pg= Nasofaringeal, P=Parietal, F=Frontal, Fp=Frontal Polar e O = Ocipital.

Fonte: (MALMIVUO; PLONSEY, 1995)

### 2.1.2 MONTAGEM DOS ELETRODOS

Para a aquisição do sinal de EEG, é possível registrar os impulsos elétricos de modo unipolar/referencial ou bipolar (MALMIVUO; PLONSEY, 1995). O sistema unipolar mede o sinal capturado através de um ponto de referência que não tenha sinal eletroencefálico, sendo por norma o lóbulo da orelha. A montagem bipolar compara a atividade cerebral entre dois pontos, subtraindo os sinais de um ponto com relação a outro ponto, eliminando sinais de origem comum.

Nesse estudo, foi utilizada a visualização parassagital unipolar contendo os sinais obtidos através da seguinte combinação de eletrodos: Fp1-F3, F3-C3, C3-P3, P3-O1, Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F4, F4-C4, C4-P4, P4-O2, Fp2-F8, F8-T4, T4-T6, T6-O2, Fz-Cz e Cz-Pz.

### 2.1.3 RESPOSTA ELÉTRICA DO ELETROENCEFALOGRAMA

Os fenômenos bioelétricos do córtex cerebral e de superfície são captados através dos eletrodos e são registrados e apresentados de forma gráfica, de acordo com a montagem dos eletrodos adotada. A resposta em frequência desse registro é comumente dividida em



sub-bandas, denominadas  $\delta$  (delta),  $\theta$  (theta),  $\alpha$  (alfa),  $\beta$  (beta) e  $\gamma$  (gamma) (HSU; YU, 2009; SHUREN; ZHONG, 2009; ERCELEBI; SUBASI, 2006). Essas sub-bandas representam faixas de frequência que variam conforme a tabela 1.

**Tabela 1:** Sub-bandas de frequências do EEG

Nome da Sub-banda	Faixas de Frequências (Hz)
$\delta$ (delta)	0 - 4
$\theta$ (theta)	4 - 8
$\alpha$ (alfa)	8 - 15
$\beta$ (beta)	15 - 30
$\gamma$ (gamma)	30 - 60

## 2.2 ESPÍCULAS E ONDAS AGUDAS NO ELETROENCEFALOGRAMA

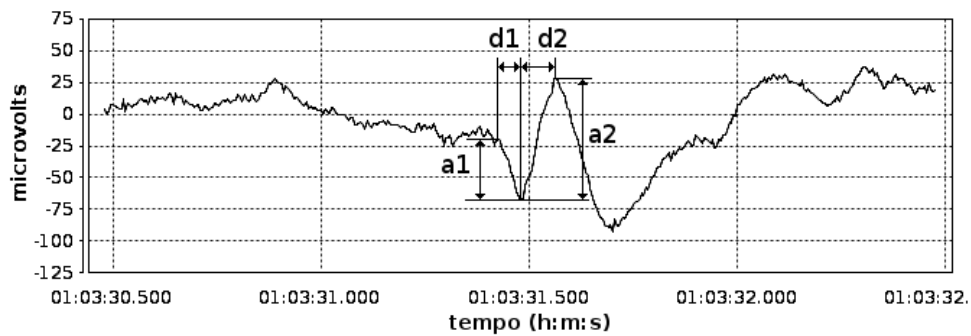
Nesta seção, serão descritas as espículas e ondas agudas. Esses dois grafoelementos patológicos são registrados e reconhecíveis através do exame de EEG. A manifestação e a caracterização das epilepsias é o ataque epiléptico, crônico e recorrente (BLUME *et al.*, 2001). A presença de espículas e ondas agudas no registro eletrográfico de um paciente com manifestação e caracterização de epilepsias define o diagnóstico de epilepsia (LAGET; SALBREUX, 1967).

Durante o registro eletrográfico de um paciente epiléptico, existe a possibilidade de observar sinais bioelétricos característicos, como a espícula e a onda aguda, podendo ainda ser seguido por uma onda lenta de despolarização dos neurônios. Estes sinais podem ocorrer isoladamente ou podem ser repetitivos por um período de tempo de alguns segundos a alguns minutos. A visualização desses sinais no EEG auxilia no diagnóstico da patologia.

Sovierzoski (2009) realizou um estudo sobre descritores morfológicos de eventos epileptiformes. Na figura 4, é possível visualizar duas regiões na realização do evento, marcadas como  $d1$  e  $d2$ , onde  $d1$  inicia na transição rápida da atividade de fundo e termina na ponta do evento, e  $d2$  inicia no término de  $d1$  e tem duração até a próxima ponta. A duração do evento é composto pelos tempos  $d1 + d2$ . As amplitudes  $a1$  e  $a2$  apresentadas na figura 4 não são usadas nesse trabalho.

A espícula (*spike*) é um transiente com pico pontiagudo, claramente distinguível da atividade de fundo do sinal de EEG, com duração ( $d1 + d2$ ) entre 20 e 70 milissegundos.

A onda aguda (*sharp wave*) é um transiente com pico pontiagudo, claramente

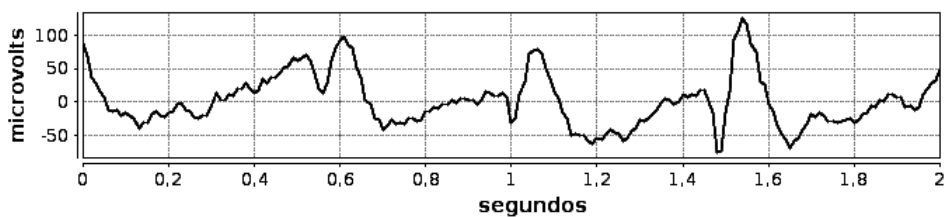


**Figura 4:** Descritor morfológico de evento epileptiforme. A duração do evento é composto pela soma dos tempos  $d1 + d2$ .

distinguível da atividade de fundo do sinal de EEG, com duração ( $d1 + d2$ ) entre 70 e 200 milissegundos.

A onda lenta (*slow wave*) é uma onda com duração superior a 1/8 de segundo, ocorrendo após a espícula ou onda aguda (iniciando-se após  $d2$ ) (CHATRIAN *et al.*, 1974; NOAHTAR *et al.*, 1999).

O complexo espícula e onda lenta (*spike-and-slow-wave complex*), ou complexo ponta e onda, é uma espícula seguida por uma onda lenta (CHATRIAN *et al.*, 1974; NOAHTAR *et al.*, 1999). O complexo ponta e onda pode ser observada na figura 5 com a ponta centrada no tempo de um segundo.

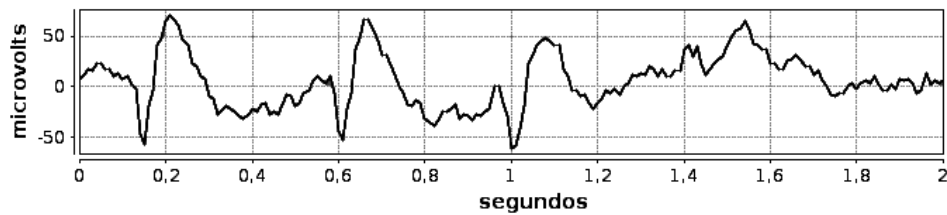


**Figura 5:** Espículas (duração inferior a 70 ms) destacando-se visualmente da atividade de fundo do EEG. A ponta da espícula está centrada na marca de 1 segundo.

O complexo onda aguda e onda lenta (*sharp-and-slow-wave-complex*) é uma onda aguda seguida por uma onda lenta (CHATRIAN *et al.*, 1974; NOAHTAR *et al.*, 1999). O complexo onda aguda e onda lenta pode ser observada na figura 6 com a onda aguda centrada no tempo de um segundo.

### 2.2.1 ARTEFATOS NO ELETROENCEFALOGRAMA

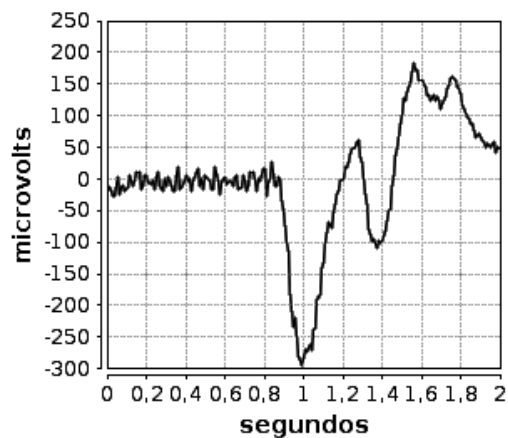
Artefato é definido como qualquer diferença de potencial de origem extra-encefálica, registrada como sinais de EEG (CHATRIAN *et al.*, 1974; NOAHTAR *et al.*, 1999).



**Figura 6:** Ondas agudas (duração inferior a 200 ms) destacando-se visualmente da atividade de fundo do EEG. A ponta da onda aguda está centrada na marca de 1 segundo.

Anghinah *et al.* (2006) apresentam várias fontes de artefatos que aparecem no sinal de EEG. Podem ser classificados como: biológicos, ambientais ou interferências externas e do equipamento. Os artefatos de origem biológica podem ser provocados por atividade muscular, movimentos mecânicos dos eletrodos e interferência do batimento cardíaco. Os artefatos ambientais são provocados por interferências conduzidas pela rede de energia elétrica e outros equipamentos no ambiente do exame e, até mesmo, fora do ambiente de exame, por interferência eletromagnética. Os artefatos do equipamento são gerados pela cadeia de circuitos eletrônicos de condicionamento e aquisição dos sinais e pelos processos de filtragem. Inclui-se também, os artefatos operacionais do exame, como marcapasso do paciente, uso de brincos e correntes, o cabelo com gel ou tinturas, a má colocação dos eletrodos e a impedância elevada de contato.

Um exemplo de artefato, a piscada palpebral, pode ser visto na figura 7, onde durante o exame de EEG o paciente piscou os olhos e o registro eletrográfico desse artefato se diferencia visualmente da atividade de fundo do EEG. O artefato tem início em 0,9 segundo e tem duração superior a 1 segundo.



**Figura 7:** Piscada palpebral destacando-se visualmente da atividade de fundo do EEG.

No reconhecimento automático de espículas é importante que o detector ou classificador utilizado seja imune aos artefatos.

### 2.3 TESTE DIAGNÓSTICO

Na área de exames clínicos, é fundamental uma ferramenta que descreva a partir de um total de indivíduos, a fração de indivíduos doentes e saudáveis. Dentre os vários métodos existentes, o mais utilizado na área de saúde são as medidas de **sensibilidade** (*sensitivity*) e **especificidade** (*specificity*), que de uma maneira dicotômica descrevem o resultado de um exame como positivo e negativo, respectivamente (ERKEL; PATTYNAMA, 1998).

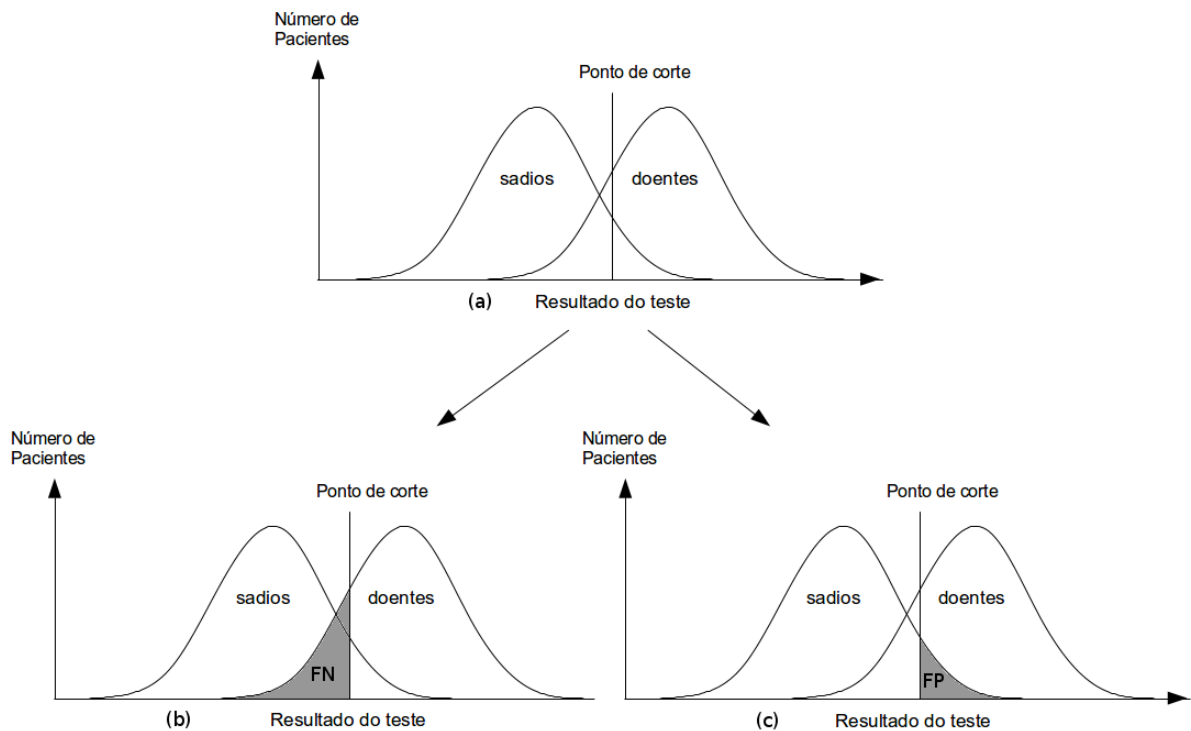
Considerando uma distribuição de indivíduos saudáveis e uma distribuição de indivíduos doentes, um exame clínico define um limiar de decisão, um ponto de corte, classificando os indivíduos dessas duas distribuições como saudáveis ou doentes. Em um caso onde esse limiar separe a totalidade dos doentes da totalidade dos saudáveis, o resultado da sensibilidade e da especificidade serão ambos de 100%. Dependendo dessas distribuições de indivíduos, as classes de indivíduos doentes e saudáveis podem se sobrepor. No caso de sobreposição das classes de pacientes, qualquer limiar de decisão conduzirá a uma classificação errônea de pacientes saudáveis como doentes e de pacientes doentes como saudáveis (ERKEL; PATTYNAMA, 1998).

Um paciente corretamente diagnosticado como doente é um verdadeiro positivo (VP) enquanto que um paciente doente classificado como saudável é um falso negativo (FN). Um paciente saudável corretamente diagnosticado como saudável é um verdadeiro negativo (VN) enquanto que um paciente saudável classificado como doente é um falso positivo (FP). A figura 8 apresenta essas situações, onde em 8(a) existem duas distribuições de pacientes, os doentes à direita e os saudáveis à esquerda. O ponto de corte definido pelo resultado do teste separa as duas classes de indivíduos. São classificados como positivos (doentes), os indivíduos localizados à direita do ponto de corte; como negativos (saudáveis), os indivíduos localizados à esquerda do ponto de corte. Porém, como visto na figura 8(b) e 8(c), devido a esse ponto de corte, existem indivíduos doentes classificados como saudáveis, assim como existem indivíduos saudáveis classificados como doentes.

Dessa maneira, a sensibilidade é definida como a razão entre o número de indivíduos corretamente diagnosticados como doentes, pelo número total de indivíduos doentes, conforme descrito pela equação 1.

$$\text{sensibilidade} = \frac{\text{VP}}{\text{FN} + \text{VP}} \quad (1)$$

Da mesma forma, a especificidade, é definida como a razão entre o número de indivíduos corretamente diagnosticados como saudáveis, pelo número total de indivíduos saudáveis,



**Figura 8:** Doentes classificados como doentes são verdadeiros positivos (VP), sadios classificados como sadios são verdadeiros negativos (VN). (a) classes de indivíduos doentes e sadios separados por um ponto de corte através de um exame clínico. (b) doentes classificados como sadios são denominados como falsos negativos (FN) (área cinza). (c) sadios classificados como doentes são denominados como falsos positivos (FP) (área cinza).

**Fonte:** modificado de (ERKEL; PATTYNAMA, 1998)

conforme descrito pela equação 2.

$$\text{especificidade} = \frac{\text{VN}}{\text{VN} + \text{FP}} \quad (2)$$

Sendo a sensibilidade e a especificidade índices de desempenho variando de zero a um, alguns autores preferem adotar o valor percentual, sendo a sensibilidade tratada como a taxa de verdadeiros positivos ou *true positive rate* (TPR) e a especificidade tratada como a taxa de verdadeiros negativos ou *true negative rate* (TNR). A TPR e a TNR são apresentadas pelas equações 3 e 4.

$$\text{TPR} = \text{sensibilidade} \cdot 100\% = \left( \frac{\text{VP}}{\text{FN} + \text{VP}} \right) \cdot 100\% \quad (3)$$

$$\text{TNR} = \text{especificidade} \cdot 100\% = \left( \frac{\text{VN}}{\text{VN} + \text{FP}} \right) \cdot 100\% \quad (4)$$

Outro índice de desempenho utilizado na literatura é a acurácia (*Accuracy* - *Acc*). Muitos autores utilizam esse termo para mensurar a razão entre as amostras classificadas corretamente pelo total de amostras (XU; SONG, 2008; OCAK, 2009; WANG *et al.*, 2009; HSU; YU, 2009). Isso se traduz no percentual de acertos do classificador, independentemente da classe em questão ser composta por amostras de doentes ou sadios. A acurácia é apresentada na equação 5.

$$Acc = \frac{VP + VN}{FN + VP + VN + FP} \quad (5)$$

A taxa de falsos positivos (*False Positive Rate* - *FPR*), apresentada pela equação 6, representa a taxa de indivíduos sadios erroneamente classificados como doentes pelo número total de indivíduos sadios.

$$FPR = (1 - \text{especificidade}) = \left[ 1 - \frac{VN}{VN + FP} = \frac{FP}{VN + FP} \right] \cdot 100\% \quad (6)$$

Exemplificando os conceitos de sensibilidade, especificidade e acurácia, consideremos a seguinte situação hipotética: em uma cidade com 1 milhão de habitantes, uma doença afeta a população na ordem de 1/100, ou seja, para cada 100 habitantes, um estará doente. Projetando para a população total da cidade, tem-se provavelmente dez mil pessoas infectadas.

Um indivíduo irá realizar um exame para essa doença, sendo que o exame possui sensibilidade de 99% e especificidade de 98%. Pelo ponto de vista da sensibilidade de 99%, o exame possivelmente identificará 9.900 das dez mil pessoas com a doença, e 100 pessoas serão diagnosticadas erroneamente como saudáveis. Por outro lado, um exame com especificidade de 98% significa que duas pessoas a cada 100 indivíduos saudáveis serão informados incorretamente que estão doentes. Considerando a população de 1 milhão de habitantes, 20 mil pessoas serão diagnosticadas erroneamente como uma doença que não possuem.

A acurácia, por sua vez, indica o percentual de acerto do exame. Nesta situação hipotética, a acurácia do exame para essa doença que afeta a cidade é de 97,99%. Isso quer dizer que o exame irá diagnosticar erroneamente 20.100 pessoas dessa cidade.

O limiar de decisão, ou ponto de corte, para definir um exame como positivo ou negativo é ajustável. O limiar que aumente a sensibilidade de exames pode ser interessante para doenças graves e o seguimento de um teste de triagem é relativamente benigno (talvez apenas a repetição do primeiro teste, ou outro que talvez seja ligeiramente mais demorado

ou custe mais caro, mas não envolva risco físico algum para o paciente). Por outro lado, aumentando a especificidade do exame, deixando assim alguns casos de fora, caso a doença não seja imediatamente perigosa ou se a continuação do teste positivo for uma biópsia cirúrgica ou outro procedimento invasivo que seja desejável evitar, se possível.

## 2.4 CURVA ROC

A curva de características de operação do receptor (*Receiver Operating Characteristics* - ROC) é um gráfico que auxilia na visualização, organização e seleção de classificadores com duas classes de eventos. A comunidade médica a usa para a interpretação da variação da sensibilidade e especificidade de testes-diagnósticos conforme o limiar de detecção usado para o exame (FAWCETT, 2006).

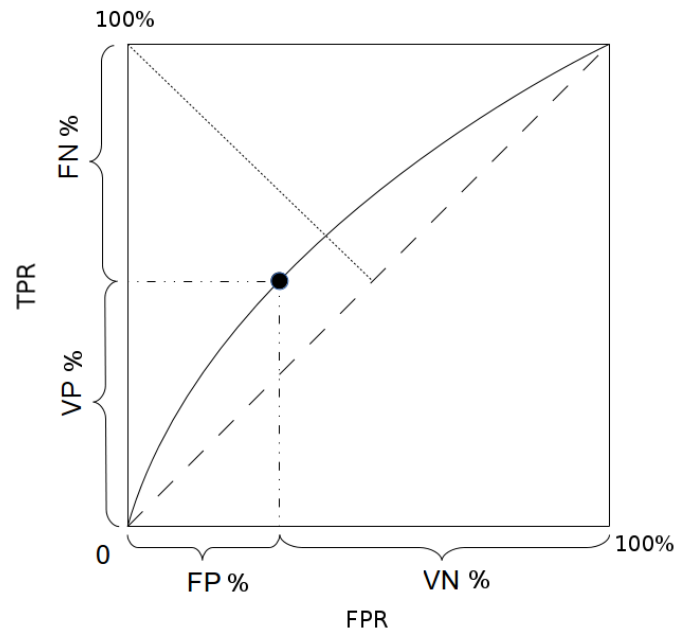
Para cada valor escolhido para o limiar de detecção ou ponto de corte (figura 8), é calculado um respectivo par ordenado com valores de sensibilidade e especificidade. Cada ponto do gráfico é o resultado de um limiar de detecção utilizado em um classificador binário. O conjunto de pontos forma uma curva em que é possível averiguar o desempenho do classificador utilizado. O eixo das ordenadas representa a TPR. O eixo das abscissas representa o FPR. Como se tratam de porcentagens, os valores dos eixos variam de zero a 100% como pode ser visto na figura 9.

Na figura 9, a diagonal tracejada representa uma classificação aleatória de classes. Se um classificador marca seus testes metade do tempo como positivo, é esperado que as outras amostras sejam classificadas como negativas, gerando um par ordenado (0,5; 0,5). Já se o classificador marca suas classes 90% como positivas, a fração de falso positivo também será de 90%, gerando um par ordenado (0,9; 0,9). Ou seja, a linha tracejada representa um classificador que possui um índice de acerto inferior a 0,5 para qualquer valor de nível de decisão (FAWCETT, 2006).

### 2.4.1 ÍNDICE DE DESEMPENHO DA CURVA ROC - AUC

A área sob a curva ROC (*Area under the ROC Curve* - AUC) é um índice de desempenho que analisa a curva ROC, determinando o quanto um classificador é melhor do que outro em uma determinada situação de pontos de corte ou níveis de decisão.

Quanto maior a taxa de verdadeiro positivo e menor a taxa de falso positivo, melhor um classificador consegue separar os sadios dos doentes. Graficamente, na curva ROC, os pontos de tal classificador estarão mais próximos do canto superior esquerdo.



**Figura 9:** Curva ROC: conjunto de pontos representando resultados de classificação variando o limiar de detecção. A sensibilidade ou TPR é representada pelo eixo das ordenadas e a fração de falsos-positivos ( $1 - \text{especificidade}$ ) ou FPR é representada pelo eixo das abscissas. A linha tracejada é o resultado de um classificador que não funciona. A linha pontilhada mede a distância do canto superior esquerdo até a curva e quanto menor, melhor o desempenho do classificador. Cada ponto da curva é formado por um par ordenado representado por (TPR; FPR).

**Fonte: Modificado de Fawcett (2006)**

No caso de um classificador ideal, o índice de desempenho AUC é igual a um (FAWCETT, 2006), sendo que  $FN=0$  e  $FP=0$ , ocorrendo a separabilidade total entre as duas classes de eventos.

## 2.5 WAVELETS

No trabalho desenvolvido por Daubechies (1990), mostrou-se que as wavelets possuem propriedades que as tornam adequadas a servirem de base para a decomposição de outras funções, assim como senos e cossenos servem de base para a decomposição de funções usando a transformada de Fourier. A diferença em relação às análises de Fourier e wavelet é que enquanto a análise de Fourier contempla a decomposição de uma função segundo às frequências que compõem essa função, a análise wavelet decompõe a função recursivamente, diminuindo a cada iteração a escala de frequências em análise, permitindo a compreensão da função não somente quanto às frequências que a compõem, mas também a compreensão da localização



temporal dessas frequências ao longo do tempo de duração dessa função. Essa característica de escala variável da análise wavelet a torna uma ferramenta interessante por identificar mais eficientemente as descontinuidades e variações bruscas de um sinal ao longo do tempo, o que não é possível realizar através da análise de Fourier.

Pelas suas propriedades únicas, as wavelets podem ser utilizadas para extrair informações de diferentes tipos de sinais, sendo utilizadas em diversas áreas: estudo de propriedades (multi)fractais, singularidades ou oscilações locais de funções, solução de equações diferenciais, reconhecimento de padrões, compressão de imagens e de sons, processamento de geometria digital, solução de vários problemas de física, biologia, medicina, astronomia, acústica, engenharia nuclear, grupo de renormalização em mecânica estatística, problemas de computação gráfica, neurofisiologia, música, ressonância magnética, identificação de voz, ótica, fractais, turbulência, previsão de terremotos, radar, visão humana, entre outros (OLIVEIRA, 2007).

A Transformada Wavelet (*Wavelet Transform* - WT) usa o conceito de janelas deslizantes sobre intervalos de sinais de diferentes durações. Intervalos de pequena duração são utilizados para obter informações de alta frequência enquanto que intervalos de longa duração são utilizados para obter informação de baixa frequência, permitindo assim compatibilizar uma resolução com o tipo de frequência estudada (OLIVEIRA, 2007).

A implementação da transformada wavelet pode ser feita de duas maneiras: a primeira, de forma analítica, é conhecida como transformada wavelet contínua (*Continuous Wavelet Transform* - CWT), a segunda, necessária para aplicações práticas utilizando computadores digitais, é conhecida como transformada wavelet discreta (*Discrete Wavelet Transform* - DWT). As transformadas contínua e discreta são apresentadas a seguir, com foco em sua implementação discreta.

### 2.5.1 TRANSFORMADA WAVELET CONTÍNUA

Mallat (2009) utilizou a expressão "o tempo é apresentado à frequência" para introduzir o conceito da transformada wavelet, onde uma função definida no domínio do tempo  $f(t)$  é transformada em outra função, definida no domínio do tempo e da frequência. Isso ocorre através da decomposição da função temporal em diferentes escalas de tempo e frequência com a utilização de uma função wavelet-mãe  $\psi(t)$ . A função  $\psi(t)$  deve possuir primariamente duas características, apresentadas pelas equações 7 e 8:

- Possuir média zero.

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (7)$$

- Possuir energia finita.

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \quad (8)$$

A função wavelet-mãe  $\psi(t)$  recebe ainda mais dois parâmetros,  $a$  e  $b$ , onde o parâmetro  $b$  permite que a função wavelet-mãe seja transladada no tempo, enquanto que o parâmetro  $a$  permite a dilatação da função  $\psi(t)$  em diferentes escalas, alterando o tamanho da janela de operação. Dessa maneira, a função wavelet-mãe  $\psi(t)$  torna-se  $\psi_{a,b}(t)$ . Normalizando a função  $\psi_{a,b}(t)$ , ou seja,  $\|\psi\| = 1$ , tem-se a definição da função wavelet-mãe apresentada pela equação 9:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad (9)$$

Dessa maneira, a transformada wavelet contínua da função  $f(t)$  com a wavelet-mãe  $\psi_{a,b}(t)$  pode ser escrita como apresentada pela equação 10:

$$Wf(a,b) = \langle f(t), \psi_{a,b}(t) \rangle = \int_{-\infty}^{\infty} f(t) \psi_{a,b}(t) dt \quad (10)$$

Sabendo-se que o parâmetro  $b$  de  $\psi_{a,b}(t)$  permite a translação da função  $\psi(t)$  sobre  $f(t)$  durante a transformação wavelet, a transformada wavelet apresentada pela equação 10 pode ser reescrita como uma convolução das duas funções  $\psi_{a,b}(t)$  e  $f(t)$ . Essa operação é apresentada pela equação 11:

$$Wf(a,b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) dt = f \star \bar{\psi}_a(b) \quad (11)$$

onde

$$\bar{\psi}_a(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{-t}{a}\right) \quad (12)$$

Como demonstrado por Mallat (2009), a transformada de Fourier de  $\bar{\psi}_a(t)$  resulta na equação 13.

$$\hat{\bar{\psi}}_a(t) = \sqrt{|a|} \hat{\psi}(a\omega) \quad (13)$$

Onde  $\hat{\psi}(0)$  resulta na equação 7, demonstrando que  $\hat{\psi}$  é na verdade uma função de transferência de um filtro passa-faixa. Logo, a convolução demonstrada pela equação 11 calcula a transformada wavelet  $Wf(a,b)$  convoluindo  $f(t)$  com um filtro passa-faixa indicado por  $\psi(t)$  dilatado por  $a$ .

A transformada wavelet contínua é de grande interesse teórico, principalmente para a obtenção e a compreensão das propriedades matemáticas das funções wavelets, porém, a sua discretização é necessária para aplicações práticas utilizando computadores digitais. A discretização da CWT é apresentada a seguir.

### 2.5.2 TRANSFORMADA WAVELET DISCRETA

A transformada wavelet discreta propõe uma forma alternativa de análise em relação à transformada wavelet contínua. A CWT caracteriza-se por realizar o cálculo da correlação entre o sinal e a função wavelet-mãe utilizando o parâmetro  $a$  como similaridade entre o sinal e a função wavelet-mãe. Ou seja, a CWT é basicamente calculada alterando-se a escala da janela de suporte da função wavelet e convoluindo-a com o sinal em análise.

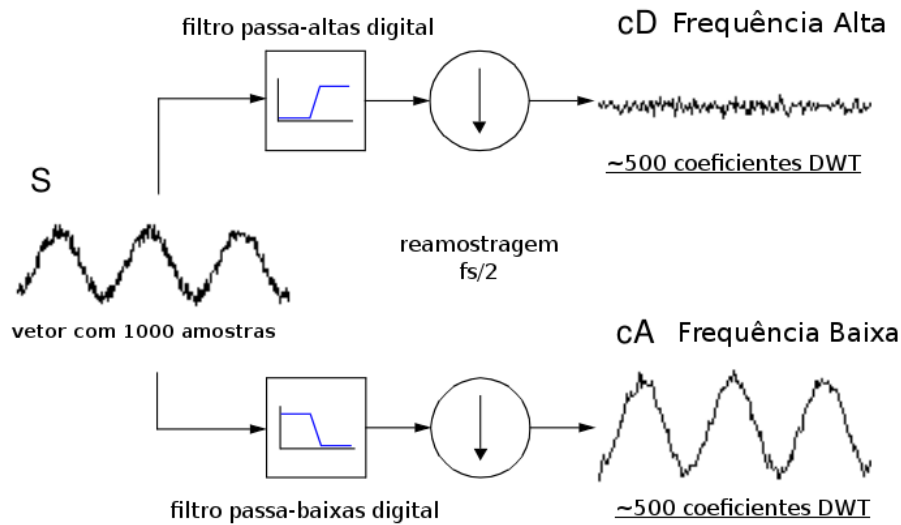
A DWT por sua vez é um cálculo recursivo, tomando como entrada um vetor  $S$  contendo  $n$  amostras obtido de um sinal  $f(t)$  amostrado a uma frequência  $f_s$ . Cada iteração origina dois vetores de tamanho  $n/2$ , denominados coeficientes de aproximação (cA) e coeficientes de detalhe (cD). Esses coeficientes são gerados processando as amostras do sinal  $f(t)$  através de filtros digitais passa-baixas e passa-altas e reamostrando a saída desses filtros pela metade de  $f_s$  (MALLAT, 1989).

Os filtros digitais utilizados pela DWT são caracterizados pelas propriedades da função wavelet utilizada, onde a função wavelet-mãe  $\psi(t)$  está relacionada ao filtro passa-altas, enquanto que a função escala  $\phi(t)$ , relaciona-se com o filtro passa-baixas.

Dessa forma, a análise da informação espectral do sinal é efetuada através de operações de filtragem, enquanto que a análise do fator escala (parâmetro  $a$  de  $\psi_{a,b}(t)$ ), é efetuada através de operações de subamostragem, reduzindo a frequência amostral pela metade a cada iteração.

A figura 10 apresenta uma iteração da transformada wavelet discreta. A partir de um vetor inicial contendo 1000 amostras, são aplicados filtros passa-altas e passa-baixas correspondentes à wavelet-mãe desejada. Após a filtragem da informação de entrada, toma-se uma em cada duas amostras, sendo a informação restante o resultado da iteração da DWT.

Os coeficientes de aproximação são o resultado da aplicação do filtro passa-baixas para essa iteração, e possuem um espectro de frequência de zero a  $f_s/4$ . Os coeficientes de



**Figura 10:** Decomposição wavelet. O sinal  $S$  é processado através de filtros digitais passa-altas e passa-baixas. O resultado de cada filtro é reamostrado gerando os coeficientes de alta frequência cD e de baixa frequência cA.

**Fonte:** Modificado de (MISITI *et al.*, 1996)

detalhe são o resultado da aplicação do filtro passa-altas para essa iteração, e possuem espectro de frequência de  $f_s/4$  a  $f_s/2$ .

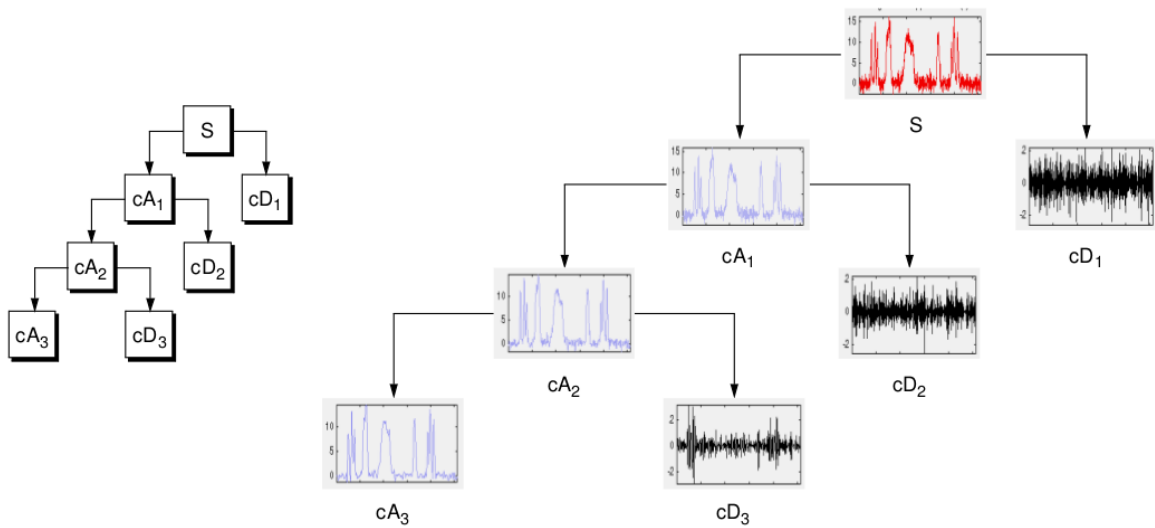
A iteração apresentada pela figura 10 pode ser repetida  $i$  vezes, onde a informação de entrada da iteração é composta pelos coeficientes de aproximação resultantes da iteração anterior  $cA_{i-1}$ , gerando os coeficientes  $cA_i$  e  $cD_i$ . Essa decomposição recursiva é apresentada pela figura 11 e é conhecida como análise multirresolução.

Conhecendo como é realizado o cálculo da DWT, é necessário conhecer os filtros que compõem as wavelets-mães. O cálculo dos filtros  $\psi(t)$  e  $\phi(t)$  é um processo longo e dispendioso computacionalmente, além de gerar uma grande quantidade de dados (MISITI *et al.*, 1996).

As wavelets-mães utilizadas nesse estudo e seus respectivos coeficientes são apresentados a seguir.

### 2.5.3 FAMÍLIAS DE WAVELETS-MÃES

O termo wavelet-mãe refere-se à função a qual gera uma função wavelet específica, representada pela função  $\psi(t)$  apresentada na equação 9. Wavelets-mães baseadas dentro de uma mesma teoria são definidas como família de wavelets-mães. Dependendo das características da aplicação, pode-se utilizar uma função wavelet-mãe mais adequada que represente melhor a decomposição do sinal.



**Figura 11:** Decomposição wavelet em múltiplas resoluções. A decomposição apresentada na figura 10 é repetida recursivamente com o coeficientes cA resultantes da iteração anterior.

**Fonte:** (MISITI *et al.*, 1996)

Como descrito por Mallat (2009), a construção de uma função wavelet  $\psi$  e sua respectiva família é possível desde que a função wavelet seja pertencente a uma base ortonormal do espaço de Hilbert  $L^2(\mathbb{R})$ . O espaço de Hilbert  $L^2(\mathbb{R})$  define que existe o produto interno para qualquer par de vetores desse espaço. E uma base ortonormal define que para uma base de vetores  $V = \{V_1, \dots, V_k\}$ , os vetores devem ser ortogonais entre si ( $V_i \cdot V_j = 0$ ) para  $i \neq j$  e ter norma um ( $V_i \cdot V_j = 1$ ) para  $i = j$ .

Assim, uma família de funções wavelet  $\psi$  pode ser construída a partir da equação 14, modificada da equação 9:

$$\left\{ \psi_{a,b}(t) = \frac{1}{\sqrt{2^j}} \psi \left( \frac{t - 2^j n}{2^j} \right) \right\}_{(j,n) \in \mathbb{Z}^2} \quad (14)$$

Mallat (2009) ainda comenta que a equação 14 permite que wavelets sejam dilatadas na relação de  $2^j$ , e que a geração de seus coeficientes pode ser conseguida através de operações de aproximação na dilatação e translação da função escala  $\phi$ .

A seguir são apresentadas as famílias de wavelets-mães utilizadas nesse estudo assim como seus coeficientes.

### Wavelet-Mãe Haar

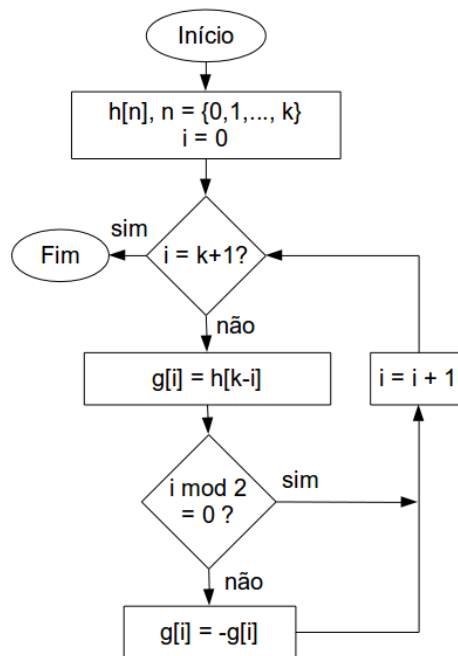
É a wavelet-mãe mais simples. É descontínua e remete a um degrau. Ainda é utilizada outra forma da wavelet Haar, normalizada por  $\sqrt{2}$ . Seus coeficientes da função escala são

definidos pela tabela 2.

**Tabela 2:** Coeficientes da wavelet-mãe Haar

Coeficiente	Haar Ortogonal	Haar
$h[0]$	1	$1/\sqrt{2}$
$h[1]$	1	$1/\sqrt{2}$

A obtenção dos coeficientes da wavelet-mãe são obtidos através dos coeficientes da função escala através do algoritmo apresentado pela figura 12.



**Figura 12:** Algoritmo de geração dos coeficientes wavelet a partir dos coeficientes escala.

**Fonte:** Modificado de (MISITI *et al.*, 1996)

### **Família Daubechies**

Ingrid Daubechies é uma das mais renomadas cientistas a estudar e definir a transformada wavelet. Ela desenvolveu as wavelets ortonormais de suporte compacto, fazendo com que a transformada wavelet discreta fosse possível. Estas funções são denominadas DBN, onde N é a série ou ordem da wavelet-mãe. DB1 é a wavelet-mãe Daubechies 1, que é a mesma wavelet que a Haar. Esse estudo avalia as wavelets-mães de DB1 até a DB38. A tabela 3 apresenta os coeficientes da função escala para as 19 primeiras séries desta família. A resolução dos coeficientes apresentados na tabela é de quatro casas decimais. A geração dos coeficientes wavelet é feita a partir dos coeficientes escala utilizando o algoritmo apresentado na figura 12.

### **Família Biortogonal**

A família de funções wavelet-mãe biortogonal utiliza duas wavelets, uma para decomposição e outra para reconstrução. Nesse estudo, como é trabalhada somente a decomposição do sinal, são apresentados os coeficientes de decomposição escala das 15 wavelets-mães que compõem essa família na tabela 4. Diferentemente de outras famílias, a geração dos coeficientes wavelet também é feita por aproximação e seus coeficientes são apresentados da mesma maneira pela tabela 4.

### **Família Reverse Biortogonal**

A família de funções wavelet-mãe biortogonal reversa é baseada na família biortogonal. A geração dos coeficientes utiliza o algoritmo apresentado na figura 12, onde os coeficientes de escala da família biortogonal reversa utilizam os coeficientes wavelet da família biortogonal como entrada. O mesmo procedimento é feito para os coeficientes escala da família biortogonal reversa, utilizando como entrada os coeficientes wavelet da família biortogonal.

**Tabela 3:** Coeficientes da função escala para as wavelets-mães da família Daubechies de DB1 a DB19.

Coef.	DB1	DB2	DB3	DB4	DB5	DB6	DB7	DB8	DB9	DB10	DB11	DB12	DB13	DB14	DB15	DB16	DB17	DB18	DB19
h[0]	7,071E-1	4,830E-1	3,327E-1	2,304E-1	1,601E-1	1,115E-1	7,785E-2	5,442E-2	3,808E-2	2,667E-2	1,869E-2	1,311E-2	9,202E-3	6,461E-3	4,539E-3	3,189E-3	2,242E-3	1,576E-3	1,109E-3
h[1]	7,071E-1	8,365E-1	8,069E-1	7,148E-1	6,038E-1	4,946E-1	3,965E-1	3,129E-1	2,438E-1	1,882E-1	1,441E-1	1,096E-1	8,286E-2	6,236E-2	4,674E-2	3,491E-2	2,599E-2	1,929E-2	1,428E-2
h[2]		2,241E-1	4,599E-1	6,309E-1	7,243E-1	7,511E-1	7,291E-1	6,756E-1	6,048E-1	5,272E-1	4,499E-1	3,774E-1	3,120E-1	2,549E-1	2,060E-1	1,651E-1	1,312E-1	1,036E-1	8,128E-2
h[3]		-1,294E-1	-1,350E-1	-2,798E-2	1,384E-1	3,153E-1	4,698E-1	5,854E-1	6,573E-1	6,885E-1	6,857E-1	6,572E-1	6,111E-1	5,543E-1	4,926E-1	4,303E-1	3,704E-1	3,147E-1	2,644E-1
h[4]			-8,544E-2	-1,870E-1	-2,423E-1	-2,263E-1	-1,439E-1	-1,583E-2	1,332E-1	2,812E-1	4,120E-1	5,159E-1	5,889E-1	6,312E-1	6,458E-1	6,374E-1	6,110E-1	5,718E-1	5,244E-1
h[5]			3,523E-2	3,084E-2	-3,224E-2	-1,298E-1	-2,240E-1	-2,840E-1	-2,933E-1	-2,498E-1	-1,623E-1	-4,476E-2	8,699E-2	2,187E-1	3,390E-1	4,403E-1	5,183E-1	5,718E-1	6,017E-1
h[6]				3,288E-2	7,757E-2	9,750E-2	7,131E-2	4,725E-4	-9,684E-2	-1,959E-1	-2,742E-1	-3,162E-1	-3,150E-1	-2,717E-1	-1,932E-1	-8,975E-2	2,731E-2	1,472E-1	2,609E-1
h[7]				-1,060E-2	-6,241E-3	2,752E-2	8,061E-2	1,287E-1	1,485E-1	1,274E-1	6,604E-2	-2,378E-2	-1,246E-1	-2,180E-1	-2,889E-1	-3,271E-1	-3,283E-1	-2,937E-1	-2,281E-1
h[8]					-1,258E-2	-3,158E-2	-3,803E-2	-1,737E-2	3,073E-2	9,306E-2	1,498E-1	1,825E-1	1,795E-1	1,384E-1	6,528E-2	-2,792E-2	-1,266E-1	-2,165E-1	-2,858E-1
h[9]					3,336E-3	5,538E-4	-1,657E-2	-4,409E-2	-6,763E-2	-7,139E-2	-4,648E-2	5,360E-3	7,295E-2	1,400E-1	1,901E-1	2,112E-1	1,973E-1	1,495E-1	7,465E-2
h[10]						4,777E-3	1,255E-2	1,398E-2	2,509E-4	-2,946E-2	-6,644E-2	-9,643E-2	-1,058E-1	-8,675E-2	-3,967E-2	2,734E-2	1,011E-1	1,671E-1	2,123E-1
h[11]							-1,077E-3	4,296E-4	8,746E-3	2,236E-2	3,321E-2	3,134E-2	1,085E-2	-2,649E-2	-7,155E-2	-1,111E-1	-1,324E-1	-1,268E-1	-9,233E-2
h[12]								-1,802E-3	-4,870E-3	-4,723E-3	3,607E-3	2,084E-2	4,155E-2	5,614E-2	5,524E-2	3,388E-2	-6,240E-3	-5,709E-2	-1,068E-1
h[13]								3,537E-4	-3,917E-4	-4,282E-3	-1,073E-2	-1,536E-2	-1,222E-2	2,380E-3	2,698E-2	5,478E-2	7,592E-2	6,489E-2	2,758E-1
h[14]									6,754E-4	1,848E-3	1,395E-3	-3,341E-3	-1,284E-2	-2,383E-2	-3,019E-2	-2,577E-2	-7,589E-3	2,231E-2	5,705E-2
h[15]									-1,175E-4	2,304E-4	1,992E-3	4,928E-3	6,711E-3	3,924E-3	-5,615E-3	-2,081E-2	-3,689E-2	-4,692E-2	-4,453E-2
h[16]										-2,520E-4	-6,859E-4	-3,086E-4	2,249E-3	7,256E-3	1,279E-2	1,508E-2	1,030E-2	-3,271E-3	-2,373E-2
h[17]										-1,165E-4	-8,930E-4	-2,180E-3	-2,762E-3	-7,462E-4	5,101E-3	1,399E-2	2,273E-2	2,667E-2	2,162E-2
h[18]										9,359E-5	2,492E-4	6,545E-6	-1,316E-3	-3,850E-3	-6,488E-3	-6,990E-3	-3,043E-3	6,262E-3	1,938E-2
h[19]										-1,326E-5	5,444E-5	3,887E-4	9,323E-4	1,062E-3	-2,418E-4	-3,644E-3	-8,603E-3	-1,305E-2	-1,399E-2
h[20]											-3,463E-5	-8,850E-5	4,925E-5	7,080E-4	1,943E-3	3,128E-3	2,968E-3	1,186E-4	-5,867E-3
h[21]											4,494E-6	-2,424E-5	-1,651E-4	-3,868E-4	-3,735E-4	4,079E-4	2,301E-3	4,943E-3	7,041E-3
h[22]												1,278E-5	3,068E-5	-4,178E-5	-3,596E-4	-9,410E-4	-1,437E-3	-1,119E-3	7,690E-4
h[23]												-1,529E-6	1,044E-5	6,876E-5	1,559E-4	1,142E-4	-3,281E-4	-1,341E-3	-2,688E-3
h[24]													-4,700E-6	-1,034E-5	2,579E-5	1,748E-4	4,395E-4	6,285E-4	3,418E-4
h[25]													5,220E-7	-4,390E-6	-2,813E-5	-6,104E-5	-2,561E-5	2,136E-4	7,358E-4
h[26]														1,725E-6	3,363E-6	-1,395E-5	-8,205E-5	-1,986E-4	-2,607E-4
h[27]														-1,787E-7	1,811E-6	1,134E-5	2,319E-5	-1,536E-7	-1,246E-4
h[28]																-6,317E-7	-1,044E-6	6,991E-6	3,741E-5
h[29]																6,133E-8	-7,364E-7	-4,506E-6	-8,521E-6
h[30]																	2,309E-7	3,017E-7	-3,333E-6
h[31]																	-2,109E-8	2,958E-7	1,769E-6
h[32]																		-8,424E-8	-7,692E-8
h[33]																		7,267E-9	-1,176E-7
h[34]																			3,069E-8
h[35]																			-2,508E-9
h[36]																			4,637E-8
h[37]																			-1,116E-8
																			8,667E-10



**Tabela 4:** Coeficientes da função escala e função wavelet para as wavelets-mães de decomposição da família Biortogonal de Bior11 a Bior68.

Coef.	Bior11	Bior13	Bior15	Bior22	Bior24	Bior26	Bior28	Bior31	Bior33	Bior35	Bior37	Bior39	Bior44	Bior55	Bior68
Coeficientes da função escala															
h[0]	7,071E-1	-8,839E-2	1,657E-2	-1,768E-1	3,315E-2	-6,905E-3	1,511E-3	-3,536E-1	6,629E-2	-1,381E-2	3,021E-3	-6,797E-4	3,783E-2	0,000E+0	1,909E-3
h[1]	7,071E-1	8,839E-2	-1,657E-2	3,536E-1	-6,629E-2	1,381E-2	-3,021E-3	1,061E+0	-1,989E-1	4,143E-2	-9,063E-3	2,039E-3	-2,385E-2	3,969E-2	-1,914E-3
h[2]		7,071E-1	-1,215E-1	1,061E+0	-1,768E-1	4,696E-2	-1,295E-2	1,061E+0	-1,547E-1	5,248E-2	-1,683E-2	5,060E-3	-1,106E-1	7,948E-3	-1,699E-2
h[3]		7,071E-1	1,215E-1	3,536E-1	4,198E-1	-1,077E-1	2,892E-2	-3,536E-1	9,944E-1	-2,679E-1	7,466E-2	-2,062E-2	3,774E-1	-5,446E-2	1,193E-2
h[4]		8,839E-2	7,071E-1	-1,768E-1	9,944E-1	-1,699E-1	5,300E-2		9,944E-1	-7,182E-2	3,133E-2	-1,411E-2	8,527E-1	3,456E-1	4,973E-2
h[5]		-8,839E-2	7,071E-1	0,000E+0	4,198E-1	4,475E-1	-1,349E-1		-1,547E-1	9,667E-1	-3,012E-1	9,913E-2	3,774E-1	7,367E-1	-7,726E-2
h[6]			1,215E-1		1,061E+0	9,667E-1	-1,638E-1		-1,989E-1	9,667E-1	-2,650E-2	1,230E-2	-1,106E-1	3,456E-1	-9,406E-2
h[7]			-1,215E-1		3,536E-1	4,475E-1	4,626E-1		6,629E-2	-7,182E-2	9,516E-1	-3,202E-1	-2,385E-2	-5,446E-2	4,208E-1
h[8]			-1,657E-2		-1,768E-1	-1,699E-1	9,516E-1			-2,679E-1	9,516E-1	2,050E-3	3,783E-2	7,948E-3	8,259E-1
h[9]			1,657E-2		0,000E+0	-1,077E-1	4,626E-1			5,248E-2	-2,650E-2	9,421E-1	0,000E+0	3,969E-2	4,208E-1
h[10]						4,696E-2	-1,638E-1			4,143E-2	-3,012E-1	9,421E-1		0,000E+0	-9,406E-2
h[11]						1,381E-2	-1,349E-1			-1,381E-2	3,133E-2	2,050E-3		0,000E+0	-7,726E-2
h[12]						-6,905E-3	5,300E-2				7,466E-2	-3,202E-1			4,973E-2
h[13]						0,000E+0	2,892E-2				-1,683E-2	1,230E-2			1,193E-2
h[14]							-1,295E-2				-9,063E-3	9,913E-2			-1,699E-2
h[15]							-3,021E-3				3,021E-3	-1,411E-2			-1,914E-3
h[16]							1,511E-3					-2,062E-2			1,909E-3
h[17]							0,000E+0					5,060E-3			0,000E+0
h[18]												2,039E-3			
h[19]												-6,797E-4			
Coeficientes da função wavelet															
h[0]	7,071E-1	0,000E+0	0,000E+0	0,000E+0	0,000E+0	0,000E+0	0,000E+0	1,768E-1	0,000E+0	0,000E+0	0,000E+0	0,000E+0	0,000E+0	0,000E+0	0,000E+0
h[1]	-7,071E-1	0,000E+0	0,000E+0	0,000E+0	0,000E+0	0,000E+0	0,000E+0	-5,303E-1	0,000E+0	0,000E+0	0,000E+0	0,000E+0	0,000E+0	-1,346E-2	0,000E+0
h[2]		7,071E-1	0,000E+0	3,536E-1	0,000E+0	0,000E+0	0,000E+0	5,303E-1	1,768E-1	0,000E+0	0,000E+0	0,000E+0	-6,454E-2	-2,695E-3	0,000E+0
h[3]		-7,071E-1	0,000E+0	-7,071E-1	0,000E+0	0,000E+0	0,000E+0	-1,768E-1	-5,303E-1	0,000E+0	0,000E+0	0,000E+0	4,069E-2	1,367E-1	0,000E+0
h[4]		0,000E+0	7,071E-1	3,536E-1	3,536E-1	0,000E+0	0,000E+0		5,303E-1	1,768E-1	0,000E+0	0,000E+0	4,181E-1	-9,350E-2	1,443E-2
h[5]		0,000E+0	-7,071E-1	0,000E+0	-7,071E-1	0,000E+0	0,000E+0		-1,768E-1	-5,303E-1	0,000E+0	0,000E+0	-7,885E-1	-4,768E-1	-1,447E-2
h[6]			0,000E+0		3,536E-1	3,536E-1	0,000E+0		0,000E+0	5,303E-1	1,768E-1	0,000E+0	4,181E-1	8,995E-1	-7,872E-2
h[7]			0,000E+0		0,000E+0	-7,071E-1	0,000E+0		0,000E+0	-1,768E-1	-5,303E-1	0,000E+0	4,069E-2	-4,768E-1	4,037E-2
h[8]			0,000E+0		0,000E+0	3,536E-1	3,536E-1		0,000E+0	5,303E-1	1,768E-1	-6,454E-2	-9,350E-2	4,178E-1	
h[9]			0,000E+0		0,000E+0	0,000E+0	-7,071E-1		0,000E+0	-1,768E-1	-5,303E-1	0,000E+0	1,367E-1	-7,589E-1	
h[10]					0,000E+0	3,536E-1			0,000E+0	0,000E+0	5,303E-1		-2,695E-3	4,178E-1	
h[11]					0,000E+0	0,000E+0			0,000E+0	0,000E+0	-1,768E-1		-1,346E-2	4,037E-2	
h[12]					0,000E+0	0,000E+0			0,000E+0	0,000E+0	0,000E+0			-7,872E-2	
h[13]					0,000E+0	0,000E+0			0,000E+0	0,000E+0	0,000E+0			-1,447E-2	
h[14]						0,000E+0			0,000E+0	0,000E+0	0,000E+0			1,443E-2	
h[15]						0,000E+0			0,000E+0	0,000E+0	0,000E+0			0,000E+0	
h[16]						0,000E+0			0,000E+0	0,000E+0	0,000E+0			0,000E+0	
h[17]						0,000E+0			0,000E+0	0,000E+0	0,000E+0			0,000E+0	
h[18]									0,000E+0	0,000E+0	0,000E+0			0,000E+0	
h[19]										0,000E+0	0,000E+0			0,000E+0	

### Família Coiflet

As propriedades da família Coiflet são: ortogonalidade, suporte compacto (poucos coeficientes), tamanho de filtro igual a  $6p$  onde  $p$  é o número da wavelet-mãe, quase simétrica possuindo  $2p$  momentos iguais a zero, sendo que a função escala tem  $2p - 1$  momentos iguais a zero. Uma wavelet com  $p$  momentos iguais a zero é definida pela equação 15. A família Coiflet é composta por 5 wavelets-mães e os coeficientes da função escala de suas respectivas wavelets-mães são apresentadas na tabela 5. A geração dos coeficientes wavelet é feita a partir dos coeficientes escala utilizando o algoritmo apresentado na figura 12.

$$\int t^k \psi(t) dt = 0, \quad k = 0, \dots, p - 1 \quad (15)$$

**Tabela 5:** Coeficientes da função escala para as wavelets-mães da família Coiflet.

Coef.	Coif1	Coif2	Coif3	Coif4	Coif5
h[0]	-7,273E-2	1,639E-2	-3,794E-3	8,923E-4	-2,121E-4
h[1]	3,379E-1	-4,146E-2	7,783E-3	-1,629E-3	3,586E-4
h[2]	8,526E-1	-6,737E-2	2,345E-2	-7,346E-3	2,178E-3
h[3]	3,849E-1	3,861E-1	-6,577E-2	1,607E-2	-4,159E-3
h[4]	-7,273E-2	8,127E-1	-6,112E-2	2,668E-2	-1,013E-2
h[5]	-1,566E-2	4,170E-1	4,052E-1	-8,127E-2	2,341E-2
h[6]		-7,649E-2	7,938E-1	-5,608E-2	2,817E-2
h[7]		-5,943E-2	4,285E-1	4,153E-1	-9,192E-2
h[8]		2,368E-2	-7,180E-2	7,822E-1	-5,204E-2
h[9]		5,611E-3	-8,230E-2	4,344E-1	4,216E-1
h[10]		-1,823E-3	3,456E-2	-6,663E-2	7,743E-1
h[11]		-7,205E-4	1,588E-2	-9,622E-2	4,380E-1
h[12]			-9,008E-3	3,933E-2	-6,204E-2
h[13]			-2,575E-3	2,508E-2	-1,056E-1
h[14]			1,118E-3	-1,521E-2	4,129E-2
h[15]			4,662E-4	-5,658E-3	3,268E-2
h[16]			-7,098E-5	3,751E-3	-1,976E-2
h[17]			-3,460E-5	1,267E-3	-9,164E-3
h[18]				-5,890E-4	6,764E-3
h[19]				-2,600E-4	2,433E-3
h[20]				6,234E-5	-1,663E-3
h[21]				3,123E-5	-6,381E-4
h[22]				-3,260E-6	3,023E-4
h[23]				-1,785E-6	1,405E-4
h[24]					-4,134E-5
h[25]					-2,132E-5
h[26]					3,735E-6
h[27]					2,064E-6
h[28]					-1,674E-7
h[29]					-9,518E-8

### Família Symlets

Esta família é similar à família Daubechies, inclusive nas suas propriedades. A família Symlet é composta por 19 séries onde os coeficientes de função escala de suas respectivas wavelets-mães são apresentadas pela tabela 6. A geração dos coeficientes wavelet é feita a partir dos coeficientes escala utilizando o algoritmo apresentado na figura 12.

**Tabela 6:** Coeficientes da função escala para as wavelets-mães da família Symlet de Sym2 a Sym20.

Coef.	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7	Sym8	Sym9	Sym10	Sym11	Sym12	Sym13	Sym14	Sym15	Sym16	Sym17	Sym18	Sym19	Sym20
h[0]	4,830E-1	3,327E-1	3,222E-2	1,954E-2	-7,801E-3	1,027E-2	1,890E-3	1,069E-3	-4,593E-4	4,893E-4	-1,791E-4	7,043E-5	4,462E-5	2,866E-5	-1,080E-5	3,791E-6	-1,513E-6	1,751E-6	-6,329E-7
h[1]	8,365E-1	8,069E-1	-1,260E-2	-2,110E-2	1,768E-3	4,010E-3	-3,029E-4	-4,732E-4	5,704E-5	1,105E-4	-1,816E-5	3,691E-5	1,933E-5	2,172E-5	-5,396E-6	-2,453E-6	7,847E-7	2,062E-6	-3,257E-7
h[2]	2,241E-1	4,599E-1	-9,922E-2	-1,753E-1	4,472E-2	-1,078E-1	-1,495E-2	-1,026E-2	4,593E-3	-6,390E-3	2,350E-3	-7,214E-4	-6,058E-4	-4,022E-4	1,655E-4	-7,607E-5	2,956E-5	-2,815E-5	1,229E-5
h[3]	-1,294E-1	-1,350E-1	2,979E-1	1,660E-2	-2,106E-2	-1,400E-1	3,809E-3	8,859E-3	-8,044E-4	-2,003E-3	3,076E-4	4,133E-4	-7,321E-5	-1,082E-4	3,657E-5	2,521E-5	-9,859E-6	-1,682E-5	4,525E-6
h[4]		-8,544E-2	8,037E-1	6,340E-1	-7,264E-2	2,886E-1	4,914E-2	6,208E-2	-2,035E-2	4,300E-2	-1,459E-2	5,675E-3	4,533E-3	3,481E-3	-1,339E-3	7,198E-4	-2,658E-4	2,762E-4	-1,174E-4
h[5]		3,523E-2	4,976E-1	7,234E-1	3,379E-1	7,678E-1	-2,722E-2	-1,823E-2	5,765E-3	3,527E-2	-2,604E-3	-1,492E-3	1,013E-3	1,526E-3	-2,221E-4	5,840E-5	4,742E-5	1,293E-4	-2,662E-5
h[6]			-2,964E-2	1,994E-1	7,876E-1	5,361E-1	-5,195E-2	-1,916E-1	4,999E-2	-1,446E-1	5,780E-2	-2,075E-2	-1,944E-2	-1,717E-2	6,938E-3	-3,932E-3	1,428E-3	-1,705E-3	7,476E-4
h[7]			-7,577E-2	-3,913E-2	4,911E-1	1,744E-2	3,644E-1	3,527E-2	-3,199E-2	-2,047E-1	1,530E-2	1,762E-2	-2,365E-3	-8,745E-3	1,360E-3	-1,905E-3	-1,888E-4	-6,179E-4	1,254E-4
h[8]				2,952E-2	-4,831E-2	-4,955E-2	7,772E-1	6,173E-1	-3,554E-2	2,377E-1	-1,704E-1	9,293E-2	6,983E-2	6,797E-2	-2,495E-2	1,240E-2	-5,240E-3	8,262E-3	-3,472E-3
h[9]					2,733E-2	-1,180E-1	6,789E-2	4,814E-1	7,179E-1	3,838E-1	7,303E-1	-7,833E-2	8,820E-3	2,590E-2	6,839E-2	-3,510E-3	9,953E-3	1,088E-3	4,319E-3
h[10]						3,491E-3	3,052E-2	-6,127E-2	2,388E-1	7,695E-1	5,720E-1	4,627E-1	-1,405E-1	-1,600E-1	-1,341E-1	7,804E-2	-1,804E-2	1,501E-2	-2,771E-2
h[11]						1,540E-2	-1,264E-2	-1,433E-1	-5,457E-2	4,717E-1	9,720E-2	7,635E-1	1,102E-1	-5,811E-2	-1,966E-1	3,072E-2	-7,262E-3	-3,261E-3	-1,691E-2
h[12]							-1,047E-3	7,607E-3	5,835E-4	-7,088E-2	-2,283E-2	3,989E-1	6,446E-1	4,753E-1	2,440E-1	-1,596E-1	1,616E-2	-3,171E-2	8,407E-2
h[13]								2,682E-3	3,170E-2	3,022E-2	-1,595E-1	6,998E-2	-2,216E-2	6,957E-1	7,600E-1	7,218E-1	-5,404E-2	-8,607E-2	-8,644E-3
h[14]									-5,421E-4	-1,153E-2	1,161E-2	3,704E-2	-3,585E-2	1,977E-1	3,932E-1	5,786E-1	4,753E-1	-1,551E-1	2,853E-2
h[15]									-3,382E-3	-1,327E-2	4,593E-2	-2,408E-2	4,918E-2	-1,244E-1	-3,532E-2	1,115E-1	7,565E-1	1,805E-1	-7,380E-2
h[16]										6,198E-4	-1,465E-3	-9,858E-3	7,554E-3	-5,975E-2	-5,763E-2	-4,108E-2	3,971E-1	6,815E-1	-3,248E-2
h[17]										1,401E-3	-8,641E-3	6,512E-3	-2,422E-2	1,386E-2	3,743E-2	4,074E-2	-3,457E-2	6,507E-1	4,015E-1
h[18]											9,563E-5	5,884E-4	-1,409E-3	-1,721E-2	4,281E-3	2,194E-2	-6,698E-2	1,424E-1	7,536E-1
h[19]												7,702E-4	-1,734E-3	7,415E-3	-2,022E-2	-2,920E-2	-3,888E-2	3,233E-2	-1,186E-1
h[20]													-3,880E-5	1,802E-4	5,296E-3	-2,754E-3	-1,941E-2	4,869E-3	1,727E-2
h[21]														1,717E-4	-1,350E-3	7,526E-3	1,004E-2	1,008E-2	-3,105E-2
h[22]																-1,135E-5	-1,709E-4	3,665E-4	3,423E-3
h[23]																	1,120E-4	-1,136E-3	-2,579E-3
h[24]																		-3,574E-5	-6,287E-5
h[25]																		6,820E-5	3,984E-4
h[26]																			1,071E-3
h[27]																			1,121E-5
h[28]																			5,512E-5
h[29]																			-2,588E-5
h[30]																			-1,607E-4
h[31]																			8,524E-4
h[32]																			-2,742E-3
h[33]																			9,502E-3
h[34]																			-5,122E-3
h[35]																			1,700E-2
h[36]																			-7,360E-6
h[37]																			2,808E-5
h[38]																			-1,386E-4
h[39]																			-4,115E-4

### Wavelet-Mãe Meyer

A função escala da wavelet Meyer discreta é baseada na função seno cardinal apresentada na equação 16. A geração dos coeficientes wavelet é feita a partir dos coeficientes escala utilizando o algoritmo apresentado na figura 12. Os coeficientes da função escala da wavelet-mãe Meyer (DMey) são apresentados pela tabela 7.

$$\text{sinc}(x) = \frac{\text{sen}(\pi x)}{\pi x} \quad (16)$$

**Tabela 7:** Coeficientes da função escala para wavelet-mãe Meyer discreta.

Coef.	Valor	Coef.	Valor	Coef.	Valor
h[0]	0,000E+0	h[21]	-1,106E-2	h[42]	-6,388E-3
h[1]	-1,010E-12	h[22]	1,527E-2	h[43]	6,046E-3
h[2]	8,519E-9	h[23]	1,742E-2	h[44]	2,203E-3
h[3]	-1,112E-8	h[24]	-3,213E-2	h[45]	-2,705E-3
h[4]	-1,080E-8	h[25]	-2,435E-2	h[46]	-6,012E-4
h[5]	6,067E-8	h[26]	6,374E-2	h[47]	6,554E-5
h[6]	-1,087E-7	h[27]	3,066E-2	h[48]	-1,631E-5
h[7]	8,201E-8	h[28]	-1,328E-1	h[49]	3,205E-6
h[8]	1,178E-7	h[29]	-3,509E-2	h[50]	7,368E-7
h[9]	-5,506E-7	h[30]	4,446E-1	h[51]	-1,490E-6
h[10]	1,131E-6	h[31]	7,446E-1	h[52]	1,131E-6
h[11]	-1,490E-6	h[32]	4,446E-1	h[53]	-5,506E-7
h[12]	7,368E-7	h[33]	-3,509E-2	h[54]	1,178E-7
h[13]	3,205E-6	h[34]	-1,328E-1	h[55]	8,201E-8
h[14]	-1,631E-5	h[35]	3,066E-2	h[56]	-1,087E-7
h[15]	6,554E-5	h[36]	6,374E-2	h[57]	6,067E-8
h[16]	-6,012E-4	h[37]	-2,435E-2	h[58]	-1,080E-8
h[17]	-2,705E-3	h[38]	-3,213E-2	h[59]	-1,112E-8
h[18]	2,203E-3	h[39]	1,742E-2	h[60]	8,519E-9
h[19]	6,046E-3	h[40]	1,527E-2	h[61]	-1,010E-12
h[20]	-6,388E-3	h[41]	-1,106E-2		

### Wavelet-Mãe Legendre

A família de wavelets-mães Legendre não é ortogonal. Seus coeficientes derivam de funções harmônicas esféricas e são apresentados na tabela 8. A geração dos coeficientes wavelet é feita a partir dos coeficientes escala utilizando o algoritmo apresentado na figura 12.

**Tabela 8:** Coeficientes da função escala para as wavelets-mães da família Legendre de Lege2 a Lege6.

Coef.	Lege2	Lege4	Lege6
h[0]	$-\sqrt{2}/2$	$-5\sqrt{2}/16$	$-63\sqrt{2}/256$
h[1]	$-\sqrt{2}/2$	$-3\sqrt{2}/16$	$-35\sqrt{2}/256$
h[2]		$-3\sqrt{2}/16$	$-30\sqrt{2}/256$
h[3]		$-5\sqrt{2}/16$	$-30\sqrt{2}/256$
h[4]			$-35\sqrt{2}/256$
h[5]			$-63\sqrt{2}/256$

Nesta seção foram apresentadas as wavelets-mães utilizadas para esse estudo. Serão abordados agora os conceitos necessários para o entendimento da convolução e dos seus efeitos gerados por ela na interpretação da extensão de borda de sinais convoluídos.

## 2.6 CONVOLUÇÃO E MÉTODOS DE EXTENSÃO DE BORDA

De acordo com Lathi (2007), a convolução de sinais é uma operação matemática que toma duas funções, ou sinais, para gerar uma terceira função como resultado da operação. Uma interpretação para essa operação, dentro da engenharia, é que essa operação descreve características intrínsecas de um sistema.

A DWT, conforme descrita na seção 2.5.2, realiza convoluções recursivas entre o sinal e os coeficientes dos filtros passa-altas e passa-baixas da função escala da wavelet-mãe, gerando os coeficientes de detalhe e aproximação de cada decomposição wavelet. No processo de convolução, a interpretação de como se comportam os dados de sinais fora da janela de convolução tem uma influência significativa no resultado final, e por consequência, na transformada wavelet (MISITI *et al.*, 2013).

O objetivo dessa seção não é explicar o processo de convolução, o qual pode ser consultado em detalhes em Lathi (2007), mas esclarecer o procedimento de extensão de bordas necessário na operação de convolução.

Existem diversos métodos de interpretação para a extensão de borda utilizada na operação de convolução. Na sequência são apresentados nove métodos de extensão de borda descritos na tabela 9.

Como exemplo, suponha-se o vetor de sinal: **[1,1 1,2 1,3]**. Este vetor de sinal será estendido de três amostras para a direita e três amostras para a esquerda como preparação para a convolução para as extensões de borda definidas na tabela 9.

**Tabela 9:** Métodos de extensão de borda.

Sigla	Nome	Descrição
PPD	<i>Periodic-Padding</i>	Extensão de borda periódica
PER	<i>Periodic Even Padding</i>	Extensão de borda periódica par
SP0	<i>Smooth-Padding of order 0</i>	Replicação da primeira e da última amostra
SP1	<i>Smooth-Padding of order 1</i>	Interpolação de borda de primeira derivação
SYMH	<i>Symmetric-Padding Half-Point</i>	Replicação simétrica de ponto intermediário
SYMW	<i>Symmetric-Padding Whole-Point</i>	Replicação simétrica de ponto cheio
ASYMH	<i>Antisymmetric-Padding Half-Point</i>	Replicação antissimétrica de ponto intermediário
ASYMW	<i>Antisymmetric-Padding Whole-Point</i>	Replicação antissimétrica de ponto cheio
ZPD	<i>Zero-Padding</i>	Os valores da borda são zero

1. PPD: O sinal é considerado periódico e as amostras são repetidas.

Exemplo aplicando extensão de borda PPD: [1,1 1,2 1,3 **1,1 1,2 1,3** 1,1 1,2 1,3].

2. PER: Esta extensão opera da mesma forma que a PPD, mas se o vetor tiver número de amostras ímpares, uma amostra extra é inserida do lado direito do vetor. A amostra extra inserida é igual ao último valor do vetor.

Exemplo aplicando extensão de borda PER: [1,2 1,3 1,3 **1,1 1,2 1,3** 1,3 1,1 1,2 1,3].

3. SP0: Replicação da primeira e da última amostra.

Exemplo aplicando extensão de borda SP0: [1,1 1,1 1,1 **1,1 1,2 1,3** 1,3 1,3 1,3].

4. SP1: Tomando a borda da direita como exemplo, adota-se como referência o valor da última amostra menos a penúltima amostra. O primeiro valor à direita a ser inserido é o valor da referência mais o valor da última amostra do vetor. O próximo valor a ser inserido é o valor de referência mais o último valor calculado. Para o lado esquerdo do vetor, o valor de referência deve ser subtraído.

Exemplo aplicando extensão de borda SP1: [0,8 0,9 1,0 **1,1 1,2 1,3** 1,4 1,5 1,6].

5. SYMH: A extensão do lado direito do vetor é feita repetindo o último valor e as amostras seguintes. Do lado esquerdo, espelha-se o vetor.

Exemplo aplicando extensão de borda SYMH: [1,3 1,2 1,1 **1,1 1,2 1,3** 1,3 1,2 1,1].

6. SYMW: Para construir essa extensão de sinal, considerando o lado esquerdo, repete-se o segundo valor e valores consecutivos para a direita.

Exemplo aplicando extensão de borda SYMW: [1,2 1,3 1,2 **1,1 1,2 1,3** 1,2 1,1 1,2].

7. ASYMH: Esta extensão de sinal opera da mesma maneira que a SYMH, mas ao repetir as amostras, inverte-se o sinal.

Exemplo aplicando extensão de borda ASYMH: [-1,3 -1,2 -1,1 **1,1 1,2 1,3** -1,3 -1,2 -1,1].

8. ASYMW: Considerando o lado direito do vetor como exemplo, utiliza-se a última amostra do vetor e seu valor multiplicado por dois é tomado como referência. O primeiro valor da borda da direita é calculado subtraindo-se o valor da referência pelo valor da penúltima amostra, e assim por diante, refazendo sempre o cálculo do valor de referência. A borda esquerda é feita de forma contrária.

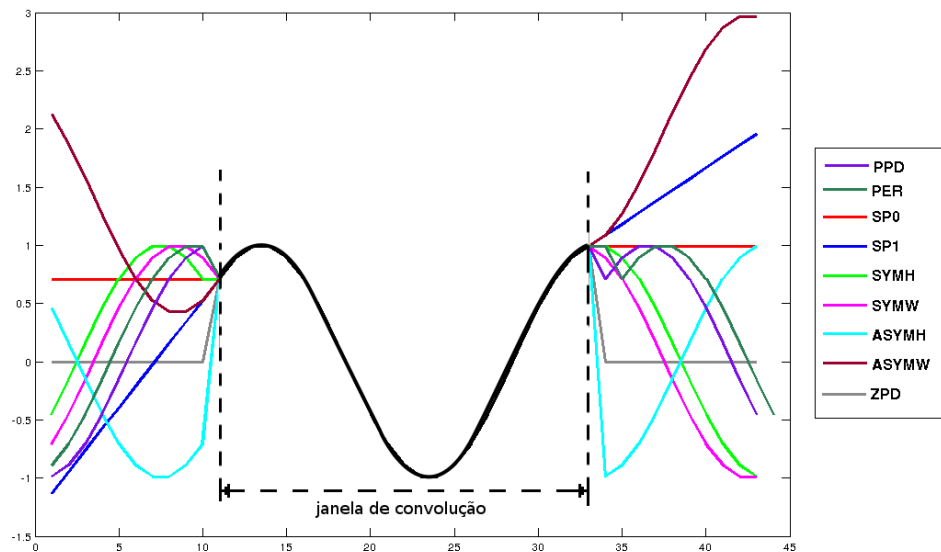
Exemplo aplicando extensão de borda ASYMW: [0,8 0,9 1,0 **1,1 1,2 1,3** 1,4 1,5 1,6].

9. ZPD: Os valores da borda são zero.

Exemplo aplicando extensão de borda ZPD: [0 0 0 **1,1 1,2 1,3** 0 0 0].

A extensão de borda ZPD é a interpretação de borda onde não são considerados existentes as informações fora da janela de convolução, ou seja, não é considerada a borda, porém o sinal em convolução recebe as amostras adicionais iguais a zero.

De modo gráfico, considerando as extensões de bordas apresentadas aplicadas a uma época de função cosseno, o resultado é apresentada pela figura 13.



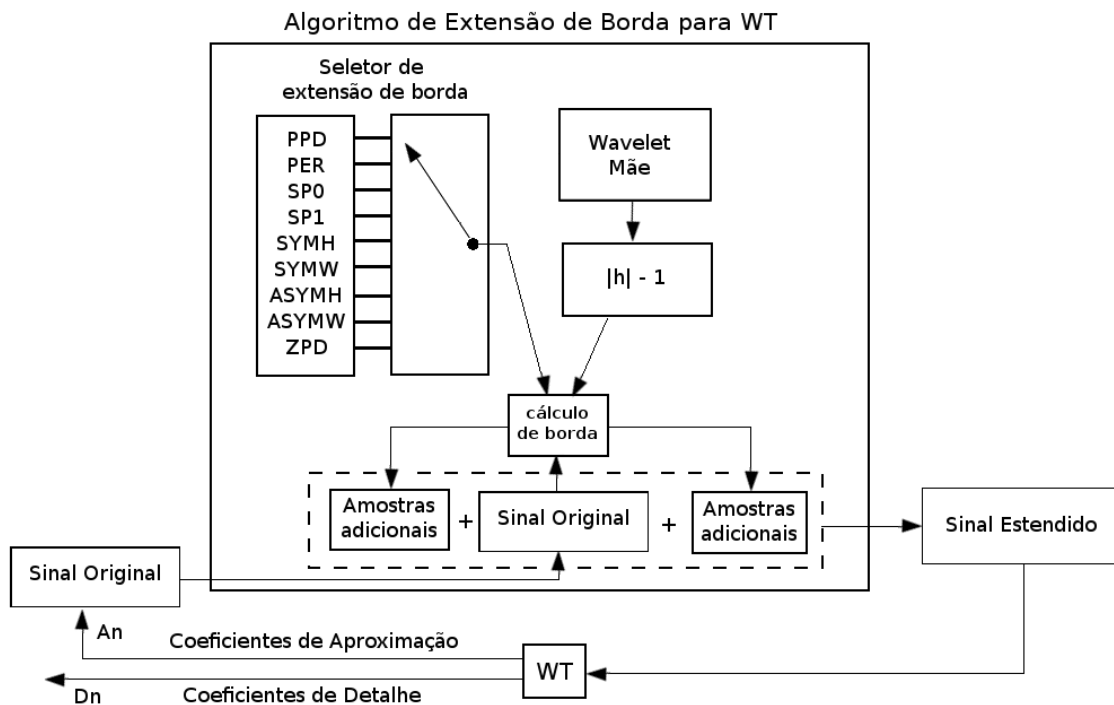
**Figura 13:** Época de uma função cosseno variando de  $-\pi/4$  a  $\pi/2$  dentro de uma janela de convolução. Fora da janela de convolução estão simuladas os nove métodos de extensão de borda do sinal.

### 2.6.1 EXTENSÃO DE BORDA NA TRANSFORMADA WAVELET DISCRETA

Como descrito na seção 2.5.2, o cálculo da transformada wavelet discreta necessita de um número finito de amostras para executar as operações recursivas de convolução do sinal em questão com os coeficientes dos filtros passa-altas e passa-baixas da wavelet-mãe. Dessa

maneira, faz-se necessária a implementação de algum tipo de extensão de borda para cada nível de decomposição wavelet conforme descrito por Mallat (1989). Ou seja, não é possível evitar a execução da extensão de borda na transformada wavelet, independente do tamanho da amostra em transformação. Mesmo em sistemas de amostragem contínua, onde não é necessária a extensão devido ao janelamento exigido pela DWT.

A geração das amostras adicionais de extensão de bordas utilizado pela DWT é executada para cada decomposição da transformada wavelet, sendo apresentada pela figura 14.



**Figura 14:** Diagrama de geração de borda para cada decomposição da transformada wavelet. O número de amostras adicionadas à esquerda e à direita do sinal a ser convoluído são iguais ao número de coeficientes da wavelet-mãe menos um.

Na figura 14, o sinal original é usado como entrada para a geração das amostras adicionais de extensão de borda da DWT. O número de amostras geradas será igual a  $|h| - 1$ , onde  $h$  representa os coeficientes da wavelet-mãe em questão e  $| \cdot |$  o operador que retorna a quantidade de coeficientes existentes para a wavelet-mãe. Tomando o número de coeficientes a serem gerados, o modelo de extensão de borda selecionado, e o sinal original como base, o cálculo de borda gera as amostras adicionais de borda que serão agregadas ao lado esquerdo e direito do sinal original. O sinal estendido é a saída do algoritmo de extensão de borda. Esse sinal estendido é então processado pela transformada wavelet gerando os coeficientes de detalhe  $D_n$  e aproximação  $A_n$ , onde  $n$  é o número da iteração. Como o algoritmo da DWT é recursivo, os coeficientes de aproximação gerados pelo primeiro processamento da DWT são usados como



entrada novamente no algoritmo de extensão de borda no lugar do sinal original. O processo se repete até atingir  $n$  decomposições wavelet desejadas.

## 2.6.2 EXTENSÃO DE BORDA NO MATLAB®

Para esse estudo, foi desenvolvida uma aplicação em Java descrita na seção 4.5. Para validar os resultados obtidos na implementação do algoritmo da DWT durante a fase inicial da pesquisa, foi utilizado o *toolbox* de DWT do software Matlab®.

Ao executar uma transformada wavelet no software Matlab® não é exigido a definição de um método de extensão de borda para o processamento da DWT. Essa definição é uma configuração do software, onde o método de extensão de borda utilizado por padrão é o SYMH (MISITI *et al.*, 2013). A entrada do comando `'dwtmode'`, sem argumentos, retorna o método de extensão de borda utilizado para a execução da DWT. A entrada do comando `'dwtmode('arg')`, onde *arg* é a abreviação do método de extensão de borda conforme descrito na tabela 9, modifica a extensão de borda padrão durante a sessão em uso. Isso causa a execução de uma mensagem de atenção para o usuário pela modificação desse parâmetro conforme visto na figura 15.

```
>> dwtmode

*****
**  DWT Extension Mode: Symmetrization (half-point)  **
*****

>> dwtmode('SP0');

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!  WARNING: Change DWT Extension Mode  !
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

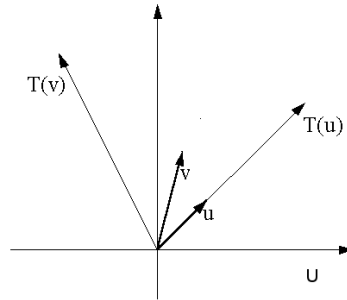
*****
**  DWT Extension Mode: Smooth Padding of order 0  **
*****
```

**Figura 15:** Entrada de comandos no software Matlab® onde na primeira linha é executado o comando `'dwtmode'` sem argumento. O software retorna o método de extensão de borda padrão em uso: SYMH. Com a entrada do comando `'dwtmode('SP0')`, é emitido um aviso de atenção e é alterado o método de extensão de borda padrão para o método de extensão de borda SP0. Essa alteração é válida para a sessão corrente.

Essa característica do Matlab® não é trivial. É importante salientar a sua necessidade para o caso de comparação de resultados na implementação de algoritmo dedicado para a transformada wavelet.

## 2.7 AUTOVETORES E AUTOVALORES

Autovetores e autovalores são conceitos da álgebra linear com aplicações em diversas áreas. Um autovetor é um vetor que sob a ação de um operador linear resulta em outro vetor com a mesma direção, e para cada autovalor  $\lambda$ , podem existir vários autovetores (APOSTOL, 1969). Esse conceito pode ser explicado pela figura 16.



**Figura 16:** O vetor  $u$  é autovetor de  $T()$ , porém  $v$  não é autovetor de  $T()$ .

Um vetor não nulo  $u$  em  $U$  é dito **autovetor** de  $T$  se existe um número real  $\lambda$  tal que a equação 17 seja verdadeira.

$$T(u) = \lambda u \quad (17)$$

O escalar  $\lambda$  é um **autovalor** de  $T$  associado a  $u$ . A conclusão é que  $u$  e  $T(u)$  são paralelos. Considerando uma matriz de características  $A$ , à qual precisamos encontrar uma projeção paralela a esse espaço com menos dimensões, substituímos  $T$  pela matriz  $A$  na equação 17 e reagrupando os termos tem-se que:

$$\begin{aligned} T(u) &= \lambda u \\ Au &= \lambda u \\ Au &= \lambda Iu \\ \lambda Iu - Au &= 0 \\ (\lambda I - A)u &= 0 \end{aligned} \quad (18)$$

Onde para a solução não nula da equação 18 deve-se ter o determinante da função característica da matriz  $(\lambda I - A)$  igual a zero conforme indicado pela equação 19:

$$F(\lambda) = \det(\lambda I - A) = 0 \quad (19)$$

Exemplificando o cálculo, considere uma matriz de características conforme a matriz  $A$  da equação 20:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 4 \\ -1 & -1 & -2 \end{bmatrix} \quad (20)$$

Aplicando  $A$  na equação 18 e resolvendo o determinante tem-se que os autovalores possíveis para a equação 17 utilizando a matriz  $A$  é conforme apresentado na equação 21.

$$\det(\lambda I - A) = (\lambda - 3)(\lambda + 1)(\lambda - 1) = 0$$

$$\lambda = [1 \quad -1 \quad 3] \quad (21)$$

Para cada autovalor de  $\lambda$  tem-se uma solução possível, um autovetor, não nulo para a equação 17. Considerando o primeiro valor de  $\lambda$ , substitui-se na equação 17 e resolvendo tem-se:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 4 \\ -1 & -1 & -2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = 1 \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \quad (22)$$

$$+u_1 + u_2 + u_3 = 0 \quad (23)$$

$$+2u_1 + 2u_2 + 4u_3 = 0 \quad (24)$$

$$-1u_1 - 1u_2 - 3u_3 = 0 \quad (25)$$

$$u = \alpha(1, -1, 0) \quad (26)$$

Onde a equação 26 é o autovetor para o autovalor  $\lambda = 1$ . Para os valores subsequentes de  $\lambda$ , se repetem os passos das equações 22 a 26.

## 2.8 ANÁLISE DE DISCRIMINANTE LINEAR

No reconhecimento de padrões, um dos grandes desafios é a redução de dimensões de classes multivariadas (DUDA; HART; STORK, 2001). Conhecido também como a maldição da dimensionalidade (*curse of dimensionality*), este termo, introduzido por Bellman em 1961, refere-se aos problemas computacionais que ocorrem em análise de dados multivariados quando a dimensão do problema cresce. Ao observar características de classes que se deseja classificar,

normalmente várias características acabam se sobrepondo, dificultando o processo de separação e a respectiva classificação de elementos. Para melhorar a classificação, extraem-se novas características não correlacionadas para aumentar a separabilidade destas classes. Os seres humanos conseguem classificar facilmente trajetórias e grupos de uma, duas ou três dimensões, mas esta capacidade se degrada facilmente para ordens maiores que quatro. Para tanto faz-se necessário uma ferramenta que diminua significativamente o número de dimensões de um problema para algo manipulável.

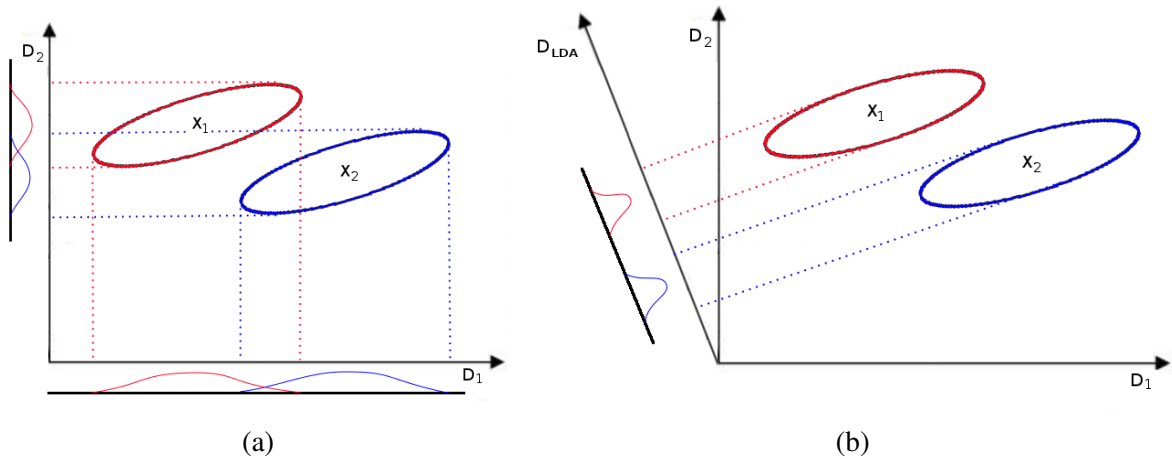
O objetivo da Análise de Discriminante Linear (*Linear Discriminant Analysis* - LDA) é reduzir a dimensionalidade de um conjunto de dados multivariado, preservando o máximo possível a informação discriminatória das classes (FISHER, 1936; MARTINEZ; KAK, 2001). O LDA, a partir de um espaço de dados multidimensional contendo  $C$  classes, rotaciona esse espaço determinando uma projeção com  $C - 1$  dimensões. A projeção determinada será a que apresenta a melhor separabilidade entre as classes.

Um exemplo da análise de discriminante linear pode ser visto na figura 17, onde na figura 17(a) são apresentadas as classes de variáveis  $X_1$  e  $X_2$  com duas características mensuráveis pelos eixos  $D_1$  e  $D_2$ . Como apresentado na figura 17(a), as projeções das classes de variáveis nos eixos  $D_1$  e  $D_2$  se sobrepõem em ambos os eixos, não permitindo a separação das duas classes através de um limiar de detecção. O LDA, com base na quantidade de classes ( $C = 2$ ) e nas características de cada classe, determinará uma nova projeção contendo  $C - 1$  dimensões. O resultado do LDA para o exemplo apresentado será unidimensional, representado por um único eixo. O novo eixo determinado pelo LDA,  $D_{LDA}$ , pode ser visto na figura 17(b) como a reta inclinada para a esquerda ao lado do eixo  $D_2$ . Verifica-se pela figura 17(b) que as projeções das duas classes de variáveis sobre o eixo  $D_{LDA}$  não estão sobrepostas, permitindo a separação total das classes através de um limiar de detecção.

Para o cálculo do LDA, supondo dois conjuntos de dados  $X_1$  e  $X_2$  representando duas classes distintas de um universo de classes  $X_c$  onde  $c = 1, 2, \dots, C$ . (Neste caso,  $C = 2$ ). Cada conjunto de dados possui  $n$  amostras, e cada amostra possui  $d$ -dimensões. O produto escalar de cada conjunto  $d$ -dimensional pela matriz de pesos  $w$ , como representado pela equação 27, gera um conjunto de dados  $Y$  transformado para um espaço amostral contendo  $C - 1$  dimensões.

$$Y = w^t X \quad (27)$$

Para encontrar a matriz de pesos  $w$ , o primeiro passo é determinar a diferença das médias das classes do conjunto  $X_c$ , como apresentada pela equação 28.



**Figura 17:** (a) Projeção das classes  $X_1$  e  $X_2$  sobre os eixos de variáveis  $D_1$  e  $D_2$ . Observa-se que as projeções estão sobrepostas não permitindo a separabilidade total das classes. (b) A análise de discriminante linear determina uma nova projeção  $D_{LDA}$  onde as classes apresentam a melhor separabilidade entre si.

$$m_i = \frac{1}{n_i} \sum_{x \in X_i} x \quad (28)$$

As médias das classes definidas pela equação 28 pode ser projetada também pela matriz  $w$ , como mostra a equação 29.

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{y \in Y_i} w^t x = w^t m_i \quad (29)$$

Onde  $\tilde{m}_i$  é a média do subconjunto projetado através da matriz  $w$ . Como o que se deseja é a maior separabilidade entre as classes, leia-se a maior separabilidade entre as médias das classes, logo a equação 30 é a função objetivo.

$$J = |\tilde{m}_1 - \tilde{m}_2| = |w^t (m_1 - m_2)| \quad (30)$$

Apesar de ser a função objetivo, a equação 30 não leva em consideração o desvio padrão de cada classe. Para isso adota-se a proposta de Fisher (DUDA; HART; STORK, 2001), que é normalizar a diferença das médias por uma medida de dispersão interna da classe (*scatter*). A dispersão interna de classe é apresentada pela equação 31.

$$\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2 \quad (31)$$

Assim, a dispersão total  $J(w)$  entre classes das amostras projetadas é igual a  $\tilde{s}_1^2 + \tilde{s}_2^2$  e a maximização do critério de Fisher é apresentada pela equação 32.

$$J(w) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \quad (32)$$

A maximização de  $J(\cdot)$  por  $w$  conduz à maior separação entre as classes em questão. Esse cálculo é feito através das matrizes de dispersão  $S_i$  e  $S_w$ , apresentadas pelas equações 33 e 34.

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^t \quad (33)$$

$$S_w = S_1 + S_2 \quad (34)$$

Essas matrizes são projetadas pela multiplicação da matriz de pesos  $w$  como apresentado pela equação 35 e da mesma maneira as médias projetadas pela equação 36.

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^t S_w w \quad (35)$$

$$(\tilde{m}_1 - \tilde{m}_2)^2 = w^t S_B w \quad (36)$$

Substituindo as equações 35 e 36 na equação 32 obtém-se a equação 37.

$$J(w) = \frac{w^t S_B w}{w^t S_w w} \quad (37)$$

A equação 37 pode ser reescrita na forma apresentada pela equação 38.

$$S_B w = \lambda S_w w \quad (38)$$

Onde a constante  $\lambda$  pode ser determinada através de autovalores e autovetores. A partir dos autovalores, seleciona-se o valor cujo argumento é máximo. O valor selecionado corresponde à coluna dos autovetores que serão multiplicados pelas características das classes de entrada. Essa multiplicação corresponde à rotação da matriz de entrada para uma projeção onde é mantida a separação máxima entre as classes.

É importante salientar que a análise de discriminante linear e obtenção da matriz de pesos que permite rotacionar a matriz de características reduzindo suas dimensões para uma projeção contendo  $C - 1$  dimensões (onde  $C$  é o número de classes) ocorre em um único

passo, uma única vez. Como o LDA rotaciona o espaço multidimensional da matriz de características para uma nova projeção com  $C - 1$  dimensões, não há perda de informação. Todas as características utilizadas para a geração da matriz de características são reagrupadas de modo a maximizar a separação entre as  $C$  classes.

## 2.9 COMENTÁRIOS

Foram apresentadas nesse capítulo os conhecimentos necessários para o entendimento do trabalho.

O próximo capítulo apresenta o **estado da arte** para a pesquisa sobre a transformada wavelet e a detecção de espículas.

### 3 ESTADO DA ARTE

Para este estudo, foi realizada uma análise de publicações dos últimos 20 anos buscando verificar o desenvolvimento no reconhecimento, detecção e classificação de espículas. O estudo foi feito em ordem cronológica a partir de 1994. Foi dada ênfase para trabalhos que utilizam a transformada wavelet seja para a extração de características do sinal ou para o uso direto na classificação dos eventos.

#### 3.1 TRANSFORMADA WAVELET APLICADA NO RECONHECIMENTO DE ESPÍCULAS

Daubechies (1988) e Mallat (1989) fizeram as primeiras publicações sobre a transformada wavelet, sendo esses dois trabalhos os mais relevantes na área e citados em praticamente todas as publicações posteriores a eles.

Kalayci, Ozdamar e Erdol (1994) utilizam, para o reconhecimento de espículas no exame de EEG, a wavelet-mãe Daubechies 20, onde 16 coeficientes são escolhidos das decomposições 2 e 3 e usados como entrada em uma rede neural com 7 neurônios na camada escondida. Os dados foram coletados de 4 pacientes com uma amostragem de 100 Hz, com 1189 épocas de sinal de 5,12 segundos para o treinamento da rede neural e 1498 épocas de sinal para teste. Os resultados obtidos foram de 81,3% de verdadeiros positivos (VP), 95% de verdadeiros negativos (VN), 18,7% de falsos positivos (FP) e 5% de falsos negativos (FN). Os autores ainda apresentam uma acurácia de 90% no teste.

Clark *et al.* (1995) fazem um estudo gráfico de épocas de EEG processadas através da transformada wavelet. Não é comentado qual wavelet-mãe é utilizada no estudo. Não comentam a fonte dos dados, mas descrevem que o sinal foi amostrado a 333 Hz e que contém artefatos, espículas e complexos ponta-e-onda. Os autores demonstram visualmente que é possível reconhecer através de determinadas decomposições, a localização dos eventos ao longo do tempo.

Kalayci e Ozdamar (1995) analisam a redução de entradas em uma rede neural



utilizando apenas 8 coeficientes de escalas diferentes de WT Daubechies 4 e 20. As amostras utilizadas para o estudo foram 15 minutos de um exame de EEG gravados de 5 pacientes diferentes, diagnosticados como epiléticos. O trabalho faz várias comparações de modelos de redes neurais, wavelets-mães e decomposições, apresentando como melhor resultado 92,0% de acurácia, 90,8% de sensibilidade e 93,3% de especificidade.

Unser (1996) faz um estudo sobre os diferentes tipos de implementação da transformada wavelet como a CWT, o *wavelet frame* (WF) e a DWT, demonstrando que as wavelets podem ser usadas como filtro casado para escalas múltiplas em sinais biomédicos.

Yong e Shengxun (1996) utilizam as wavelets e suas decomposições para capturar e separar as regiões de frequências  $\alpha$ ,  $\theta$  e  $\beta$  do sinal gravado de um exame de EEG. Para esse estudo, utilizaram os sinais de 31 pacientes, mas não descrevem como foi realizada a digitalização do sinal.

Blanco *et al.* (1998) aplicam a transformada wavelet de pacote (*Wavelet Packet Transform* - WPT). A WPT é aplicada a um sinal de EEG contendo todas as fases de uma crise epilética do início ao fim. Os autores usam a WPT para separar as frequências  $\gamma$  e  $\theta$ , como um filtro. O sinal filtrado, amostrado em 102,4 Hz, não apresenta mais os espasmos musculares no sinal, permitindo a visualização do sinal durante a crise. O artigo descreve matematicamente a WPT e a função de custo de informação (*Information Cost Function* - ICF) utilizado para detectar a presença de eventos durante o sinal. Os autores não descrevem os valores de desempenho nem de valores para a técnica utilizada.

Geva e Kerem (1998) utilizam uma abordagem diferente quanto à interpretação dos dados. Fazem o estudo da wavelet-mãe Mallat e a reunião das características do sinal através do agrupamento fuzzy não supervisionado em ratos epiléticos. Apresentam os resultados e como foi feita o agrupamento a partir das características extraídas da wavelet. Os descritores utilizados para o agrupamento foram a variância, a simetria (*skewness*) e a curtose (*kurtosis*) das distribuições dos dados. Os agrupamentos, com exceção de um agrupamento, apresentaram sobre-posição.

Teixeira (2001) analisa diversos tipos de wavelets e compara a acurácia de cada família em detectar o complexo ponta-e-onda de um sinal de EEG. A detecção é comparada através da exatidão no tempo e do nível de correlação dos níveis de detalhe com o complexo ponta-e-onda utilizado. As famílias de wavelet que apresentaram melhores resultados foram as bior3.1 e a sym5.

Wilson e Emerson (2002) apresentam um artigo de revisão sobre detecção de espículas

em sinais de EEG. Eles analisam trabalhos de diversos autores, muitos deles comentados neste texto, fazendo um comparativo sobre qual seria a melhor forma de comparar os resultados de trabalhos e extrair valores comparáveis. Os autores definem que a especificidade, a sensibilidade e a curva ROC são as características de desempenho comparáveis entre técnicas para determinar se um método é melhor do que outro. Nesse trabalho, ainda é feita a comparação de sensibilidade e especificidade entre especialistas. A conclusão dos autores é de que há incoerência entre os próprios especialistas que marcam as espículas, obtendo sensibilidade de pouco mais de 60% entre dois especialistas.

Adeli, Zhou e Dadmehr (2003) aplicam a transformada wavelet contínua e a discreta em dois sinais de EEG com espículas. Os autores demonstram o resultado da decomposição da espícula em até 6 níveis. Foram utilizados 64 eletrodos no escalpo amostrado em 256 Hz em um arquivo de 1s. Foi utilizada uma base de dados pública e os autores concluem que para detectar espículas a melhor wavelet é a Daubechies 4, mas não definem uma decomposição específica.

Yamaguchi (2003) aplicou tanto a transformada wavelet contínua como a transformada wavelet discreta em 82 segundos de sinal com eventos epileptiformes amostrados a 100 Hz. A partir dos resultados obtidos, comenta que é possível extrair dados de tempo e frequência das espículas, e que a melhor wavelet-mãe para esse fim é a Daubechies 8 com 5 níveis de decomposição. Comenta ainda que, durante os eventos epileptiformes, as ondas  $\alpha$  e  $\beta$  não estão presentes no sinal.

Ramanan, Kalpakam e Sahambi (2004) propõem técnicas para detectar e remover artefatos de sinais normais e epileptiformes. Os artefatos são as piscadas palpebrais. A utilização da transformada wavelet discreta com a wavelet-mãe Haar possibilitou a detecção do momento de aparição desses artefatos através dos níveis 7 e 8 de decomposição. Sugerem que é possível filtrar esses artefatos através da mesma técnica.

Nenadic e Burdick (2005) utilizam o eletroencefalograma de um macaco com sinais de epilepsia. Os autores fazem uma análise estatística do ruído neural e propõem uma técnica automatizada não supervisionada para a detecção de espículas. Eles mostram através da curva ROC que o uso de limiares de detecção sem um pós-processamento não é eficiente e propõem o uso do método de detecção wavelet (*Wavelet Detection Method* - WDM). O método de detecção proposto separa o sinal em 5 classes e a classificação é feito por *K-means*. Não são apresentados valores de acurácia, falso positivo ou negativo neste estudo, sendo apresentado apenas a separabilidade dos eventos e sua detecção estatística.

Ercelebi e Subasi (2006) apresentam uma técnica de transformada wavelet denominada *Lifting Base Wavelet Transform* (LBWT) que reduz 50% a carga de processamento e fazem um

estudo das wavelets-mães Haar, Daubechies 4, 6 e 8 com uma rede neural para detectar ondas  $\alpha$ ,  $\beta$ ,  $\theta$  e  $\delta$  em sinais de EEG.

Sadati, Mohseni e Maghsoudi (2006) utilizaram 500 épocas de 23,6 segundos de duração, onde 200 épocas foram registradas em pacientes epiléticos através de EEG de escalpo, 200 épocas foram capturadas sem a presença de eventos epileptiformes e 100 épocas foram registradas através de exame de EEG intracraniano. Essa base é pública, primeiramente formalizada por Andrzejak *et al.* (2001) e disponível em Andrzejak, Schindler e Rummel (2012a). Essas épocas foram processadas por WT DB4. A energia das cinco decomposições da transformada wavelet foi utilizada como entrada para uma lógica fuzzy experimental. Obtiveram uma acurácia de 85,6%.

Brechet *et al.* (2007) comprimiram os sinais biomédicos para transmissão. Não satisfeitos com os resultados das transformadas wavelets conhecidas, eles projetam uma função wavelet-mãe que condiz melhor com o sinal a ser comprimido. O artigo explica o processo de cálculo da função wavelet-mãe e apresenta resultados melhores de compactação do que as wavelets por eles analisadas.

Adeli, Ghosh-Dastidar e Dadmehr (2007) fazem um estudo do caos em cada decomposição da wavelet Daubechies 4 representando frequências  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  e  $\gamma$  em sinais de EEG. Aparentemente é possível separar as classes de eventos epileptiformes, mas não confirmam a separabilidade.

Fatourechi, Birch e Ward (2007) comentam que as características interessantes para análise de eventos epileptiformes se encontram em frequências abaixo de 4 Hz e que faltam estudos nessa área. Afirmam também que analisar apenas um canal de EEG ou dois não são suficientes e conclusivos na detecção e classificação de espículas. Um algoritmo deve analisar vários canais ao mesmo tempo para chegar a uma conclusão mais robusta.

Inuso *et al.* (2007) descrevem a decomposição de um sinal de EEG contendo padrões epileptiformes usando Daubechies 4 para identificação e filtragem de artefatos. Através da análise de entropia de Reyni e do descritor curtose, conseguem identificar os artefatos, mas apenas comentam que é possível retirar os artefatos em um pré-processamento.

Indiradevi *et al.* (2008) aplicam a transformada wavelet discreta utilizando Daubechies 4 e aplicam um limiar de detecção variável nas decomposições 4 e 5. Foram utilizadas 684 épocas contendo espículas captadas em EEG de 22 pacientes epiléticos e 1644 épocas contendo transientes não epileptiformes. Todas as épocas possuem 3 segundos de duração, foram amostradas em 256 Hz e filtradas com um filtro passa-baixas com frequência de corte de

100 Hz. A sensibilidade e a especificidade ficaram em 91,7% e 89,3%, respectivamente.

Xu e Song (2008) utilizaram várias wavelets-mães para determinar um classificador usando LDA para uma interface cérebro computador. Não é objetivo caracterizar espículas, mas eles testam o reconhecimento de padrões com as wavelets Daubechies 10, Dmeyer, Coiflet5 e Rbio1.3, obtendo acurácia de 90%, 90%, 89,29% e 87,86% respectivamente.

Ocak (2009), utilizando a base Andrzejak *et al.* (2001), separa as épocas em grupos. Para cada época, é calculada a entropia aproximada (*Approximate Entropy* - ApEn) da época decomposta em 3 níveis de DWT na detecção de espículas. Não é comentado qual wavelet-mãe foi utilizada, apenas que a primeira decomposição foi a que apresentou o melhor resultado. Ele consegue uma sensibilidade de 93,7% e especificidade de 96%.

Wang *et al.* (2009) implementam um classificador utilizando uma DWT com wavelet-mãe DB4 e aplicam a entropia aproximada em suas decomposições para classificar através de limiar de decisão. Foram utilizadas 100 épocas não contendo eventos epileptiformes e 133 épocas contendo espículas gravadas de pacientes saudáveis e epiléticos, respectivamente, amostradas a 128 Hz. Foi obtida uma sensibilidade de 0,984.

Yu (2009) utilizou a wavelet Mallat e Daubechies 4 com limiares de corte nos coeficientes das decomposições, recompondo o sinal de EEG para filtrá-lo e eliminar artefatos. O artigo descreve as características de cada artefato e os limiares utilizados.

Ubeyli (2009), utilizando a base Andrzejak *et al.* (2001), aplicou a DWT com Daubechies 2 com 4 níveis de detalhes e um de aproximação sobre essas épocas. De cada decomposição, extraiu 4 descritores (valor máximo, valor mínimo, média e desvio padrão) e utilizou-os em uma rede neural artificial MLP (*Multilayer Perceptron*) com uma camada escondida. A conclusão do autor é que como não há método analítico para determinar a melhor wavelet-mãe para um caso particular de dados, a função wavelet-mãe ótima deve ser determinada experimentalmente. Os resultados obtidos na classificação dos eventos foram de 94,5% de sensibilidade e 96% de especificidade.

Hsu e Yu (2009) utilizam a wavelet-mãe Daubechies 4 com 4 níveis de decomposição em um sinal amostrado em 120 Hz. São separadas as frequências  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  e  $\gamma$ , e que estas frequências são interessantes para a análise de sinais de EEG. A seleção das características do sinal influenciam muito o desempenho do classificador. Implementam uma máquina de suporte de vetor (*Support Vector Machine* - SVM) e algoritmos genéticos (GA) como classificador e obtêm uma acurácia de 0,812 com SVM, e 0,918 com SVM e GA.

Shuren e Zhong (2009) obtêm, a partir de um sinal de EEG com eventos epileptiformes

decomposto em 4 níveis utilizando a wavelet-mãe Daubechies 4, os sinais  $\delta$ ,  $\theta$ ,  $\alpha$  e  $\beta$ . Utilizam uma rede neural para classificar os eventos obtendo uma sensibilidade de 83,5%. Não é descrito a especificidade. Não informam o tipo de rede neural utilizada.

Zandi *et al.* (2010) implementam um alarme por canal de EEG, onde o sinal obtido é decomposto através de transformada wavelet discreta de pacote. São analisadas a função densidade de probabilidade (*Probability Density Function* - PDF), a energia de cada decomposição e a frequência do sinal. Comentam uma sensibilidade de 90%, porém não comentam qual wavelet-mãe foi utilizada e nem quantas decomposições foram utilizados.

Vavadi, Ayatollahi e Mirzaei (2010) decompõem o sinal de EEG disponível em Andrzejak *et al.* (2001) utilizando a wavelet-mãe Daubechies 3 nas sub-bandas  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  e  $\gamma$ , utilizando a entropia aproximada, verificam estatisticamente que é possível separar épocas de sinais contendo eventos epileptiformes de épocas de sinal de EEG sem eventos epileptiformes com uma eficiência acima de 99%, mas não implementam nenhum classificador.

Subasi e Gursoy (2010) apresentam um estudo comparativo de técnicas de redução de dimensionalidade aplicadas nas características extraídas de sinais de EEG. A extração de características é feita através das decomposições D3, D4, D5 e A5 da WT DB4 com os descritores média dos valores absolutos dos coeficientes de cada decomposição, energia média, desvio padrão e a relação de valores médios entre as decomposições adjacentes. Essas características são reduzidas através de análise de componentes principais (*Principal Component Analysis* - PCA), análise de componentes independentes (*Independent Component Analysis* - ICA) e LDA. A base de dados é obtida de Andrzejak *et al.* (2001). Ao final da redução de características, é aplicado um classificador SVM.

Veneri *et al.* (2011) utilizam a entropia das bandas de transformada wavelet com Coiflet e Daubechies 1 para retirar o sinal de espículas do sinal de EEG. Os autores conseguiram filtrar 90% das espículas através do ceifamento de coeficientes que excedem o limiar estabelecido por eles.

Song e Lio (2011) comparam a acurácia na detecção de espículas em sinais de EEG do sistema deles utilizando a entropia amostrada (*Sample Entropy* - SampEn) e *Extreme Learning Machine* - (ELM) com outros dois artigos utilizando wavelet/fuzzy e wavelet/ANN. Obtém uma acurácia de 96,18% contra 94,83% da técnica wavelet/ANN e 85,90% da técnica wavelet/fuzzy. Ou seja, uma sensibilidade de 97,3% e especificidade de 99,01%. As épocas de sinal de EEG utilizadas são as mesmas adotadas por Ocak (2009).

Mirzaei, Ayatollahi e Vavadi (2011) aplicam a DWT com Daubechies 5 com 4 níveis

para separar as regiões  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  e  $\gamma$  de sinais de EEG, e determinam um histograma de cada subnível e verificam que há diferenças no histograma de pessoas saudáveis e doentes. Aplicam a média e o desvio padrão nesses histogramas e mostram que é possível diferenciar sinais de EEG contendo eventos epileptiformes de sinais de EEG normais com uma eficiência acima de 99,95%, mas não implementam nenhum classificador.

Gandhi, Panigrahi e Anand (2011) comentam que a maior parte dos estudos analisa somente a família de wavelets-mães Daubechies entre duas e quatro decomposições obtendo uma acurácia entre 70% e 90%. Eles propõem trabalhar com realizações de espículas amostradas em 400 Hz. Os autores utilizam 300 épocas de cinco segundos de duração contendo realizações de espículas e 300 épocas de sinal de EEG com duração de cinco segundos contendo EEG de fundo. O estudo analisa as famílias de wavelets-mães Daubechies, Haar, Coiflets e Biortogonal com seis decomposições. O estudo analisa 18 descritores para extração de características das decomposições wavelets. Como classificador utilizam uma rede neural probabilística (PNN). Concluem que a melhor wavelet, considerando acurácia e carga computacional para o cálculo, é a Coiflet de primeira ordem (Coif1). Os melhores descritores são a entropia, a energia e o desvio padrão. O estudo consegue uma sensibilidade e especificidade iguais a 99,3%. Além disso, concluem que a seleção de uma wavelet através de parametrização infere resultados melhores do que uma seleção aleatória.

Balasubramanian e Obeid (2011) comparam um método de classificação de espículas, obtidas de sinais de EEG, dito como padrão ouro baseado em análise dos componentes principais (*Principal Component Analysis* - PCA) (dentro de certas condições) com outro método *online* de classificação baseado em lógica fuzzy. Os resultados são os mesmos obtidos com o PCA com a vantagem de que as espículas podem estar defasadas no tempo e não precisam ser pré-selecionadas.

Birot *et al.* (2013) propõem que existem características em sinais de EEG gravados de pacientes epiléticos na faixa de 256 Hz a 600 Hz que são significativos na detecção de eventos. Essas características são nomeados de ondas rápidas (*fast-ripples*). Não considerando as espículas, e sim somente as características denominadas de ondas rápidas extraídas da transformada wavelet com DB4 na primeira decomposição, os autores conseguem através da curva ROC um índice de desempenho AUC de 0,98.

Halford *et al.* (2013) apresentam um trabalho significativo sobre o estado da arte das abordagens feitas sobre a detecção de espículas. Os autores fazem um estudo detalhado na base de dados denominada EEGnet (WATERS; DEAN; HAL, 2013) e com a ajuda de 11 especialistas em exames de EEG calculam a especificidade e a sensibilidade dos especialistas e das técnicas

mais comumente utilizadas pelos pesquisadores na análise da epilepsia. São analisados 9319 eventos e catalogados na base de dados. Um objetivo é verificar se a base de dados utilizada é grande o suficiente para permitir a treinamento de um algoritmo que vise identificar eventos epileptiformes. Outros objetivos foram especificados neste trabalho para verificar qual método identifica melhor a patologia, dentre eles: (1) uma abordagem por descritores morfológicos; (2) uma abordagem no do domínio da frequência; e (3) uma abordagem pela transformada wavelet. Concluem que a base de dados ainda não é grande o suficiente e que entre as abordagens escolhidas, a transformada wavelet foi a que apresentou o melhor desempenho. Além disso, verificam a sensibilidade, a especificidade e taxa de falso-positivo por minuto entre os 11 neurologistas que selecionaram e identificaram os eventos epileptiformes. O resultado é apresentado na tabela 10. Cada neurologista levou em torno de oito horas para analisar e classificar os eventos. Levanta-se a questão de que os próprios neurologistas não estão de acordo com o que seria realmente um evento epileptiforme ou o que definiria um evento como sendo epileptiforme ou não. Ainda é levantada a questão que o neurologista que pré-selecionou os eventos para a partir daí calcular a sensibilidade e a especificidade dos neurologistas também não estivesse de acordo com os outros neurologistas.

**Tabela 10:** Estudo de Halford *et al.* (2013), com sensibilidade, especificidade e falsos-positivos por minuto entre 11 neurologistas na análise e seleção de 9319 eventos epileptiformes.

**Fonte: (HALFORD *et al.*, 2013)**

Especialista	1	2	3	4	5	6	7	8	9	10	11
Sensibilidade	0,93	0,76	0,43	0,90	0,55	0,68	0,49	0,89	0,22	0,82	0,76
Especificidade	0,97	0,98	0,99	0,93	1,00	0,99	0,98	0,98	0,99	0,97	0,98
FP por Minuto	1,58	0,98	0,34	3,40	0,20	0,36	1,04	1,20	0,34	1,28	1,10

Acharya *et al.* (2013) fazem uma revisão de métodos de extração de características e métodos de classificação encontrados na literatura. Os métodos de extração de características revisados são divididos em: (1) domínio do tempo: predição linear, PCA, ICA e LDA; (2) domínio da frequência: métodos paramétrico e não paramétrico; (3) domínio tempo-frequência: WT e transformada Hilbert-Huang (*Hilbert-Huang Transform - HHT*); e (4) análises não lineares de dados: espectro de ordem superior (*Higher Order Spectra - HOS*), cumulante de ordem superior (*Higher Order Cumulants*), RP (*Recurrence Plot*), RQA (*Recurrence Quantification Analysis*), ApEn, SampEn, FD (*Fractal Dimension*), CD (*Correlation Dimension*), H (*Hurst Exponent*) e LLE (*Largest Lyapunov Exponent*). A conclusão é que as análises não lineares são as que produzem os melhores resultados em termos de sensibilidade e especificidade, porém, a base de dados utilizada Andrzejak *et al.* (2001), pode ser considerada de fácil detecção e não abrange a aleatoriedade dos eventos epileptiformes.

Kumar, Dewal e Anand (2014) utilizaram a base Andrzejak *et al.* (2001). Não é especificado qual wavelet-mãe é utilizada nesse estudo, porém os sinais da base de dados são decompostos em cinco níveis e é aplicado a entropia aproximada tanto nos níveis de detalhes quanto nos níveis de aproximação da WT. Obtiveram sensibilidade e especificidade de 100% nos eventos classificados como espícula e não-espícula.

Gajic *et al.* (2014) utilizam a base de dados Andrzejak *et al.* (2001), onde realizam cinco decomposições desses sinais utilizando o método WP (*Wavelet Packet*) e DB4. São extraídas das decomposições de detalhes e de aproximação: a energia, a entropia e o desvio padrão. É utilizado o PCA para reduzir o número de características para três dimensões. Após a redução de características são aplicados dois classificadores quadráticos, onde o primeiro identifica o EEG normal, e dos resultados não identificados pelo primeiro classificador como EEG normal, o segundo classificador classifica entre eventos epilépticos e não epilépticos. É obtida uma sensibilidade de 98% e especificidade de 99,99%.

Nunes *et al.* (2014) também utilizam a base Andrzejak *et al.* (2001). Comparam as wavelets Haar, Coiflet 2 a 4, Symlet 2 a 4, Daubechies 2 a 4, todas com cinco decomposições. Como descritores foram extraídos: o valor mínimo, o valor máximo, o valor médio, o desvio padrão, a potência, a média absoluta e a variância, totalizando 40 dimensões. É utilizado PCA para redução de dimensões. São comparados quatro tipos de classificadores: SVM, classificador bayesiano, rede neural artificial MLP e OPF (*Optimum Path Forest*). Concluem que o OPF em conjunto com a wavelet-mãe Coiflet 4 apresenta a melhor acurácia, igual a 89,2%.

Correa *et al.* (2015) utilizam EEG intracraniano de 21 pacientes, e analisando 196 épocas de uma hora de duração, conseguem sensibilidade e especificidade iguais a 85,4% e 83,17% respectivamente. As épocas foram amostradas a 256 Hz e decompostas por DB2 em cinco decomposições, mas somente as decomposições três, quatro e cinco foram utilizadas pois representam as ondas  $\beta$ ,  $\alpha$  e  $\theta$ . Os autores calculam a energia de cada decomposição e definem limiares de detecção para cada nível de decomposição. A metodologia é bastante semelhante com os estudos de Ocak (2009) e Wang *et al.* (2009), que ao invés da energia calculam a entropia aproximada. Comentam que os valores obtidos são melhores do que os encontrados na literatura pois abrangem um número maior de voluntários do que outros estudos e salientam que o tempo de processamento é um fator que deve ser considerado no desenvolvimento de algoritmos.

Foram apresentados em ordem cronológica os principais trabalhos que relatam desenvolvimentos com a transformada wavelet e a detecção de espículas. Os resultados são comentados agora com relação a abordagens específicas.



## 3.2 ABORDAGENS DO ESTADO DA ARTE

Esta seção discute os trabalhos apresentados na seção 3.1 em relação a tópicos específicos: a espícula e o registro do sinal, wavelets-mães na detecção de espícula, extração de características da transformada wavelet, classificadores utilizados e resultados obtidos na detecção de espículas.

### 3.2.1 ESTADO DA ARTE - A ESPÍCULA E O REGISTRO DO SINAL

Estudos recentes contendo a transformada wavelet comentam sobre a variabilidade da duração do sinal da espícula de 20 ms a 200 ms como Birot *et al.* (2013) e Halford *et al.* (2013), mas não correlacionam a transformada wavelet com essa variabilidade.

Quanto à frequência de amostragem do sinal, a maior parte dos estudos é realizada com exames de EEG abaixo de 200 Hz (KALAYCI; OZDAMAR; ERDOL, 1994; KALAYCI; OZDAMAR, 1995; YAMAGUCHI, 2003; SONG; LIO, 2011; ACHARYA *et al.*, 2013; GAJIC *et al.*, 2014; NUNES *et al.*, 2014).

Muitos estudos utilizam a base de dados de Andrzejak, Schindler e Rummel (2012b), amostrada em 173,5 Hz.

Dois trabalhos foram encontrados com frequências de amostragem elevadas como 400 Hz (GANDHI; PANIGRAHI; ANAND, 2011) e 600 Hz (BIROT *et al.*, 2013). Esse último estudo comenta que há características no sinal de EEG, denominadas como oscilações de alta frequência, que podem ser relevantes para a detecção de espículas na faixa de 250 a 600 Hz, porém, ainda não estudadas.

### 3.2.2 ESTADO DA ARTE - WAVELETS-MÃES NA DETECÇÃO DA ESPÍCULA

Não há uma concordância entre os autores sobre qual família de wavelets é a mais indicada para a detecção de espículas.

Existe uma preferência dos autores para a família Daubechies (KALAYCI; OZDAMAR; ERDOL, 1994; KALAYCI; OZDAMAR, 1995; EHRENTREICH; SUEMMCHEN, 2001; YAMAGUCHI, 2003; ERCELEBI; SUBASI, 2006; SADATI; MOHSENI; MAGHSOUDI, 2006; INDIRADEVI *et al.*, 2008; HSU; YU, 2009; MIRZAEI; AYATOLLAHI; VAVADI, 2011).

Outras famílias também são referenciadas como significativas, entre elas as wavelets-

mães Biortogonal, Symlet, Coiflet e até mesmo a Haar (TEIXEIRA, 2001; XU; SONG, 2008; ERCELEBI; SUBASI, 2006; RAMANAN; KALPAKAM; SAHAMBI, 2004).

Ubeyli (2009) comenta não ser possível afirmar que uma família é melhor do que outra e é necessário avaliar experimentalmente qual família de wavelets gera resultados melhores para um determinado tipo de sinal.

Gandhi, Panigrahi e Anand (2011) comentam que a escolha de uma wavelet-mãe através de parametrização infere resultados melhores do que uma seleção aleatória.

### 3.2.3 ESTADO DA ARTE - EXTRAÇÃO DE CARACTERÍSTICAS DA TRANSFORMADA WAVELET

Para utilizar a transformada wavelet como entrada para algum classificador, os coeficientes gerados pelas WT são interpretados através de descritores. Os descritores são calculados a partir de cada sub-banda da transformada wavelet ou de apenas algumas sub-bandas selecionadas. São encontrados na literatura descritores como: valor máximo, valor mínimo, média, variância (GEVA; KEREM, 1998; SADATI; MOHSENI; MAGHSOUDI, 2006; UBEYLI, 2009; GANDHI; PANIGRAHI; ANAND, 2011), ICF (BLANCO *et al.*, 1998), PDF (ZANDI *et al.*, 2010), análises de entropia dos coeficientes como ApEn, SampEn, *Kurtosis*, Reyni e bandas (WANG *et al.*, 2009; VAVADI; AYATOLLAHI; MIRZAEI, 2010; SONG; LIO, 2011; INUSO *et al.*, 2007; VENERI *et al.*, 2011; GANDHI; PANIGRAHI; ANAND, 2011; ACHARYA *et al.*, 2013) e análise estatística como LDA (XU; SONG, 2008; SUBASI; GURSOY, 2010; ACHARYA *et al.*, 2013) e histogramas (MIRZAEI; AYATOLLAHI; VAVADI, 2011).

Gandhi, Panigrahi e Anand (2011) comentam que a seleção de descritores conforme a característica do sinal estudado influencia de forma significativa o desempenho de um classificador.

### 3.2.4 ESTADO DA ARTE - CLASSIFICADORES

Uma vez definida a taxa de amostragem dos sinais, a família de wavelets-mães, a quantidade de níveis de decomposição da WT e características extraídas dessa seleção, as técnicas de classificação são as mais variadas indo de classificadores de limiar (TEIXEIRA, 2001; INDIRADEVI *et al.*, 2008; YU, 2009; WANG *et al.*, 2009; OCAK, 2009; CORREA *et al.*, 2015), à implementação de algum tipo de rede neural artificial (KALAYCI; OZDAMAR; ERDOL, 1994; KALAYCI; OZDAMAR, 1995; UBEYLI, 2009; SHUREN; ZHONG, 2009;

SONG; LIO, 2011; GANDHI; PANIGRAHI; ANAND, 2011; NUNES *et al.*, 2014), o uso de lógica fuzzy (GEVA; KEREM, 1998; SADATI; MOHSENI; MAGHSOUDI, 2006; SONG; LIO, 2011) e outras técnicas de classificação como *K-Means* (NENADIC; BURDICK, 2005), ELM (SONG; LIO, 2011), SVM (HSU; YU, 2009; NUNES *et al.*, 2014) e PCA (BALASUBRAMANIAN; OBEID, 2011).

### 3.2.5 ESTADO DA ARTE - RESULTADOS OBTIDOS NA DETECÇÃO DE ESPÍCULAS

Wilson e Emerson (2002) definem que a especificidade, a sensibilidade e a curva ROC são as características comparáveis entre técnicas para determinar se o desempenho de um método é melhor do que outro. Desse modo, os melhores resultados encontrados na literatura para sensibilidade e especificidade são de Gandhi, Panigrahi e Anand (2011), com sensibilidade e especificidade iguais a 99,3% seguido pelo estudo de Song e Lio (2011) com uma sensibilidade de 97,3% e especificidade de 99,01%. Considerando estudos que contemplam a curva ROC, o melhor resultado do índice de desempenho AUC é de 0,9860 (BIROT *et al.*, 2013).

Halford *et al.* (2013) descrevem que não há concordância entre os próprios especialistas, como apresentado na tabela 10. Comentam ainda que, mesmo utilizando uma base de dados de 9319 eventos epileptiformes, não é suficiente para verificar o funcionamento de um classificador.

## 3.3 COMENTÁRIOS

Analisando os trabalhos publicados dos últimos 20 anos observa-se que, com o passar do tempo, não há melhorias significativas na área de pesquisa da detecção de espículas. Percebe-se que há o uso abrangente de técnicas de redes neurais e que são implementados novos descritores na tentativa de obter melhorias na detecção, porém, cada trabalho apresenta resultados diferentes, principalmente pelo uso de bases de dados diferentes. Não existe um padrão ouro para a espícula e sua variabilidade muda de paciente para paciente, dificultando avanços na implementação de um algoritmo de fato para a detecção de espículas.

Os índices de desempenho de sensibilidade e especificidade encontrados variam de 83,5% a 99,3% para as diversas técnicas apresentadas, porém, como apresentado por Wilson e Emerson (2002), representam apenas um valor para o limiar de detecção do classificador, assim como os resultados derivam de bases com frequências de amostragem diferentes e número de amostrais diferentes.

O próximo capítulo apresenta os materiais e ferramentas utilizados para o desenvolvimento da pesquisa como a obtenção da base de sinais, a proposição da arquitetura do classificador de espículas e a geração da matriz de características extraída da base de sinais para o processamento da análise de discriminante linear.

## 4 MATERIAIS E FERRAMENTAS

Neste capítulo serão abordados os materiais e as ferramentas utilizadas para o desenvolvimento da pesquisa. Primeiramente é comentado sobre o banco de dados de sinais de EEG utilizado e sua obtenção. A seguir, é apresentada a estrutura do classificador de espículas. Da mesma forma, são apresentados os descritores de sinal utilizados para esse detector de espículas, onde foi desenvolvido um descritor denominado energia centrada. Na sequência do capítulo é apresentada a análise de discriminante linear e como é montada a matriz de características necessária para o processamento de discriminante linear.

### 4.1 BANCO DE SINAIS DE EEG

Esse estudo foi aprovado pelo comitê de ética do Hospital Pequeno Príncipe, em Curitiba, no estado do Paraná, onde os sinais de EEG foram obtidos.

Os sinais foram gravados durante um exame de EEG com duração de 24 horas de um paciente adolescente, do sexo feminino, diagnosticado com epilepsia focal sintomática. O exame foi realizado utilizando EEG de escalpo, na montagem 10-20. O paciente não estava fazendo uso de medicamentos e não estava sedado. Os sinais foram amostrados a uma frequência de 1000 Hz.

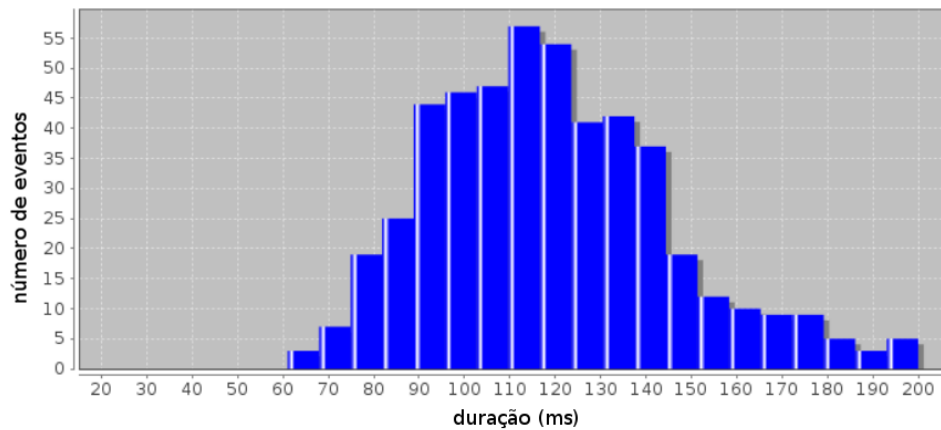
Os sinais foram obtidos sem nenhum tratamento de sinal feito pelo equipamento de EEG. Para esses sinais, foram aplicados filtros notch de 60 Hz e suas harmônicas, além de um filtro passa-baixas com frequência de corte em 350 Hz. O filtro utilizado foi do tipo FIR, com 21 coeficientes e janelamento por Blackman. Esse filtro é equivalente a um filtro passa-baixas Butterworth de segunda ordem com atenuação de -40 dB/década.

O neurologista marcou 494 espículas com duração entre 60 e 200 ms (ondas agudas). Esses eventos, classificados como eventos epileptiformes ou eventos espícula, foram separados do exame original em épocas de 2 segundos, com a ponta do evento centrado na época. Adicionalmente, foram extraídos aleatoriamente do mesmo exame 1500 épocas contendo EEG

de fundo, sendo estas classificadas como eventos não epileptiformes ou eventos não-espícula. As épocas de eventos não epileptiformes foram extraídos de canais diferentes do exame de EEG contendo diferentes padrões de EEG, como fusos de sono, artefatos, etc. que não deveriam ser reconhecidos como eventos espícula.

As épocas extraídas do exame foram selecionadas no período em que o paciente estava em sono.

A distribuição dos eventos espícula pode ser visto figura 18. Observa-se que há uma incidência maior de espículas com duração entre 90 e 140 milissegundos de duração. Porém, há incidência de espículas variando de 60 a 200 milissegundos.

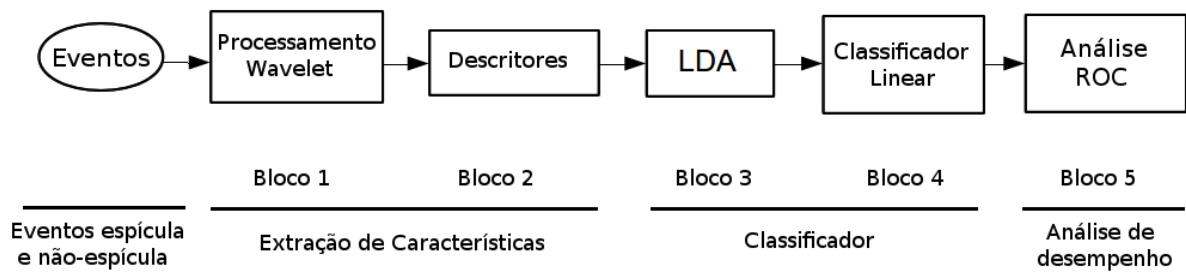


**Figura 18:** Histograma de distribuição dos eventos espícula da base de sinais.

#### 4.2 CLASSIFICADOR DE EVENTOS EPILEPTIFORMES

É proposto um detector de espículas cujo diagrama em blocos é apresentado pela figura 19.

Os eventos são processados por uma combinação de wavelets-mães no bloco 1. No bloco 2 são calculados descritores para o resultado das decomposições wavelet executadas no bloco 1. O resultado do bloco 2 é uma matriz multidimensional de características que será processada por LDA no bloco 3. O LDA rotaciona a matriz de características em um único passo, determinando uma projeção unidimensional onde as características discriminantes mantém a melhor separabilidade entre as classes. No bloco 4, as distribuições unidimensionais resultantes são separadas por um classificador linear. Após a classificação, os resultados são analisados através da AUC de uma curva ROC no bloco 5, de modo a comparar diferenças de desempenho entre configurações diferentes de combinação de wavelets, níveis de decomposição e descritores.



**Figura 19:** O diagrama em blocos do classificador proposto é dividido em 5 etapas principais: (1) decomposição do sinal de EEG pela WT; (2) extração de características por descritores; (3) processamento LDA; (4) seleção entre evento epileptiforme e evento não epileptiforme por um classificador linear; e (5) cálculo de desempenho do classificador através da análise ROC.

### 4.3 DESCRITORES

Os descritores são utilizados para converter dados de um pré-processamento em informações computáveis para um classificador na fase de pós-processamento. Como descritores comumente encontrados na literatura tem-se: o valor máximo, o valor mínimo, o desvio padrão e a energia. Os descritores utilizados para esse estudo, assim como as abreviações utilizadas para eles, são apresentados na tabela 11. Os descritores entropia aproximada e energia centrada do vetor são descritos nas seções 4.3.1 e 4.3.2 respectivamente.

**Tabela 11:** Descritores estatísticos utilizados.

abreviação	nome
ApEn	Entropia Aproximada
<i>avg_pow</i>	Potência Média
<i>centred</i>	Energia Centrada do Vetor
<i>energy</i>	Energia
<i>rms</i>	Valor RMS
<i>range</i>	Valor Máximo - Valor Mínimo
<i>max</i>	Valor Máximo
<i>min</i>	Valor Mínimo
<i>mean</i>	Média
<i>dev</i>	Desvio Padrão
<i>var</i>	Variância
<i>median</i>	Mediana
<i>mode</i>	Moda
<i>norm</i>	Norma

Para exemplificar cada descritor da tabela 11, tome-se como base o vetor  $V$  descrito pela equação 39:

$$V = [1 \ 2 \ 1 \ 3 \ 3 \ 5 \ 1 \ 7 \ 4 \ 2 \ 5 \ 7 \ 2 \ 2 \ 2 \ 15] \quad (39)$$

Os descritores *max* e *min*, retornam os valores 15 e 1 respectivamente, sendo os valores máximo e mínimo do vetor  $V$ .

O descritor *range* por sua vez é a subtração dos resultados dos descritores *max* e *min*, retornando  $range(V) = 14$ .

Os descritores *mean*, *var* e *dev* são apresentados pelas equações 40, 41 e 42 e representam o valor médio, a variância e o desvio padrão.

$$mean(V) = \bar{V} = \frac{\sum_{i=1}^n V_i}{n} = 3,88 \quad (40)$$

$$var(V) = \frac{\sum_{i=1}^n (V_i - \bar{V})^2}{n - 1} = 12,65 \quad (41)$$

$$dev(V) = \sqrt{var(V)} = 3,56 \quad (42)$$

Onde  $n$  é o tamanho do vetor  $V$ .

Os descritores *energy*, *avg\_pow* e *rms* calculam a energia, a potência média e a raiz do valor quadrático médio, sendo apresentados pelas equações 43, 44 e 45.

$$energy(V) = \sum_{i=1}^n V_i^2 = 430 \quad (43)$$

$$avg\_pow(V) = \frac{energy(V)}{n} = 26,88 \quad (44)$$

$$rms(V) = \sqrt{avg\_pow(V)} = 5,18 \quad (45)$$

O descritor *norm*, calcula a norma euclidiana, apresentada pela equação 46.

$$norm(V) = \sqrt{\sum_{i=1}^n V_i^2} = 20,74 \quad (46)$$

Para o cálculo dos descritores *median* e *mode*, ordena-se o vetor  $V$  em ordem crescente, resultando o vetor  $V_{ord}$  como apresentado pela equação 47.



$$V_{ord} = [1 \ 1 \ 1 \ 2 \ 2 \ 2 \ 2 \ 2 \ 3 \ 3 \ 4 \ 5 \ 5 \ 7 \ 7 \ 15] \quad (47)$$

A partir do vetor  $V_{ord}$ , o descritor *mode* retorna o valor que possui a maior frequência de aparição no vetor  $V$ , neste caso,  $mode(V) = 2$ . Um vetor pode ter mais de um valor resultante para o descritor *mode*, e para esse estudo, o resultado do descritor será a moda com menor valor.

O descritor *median*, caso o tamanho do vetor seja ímpar, retorna o valor central do vetor. Caso o tamanho do vetor seja par, o descritor retorna a média dos valores centrais do vetor ordenado. No caso do vetor de exemplo ordenado apresentado na equação 47, o retorno do descritor *median* é igual a 2,5.

#### 4.3.1 DESCRITOR ENTROPIA APROXIMADA

A entropia aproximada (ApEn) é um método matemático apresentado por Pincus (1990) que determina a entropia de um vetor de amostras. Quanto mais baixo o valor resultante, menor é a entropia e menos caótico é o comportamento das amostras. Em contrapartida, quanto maior o valor da ApEn, menos predizível é o comportamento do sinal. A fórmula para o cálculo da entropia aproximada é apresentada pela equação 48.

$$ApEn(S_n, m, r) = \ln \left( \frac{C_m(r)}{C_{m+1}} \right) \quad (48)$$

Onde  $S_n$  é o vetor de sinal para o qual a entropia aproximada será calculada,  $m$  é o comprimento do padrão repetitivo a ser considerado internamente em  $S_n$ , e  $r$ , o critério de diferença entre qualquer par do padrão de repetição e os pares em comparação. Como mostrado pelas equações 49, 50, 51 e 52,  $C_m(r)$  na equação 48 é a média de padrões repetitivos encontrados internamente em  $S_n$  que obedecem ao critério de diferença  $r$ .  $N$  é o tamanho de  $S_n$ .

$$C_m = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} C_i^m(r) \quad (49)$$

$$C_i^m(r) = \frac{1}{N - m + 1} \sum d[x(i), x(j)] < r \quad (50)$$

$$d[x(i), x(j)] = \max |x(i) - x(j)| \quad (51)$$

onde  $d$  representa a distância euclidiana entre os vetores  $x(i)$  e  $x(j)$ .

$$x(i) = [S(i), S(i+1), \dots, S(i+m-1)] \quad (52)$$

onde  $x(i)$  é o padrão de repetição construído com as amostras do vetor  $S_n$ , com tamanho total  $m$  e  $i = 1, 2, \dots, N - m + 1$ .

Para este estudo, o valor dois foi atribuído a  $m$ . A variável  $r$  é a variância das amostras de  $S_n$  multiplicado por 0,15. Os valores para  $m$  e  $r$  adotados são os valores sugeridos por Pincus (1990) para refletir a diferença de entropia entre os segmentos de comparação do vetor. As amostras do vetor  $S_n$  são os coeficientes de uma decomposição wavelet.

Exemplificando o uso do descritor ApEn, a sua aplicação  $ApEn(V)$  onde  $V$  é o vetor apresentado na equação 39, teremos como resultado o valor 0,06.

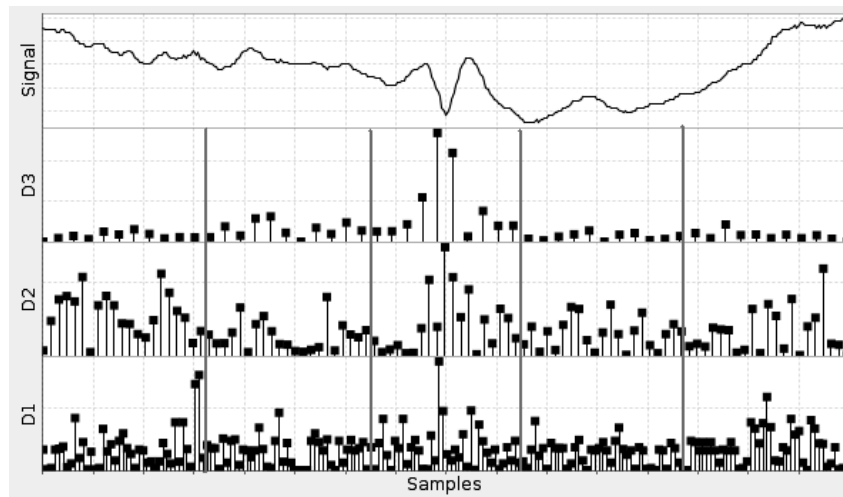
#### 4.3.2 DESCRITOR ENERGIA CENTRADA

A energia centrada foi desenvolvida para esse estudo. Considerando que um evento epileptiforme esteja centrado na janela amostral, o objetivo desse descritor é ressaltar coeficientes próximo ao evento comparativamente ao restante dos coeficientes obtidos da mesma janela amostral. Esse comportamento é interessante principalmente quando se espera encontrar eventos epileptiformes em uma janela deslizante sobre um canal do exame de EEG, onde o evento, a cada passo da janela, estará em uma posição diferente. Isso permite detectar o eventos quando este estiver centrado na janela.

A figura 20 apresenta uma espícula centrada em uma janela de amostras. D1, D2 e D3 são três níveis de coeficientes obtidos por WT utilizando Daubechies série 3. É possível observar coeficientes com valores maiores na mesma região do evento do que coeficientes distantes do evento. Para determinar a energia centrada, cada nível de decomposição é dividido em 5 grupos, denotados pelas linhas verticais visíveis na figura 20. É calculada dessa maneira a razão entre a energia do grupo central de uma decomposição e a energia total da decomposição.

A equação 53 determina a energia centrada de um vetor  $S_i$  com  $N$  amostras onde  $i = 1, 2, \dots, N$ .  $S_i$  é qualquer nível de decomposição de um sinal processado por uma WT.

$$centred = \frac{\sum_{i=2\frac{N}{5}+1}^{3\frac{N}{5}} S_i^2}{\sum_{i=1}^N S_i^2} \quad (53)$$



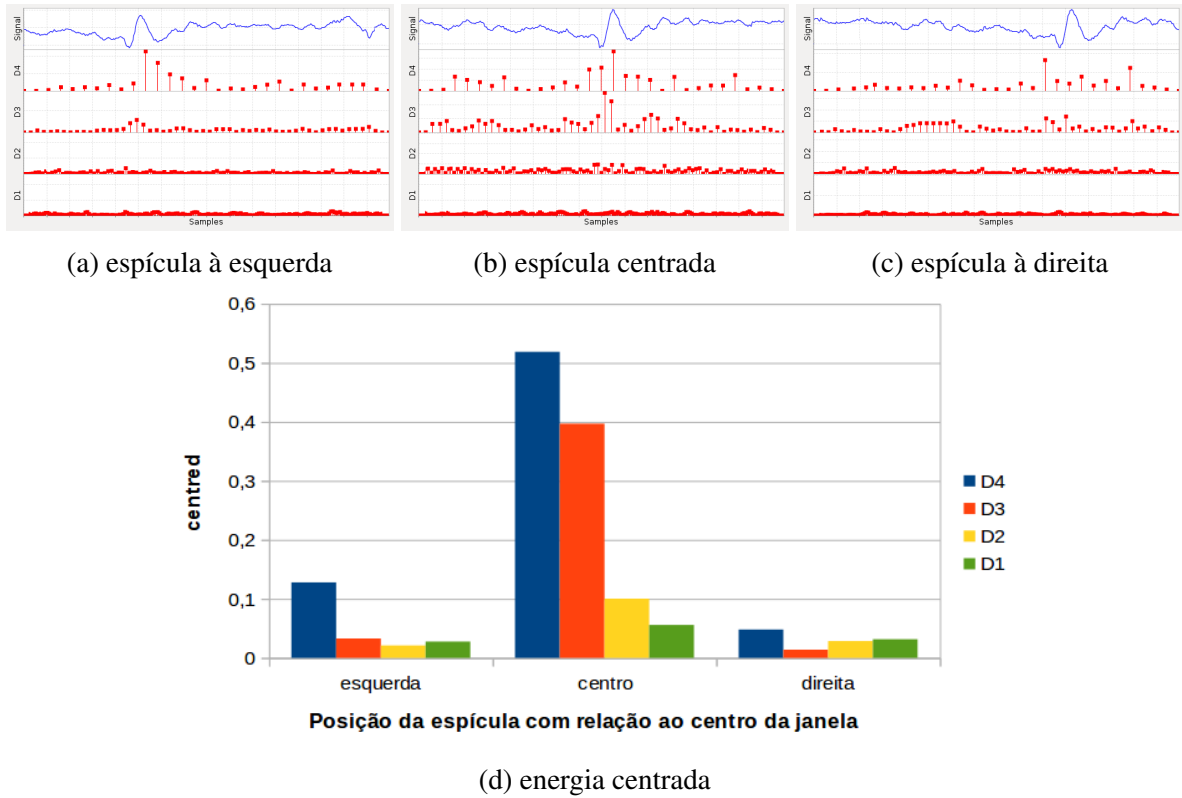
**Figura 20:** O sinal possui uma espícula centrada na janela de amostragem. D1, D2 e D3 são três níveis de decomposição de uma WT usando daubechies série três. Para calcular o descritor **energia centrada**, os coeficientes de um nível de decomposição são divididos em cinco grupos como mostrado pelas linhas verticais no gráfico.

Exemplificando o uso do descritor *centred*, a sua aplicação  $centred(V)$  onde  $V$  é o vetor apresentado na equação 39, tem-se como resultado o valor 0,16.

A ideia desse descritor é ressaltar eventos cujos coeficientes apresentam maior valor do que seus vizinhos, dessa maneira identificando eventos que possivelmente tenham sido identificados pela WT comparado com coeficientes resultantes da atividade de fundo do EEG.

A figura 21 exemplifica uma janela deslizante sobre um canal de exame de EEG onde a energia centrada foi aplicada sobre a mesma espícula, porém, à direita do centro (figura 21(c)), centrada na janela (figura 21(b)), e à esquerda do centro da janela deslizante (figura 21(a)). A figura 21(d) apresenta os valores da energia centrada para as diferentes posições da espícula com relação ao centro da janela para cada nível de decomposição da transformada wavelet. Para esse exemplo foi utilizada a wavelet-mãe Daubechies 4 com quatro níveis de decomposição.

É possível observar na figura 21(d) que os valores resultantes da energia centrada são maiores em valor para o caso onde a espícula está centrada na janela do que quando a espícula está à esquerda ou à direita do centro. Isso se repete para cada nível de decomposição wavelet.



**Figura 21:** Energia centrada para a mesma espícula para o caso de uma janela deslizante onde a espícula se encontra em (a) à esquerda do centro, (b) centrada na janela, e (c) à direita do centro. Em (d) são apresentados os valores resultantes para a energia centrada para cada posição da espícula. A transformada wavelet utilizada é a Daubechies 4 com quatro níveis de decomposição.

#### 4.4 MATRIZ DE CARACTERÍSTICAS

Para que o algoritmo LDA do classificador proposto no bloco 4 da figura 19 determine a projeção que maximize a separabilidade entre as classes, é necessário primeiramente construir a matriz de características  $\omega$  para cada classe  $X$ . Considerando uma classe de eventos  $EEG$  contendo  $N$  eventos  $S_n$  onde  $n = 1, 2, \dots, N$ . Cada evento é processado por uma transformada wavelet resultando a matriz  $WT$ . Fazendo  $WT = WT(EEG)$  ser uma matriz  $N \times d$  onde  $d$  é a  $d$ -ésima decomposição de uma transformada wavelet. A matriz  $WT$  é apresentada pela equação 54.

$$WT(EEG) = \begin{bmatrix} D_1(S_1) & D_2(S_1) & \cdots & D_d(S_1) \\ D_1(S_2) & D_2(S_2) & \cdots & D_d(S_2) \\ \vdots & \vdots & \ddots & \vdots \\ D_1(S_N) & D_2(S_N) & \cdots & D_d(S_N) \end{bmatrix} \quad (54)$$

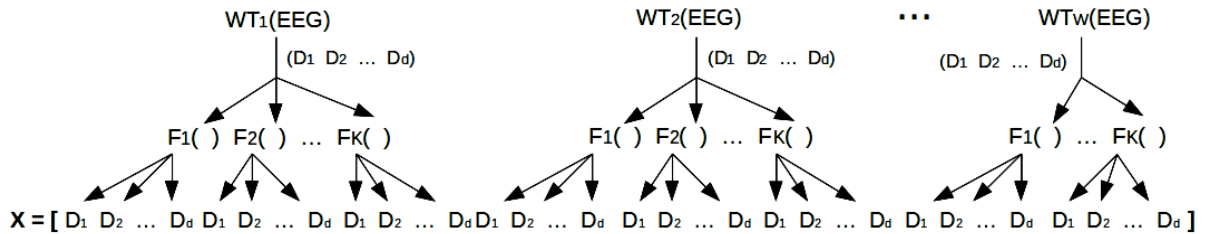
Após a transformada wavelet, um descritor com função  $F()$  é aplicado na matriz  $WT$ , resultando em uma matriz de características  $F = F(WT)$  com tamanho  $N \times d$ . Como  $K$  descritores diferentes podem ser calculados sobre o mesmo resultado de uma transformada wavelet  $WT$ , a matriz  $F$  torna-se  $F_k(WT)$ , onde  $k = 1, 2, \dots, K$ . Dessa maneira,  $F$  é uma matriz com tamanho  $N \times d.K$ , conforme apresenta a equação 55.

$$F = F_k(WT) = \left[ F_1(WT(EEG)) \quad F_2(WT(EEG)) \quad \dots \quad F_K(WT(EEG)) \right] \quad (55)$$

Seguindo a mesma lógica, para combinar  $W$  wavelets diferentes, necessita-se calcular a matriz  $F_k(WT_w)$  onde  $w = 1, 2, \dots, W$ . A matriz de características final  $X$  é apresentada na equação 56 com dimensão final igual a  $N \times d.K.W$ . Essa matriz de características  $X$  será calculada para cada classe de eventos  $EEG$  sob análise.

$$X = \left[ F_k(WT_1(EEG)) \quad F_k(WT_2(EEG)) \quad \dots \quad F_k(WT_W(EEG)) \right] \quad (56)$$

A figura 22 representa graficamente a construção da matriz  $X$ .



**Figura 22:** Um evento de EEG processado por  $W$  wavelets diferentes. Cada wavelet decompõe o evento  $EEG$  em  $d$  decomposições. As características são extraídas do processamento wavelet por  $K$  descritores. O número final de colunas (dimensões) é de  $W.d.K$ .

Considerando que as classes para esse estudo são formadas por eventos espícula e não-espícula, tem-se como resultado duas matrizes de características  $X_1$  e  $X_2$  compondo o conjunto de matrizes de características  $X_c$ .

#### 4.5 SOFTWARE

Dois programas diferentes foram desenvolvidos para esse estudo. O primeiro, desenvolvido em conjunto com o neurologista do Hospital Pequeno Príncipe, em Curitiba - PR, faz a leitura, filtragem e seleção dos eventos de EEG amostrados a 1.000 Hz pelo equipamento de eletroencefalografia do Hospital. Esses eventos formam o banco de sinais para esse estudo

apresentado na seção 4.1.

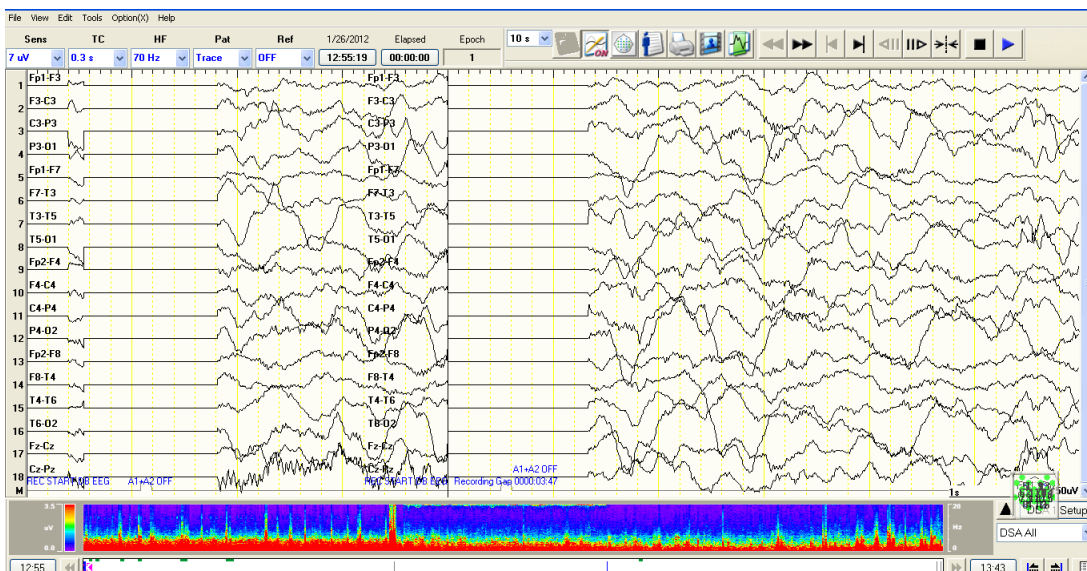
O segundo programa utiliza os sinais do banco de sinais para fazer as transformações wavelet, extrair as características, aplicar o algoritmo do LDA e obter os índices de desempenho do classificador linear.

Os programas desenvolvidos foram escritos em Java, usando a versão 1.7 do JDK (*Java Development Kit*), e são apresentados a seguir.

#### 4.5.1 PROGRAMA DE SELEÇÃO DE ESPÍCULAS

O equipamento de eletroencefalografia utilizado registra exames com frequência de amostragem até 1.000 Hz. Esta frequência de amostragem é necessária para verificar um dos objetivos desse trabalho, conforme descrito na seção 1.3. Esse aparelho gera arquivos armazenados no computador onde está instalado, sendo que esses arquivos são gravados em um formato proprietário com extensão ".EEG". Analisando o cabeçalho desses arquivos, a estrutura em que a informação estava armazenada no arquivo, e de posse de uma tela do programa de leitura proprietário do aparelho de EEG, foi possível criar o software necessário para traduzir os dados em formato binário do arquivo gerado pelo aparelho de eletroencefalografia em informação visual para o neurologista selecionar os eventos espícula.

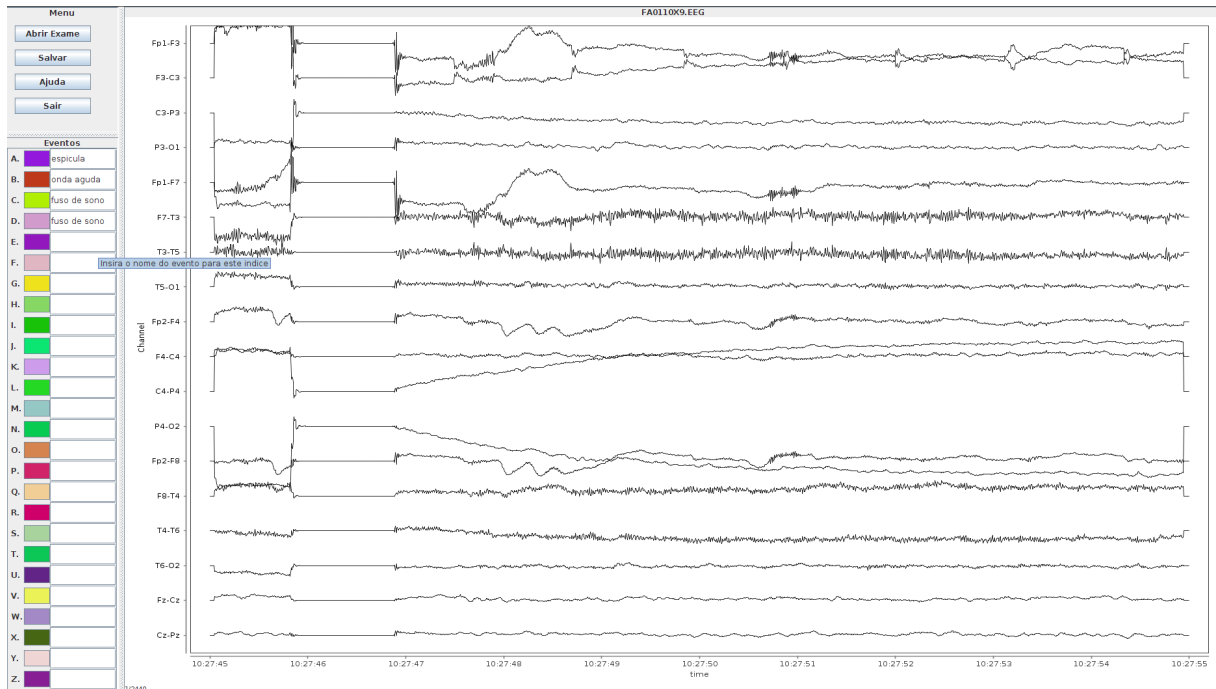
A tela do software proprietário fornecida pelo neurologista é apresentada pela figura 23



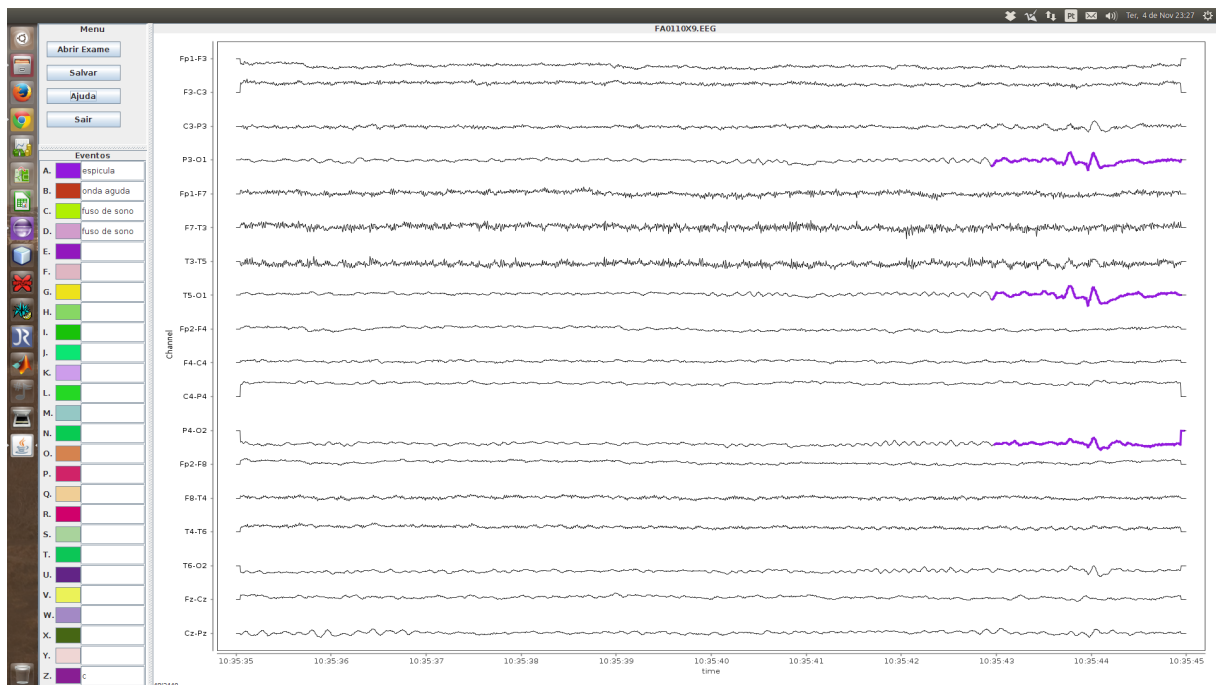
**Figura 23:** Tela do programa proprietário do equipamento de eletroencefalografia.

A tela do programa em Java para seleção e exportação de eventos de EEG é apresentada na figura 24. Na figura 25 foram selecionadas pelo neurologista sobre o exame de EEG três

espículas em canais diferentes do exame.



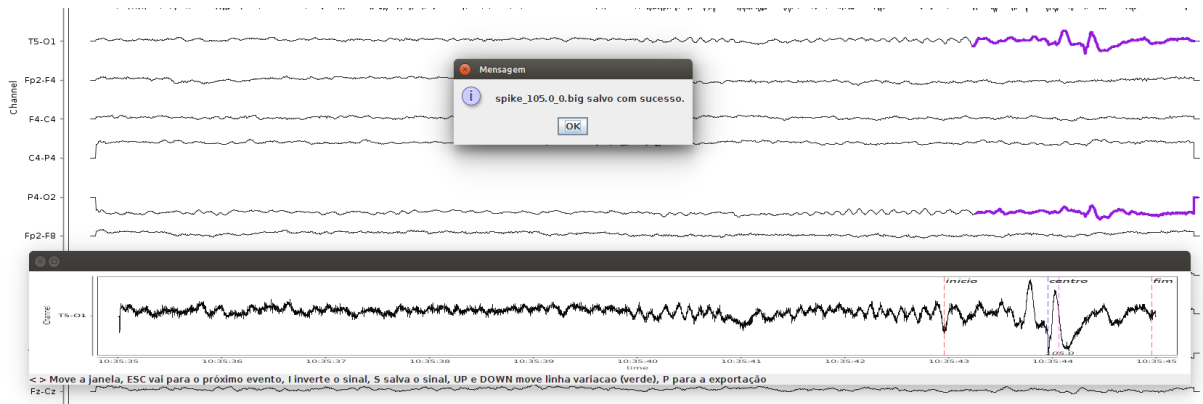
**Figura 24:** Tela do programa de seleção e exportação de eventos de EEG.



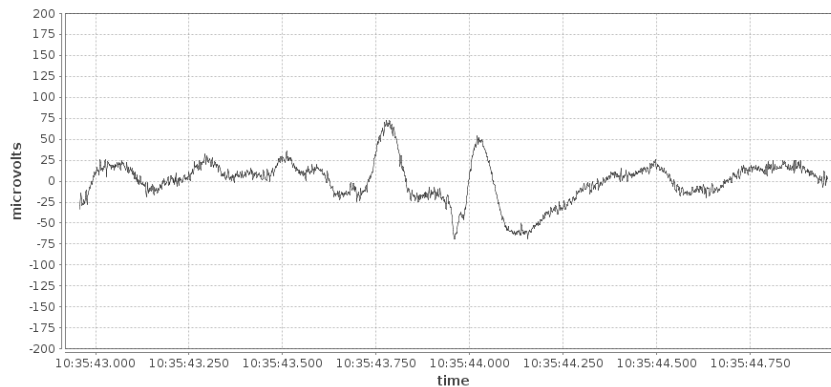
**Figura 25:** Três espículas selecionadas pelo neurologista.

Após a seleção dos eventos, a exportação é feita de forma semi-automática. O programa, definindo a duração em segundos da janela de amostra, captura os eventos um-a-um e pede ao usuário para definir onde ficará o centro da janela. A janela é deslocada para o

centro do evento selecionado pelo usuário e ocorre a exportação como apresentado na figura 26. O evento exportado pode ser visto na figura 27.



**Figura 26:** Janela de seleção e exportação de eventos selecionados. O programa abre em uma janela separada os eventos selecionados, um de cada vez, e permite centrar o evento na janela de amostra. Ao concluir a seleção e centragem do evento, o evento é salvo em arquivo para processamento posterior.



**Figura 27:** Espícula separada do sinal do exame original pelo programa desenvolvido em conjunto com o neurologista.

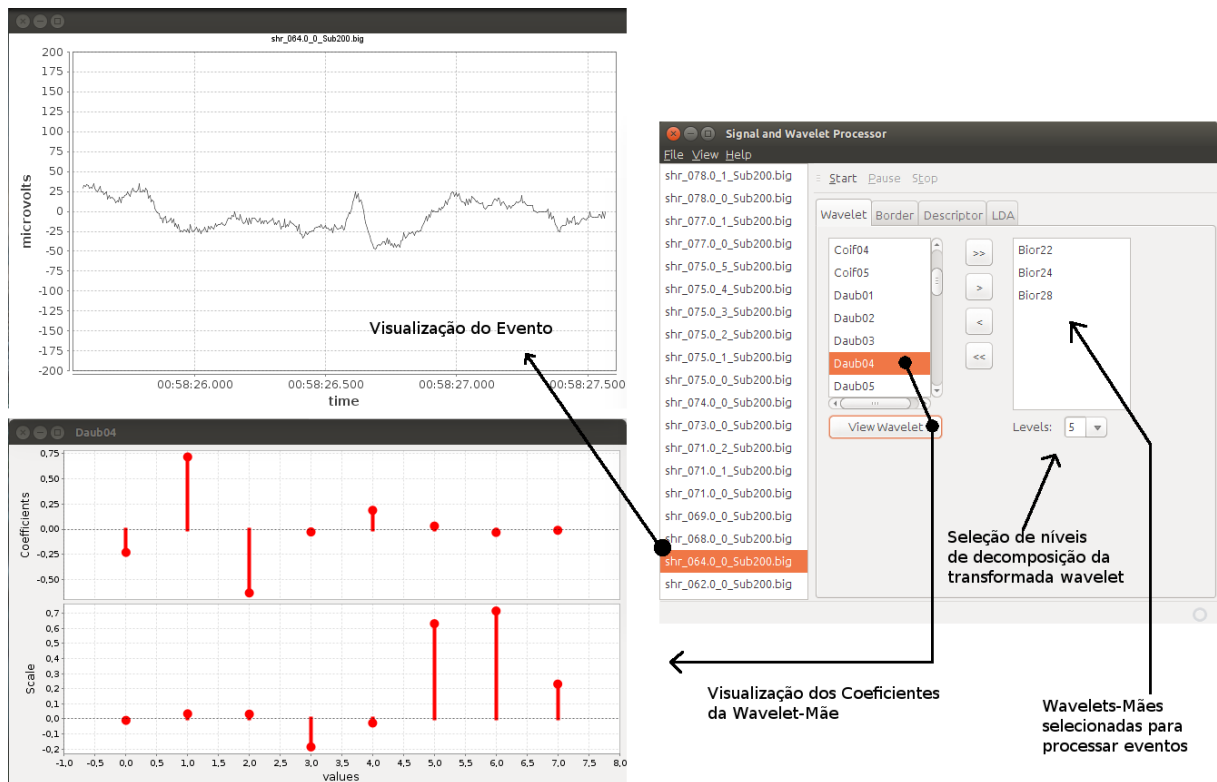
#### 4.5.2 PROGRAMA DE PROCESSAMENTO DE ESPÍCULAS

O programa de processamento de espículas foi desenvolvido para, a partir dos eventos selecionados, aplicar a transformada wavelet discreta de modo configurável. É possível selecionar quais eventos serão processados, com quais wavelets-mães e o número de decomposições da transformada wavelet. É ainda possível selecionar um ou mais métodos de extensão de bordas para ser aplicado durante a execução da transformada wavelet. Após o processamento da transformada wavelet o programa aplica um ou mais descritores aos coeficientes obtidos das transformações.



A figura 28 apresenta a tela principal do programa onde um evento foi selecionado e visualizado. O botão 'View Wavelet' abre uma nova janela onde é possível visualizar os coeficientes de função e escala utilizados pela wavelet-mãe selecionada. Nessa tela também é possível selecionar quais wavelets-mães serão utilizadas para o processamento wavelet dos sinais carregados no programa assim como os níveis de decomposição que serão calculados.

O botão 'Start' inicia os cálculos de transformação de modo *multi-thread*. *Multi-thread* significa que para cada sinal selecionado, será criado um processo adicional, além do programa principal, para executar a transformada wavelet.



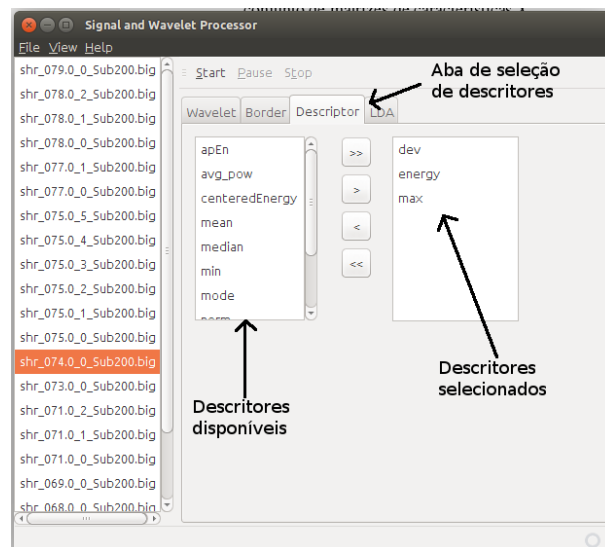
**Figura 28:** Tela principal do programa de processamento wavelet onde é visualizado um evento selecionado assim como os coeficientes da wavelet-mãe selecionada. É possível selecionar quais wavelets-mães e quantas decomposições serão utilizadas para o processamento dos eventos.

A figura 29 apresenta a aba de seleção de métodos de extensão de borda para a transformada wavelet. O programa implementa nove métodos, sendo possível selecionar um ou mais métodos para a transformação wavelet dos sinais carregados.

A figura 30 apresenta a aba de seleção de descritores. Após a transformação wavelet, os descritores selecionados serão aplicados aos coeficientes das decomposições wavelets calculadas para todas as wavelets-mães selecionadas na tela apresentada pela figura 28.



**Figura 29:** Aba de seleção de métodos de extensão de borda da transformada wavelet.



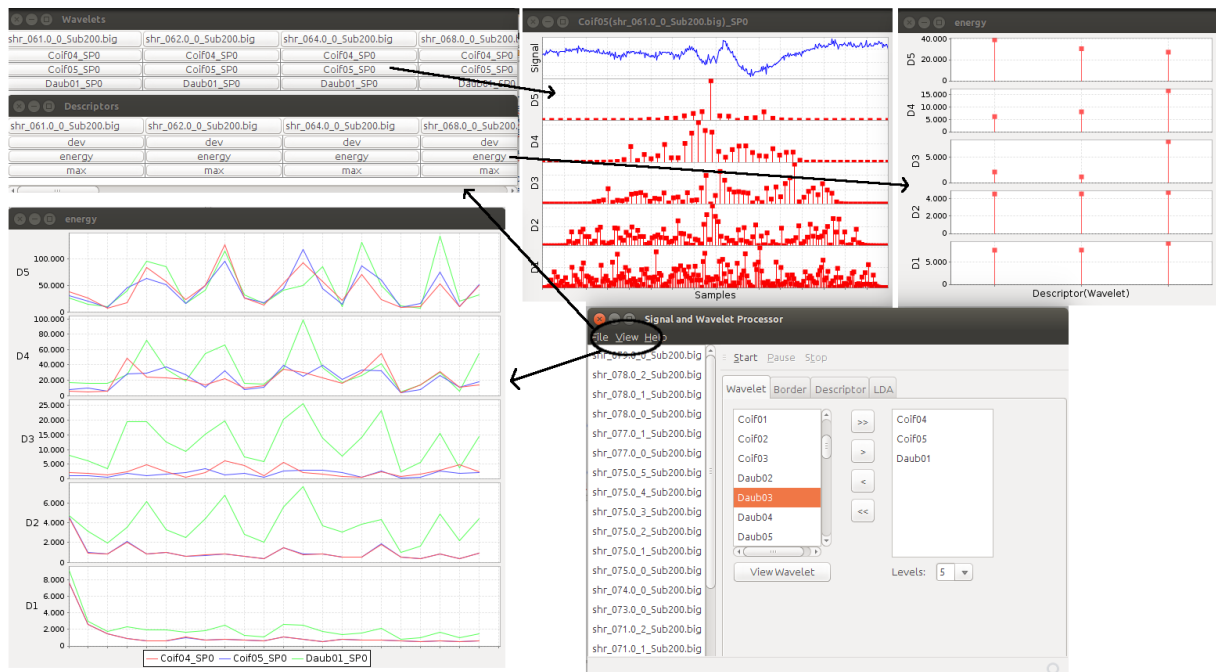
**Figura 30:** Aba de seleção de descritores a serem aplicados nos níveis de decomposição das wavelets-mães selecionadas.

Ao término dos cálculos, clicando no botão 'View', são abertas três janelas como apresentado pela figura 31. A primeira janela é uma matriz de botões onde cada coluna representa um evento carregado pelo sistema e cada linha representa uma transformada wavelet associada a uma extensão de borda. Clicando-se no botão de um elemento dessa matriz, é aberta uma janela onde o evento é apresentado no topo da janela em azul e os coeficientes gerados pela transformada wavelet são apresentados abaixo do sinal em vermelho, separados em níveis de decomposição.

A segunda janela aberta também é uma matriz de botões onde cada coluna representa

um evento específico carregado pelo programa e cada linha representa um descritor. Ao clicar em um botão dessa matriz, abre-se uma janela onde são apresentados campos em igual número ao número de decomposições wavelet selecionados. Cada campo apresenta vetores correspondentes ao resultado desse descritor aplicado a esse determinado nível de decomposição wavelet para cada wavelet-mãe selecionada.

A terceira janela aberta apresenta o resultado de cada descritor aplicado no resultado da transformada wavelet para todos os eventos selecionados, permitindo ver variações, como por exemplo, entre eventos espícula e não-espícula.

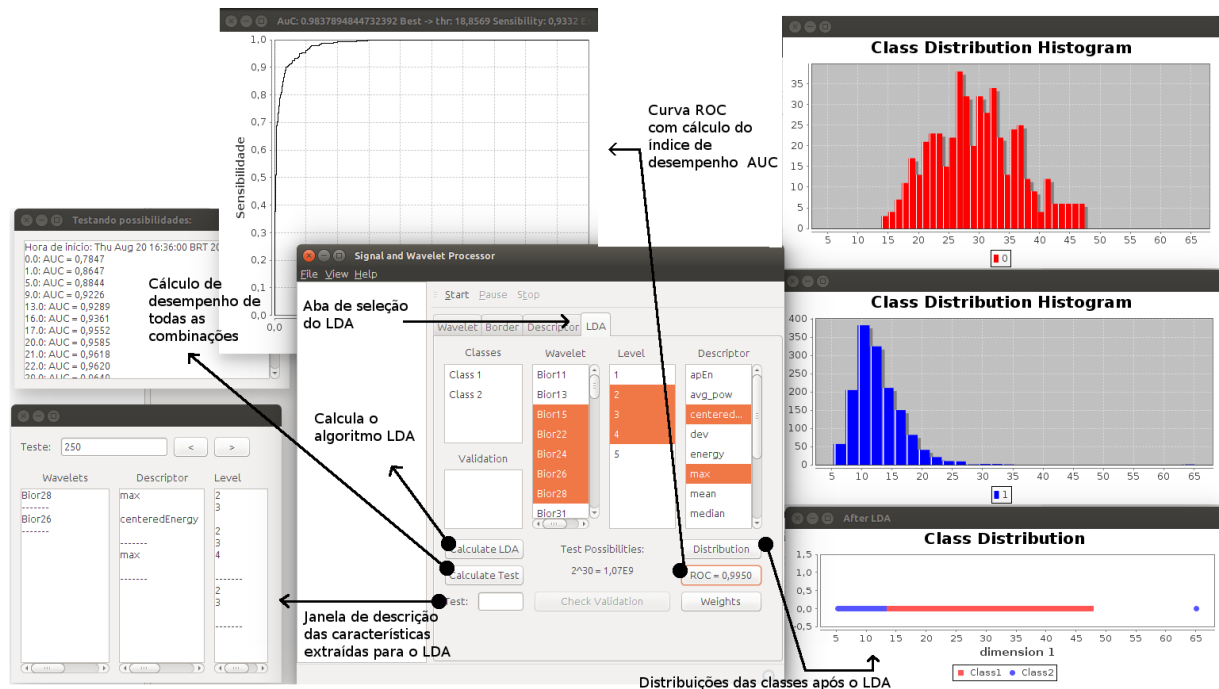


**Figura 31:** Janelas de resultado após o cálculo da transformada wavelet e descritores. É possível visualizar os coeficientes wavelet obtidos para cada sinal, assim como o resultado dos descritores para cada wavelet.

Após o cálculo da transformada wavelet, é possível exportar os resultados para planilhas de cálculo para processamento posterior.

A última tela do programa permite realizar a análise de discriminante linear. Esta tela é apresentada pela figura 32. Clicando sobre o campo classes, o usuário seleciona uma planilha exportada pelo programa contendo dados de eventos espícula ou não espícula, embora nesse caso o algoritmo aceite qualquer número de classes distintas. Nos campos 'wavelet', 'level', e 'descriptor', é possível selecionar quais características serão utilizadas pela algoritmo de LDA. Clicando sobre o botão 'Calculate Test' todas as combinações possíveis para a seleção serão calculadas.

Clicando sobre os botões 'Distribution' e 'ROC', são abertas as telas de distribuição de classes após o processamento do LDA e a curva ROC, respectivamente. É apresentado ainda o índice de desempenho AUC, o melhor limiar para as distribuições e os valores de sensibilidade e especificidade para o melhor limiar de classificação.



**Figura 32:** Janela para seleção de classes e características para o processamento do algoritmo de LDA. São apresentadas as telas de distribuição das classes, a curva ROC e telas auxiliares de teste.

Nesta seção foram apresentados os programas desenvolvidos que aplicam as ferramentas necessárias para o desenvolvimento desse estudo.

Durante a fase de implementação, os resultados iniciais da transformada wavelet e extensão de borda foram aferidos com o programa Matlab®. Ao término dessa fase, o programa Matlab® não foi mais utilizado.

**Comentário:** Após o processamento da transformada wavelet, a exportação de planilhas gera um número proporcional de arquivos para cada evento processado. O Matlab® sendo utilizado no mesmo computador onde roda o programa, precisa de mais de 1 minuto para carregar uma única planilha, enquanto que o programa escrito em Java carrega as mais de mil planilhas em menos de 30 segundos.

## 4.6 COMENTÁRIOS

Neste capítulo foram apresentadas os métodos e ferramentas adotados para o desenvolvimento dessa pesquisa. O classificador proposto na figura 19 é a base para todos os resultados a serem apresentados no capítulo 5. A extração de características dos eventos epileptiformes contidos na base de sinais é feita através da transformada wavelet e subsequente cálculo dos descritores para cada decomposição da transformada. É construída a matriz de características e aplicado no processamento do LDA. O resultado do algoritmo LDA é então classificado linearmente através do limiar definido através da curva ROC como tendo os melhores pares de sensibilidade e especificidade.

O próximo capítulo apresenta como as ferramentas foram aplicadas no conjunto de dados de EEG contidos na base de eventos e os resultados obtidos.

## 5 MÉTODOS E RESULTADOS

Este capítulo apresenta os resultados do estudo da transformada wavelet na detecção de espículas. Primeiramente é feita uma avaliação comparativa do classificador proposto na seção 4.2 com resultados de estudos encontrados na literatura. Em seguida são apresentados os resultados obtidos dos objetivos definidos na seção 1.3, como a extensão de borda inerente ao cálculo da transformada wavelet, a análise da quantidade de níveis de decomposição, a melhor wavelet-mãe para detecção de espículas, assim como os descritores que melhor discriminam eventos espícula de eventos não-espícula. Por fim é feita uma análise da influência da taxa de amostragem na detecção de espículas.

### 5.1 AVALIAÇÃO COMPARATIVA DE RESULTADOS

O modelo proposto para o classificador de espículas na seção 4.2, figura 19, não possui o bloco de predição, normalmente existente nos sistemas de reconhecimento de padrões (DUDA; HART; STORK, 2001). Como proposto nesse estudo, a redução de dimensionalidade provida pelo algoritmo LDA torna possível o uso de um classificador linear sem passar pela fase de predição. A fase de predição é normalmente encontrada em estudos na literatura implementada através de alguma abordagem de rede neural, lógica fuzzy ou outro sistema de predição descrito na seção 3.2.4.

O bloco de predição é necessário em sistemas onde a dimensionalidade das características, após a sua redução, não permite a detecção ou classificação direta de elementos, exigindo uma classificação preditiva de classes. O bloco de predição ainda é necessário caso hajam mais do que duas classes de elementos, o que não é o caso desse estudo de detecção de espículas.

O objetivo da avaliação comparativa de resultados é verificar se as ferramentas desenvolvidas para esse estudo, ao serem comparadas com outros estudos, são capazes de reproduzir experimentos obtendo resultados semelhantes.

O estudo publicado por Wang *et al.* (2009) foi utilizado como referência. Nesse estudo, os autores utilizam uma base com 100 épocas de fundo de EEG e 64 épocas com realizações de espículas. Cada época foi transformada pela wavelet-mãe DB4 nas três primeiras decomposições: D1, D2, e D3. De cada decomposição é calculada a entropia aproximada e para cada decomposição é aplicado um limiar de detecção. O limiar de detecção para cada decomposição é calculado pelo critério de Neymann-Pearson. O estudo de referência considera como positivo épocas que ultrapassem o limiar de detecção para as três decomposições.

**Comentário:** O estudo de Wang *et al.* (2009) foi selecionado por apresentar uma arquitetura onde o método de detecção é similar ao método de detecção do classificador proposto, utilizando um descritor conhecido e de fácil implementação.

Para obter uma base de dados comparável ao estudo de Wang *et al.* (2009), a base de eventos apresentada na seção 4.1 foi reamostrada a 200 Hz com aplicação de filtro passa-baixas de 70 Hz. Foram separadas aleatoriamente 64 épocas de espículas e 100 épocas de fundo de EEG da base de eventos. Diferentemente da técnica utilizada no estudo de Wang *et al.* (2009), após o cálculo da entropia aproximada, foi gerada a matriz de características e reduzido a uma dimensão pelo algoritmo do LDA. O resultado do processamento LDA é submetido a um classificador linear.

Os resultados do experimento da literatura e do classificador proposto são apresentados na tabela 12 em termos de sensibilidade e especificidade.

**Tabela 12:** Avaliação comparativa de ferramentas e base de dados de eventos utilizando como referência o trabalho de Wang *et al.* (2009).

	(WANG <i>et al.</i> , 2009)	Classificador Proposto
sensibilidade	0,984	0,982
especificidade	0,940	0,970

Observa-se que o classificador proposto na seção 4.2, figura 19, apresenta resultado similar quanto a sensibilidade e resultado superior quanto a especificidade. Como validação, o classificador proposto repetiu o experimento atingindo valores similares aos apresentados na literatura.

**Comentário:** A similaridade dos resultados obtidos pelo classificador proposto com os resultados do estudo de Wang *et al.* (2009) indica que o classificador proposto conseguiu obter resultados esperados.

## 5.2 AVALIAÇÃO DOS MÉTODOS DE EXTENSÃO DE BORDA

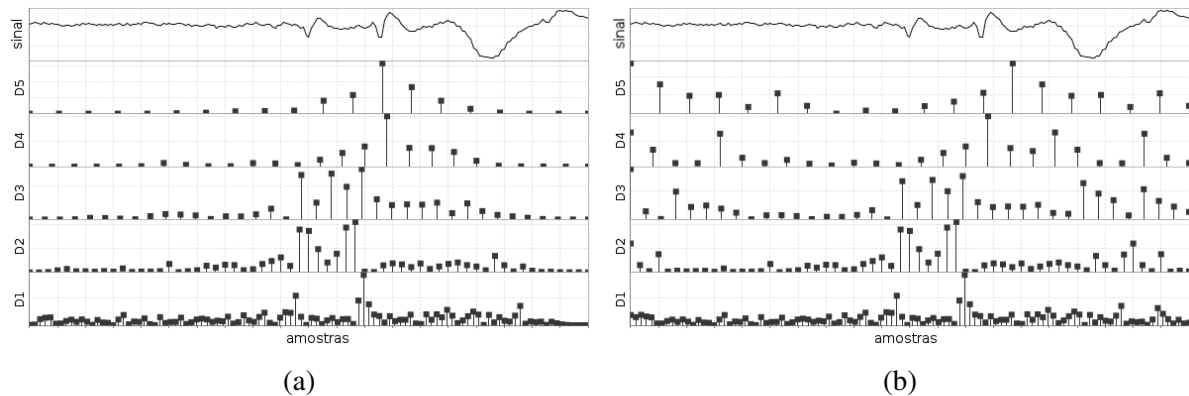
Na implementação da transformada wavelet, foi observado um efeito no resultado da transformada wavelet causado pela maneira de como se interpreta o sinal de EEG fora da janela de amostragem. Durante a transformação wavelet, essa interpretação tem influência significativa no resultado final (MISITI *et al.*, 2013). Pelo fato da transformada discreta wavelet necessitar de um número de amostras finita para seu cálculo, faz-se necessário a implementação de algum tipo de extensão de borda para cada decomposição wavelet calculada recursivamente conforme descrito por Mallat (1989). Ou seja, não é possível evitar a execução da extensão de borda na transformada wavelet, independente do tamanho da amostra em transformação. A teoria sobre convolução e métodos de extensão de borda foram apresentados na seção 2.6.

**Comentário:** O estudo sobre extensões de borda e seus efeitos na transformada wavelet é amplamente conhecido na área de processamento de imagens (SU; LIU; LI, 2012; MONTANARI *et al.*, 2015), inclusive sendo sugerido métodos para evitar os seus efeitos (KATUNIN, 2012). Entretanto, no processamento de sinais unidimensionais, especificamente no processamento de eventos de EEG, não foram encontrados trabalhos onde seja comentado sobre o método de extensão de borda usado na transformada wavelet.

Ao decompor um sinal pela transformada wavelet, para cada decomposição wavelet, o sinal a ser convoluído recebe amostras adicionais à esquerda e à direita, onde a quantidade de valores é igual ao número de coeficientes menos um do filtro passa-altas de sua respectiva wavelet-mãe. O algoritmo de extensão de bordas é executado para cada decomposição da transformada wavelet, sendo que mesmo para amostras de longa duração, os efeitos de distorção inseridos pela extensão de borda aumentam e se propagam pelo sinal transformado conforme aumenta o número de decomposições wavelet. O algoritmo de extensão de bordas para a DWT foi abordado na seção 2.6.1.

A figura 33 apresenta um exemplo do efeito da extensão de borda, onde a mesma realização de espícula é transformada pela wavelet-mãe DB4 em cinco níveis de decomposição. Na figura 33(a) é utilizada a extensão de borda SP0 e na figura 33(b) é utilizada a extensão de borda PPD. É possível observar que utilizando a extensão de borda SP0 (figura 33(a)), os coeficientes dos níveis de decomposição são maiores na região central da janela onde se encontra a espícula, e os coeficientes às margens da janela se mantém com valores baixos. Utilizando a extensão de borda PPD (figura 33(b)), observa-se que a partir da decomposição D2, os coeficientes nas laterais possuem valores similares aos coeficientes centrais e se propagam para o centro da janela conforme aumenta o nível de decomposição.





**Figura 33:** Cinco decomposições da transformada wavelet DB4 da mesma espícula com extensão de borda (a) SP0 e (b) PPD.

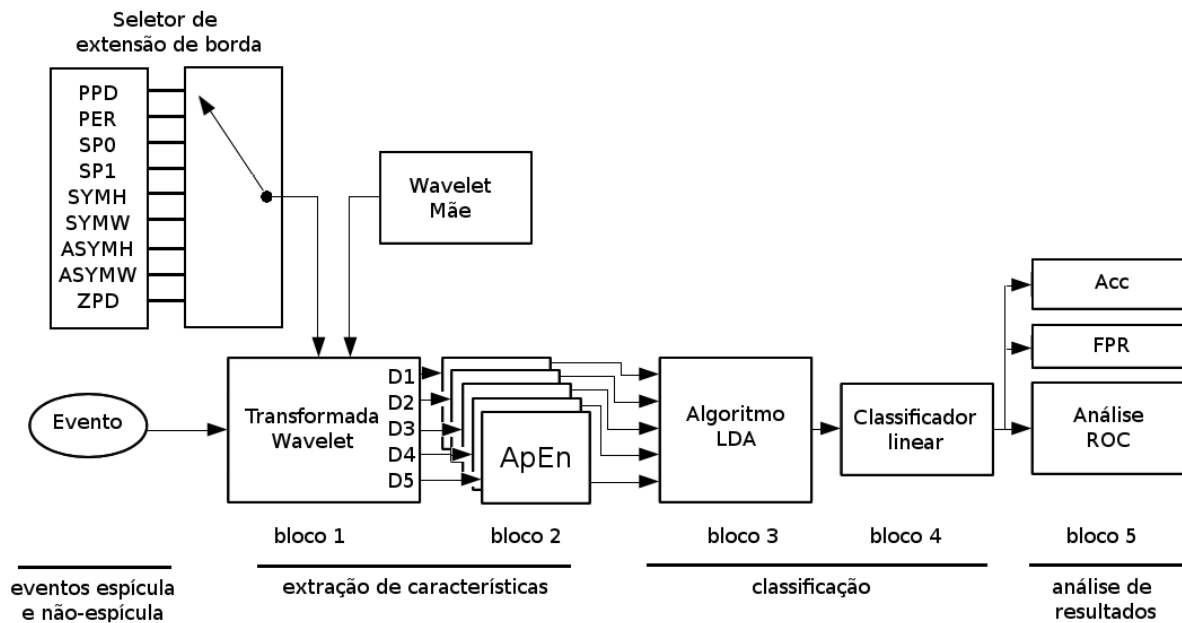
Considerando essas diferenças de comportamento, o objetivo desse estudo visa avaliar os efeitos de diferentes métodos de extensão de borda do sinal de EEG na transformada wavelet com relação à detecção de espículas.

**Comentário:** Um estudo preliminar desse efeito foi publicado em Pacola *et al.* (2012b).

O classificador proposto na seção 4.2, figura 19, foi modificado como apresentado no diagrama em blocos da figura 34. Todas as transformações wavelet utilizaram cinco decomposições, e, para cada decomposição, foi calculada a entropia aproximada. A entropia aproximada foi utilizada nesse estudo por ser um descritor utilizado com frequência nos estudos encontrados na literatura como descrito na seção 3.2.3.

Foram selecionadas 19 wavelets a partir de estudos encontrados na literatura (KALAYCI; OZDAMAR, 1995; ERCELEBI; SUBASI, 2006; SADATI; MOHSENI; MAGHSOUDI, 2006; ADELI; GHOSH-DASTIDAR; DADMEHR, 2007; INUSO *et al.*, 2007; INDIRADEVI *et al.*, 2008; WANG *et al.*, 2009; HSU; YU, 2009; SHUREN; ZHONG, 2009; BIROT *et al.*, 2013; GAJIC *et al.*, 2014; UBEYLI, 2009; NUNES *et al.*, 2014; VAVADI; AYATOLLAHI; MIRZAEI, 2010; VENERI *et al.*, 2011; MIRZAEI; AYATOLLAHI; VAVADI, 2011), sendo elas: Coiflet 2, 3, 4, e 5, Daubechies 2, 3, 4, 5, 6, 8, 10, e 20, discrete Meyer, Reverse Bior 13, Haar, e Symlet 2, 3, 4, e 5. Essas wavelets são mencionadas nesses estudos como sendo as melhores para a detecção de espículas em sinais de EEG.

As wavelets-mães selecionadas foram divididas em dois grupos de acordo com o número de coeficientes: grupo 1 contendo wavelets-mães com até oito coeficientes e grupo 2 contendo wavelets-mães com mais de oito coeficientes. Os grupos e o número de coeficientes pertencentes às wavelets analisadas nesse estudo são apresentados na tabela 13.



**Figura 34:** Diagrama em blocos do classificador proposto para o estudo dos métodos de extensão de bordas. O classificador proposto na seção 4.2, figura 19, recebe os blocos adicionais de seleção de borda para cada wavelet-mãe utilizada.

**Comentário:** As wavelets-mães com até oito coeficientes são as mais utilizadas nos estudos citados.

Para verificar diferenças de classificação causadas pela extensão de borda na detecção de espículas, foram utilizados os métodos de extensão de borda da tabela 9 para cada uma das 19 wavelets-mães selecionadas.

Para utilizar a extensão de borda PER, por exemplo, os eventos espícula e não-espícula contidos na base de dados foram transformados pela wavelet-mãe Haar em cinco níveis de decomposição adotando a extensão de borda PER como padrão. Em seguida foi calculado o descritor entropia aproximada e o algoritmo LDA com posterior classificação linear das distribuições de espícula e não-espícula resultante.

O processamento foi repetido para cada uma das 19 wavelets-mães contidas na tabela 13, tendo-se assim a contribuição de cada wavelet-mãe para a extensão de borda PER.

Em seguida foram repetidos os mesmos cálculos, porém, adotando-se como padrão as extensões de borda PPD, ZPD, SP1, SP0, SYMH, SYMW, ASYMW e ASYMH sequencialmente para as wavelets selecionadas.

Para comparar as diferenças de desempenho na classificação, foram utilizados três índices de desempenho: a área sob a curva ROC (AUC), acurácia (Acc), e taxa de falsos

**Tabela 13:** Número de coeficientes de cada wavelet-mãe

Grupo	Número de Coeficientes	Wavelet-Mãe
1	2	Haar
	4	DB2, Sym2
	6	DB3, Sym3, RBior13
	8	DB4, Sym4
	10	DB5, Sym5
2	12	DB6, Coif2
	16	DB8
	18	Coif3
	20	DB10
	24	Coif4
	30	Coif5
	40	DB20
	62	Dmey

positivos (FPR). Os índices obtidos para cada borda são apresentados na tabela 14.

**Tabela 14:** Índices de desempenho AUC, Acc, e FPR para os métodos de extensão de borda das 19 wavelets-mães selecionadas. As extensões de borda para cada wavelet-mãe foram organizadas por ordem de classificação onde a extensão de borda que apresentou o melhor resultado referente ao índice de desempenho AUC está na primeira posição.

Wavelet	Coef.	Rótulo	Ordem de Classificação								
			1º	2º	3º	4º	5º	6º	7º	8º	9º
Haar	2	borda	SP0	SYMH	ASYMW	SP1	ASYMH	PPD	PER	SYMW	ZPD
		AUC	0,9332	0,9332	0,9328	0,9328	0,9327	0,9325	0,9325	0,9325	0,8848
		Acc	0,8476	0,8476	0,8476	0,8476	0,8471	0,8441	0,8441	0,8481	0,8156
		FPR	0,1526	0,1526	0,1526	0,1526	0,1532	0,1559	0,1559	0,1519	0,1559
DB2	4	borda	SP0	SYMW	SYMH	SP1	ASYMW	PPD	PER	ZPD	ASYMH
		AUC	0,9300	0,9282	0,9251	0,9226	0,9199	0,8631	0,8631	0,8501	0,8326
		Acc	0,8541	0,8552	0,8483	0,8501	0,8471	0,7895	0,7975	0,7689	0,7705
		FPR	0,1459	0,1452	0,1518	0,1499	0,1532	0,2105	0,2025	0,2312	0,2298
Sym2	4	borda	SP0	SYMW	SYMH	SP1	ASYMW	PER	PPD	ZPD	ASYMH
		AUC	0,9300	0,9282	0,9251	0,9226	0,9199	0,8631	0,8631	0,8501	0,8326
		Acc	0,8541	0,8552	0,8483	0,8501	0,8471	0,7975	0,7895	0,7689	0,7705
		FPR	0,1459	0,1452	0,1518	0,1499	0,1532	0,2025	0,2105	0,2312	0,2298
DB3	6	borda	SP1	SP0	ASYMW	SYMH	SYMW	ZPD	PER	PPD	ASYMH
		AUC	0,9059	0,9057	0,9052	0,8980	0,8904	0,8687	0,8431	0,8298	0,7376
		Acc	0,8286	0,8239	0,8251	0,8265	0,8075	0,7854	0,7709	0,7454	0,6897
		FPR	0,1719	0,1761	0,1752	0,1739	0,1925	0,2146	0,2292	0,2551	0,3105
RBior13	6	borda	SP0	SYMH	SYMW	ASYMW	SP1	ZPD	PPD	PER	ASYMH
		AUC	0,9187	0,9185	0,9055	0,8988	0,8902	0,8712	0,8485	0,8474	0,7873
		Acc	0,8456	0,8461	0,8315	0,8275	0,8176	0,7895	0,7629	0,7669	0,7130
		FPR	0,1546	0,1539	0,1686	0,1726	0,1825	0,2105	0,2372	0,2332	0,2874
Sym3	6	borda	SP1	SP0	ASYMW	SYMH	SYMW	ZPD	PER	PPD	ASYMH
		AUC	0,9059	0,9057	0,9052	0,8980	0,8904	0,8687	0,8431	0,8298	0,7376
		Acc	0,8286	0,8239	0,8251	0,8265	0,8075	0,7854	0,7709	0,7454	0,6897
		FPR	0,1719	0,1761	0,1752	0,1739	0,1925	0,2146	0,2292	0,2551	0,3105
DB4	8	borda	SP0	SYMH	ASYMW	SYMW	SP1	ZPD	PPD	PER	ASYMH
		AUC	0,9055	0,8987	0,8938	0,8916	0,8854	0,8551	0,8150	0,8093	0,7804
		Acc	0,8376	0,8216	0,8275	0,8123	0,8115	0,7649	0,7353	0,7269	0,7063
		FPR	0,1626	0,1785	0,1726	0,1859	0,1885	0,2352	0,2652	0,2733	0,2938

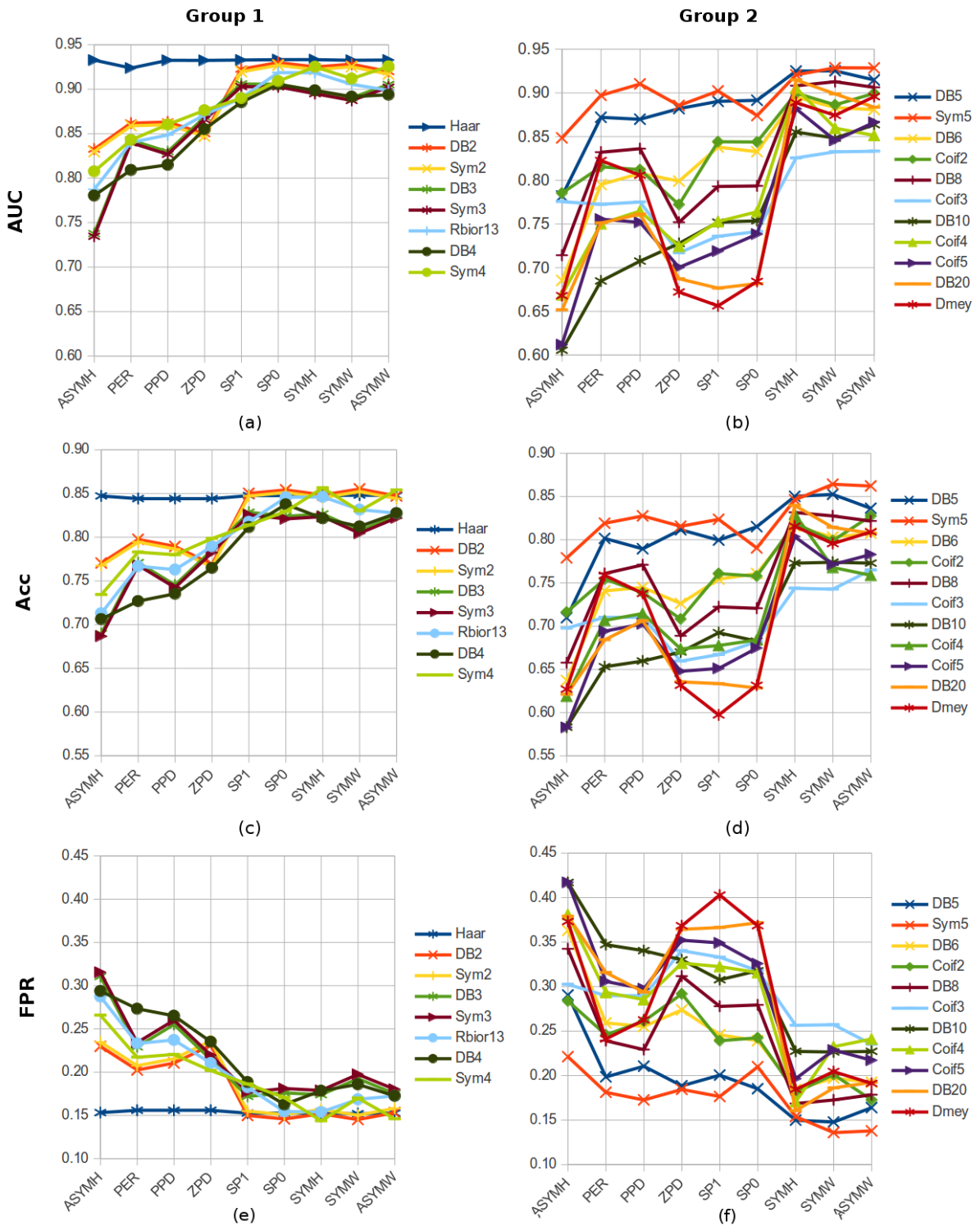
*Continua na próxima página*

Tabela 14 – Continuação da página anterior

Wavelet	Coef.	Rótulo	Ordem de Classificação								
			1º	2º	3º	4º	5º	6º	7º	8º	9º
Sym4	8	borda	ASYMW	SYMH	SYMW	SP0	SP1	ZPD	PPD	PER	ASYMH
		AUC	0,9257	0,9252	0,9119	0,9091	0,8900	0,8763	0,8603	0,8576	0,8076
		Acc	0,8541	0,8564	0,8300	0,8295	0,8139	0,7985	0,7800	0,7829	0,7343
		FPR	0,1459	0,1437	0,1700	0,1706	0,1862	0,2019	0,2205	0,2172	0,2658
DB5	10	borda	SYMW	SYMH	ASYMW	SP0	SP1	ZPD	PER	PPD	ASYMH
		AUC	0,9252	0,9250	0,9148	0,8915	0,8903	0,8820	0,8723	0,8698	0,7819
		Acc	0,8525	0,8501	0,8363	0,8150	0,7995	0,8115	0,8015	0,7895	0,7097
		FPR	0,1478	0,1499	0,1640	0,1852	0,2005	0,1885	0,1985	0,2105	0,2905
Sym5	10	borda	SYMW	ASYMW	SYMH	PPD	SP1	PER	ZPD	SP0	ASYMH
		AUC	0,9288	0,9285	0,9207	0,9102	0,9022	0,9015	0,8855	0,8740	0,8484
		Acc	0,8642	0,8621	0,8461	0,8275	0,8236	0,8190	0,8156	0,7904	0,7789
		FPR	0,1359	0,1379	0,1539	0,1726	0,1765	0,1812	0,1845	0,2099	0,2212
DB6	12	borda	SYMH	SYMW	ASYMW	SP1	SP0	PPD	PER	ZPD	ASYMH
		AUC	0,8952	0,8832	0,8808	0,8379	0,8327	0,8079	0,8037	0,7990	0,6853
		Acc	0,8156	0,8030	0,8075	0,7544	0,7609	0,7448	0,7408	0,7263	0,6371
		FPR	0,1845	0,1972	0,1925	0,2458	0,2392	0,2552	0,2592	0,2738	0,3631
Coif2	12	borda	SYMH	ASYMW	SYMW	SP1	SP0	PER	PPD	ASYMH	ZPD
		AUC	0,8997	0,8995	0,8863	0,8441	0,8438	0,8319	0,8121	0,7853	0,7722
		Acc	0,8190	0,8275	0,7990	0,7609	0,7579	0,7544	0,7388	0,7163	0,7083
		FPR	0,1812	0,1726	0,2012	0,2392	0,2425	0,2458	0,2612	0,2838	0,2918
DB8	16	borda	SYMW	SYMH	ASYMW	PER	PPD	SP0	SP1	ZPD	ASYMH
		AUC	0,9128	0,9081	0,9063	0,8377	0,8361	0,7935	0,7929	0,7519	0,7141
		Acc	0,8275	0,8315	0,8216	0,7609	0,7709	0,7207	0,7223	0,6888	0,6577
		FPR	0,1726	0,1686	0,1785	0,2392	0,2292	0,2794	0,2778	0,3117	0,3424
Coif3	18	borda	ASYMW	SYMW	SYMH	ASYMH	PPD	PER	SP0	SP1	ZPD
		AUC	0,8332	0,8326	0,8256	0,7755	0,7751	0,7751	0,7413	0,7359	0,7170
		Acc	0,7649	0,7428	0,7438	0,6977	0,7103	0,7103	0,6822	0,6672	0,6597
		FPR	0,2352	0,2572	0,2565	0,3025	0,2898	0,2898	0,3178	0,3331	0,3404
DB10	20	borda	ASYMW	SYMH	SYMW	SP0	SP1	ZPD	PPD	PER	ASYMH
		AUC	0,8636	0,8552	0,8482	0,7536	0,7521	0,7277	0,7075	0,7041	0,6055
		Acc	0,7729	0,7729	0,7739	0,6827	0,6924	0,6700	0,6597	0,6531	0,5831
		FPR	0,2272	0,2272	0,2265	0,3178	0,3077	0,3300	0,3404	0,3471	0,4170
Coif4	24	borda	SYMH	SYMW	ASYMW	PPD	SP0	PER	SP1	ZPD	ASYMH
		AUC	0,9055	0,8596	0,8514	0,7648	0,7640	0,7562	0,7528	0,7244	0,6690
		Acc	0,8295	0,7679	0,7589	0,7147	0,6842	0,7067	0,6776	0,6737	0,6190
		FPR	0,1706	0,2325	0,2412	0,2854	0,3158	0,2935	0,3225	0,3264	0,3811
Coif5	30	borda	SYMH	ASYMW	SYMW	PER	PPD	SP0	SP1	ZPD	ASYMH
		AUC	0,8818	0,8664	0,8456	0,7578	0,7518	0,7384	0,7189	0,7003	0,6119
		Acc	0,8035	0,7829	0,7713	0,6942	0,7027	0,6747	0,6511	0,6479	0,5829
		FPR	0,1965	0,2172	0,2287	0,3058	0,2976	0,3258	0,3491	0,3522	0,4171
DB20	40	borda	SYMH	SYMW	ASYMW	PER	PPD	ZPD	SP0	SP1	ASYMH
		AUC	0,9148	0,8988	0,8836	0,7614	0,7613	0,6873	0,6818	0,6766	0,6517
		Acc	0,8396	0,8142	0,8075	0,6842	0,7063	0,6356	0,6285	0,6336	0,6210
		FPR	0,1606	0,1862	0,1925	0,3158	0,2938	0,3644	0,3718	0,4028	0,3791
Dmey	62	borda	ASYMW	SYMH	SYMW	PER	PPD	SP0	ZPD	ASYMH	SP1
		AUC	0,8957	0,8889	0,8743	0,8229	0,8058	0,6842	0,6721	0,6679	0,6563
		Acc	0,8090	0,8156	0,7955	0,7589	0,7378	0,6316	0,6316	0,6270	0,5973
		FPR	0,1912	0,1845	0,2045	0,2412	0,2625	0,3684	0,3684	0,3731	0,4028

A figura 35 apresenta os resultados da tabela 14. Como visto na figura 35, a influência de cada método de extensão de borda no desempenho do classificador varia conforme o tamanho da wavelet usada, permanecendo praticamente constante para wavelets com até oito coeficientes (grupo 1), e com bastante variação para wavelets com mais de oito coeficientes (grupo 2).

Considerando os resultados dos índices de desempenho AUC, acurácia e FPR, é possível visualizar dois aspectos diferentes sobre as extensões de borda: As melhores extensões



**Figura 35:** Comportamento dos índices de desempenho AUC, Acc e FPR para os métodos de extensão de borda relativo aos grupos 1 e 2.

de borda para wavelets com até 8 coeficientes são SP0, SP1, SYMH, SYMW, e ASYMW, enquanto que, para wavelets que possuem mais de oito coeficientes, as melhores extensões de borda são SYMH, SYMW, e ASYMW.

As extensões de borda PER, PPD, ZPD, e ASYMH, apresentam os piores resultados na detecção de espículas em ambos os grupos, independente do índice de desempenho.

Como visto na figura 35, o desempenho das extensões de borda SP0 e SP1 é elevado para wavelets com número pequeno de coeficientes e se deteriora com o aumento do número de coeficientes da wavelet.

Das figuras 35a e 35b, a wavelet que apresentou o melhor desempenho, considerando cinco níveis de decomposição, ApEn e extensão de borda SP0, foram as wavelets Haar, com AUC igual a 0,9332 e DB2 e Sym2, ambas com AUC igual a 0,9300 como observado na tabela 14.

Das figuras 35c e 35d, o melhor resultado foi obtido usando a wavelet Sym5 com extensão de borda SYMW (0,8642), seguida pelas wavelets DB2 e Sym2 com extensão de borda SP0, e Sym4 com extensão de borda ASYMW, esses últimos três casos com Acc igual a 0,8541, como observado na tabela 14.

Das figuras 35e e 35f, considerando o índice de desempenho FPR, as wavelets que apresentaram o melhor resultado foram a Sym5 com extensão de borda SYMW (0,1359), seguida pelas wavelets DB2 e Sym2 com extensão de borda SP0, e Sym4 com extensão de borda ASYMW, com índices iguais a 0,1459, como observado na tabela 14.

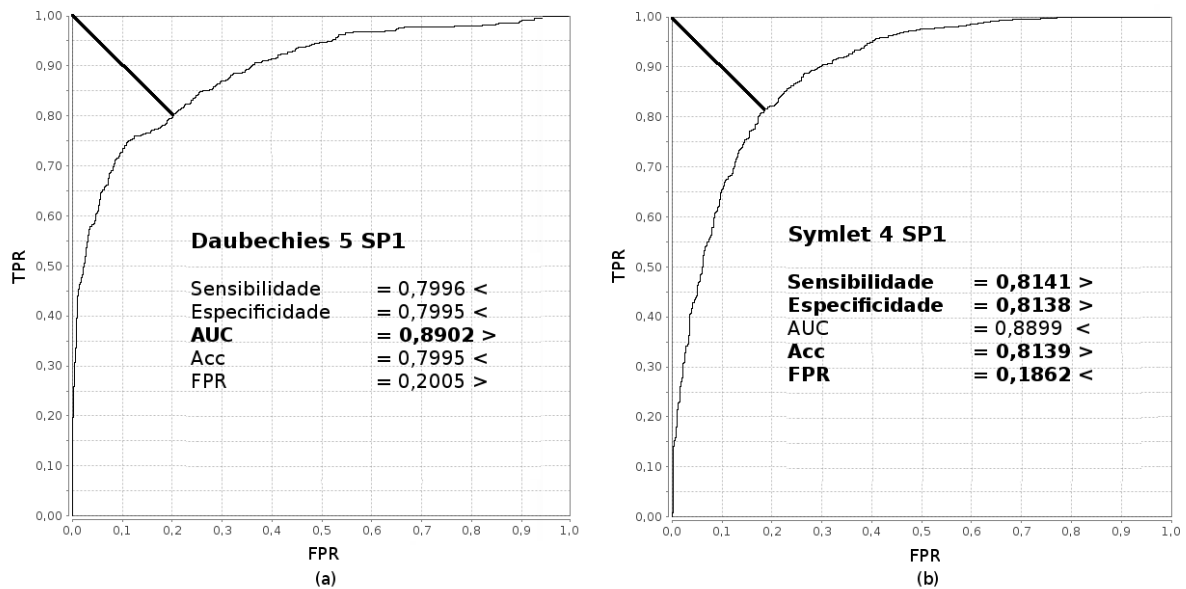
**Comentário:** Cada método de extensão de borda influencia no desempenho da wavelet-mãe utilizada no classificador. A comparação de índices de desempenho pode sofrer vieses. O índice de desempenho AUC é um índice de desempenho global do classificador, determinada pela curva ROC. A acurácia (Acc) e o FPR são índices de desempenho pontuais do classificador para um determinado limiar de detecção. Isso gera diferenças de desempenho como observado na classificação existente na tabela 14, para a wavelet-mãe DB8, onde a extensão de borda SYMW apresenta melhor resultado considerando o índice de desempenho AUC, e, considerando a acurácia, a melhor extensão de borda para essa wavelet-mãe é a SYMH. Dessa maneira faz-se necessária uma análise dos índices de desempenho.

### 5.2.1 COMPARAÇÃO DE ÍNDICES DE DESEMPENHO

O uso de diversos índices de desempenho para determinar a melhor wavelet-mãe na detecção de espículas pode levar a conclusões diferentes. Pares de valores de sensibilidade e especificidade são gerados para cada limiar de detecção do classificador. Cada par, como visto na seção 2.4, figura 9, determina um ponto da curva ROC. A acurácia e o FPR levam em consideração um par de valores de sensibilidade e especificidade obtidos do classificador.

Sabendo que o limiar de detecção é configurável, o valor da acurácia e do FPR podem não representar o melhor par de especificidade e sensibilidade, muito menos variações de detecção inerentes ao classificador em outras regiões do limiar de detecção. O índice de desempenho AUC, por sua vez, leva em consideração todos os pares possíveis de valores de sensibilidade e especificidade usados para determinar a curva ROC, sendo menos vulnerável a uma configuração específica do limiar de detecção do classificador linear.

Isso é apresentado na figura 36, onde, por exemplo, a wavelet DB5 produz um valor de AUC maior do que quando usando a wavelet Sym4. Por outro lado, utilizando apenas os melhores pares de sensibilidade e especificidade de cada curva, denotada pela linha enfatizada tocando a curva, os resultados respectivos de Acc e FPR apresentam conclusões contrárias, onde a wavelet Sym4 é melhor do que a DB5 para a detecção de espículas.



**Figura 36:** Curvas ROC das wavelets (a) DB5 com borda SP1 e (b) Sym4 com borda SP1.

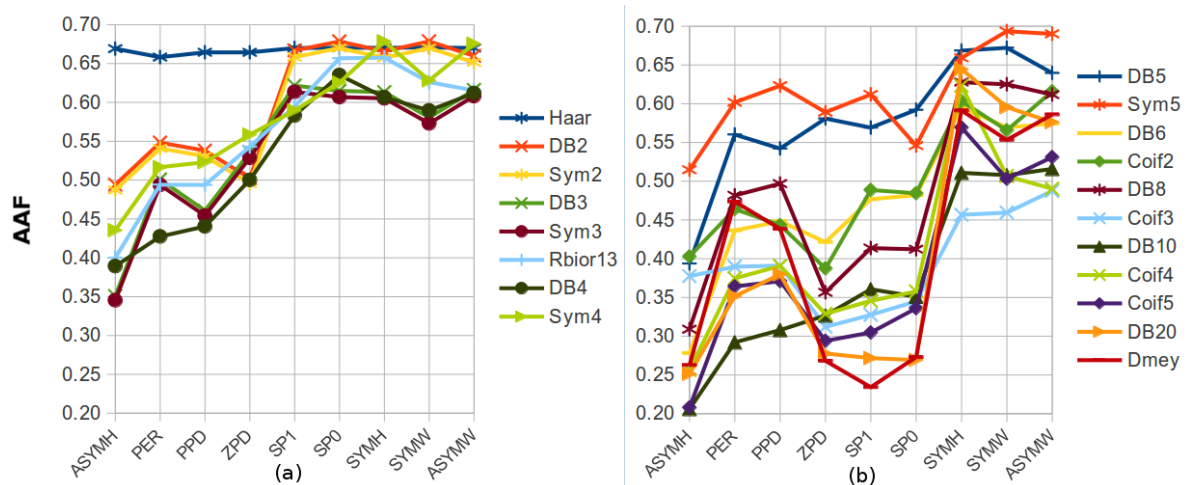
Considerando esse comportamento, um novo índice de desempenho é proposto para comparar resultados, este agora considerando AUC, Acc e FPR juntos. Esse índice, AAF (acrônimo para AUC, Acc e FPR), é apresentado pela equação 57 e corresponde a um produto multicritério dos índices que o precedem, apresentando um comportamento similar a cada índice individualmente.

$$AAF = AUC \cdot Acc \cdot (1 - FPR) \quad (57)$$

A proposição desse índice é representar com um único índice o comportamento dos índices AUC, Acc e FPR. O índice AAF é influenciado pelo comportamento individual de cada índice por ser um produto resultante de sua multiplicação. O uso desse índice visa evitar

a comparação entre índices de desempenho e facilitar o reconhecimento do desempenho do classificador.

O índice de desempenho AAF foi calculado, semelhante à figura 35, sendo o resultado apresentado na figura 37 confirmando a tendência individual dos índices de desempenho AUC, Acc e FPR.



**Figura 37:** Comportamento do índice multicritério AAF para os métodos de extensão de borda relativo a (a) grupo 1 e (b) grupo 2.

Agora, considerando o índice de desempenho AAF, a extensão de borda SP0 se mostrou melhor quando usada por wavelets com até oito coeficientes. A extensão de borda SYMH, o método padrão utilizado pelo software Matlab®, degrada os resultados em até 3,64% em relação ao método de extensão de borda SP0. Conceitos sobre como o Matlab® interpreta a extensão de borda durante a DWT foram abordados na seção 2.6.2.

Considerando wavelets com mais de oito coeficientes, a extensão de borda SYMH se mostrou melhor para a detecção de espículas para wavelets DB6, Coif2, Coif3, Coif4, COif5 e DB20.

A extensão de borda SYMW apresentou os melhores resultados para wavelets DB5, Sym5, e DB8, enquanto que a extensão de borda ASYMW se mostrou própria para wavelets DB10 e Dmey.

A extensão de borda ASYMH apresentou o pior resultado, independente do número de coeficientes da wavelet, degradando os resultados em até 44,11%.

A wavelet-mãe Haar, apresentada nas figuras 35 e 37 não sofre influência da extensão de borda usada, uma vez que o pior resultado obtido foi com a extensão de borda ZPD, a qual degradou o desempenho do classificador em 0,09%.



**Contribuição:** A extensão de borda SP0 se mostrou mais adequada para uso de wavelets com até oito coeficientes para o reconhecimento de espículas. Isso é importante pois wavelets com até oito coeficientes são as mais utilizadas em estudos encontrados na literatura.

**Contribuição:** As extensões de bordas PER, PPD, ZPD e ASYMH sempre apresentaram os piores resultados para a classificação de espículas, não sendo indicada para uso nessa aplicação. Isso é importante pois a extensão de borda ZPD é a que exige menor esforço de implementação, onde os valores adicionais de extensão de borda são iguais a zero.

**Contribuição:** A extensão de borda padrão utilizada pelo Matlab® é a SYMH (seção 2.6.2). Porém, para a detecção de espículas, essa borda não é a melhor para qualquer tipo de wavelet-mãe usada, podendo degradar o desempenho de classificação em até 3,64% para a base de dados utilizada de eventos espícula e não-espícula.

**Comentário:** Não é objetivo desse estudo avaliar o Matlab®, e sim a influência das extensões de borda na detecção de espículas. Porém, como observado durante a fase de validação do algoritmo da transformada wavelet que não há necessidade de configurar uma extensão de borda para executar a DWT no Matlab® por existir uma configuração padronizada, e por não ter sido encontrado na literatura comentário sobre métodos de extensão de borda aplicados na detecção de espícula, torna relevante a contribuição desse trabalho.

**Comentário:** Não foi encontrado na literatura nenhum trabalho que abordasse o estudo da influência da extensão de borda na transformada wavelet com relação à detecção de espículas. Tampouco foram encontrados comentários sobre distorções encontradas no resultado das decomposições wavelet.

A conclusão desse estudo é que a extensão de borda de sinais na transformada wavelet influencia o resultado de classificação de acordo com o número de coeficientes da wavelet utilizada. Inclusive que índices de desempenho diferentes podem apresentar resultados contrários, sendo o índice AUC menos influenciado por essas variações de extensão de bordas.

### 5.3 ANÁLISE DE FAMÍLIAS DE WAVELETS-MÃES E DESCRITORES

Nesta seção são analisadas as wavelets-mães e descritores que em conjunto obtém os melhores índices de desempenho AUC na detecção de espículas. O classificador para os experimentos é o proposto na seção 4.2, figura 19.

Primeiramente é feita uma análise dos níveis de decomposição wavelet e sua influência na detecção de espículas. Em seguida as wavelets-mães de uma mesma família são agrupadas

para verificar o seu desempenho. Após essa análise, são feitas comparações dos resultados entre as famílias de wavelets-mães e quais descritores apresentam melhor característica discriminante entre eventos espícula e não espícula. Ao final tem-se a fase completa da extração de características com o melhor agrupamento de wavelets-mães em conjunto com o melhor agrupamento de descritores para a detecção de espículas.

### 5.3.1 ANÁLISE DE DECOMPOSIÇÕES WAVELET

Como apresentado no capítulo 3, a maior parte dos estudos encontrados na literatura, no contexto de detecção de espículas utilizando a transformada wavelet, fazem uso de três a cinco níveis de decomposição com eventos amostrados até 200 Hz. Os níveis de decomposição são utilizados individualmente ou em conjunto, mas não há concordância entre os autores sobre qual nível de decomposição é o mais adequado para a detecção de espículas, ou se há um conjunto de decomposições mais adequado para esse fim. Desse modo, é verificado agora quantas decomposições, e quais decomposições, apresentam os melhores índices de desempenho AUC na detecção de espículas.

Para esse estudo foram analisadas 98 wavelets-mães, apresentadas na seção 2.5.2. Foram consideradas as combinações de cinco níveis de decomposição para cada transformada wavelet, gerando um espaço de teste de 31 combinações ( $2^5 - 1$ ) para cada wavelet-mãe. O estudo consiste em processar os eventos espícula e não-espícula da base de dados no classificador proposto para cada combinação de decomposições para cada wavelet-mãe apresentada, num total de  $98 * (2^5 - 1) = 3038$  possibilidades. É calculada a entropia aproximada de cada decomposição wavelet e o resultado é processado pelo algoritmo de LDA. O resultado é processado por um classificador linear de onde é determinada a curva ROC. Da curva ROC obtém-se o índice de desempenho AUC.

A tabela 15 apresenta o número de combinações de decomposições wavelet e quais decomposições foram utilizadas para cada possibilidade.

As wavelets-mães utilizadas foram reunidas em grupos e são apresentadas pela tabela 16.

As tabelas 17 e 18 apresentam os resultados do processamento para as combinações de níveis de decomposição para a família de wavelets-mães Daubechies (grupo 1), enquanto que as tabelas 19, 20, 21, e 22, apresentam os resultados para as combinações de decomposições wavelet para os grupos 2, 3, 4, e 5, respectivamente.

**Tabela 15:** Combinações de decomposições wavelet. A letra 'D' seguida de cinco caracteres indica as decomposições utilizadas para a combinação, onde o número indica o nível de decomposição utilizado e o caractere '-' indica a não utilização de um nível de decomposição.

Combinação	Decomposições	Combinação	Decomposições
1	D1----	17	D1---5
2	D-2---	18	D-2--5
3	D12---	19	D12--5
4	D--3--	20	D--3-5
5	D1-3--	21	D1-3-5
6	D-23--	22	D-23-5
7	D123--	23	D123-5
8	D---4-	24	D---45
9	D1--4-	25	D1--45
10	D-2-4-	26	D-2-45
11	D12-4-	27	D12-45
12	D--34-	28	D--345
13	D1-34-	29	D1-345
14	D-234-	30	D-2345
15	D1234-	31	D12345
16	D-----5		

**Tabela 16:** Grupos de wavelets-mães

Grupo	Família	Wavelets-mães	Quantidade
1	Daubechies	DB1, DB2, DB3, DB4, DB5, DB6, DB7, DB8, DB9, DB10, DB11, DB12, DB13, DB14, DB15, DB16, DB17, DB18, DB19, DB20, DB21, DB22, DB23, DB24, DB25, DB26, DB27, DB28, DB29, DB30, DB31, DB32, DB33, DB34, DB35, DB36, DB37, DB38	38
2	Biortogonal	Bior11, Bior13, Bior15, Bior22, Bior24, Bior26, Bior28, Bior31, Bior33, Bior35, Bior37, Bior39, Bior44, Bior55, Bior68	15
3	Reverse Biortogonal	RBior11, RBior13, RBior15, RBior22, RBior24, RBior26, RBior28, RBior31, RBior33, RBior35, RBior37, RBior39, RBior44, RBior55, RBior68	15
4	Symlet	Sym2, Sym3, Sym4, Sym5, Sym6, Sym7, Sym8, Sym9, Sym10, Sym11, Sym12, Sym13, Sym14, Sym15, Sym16, Sym17, Sym18, Sym19, Sym20	19
5	Coiflet, Legendre, Haar, Dmeyer	Coif1, Coif2, Coif3, Coif4, Coif5, Lege2, Lege4, Lege6, Haar, Haar Ortogonal, Dmeyer	11
<b>Total de Wavelets-Mães</b>			<b>98</b>

**Tabela 17:** Índice de desempenho AUC para as combinações de decomposições apresentadas na tabela 15 para as wavelets do grupo 1 apresentadas na tabela 16. Nesta tabela são apresentadas as wavelets da família Daubechies de DB1 a DB19. As combinações de níveis de decomposições estão ordenadas em ordem decrescente de índice AUC de acordo com a wavelet-mãe DB2.

Decomposições	Combinação	DB1	DB2	DB3	DB4	DB5	DB6	DB7	DB8	DB9	DB10	DB11	DB12	DB13	DB14	DB15	DB16	DB17	DB18	DB19
D12345	31	0,9332	0,9300	0,9057	0,9055	0,8915	0,8327	0,7725	0,7935	0,7862	0,7536	0,7485	0,7434	0,7306	0,7212	0,6997	0,6869	0,7022	0,7174	0,6733
D1234-	15	0,9323	0,9279	0,8927	0,8756	0,8870	0,8328	0,7694	0,7914	0,7583	0,7343	0,7279	0,7185	0,6912	0,6914	0,6793	0,6559	0,6773	0,6806	0,6608
D-2345	30	0,9248	0,9278	0,9032	0,9053	0,8895	0,8284	0,7693	0,7876	0,7761	0,7427	0,7355	0,7260	0,7170	0,7131	0,6976	0,6818	0,6884	0,7024	0,6627
D-234-	14	0,9234	0,9260	0,8892	0,8733	0,8843	0,8284	0,7642	0,7854	0,7437	0,7215	0,7123	0,6966	0,6768	0,6796	0,6738	0,6473	0,6605	0,6595	0,6472
D1-345	29	0,9067	0,9214	0,8989	0,8915	0,8815	0,8044	0,7120	0,7545	0,7381	0,7201	0,7038	0,7131	0,6939	0,6908	0,6428	0,6527	0,6679	0,6765	0,6345
D1-34-	13	0,9052	0,9183	0,8813	0,8541	0,8754	0,8044	0,7048	0,7497	0,6949	0,6893	0,6677	0,6788	0,6321	0,6433	0,6093	0,6044	0,6339	0,6274	0,6147
D--345	28	0,8794	0,9167	0,8927	0,8865	0,8743	0,7883	0,6885	0,7371	0,7143	0,6910	0,6741	0,6849	0,6714	0,6694	0,6318	0,6440	0,6487	0,6655	0,6082
D--34-	12	0,8762	0,9137	0,8720	0,8434	0,8670	0,7883	0,6756	0,7321	0,6640	0,6566	0,6221	0,6395	0,5826	0,6101	0,5777	0,5775	0,5992	0,5805	0,5756
D12-45	27	0,8861	0,9067	0,8842	0,8631	0,8737	0,7757	0,7504	0,7477	0,7786	0,7254	0,7433	0,7206	0,7218	0,7128	0,6893	0,6787	0,6929	0,7102	0,6595
D-2-45	26	0,8749	0,9044	0,8815	0,8616	0,8703	0,7678	0,7438	0,7357	0,7653	0,7098	0,7302	0,7022	0,7069	0,7062	0,6843	0,6728	0,6768	0,6930	0,6483
D12-4-	11	0,8835	0,9015	0,8635	0,8117	0,8675	0,7754	0,7434	0,7436	0,7517	0,7009	0,7228	0,6929	0,6837	0,6828	0,6673	0,6458	0,6678	0,6723	0,6414
D-2-4-	10	0,8716	0,8998	0,8593	0,8068	0,8633	0,7676	0,7361	0,7300	0,7340	0,6828	0,7076	0,6689	0,6671	0,6702	0,6595	0,6348	0,6478	0,6491	0,6247
D123-5	23	0,9296	0,8930	0,8728	0,8941	0,8036	0,8049	0,7619	0,7743	0,7709	0,7525	0,7335	0,7394	0,7276	0,7123	0,6982	0,6846	0,6881	0,7174	0,6735
D-23-5	22	0,9205	0,8891	0,8697	0,8936	0,7985	0,7997	0,7576	0,7660	0,7587	0,7409	0,7179	0,7231	0,7133	0,7013	0,6958	0,6796	0,6725	0,7023	0,6628
D123--	7	0,9287	0,8888	0,8426	0,8534	0,7939	0,8044	0,7592	0,7724	0,7416	0,7335	0,7104	0,7147	0,6878	0,6809	0,6789	0,6538	0,6623	0,6808	0,6608
D-23--	6	0,9189	0,8853	0,8398	0,8505	0,7886	0,7992	0,7536	0,7641	0,7240	0,7203	0,6904	0,6932	0,6724	0,6661	0,6728	0,6448	0,6420	0,6596	0,6472
D1--45	25	0,8041	0,8845	0,8714	0,8338	0,8566	0,7200	0,6572	0,6865	0,7194	0,6822	0,6885	0,6827	0,6837	0,6732	0,6320	0,6419	0,6509	0,6648	0,6184
D1--4-	9	0,7974	0,8772	0,8432	0,7676	0,8483	0,7198	0,6441	0,6771	0,6700	0,6343	0,6467	0,6334	0,6203	0,6105	0,5838	0,5853	0,6099	0,6096	0,5850
D---45	24	0,6857	0,8743	0,8609	0,8215	0,8453	0,6895	0,6226	0,6487	0,6818	0,6391	0,6592	0,6555	0,6672	0,6643	0,6394	0,6522	0,6381	0,6900	0,6481
D1-3-5	21	0,9005	0,8736	0,8609	0,8787	0,7659	0,7670	0,6967	0,7225	0,7139	0,7176	0,6855	0,7066	0,6889	0,6795	0,6395	0,6501	0,6535	0,6766	0,6346
D---4-	8	0,7084	0,8723	0,8355	0,7455	0,8368	0,6905	0,5997	0,6304	0,6109	0,5411	0,5797	0,5591	0,5328	0,5414	0,5027	0,5100	0,5400	0,5228	0,5268
D1-3--	5	0,8989	0,8671	0,8201	0,8273	0,7512	0,7664	0,6893	0,7182	0,6709	0,6883	0,6467	0,6700	0,6299	0,6304	0,6050	0,6004	0,6170	0,6274	0,6148
D--3-5	20	0,8695	0,8633	0,8515	0,8718	0,7440	0,7483	0,6684	0,6980	0,6851	0,6884	0,6552	0,6791	0,6652	0,6524	0,6299	0,6440	0,6370	0,6654	0,6081
D--3--	4	0,8692	0,8587	0,8067	0,8133	0,7275	0,7478	0,6557	0,6932	0,6312	0,6559	0,5876	0,6287	0,5759	0,5831	0,5713	0,5724	0,5735	0,5806	0,5756
D12--5	19	0,8656	0,8074	0,8142	0,8328	0,7575	0,7281	0,7316	0,7182	0,7621	0,7211	0,7267	0,7145	0,7163	0,7026	0,6881	0,6750	0,6782	0,7100	0,6595
D-2--5	18	0,8532	0,7994	0,8084	0,8294	0,7482	0,7174	0,7233	0,7025	0,7451	0,7035	0,7107	0,6977	0,7008	0,6933	0,6826	0,6689	0,6611	0,6932	0,6481
D12---	3	0,8606	0,7921	0,7379	0,7422	0,7411	0,7241	0,7267	0,7147	0,7309	0,6972	0,7030	0,6859	0,6788	0,6707	0,6672	0,6417	0,6491	0,6724	0,6412
D-2---	2	0,8481	0,7851	0,7254	0,7276	0,7313	0,7136	0,7187	0,6977	0,7087	0,6780	0,6828	0,6625	0,6608	0,6543	0,6590	0,6312	0,6267	0,6492	0,6248
D----5	16	0,7794	0,7761	0,8216	0,7927	0,6168	0,5311	0,5791	0,5799	0,6623	0,6390	0,6493	0,6633	0,6650	0,6605	0,6530	0,6607	0,6576	0,6886	0,6494
D1---5	17	0,7667	0,6850	0,7769	0,7925	0,6753	0,6328	0,6210	0,6333	0,6955	0,6739	0,6683	0,6729	0,6769	0,6638	0,6295	0,6399	0,6386	0,6650	0,6188
D1----	1	0,7604	0,6357	0,6408	0,6490	0,6364	0,6241	0,6111	0,6202	0,6298	0,6263	0,6178	0,6137	0,6160	0,5937	0,5821	0,5732	0,5886	0,6096	0,5843

**Tabela 18:** Índice de desempenho AUC para as combinações de decomposições apresentadas na tabela 15 para as wavelets do grupo 1 apresentadas na tabela 16. Nesta tabela são apresentadas as wavelets da família Daubechies de DB20 a DB38. As combinações de níveis de decomposições estão ordenadas em ordem decrescente de índice AUC de acordo com a wavelet-mãe DB2 apresentada na tabela 17 (complementa a tabela 17).

Decomposições	Combinação	DB20	DB21	DB22	DB23	DB24	DB25	DB26	DB27	DB28	DB29	DB30	DB31	DB32	DB33	DB34	DB35	DB36	DB37	DB38
D12345	31	0,6818	0,6871	0,6741	0,6705	0,6814	0,6770	0,6524	0,6637	0,6526	0,6422	0,6325	0,6313	0,6472	0,6411	0,6333	0,6545	0,6431	0,6171	0,5962
D1234-	15	0,6346	0,6609	0,6504	0,6590	0,6605	0,6485	0,6487	0,6279	0,6353	0,6233	0,6100	0,6162	0,6340	0,6330	0,6149	0,6369	0,6143	0,6107	0,5820
D-2345	30	0,6813	0,6830	0,6709	0,6615	0,6769	0,6732	0,6475	0,6599	0,6498	0,6390	0,6222	0,6275	0,6446	0,6373	0,6242	0,6384	0,6313	0,6067	0,5885
D-234-	14	0,6341	0,6519	0,6478	0,6488	0,6554	0,6433	0,6426	0,6232	0,6309	0,6184	0,6008	0,6138	0,6307	0,6288	0,6053	0,6220	0,5953	0,5997	0,5708
D1-345	29	0,6555	0,6507	0,6173	0,6284	0,6279	0,6339	0,6116	0,6437	0,6073	0,6078	0,6179	0,5976	0,6116	0,5891	0,6003	0,6098	0,6262	0,5815	0,5793
D1-34-	13	0,5696	0,6057	0,5726	0,6093	0,5778	0,5808	0,6016	0,5684	0,5773	0,5800	0,5840	0,5721	0,5821	0,5727	0,5717	0,5836	0,5864	0,5692	0,5534
D--345	28	0,6627	0,6400	0,6154	0,6075	0,6227	0,6402	0,5984	0,6526	0,5997	0,6007	0,6124	0,5957	0,6040	0,5816	0,5988	0,5960	0,6364	0,5713	0,5845
D--34-	12	0,5362	0,5731	0,5492	0,5810	0,5490	0,5642	0,5844	0,5506	0,5547	0,5325	0,5714	0,5600	0,5658	0,5473	0,5328	0,5568	0,5225	0,5468	0,5249
D12-45	27	0,6818	0,6829	0,6680	0,6616	0,6790	0,6677	0,6302	0,6617	0,6436	0,6392	0,6169	0,6247	0,6384	0,6384	0,6300	0,6488	0,6421	0,6103	0,5922
D-2-45	26	0,6813	0,6793	0,6638	0,6517	0,6738	0,6640	0,6240	0,6569	0,6394	0,6375	0,6089	0,6212	0,6340	0,6339	0,6197	0,6317	0,6298	0,5982	0,5845
D12-4-	11	0,6337	0,6538	0,6440	0,6462	0,6573	0,6388	0,6233	0,6235	0,6258	0,6202	0,5906	0,6074	0,6239	0,6290	0,6119	0,6325	0,6118	0,6052	0,5794
D-2-4-	10	0,6335	0,6465	0,6403	0,6347	0,6514	0,6322	0,6155	0,6164	0,6211	0,6170	0,5806	0,6040	0,6186	0,6240	0,5994	0,6166	0,5928	0,5931	0,5658
D123-5	23	0,6764	0,6832	0,6739	0,6656	0,6804	0,6770	0,6506	0,6636	0,6518	0,6276	0,6266	0,6312	0,6437	0,6371	0,6324	0,6536	0,6412	0,6162	0,5933
D-23-5	22	0,6744	0,6793	0,6709	0,6550	0,6763	0,6732	0,6457	0,6599	0,6489	0,6243	0,6185	0,6276	0,6408	0,6337	0,6239	0,6377	0,6299	0,6046	0,5830
D123--	7	0,6295	0,6574	0,6504	0,6526	0,6586	0,6485	0,6460	0,6279	0,6338	0,6069	0,6027	0,6164	0,6291	0,6271	0,6145	0,6370	0,6118	0,6090	0,5759
D-23--	6	0,6269	0,6483	0,6479	0,6387	0,6538	0,6433	0,6404	0,6232	0,6295	0,6017	0,5953	0,6140	0,6247	0,6231	0,6048	0,6222	0,5913	0,5961	0,5643
D1--45	25	0,6568	0,6433	0,6214	0,6106	0,6280	0,6299	0,5862	0,6479	0,6045	0,6079	0,6082	0,5816	0,5996	0,5803	0,5948	0,6022	0,6233	0,5742	0,5745
D1--4-	9	0,5652	0,5751	0,5554	0,5829	0,5684	0,5555	0,5647	0,5561	0,5570	0,5773	0,5551	0,5436	0,5611	0,5615	0,5608	0,5727	0,5808	0,5593	0,5500
D---45	24	0,6686	0,6699	0,6626	0,5994	0,6467	0,6775	0,5954	0,6962	0,6381	0,6031	0,6269	0,6387	0,6124	0,5826	0,6543	0,6226	0,6498	0,6072	0,6027
D1-3-5	21	0,6529	0,6495	0,6171	0,6242	0,6300	0,6337	0,6101	0,6439	0,6069	0,5974	0,6151	0,5974	0,6051	0,5816	0,5997	0,6064	0,6228	0,5812	0,5779
D---4-	8	0,5018	0,5141	0,5598	0,5013	0,5258	0,5476	0,5243	0,5524	0,5381	0,5180	0,5188	0,5755	0,5341	0,5410	0,5551	0,5663	0,5521	0,5843	0,5511
D1-3--	5	0,5611	0,6056	0,5723	0,6043	0,5776	0,5807	0,5969	0,5684	0,5756	0,5607	0,5764	0,5719	0,5767	0,5602	0,5721	0,5816	0,5834	0,5693	0,5497
D--3-5	20	0,6655	0,6390	0,6153	0,6094	0,6310	0,6398	0,5971	0,6526	0,6018	0,6243	0,6133	0,5968	0,5970	0,5907	0,5993	0,6006	0,6533	0,5775	0,5869
D--3--	4	0,5168	0,5726	0,5494	0,5712	0,5509	0,5643	0,5753	0,5505	0,5552	0,5187	0,5676	0,5610	0,5565	0,5370	0,5320	0,5572	0,5268	0,5383	0,5177
D12--5	19	0,6768	0,6782	0,6679	0,6574	0,6777	0,6677	0,6270	0,6617	0,6419	0,6249	0,6097	0,6243	0,6342	0,6349	0,6290	0,6480	0,6400	0,6096	0,5897
D-2--5	18	0,6751	0,6751	0,6637	0,6460	0,6725	0,6639	0,6199	0,6569	0,6383	0,6220	0,6040	0,6208	0,6307	0,6309	0,6190	0,6309	0,6286	0,5969	0,5798
D12---	3	0,6287	0,6490	0,6440	0,6398	0,6551	0,6386	0,6203	0,6235	0,6242	0,6052	0,5817	0,6074	0,6157	0,6241	0,6113	0,6326	0,6093	0,6041	0,5739
D-2---	2	0,6263	0,6421	0,6407	0,6244	0,6487	0,6322	0,6125	0,6165	0,6198	0,6003	0,5729	0,6044	0,6092	0,6188	0,5988	0,6166	0,5891	0,5894	0,5584
D----5	16	0,6746	0,6791	0,6629	0,6563	0,6703	0,6776	0,6386	0,6960	0,6582	0,6536	0,6774	0,6490	0,6548	0,6527	0,6630	0,6523	0,6822	0,6551	0,6553
D1----5	17	0,6553	0,6401	0,6216	0,6075	0,6310	0,6294	0,5853	0,6480	0,6023	0,5987	0,6122	0,5816	0,5940	0,5761	0,5932	0,5999	0,6200	0,5736	0,5754
D1----	1	0,5548	0,5706	0,5550	0,5736	0,5659	0,5554	0,5602	0,5562	0,5526	0,5594	0,5431	0,5427	0,5532	0,5519	0,5595	0,5716	0,5779	0,5596	0,5464

**Tabela 19:** Índice de desempenho AUC para as combinações de decomposições apresentadas na tabela 15 para as wavelets do grupo 2 apresentadas na tabela 16. Nesta tabela são apresentadas as wavelets da família Biortogonal. As combinações de níveis de decomposições estão ordenadas em ordem decrescente de índice AUC de acordo com a wavelet-mãe Bior15.

Decomposições	Combinação	Bior11	Bior13	Bior15	Bior22	Bior24	Bior26	Bior28	Bior31	Bior33	Bior35	Bior37	Bior39	Bior44	Bior55	Bior68
D12345	31	0,9332	0,9409	0,9434	0,9015	0,8781	0,8230	0,7594	0,8521	0,8649	0,7751	0,7459	0,7410	0,8783	0,8392	0,7436
D1234-	15	0,9323	0,9401	0,9428	0,8900	0,8706	0,8212	0,7529	0,7603	0,8157	0,7663	0,7459	0,7350	0,8650	0,8375	0,7421
D123-5	23	0,9296	0,9390	0,9390	0,8539	0,8520	0,8182	0,7479	0,8401	0,8395	0,7271	0,6956	0,7041	0,8365	0,8164	0,7353
D123--	7	0,9287	0,9378	0,9384	0,8197	0,8419	0,8160	0,7407	0,7360	0,7723	0,7161	0,6956	0,6993	0,8152	0,8135	0,7332
D-2345	30	0,9248	0,9327	0,9290	0,9007	0,8755	0,8226	0,7593	0,8479	0,8616	0,7715	0,7442	0,7402	0,8750	0,8330	0,7299
D-234-	14	0,9234	0,9320	0,9281	0,8885	0,8676	0,8206	0,7529	0,7549	0,8146	0,7623	0,7441	0,7342	0,8611	0,8315	0,7284
D12-45	27	0,8861	0,9257	0,9270	0,8950	0,8148	0,7601	0,7103	0,8339	0,8447	0,7534	0,7237	0,7108	0,8452	0,7676	0,7212
D12-4-	11	0,8835	0,9235	0,9267	0,8793	0,8018	0,7559	0,6996	0,7183	0,7768	0,7454	0,7236	0,7023	0,8270	0,7664	0,7201
D-23-5	22	0,9205	0,9303	0,9247	0,8533	0,8492	0,8177	0,7479	0,8364	0,8350	0,7222	0,6906	0,7036	0,8318	0,8101	0,7208
D-23--	6	0,9189	0,9294	0,9241	0,8190	0,8392	0,8154	0,7407	0,7301	0,7696	0,7109	0,6905	0,6987	0,8078	0,8070	0,7189
D12--5	19	0,8656	0,9192	0,9185	0,8123	0,7586	0,7447	0,6834	0,8176	0,8062	0,6928	0,6611	0,6467	0,7749	0,7310	0,7077
D12---	3	0,8606	0,9157	0,9184	0,7430	0,7287	0,7401	0,6703	0,6736	0,6924	0,6816	0,6611	0,6401	0,7314	0,7265	0,7033
D-2-45	26	0,8749	0,9105	0,9077	0,8946	0,8144	0,7598	0,7101	0,8311	0,8413	0,7514	0,7227	0,7087	0,8417	0,7551	0,7060
D-2-4-	10	0,8716	0,9074	0,9075	0,8784	0,8013	0,7554	0,6996	0,7156	0,7748	0,7425	0,7230	0,7003	0,8232	0,7533	0,7044
D1-345	29	0,9067	0,9179	0,9073	0,8864	0,8607	0,7805	0,7281	0,8418	0,8561	0,7517	0,7189	0,7218	0,8651	0,8224	0,6942
D1-34-	13	0,9052	0,9159	0,9067	0,8693	0,8505	0,7770	0,7183	0,7294	0,7992	0,7371	0,7187	0,7148	0,8512	0,8206	0,6910
D-2--5	18	0,8532	0,9025	0,8992	0,8123	0,7585	0,7446	0,6836	0,8155	0,8010	0,6896	0,6576	0,6450	0,7656	0,7142	0,6914
D-2---	2	0,8481	0,8974	0,8992	0,7434	0,7288	0,7399	0,6706	0,6711	0,6879	0,6779	0,6575	0,6387	0,7177	0,7091	0,6878
D1-3-5	21	0,9005	0,9130	0,8960	0,8273	0,8252	0,7706	0,7119	0,8263	0,8226	0,6830	0,6504	0,6735	0,8126	0,7931	0,6828
D1-3--	5	0,8989	0,9103	0,8956	0,7776	0,8114	0,7664	0,7024	0,6919	0,7388	0,6632	0,6504	0,6698	0,7895	0,7886	0,6772
D--345	28	0,8794	0,9011	0,8752	0,8839	0,8573	0,7792	0,7275	0,8385	0,8532	0,7487	0,7187	0,7213	0,8563	0,8098	0,6659
D--34-	12	0,8762	0,8978	0,8745	0,8665	0,8468	0,7755	0,7180	0,7251	0,7993	0,7339	0,7185	0,7141	0,8418	0,8084	0,6624
D--3-5	20	0,8695	0,8946	0,8609	0,8244	0,8210	0,7696	0,7117	0,8240	0,8197	0,6770	0,6479	0,6725	0,7988	0,7802	0,6434
D--3--	4	0,8692	0,8928	0,8609	0,7791	0,8084	0,7655	0,7028	0,6914	0,7385	0,6572	0,6497	0,6707	0,7713	0,7760	0,6380
D1--45	25	0,8041	0,8375	0,8108	0,8739	0,7671	0,6442	0,6395	0,8189	0,8252	0,7121	0,6771	0,6717	0,8178	0,7136	0,6528
D1--4-	9	0,7974	0,8217	0,8107	0,8482	0,7438	0,6272	0,6238	0,6507	0,7384	0,6932	0,6762	0,6636	0,7970	0,7123	0,6479
D1---5	17	0,7667	0,7849	0,7576	0,7425	0,6437	0,5239	0,5229	0,7962	0,7646	0,5624	0,5333	0,5169	0,7088	0,6375	0,6262
D1----	1	0,7604	0,7577	0,7571	0,5347	0,5354	0,5341	0,5317	0,5381	0,5434	0,5392	0,5369	0,5289	0,6348	0,6302	0,6111
D---4-	8	0,7084	0,7521	0,7351	0,8530	0,7512	0,6388	0,6346	0,6701	0,7424	0,6939	0,6782	0,6676	0,7856	0,6763	0,6015
D---45	24	0,6857	0,7755	0,7271	0,8713	0,7664	0,6418	0,6343	0,8209	0,8229	0,7100	0,6774	0,6700	0,8044	0,6775	0,6055
D----5	16	0,7794	0,7576	0,5608	0,8067	0,6727	0,5532	0,5693	0,8710	0,7895	0,5953	0,5258	0,5304	0,6805	0,5201	0,5096

**Tabela 20:** Índice de desempenho AUC para as combinações de decomposições apresentadas na tabela 15 para as wavelets do grupo 3 apresentadas na tabela 16. Nesta tabela são apresentadas as wavelets da família Reverse Biortogonal. As combinações de níveis de decomposições estão ordenadas em ordem decrescente de índice AUC de acordo com a wavelet-mãe RBior35.

Decomposições	Combinação	RBior11	RBior13	RBior15	RBior22	RBior24	RBior26	RBior28	RBior31	RBior33	RBior35	RBior37	RBior39	RBior44	RBior55	RBior68
D12345	31	0,9332	0,9187	0,8727	0,9452	0,8904	0,7996	0,7284	0,9579	0,9296	0,8491	0,7567	0,7552	0,8821	0,8265	0,7385
D1234-	15	0,9323	0,8784	0,8597	0,9413	0,8767	0,7994	0,7228	0,9568	0,9281	0,8476	0,7499	0,7434	0,8703	0,8255	0,7385
D-2345	30	0,9248	0,9171	0,8650	0,9418	0,8881	0,7889	0,7186	0,9469	0,9254	0,8402	0,7466	0,7432	0,8799	0,8216	0,7245
D-234-	14	0,9234	0,8754	0,8531	0,9383	0,8749	0,7888	0,7118	0,9459	0,9238	0,8384	0,7389	0,7311	0,8684	0,8209	0,7243
D123-5	23	0,9296	0,9070	0,8317	0,9298	0,8661	0,7911	0,7240	0,9579	0,9129	0,8157	0,7508	0,7386	0,8510	0,8113	0,7334
D12-45	27	0,8861	0,8791	0,8409	0,9311	0,8451	0,7605	0,7167	0,9556	0,8830	0,8137	0,7449	0,7486	0,8366	0,7640	0,7178
D123--	7	0,9287	0,8581	0,8155	0,9162	0,8421	0,7911	0,7186	0,9567	0,9069	0,8136	0,7458	0,7280	0,8305	0,8099	0,7333
D12-4-	11	0,8835	0,8175	0,8252	0,9262	0,8234	0,7602	0,7110	0,9546	0,8778	0,8112	0,7366	0,7371	0,8206	0,7603	0,7178
D1-345	29	0,9067	0,9119	0,8572	0,9313	0,8709	0,7491	0,6677	0,9292	0,9268	0,8103	0,6871	0,6946	0,8667	0,8061	0,6815
D1-34-	13	0,9052	0,8638	0,8411	0,9258	0,8548	0,7487	0,6633	0,9276	0,9252	0,8081	0,6673	0,6818	0,8528	0,8058	0,6817
D-23-5	22	0,9205	0,9055	0,8203	0,9240	0,8626	0,7804	0,7143	0,9467	0,9076	0,8063	0,7383	0,7250	0,8471	0,8048	0,7198
D-23--	6	0,9189	0,8542	0,8057	0,9095	0,8389	0,7804	0,7082	0,9456	0,9008	0,8040	0,7332	0,7144	0,8269	0,8034	0,7198
D-2-45	26	0,8749	0,8754	0,8324	0,9257	0,8423	0,7477	0,7033	0,9417	0,8738	0,8019	0,7316	0,7353	0,8321	0,7552	0,7010
D-2-4-	10	0,8716	0,8109	0,8175	0,9206	0,8205	0,7477	0,6967	0,9406	0,8682	0,7984	0,7231	0,7239	0,8151	0,7513	0,7009
D--345	28	0,8794	0,9086	0,8423	0,9244	0,8617	0,7134	0,6327	0,8991	0,9199	0,7786	0,6433	0,6539	0,8579	0,7897	0,6360
D--34-	12	0,8762	0,8550	0,8250	0,9188	0,8447	0,7125	0,6237	0,8958	0,9179	0,7760	0,6148	0,6315	0,8435	0,7897	0,6360
D1-3-5	21	0,9005	0,8983	0,8110	0,9072	0,8389	0,7290	0,6525	0,9290	0,9057	0,7660	0,6777	0,6677	0,8282	0,7865	0,6728
D1-3--	5	0,8989	0,8380	0,7907	0,8891	0,8104	0,7289	0,6497	0,9271	0,8992	0,7639	0,6618	0,6543	0,8021	0,7862	0,6725
D12--5	19	0,8656	0,8397	0,7547	0,8966	0,7892	0,7449	0,7067	0,9552	0,8252	0,7592	0,7376	0,7296	0,7800	0,7298	0,7088
D12---	3	0,8606	0,7409	0,7265	0,8710	0,7378	0,7448	0,7027	0,9541	0,7989	0,7559	0,7317	0,7185	0,7360	0,7227	0,7087
D1--45	25	0,8041	0,8630	0,8111	0,8995	0,8029	0,6839	0,6424	0,8867	0,8600	0,7531	0,6648	0,6772	0,8059	0,7126	0,6475
D1--4-	9	0,7974	0,7708	0,7899	0,8895	0,7750	0,6825	0,6357	0,8832	0,8546	0,7492	0,6417	0,6620	0,7854	0,7087	0,6480
D-2--5	18	0,8532	0,8354	0,7340	0,8836	0,7827	0,7320	0,6942	0,9404	0,8118	0,7431	0,7217	0,7149	0,7732	0,7144	0,6924
D-2---	2	0,8481	0,7313	0,7070	0,8506	0,7266	0,7321	0,6893	0,9394	0,7823	0,7396	0,7158	0,7043	0,7256	0,7073	0,6925
D--3-5	20	0,8695	0,8945	0,7884	0,8934	0,8262	0,6877	0,6151	0,8970	0,8957	0,7300	0,6295	0,6141	0,8139	0,7641	0,6190
D--3--	4	0,8692	0,8305	0,7652	0,8743	0,7940	0,6876	0,6076	0,8949	0,8872	0,7277	0,6034	0,5863	0,7865	0,7641	0,6187
D---45	24	0,6857	0,8574	0,7891	0,8813	0,7877	0,6179	0,5903	0,7148	0,8341	0,7024	0,6054	0,6274	0,7852	0,6716	0,5697
D---4-	8	0,7084	0,7622	0,7701	0,8739	0,7583	0,6140	0,5755	0,7280	0,8307	0,6966	0,5716	0,5930	0,7645	0,6701	0,5701
D1---5	17	0,7667	0,8061	0,6834	0,8076	0,7243	0,6362	0,6054	0,8778	0,7395	0,6613	0,6480	0,6432	0,7143	0,6515	0,6267
D1----	1	0,7604	0,6046	0,6336	0,7202	0,6396	0,6365	0,6037	0,8726	0,6880	0,6537	0,6288	0,6280	0,6447	0,6434	0,6249
D----5	16	0,7794	0,8482	0,6336	0,8078	0,6951	0,5394	0,5649	0,7649	0,6967	0,5374	0,5851	0,5874	0,6788	0,5285	0,5356

**Tabela 21:** Índice de desempenho AUC para as combinações de decomposições apresentadas na tabela 15 para as wavelets do grupo 4 apresentadas na tabela 16. Nesta tabela são apresentadas as wavelets da família Symlet. As combinações de níveis de decomposições estão ordenadas em ordem decrescente de índice AUC de acordo com a wavelet-mãe Sym2.

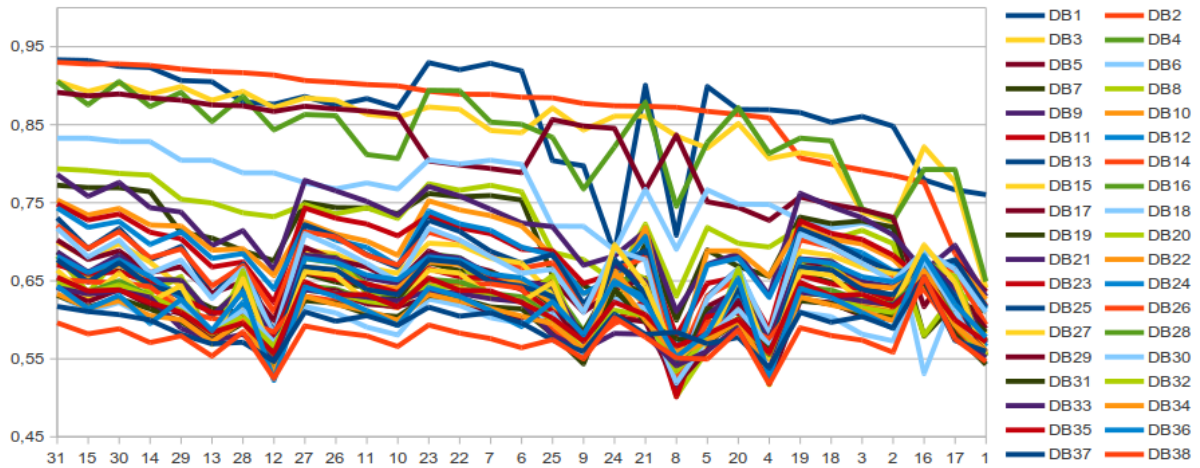
Decomposições	Combinação	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7	Sym8	Sym9	Sym10	Sym11	Sym12	Sym13	Sym14	Sym15	Sym16	Sym17	Sym18	Sym19	Sym20
D12345	31	0,9300	0,9057	0,9091	0,8740	0,8189	0,7699	0,7662	0,7651	0,7312	0,7585	0,6938	0,7277	0,7406	0,6967	0,7219	0,6962	0,7166	0,7066	0,6982
D1234-	15	0,9279	0,8927	0,8874	0,8712	0,8183	0,7702	0,7664	0,7619	0,7306	0,7588	0,6901	0,7263	0,7397	0,6965	0,7201	0,6959	0,7166	0,7056	0,6959
D-2345	30	0,9278	0,9032	0,9081	0,8714	0,8155	0,7567	0,7573	0,7593	0,7178	0,7478	0,6794	0,7223	0,7299	0,6839	0,7095	0,6886	0,7104	0,7013	0,6874
D-234-	14	0,9260	0,8892	0,8851	0,8684	0,8147	0,7567	0,7574	0,7569	0,7172	0,7477	0,6752	0,7211	0,7288	0,6838	0,7070	0,6886	0,7110	0,7001	0,6841
D1-345	29	0,9214	0,8989	0,8977	0,8538	0,7918	0,7183	0,7194	0,7275	0,6621	0,7224	0,6565	0,6749	0,6844	0,6466	0,6887	0,6404	0,6695	0,6611	0,6615
D1-34-	13	0,9183	0,8813	0,8697	0,8506	0,7919	0,7179	0,7193	0,7244	0,6615	0,7209	0,6535	0,6724	0,6825	0,6447	0,6873	0,6386	0,6702	0,6567	0,6573
D--345	28	0,9167	0,8927	0,8932	0,8452	0,7817	0,6839	0,6924	0,7128	0,6191	0,6979	0,6241	0,6612	0,6620	0,6154	0,6641	0,6244	0,6559	0,6466	0,6358
D--34-	12	0,9137	0,8720	0,8617	0,8406	0,7818	0,6835	0,6924	0,7104	0,6178	0,6967	0,6214	0,6571	0,6593	0,6108	0,6615	0,6206	0,6555	0,6421	0,6220
D12-45	27	0,9067	0,8842	0,8940	0,8617	0,7606	0,7575	0,7318	0,7404	0,7222	0,7403	0,6873	0,7068	0,7256	0,6934	0,7025	0,6845	0,7137	0,6805	0,6939
D-2-45	26	0,9044	0,8815	0,8931	0,8600	0,7524	0,7447	0,7194	0,7326	0,7087	0,7271	0,6728	0,7039	0,7127	0,6810	0,6909	0,6770	0,7080	0,6692	0,6825
D12-4-	11	0,9015	0,8635	0,8664	0,8574	0,7593	0,7578	0,7319	0,7393	0,7210	0,7400	0,6837	0,7052	0,7244	0,6931	0,6999	0,6846	0,7137	0,6791	0,6917
D-2-4-	10	0,8998	0,8593	0,8643	0,8553	0,7509	0,7445	0,7194	0,7316	0,7072	0,7269	0,6685	0,7022	0,7117	0,6806	0,6878	0,6772	0,7082	0,6682	0,6796
D123-5	23	0,8930	0,8728	0,8890	0,7952	0,8009	0,7498	0,7609	0,7415	0,7260	0,7363	0,6801	0,7106	0,7155	0,6832	0,7059	0,6854	0,6729	0,6966	0,6701
D-23-5	22	0,8891	0,8697	0,8871	0,7930	0,7966	0,7327	0,7526	0,7314	0,7118	0,7155	0,6637	0,7027	0,7050	0,6680	0,6880	0,6776	0,6651	0,6899	0,6518
D123--	7	0,8888	0,8426	0,8346	0,7879	0,7992	0,7496	0,7609	0,7371	0,7249	0,7339	0,6764	0,7093	0,7144	0,6824	0,7035	0,6850	0,6732	0,6961	0,6615
D-23--	6	0,8853	0,8398	0,8292	0,7846	0,7949	0,7320	0,7526	0,7268	0,7104	0,7140	0,6600	0,7019	0,7037	0,6674	0,6848	0,6777	0,6651	0,6896	0,6422
D1--45	25	0,8845	0,8714	0,8772	0,8316	0,6966	0,6813	0,6517	0,6774	0,6370	0,6835	0,6357	0,6311	0,6568	0,6340	0,6536	0,6087	0,6570	0,6134	0,6514
D1--4-	9	0,8772	0,8432	0,8393	0,8255	0,6966	0,6805	0,6516	0,6755	0,6363	0,6812	0,6303	0,6250	0,6549	0,6320	0,6517	0,6063	0,6574	0,6056	0,6462
D---45	24	0,8743	0,8609	0,8710	0,8218	0,6620	0,6266	0,5889	0,6447	0,5792	0,6377	0,5850	0,5979	0,6037	0,5787	0,6202	0,5763	0,6240	0,5615	0,6073
D1-3-5	21	0,8736	0,8609	0,8688	0,7387	0,7653	0,6910	0,7024	0,6942	0,6532	0,6837	0,6343	0,6507	0,6466	0,6185	0,6623	0,6203	0,5942	0,6442	0,6232
D---4-	8	0,8723	0,8355	0,8333	0,8156	0,6609	0,6292	0,5899	0,6442	0,5761	0,6334	0,5738	0,5830	0,5979	0,5649	0,5989	0,5554	0,6240	0,5221	0,5742
D1-3--	5	0,8671	0,8201	0,7928	0,7306	0,7635	0,6900	0,7024	0,6905	0,6515	0,6816	0,6327	0,6487	0,6454	0,6175	0,6605	0,6186	0,5941	0,6400	0,6144
D--3-5	20	0,8633	0,8515	0,8611	0,7169	0,7509	0,6468	0,6707	0,6709	0,6050	0,6390	0,5958	0,6238	0,6235	0,5820	0,6309	0,6030	0,5672	0,6310	0,5930
D--3--	4	0,8587	0,8067	0,7729	0,7067	0,7497	0,6454	0,6706	0,6675	0,6019	0,6360	0,5950	0,6210	0,6211	0,5775	0,6270	0,5957	0,5657	0,6226	0,5642
D12--5	19	0,8074	0,8142	0,8536	0,7579	0,7303	0,7315	0,7248	0,7068	0,7156	0,7132	0,6713	0,6856	0,6924	0,6765	0,6741	0,6709	0,6693	0,6681	0,6633
D-2--5	18	0,7994	0,8084	0,8519	0,7525	0,7195	0,7146	0,7135	0,6930	0,7007	0,6875	0,6571	0,6792	0,6799	0,6602	0,6538	0,6636	0,6617	0,6560	0,6439
D12---	3	0,7921	0,7379	0,7634	0,7473	0,7274	0,7315	0,7246	0,7054	0,7128	0,7113	0,6681	0,6849	0,6920	0,6755	0,6703	0,6709	0,6694	0,6676	0,6538
D-2---	2	0,7851	0,7254	0,7561	0,7407	0,7173	0,7135	0,7136	0,6915	0,6979	0,6866	0,6533	0,6783	0,6791	0,6597	0,6495	0,6635	0,6616	0,6562	0,6323
D----5	16	0,7761	0,8216	0,8254	0,6138	0,5199	0,5460	0,5423	0,5566	0,5339	0,5970	0,5803	0,6025	0,5919	0,6276	0,6123	0,6481	0,6269	0,6589	0,6538
D1---5	17	0,6850	0,7769	0,8151	0,6456	0,6291	0,6260	0,6158	0,6222	0,6181	0,6328	0,6019	0,5921	0,5937	0,5958	0,6014	0,5762	0,5688	0,5898	0,6043
D1----	1	0,6357	0,6408	0,6257	0,6245	0,6242	0,6265	0,6150	0,6204	0,6149	0,6321	0,5990	0,5886	0,5919	0,5969	0,5970	0,5690	0,5683	0,5846	0,5918



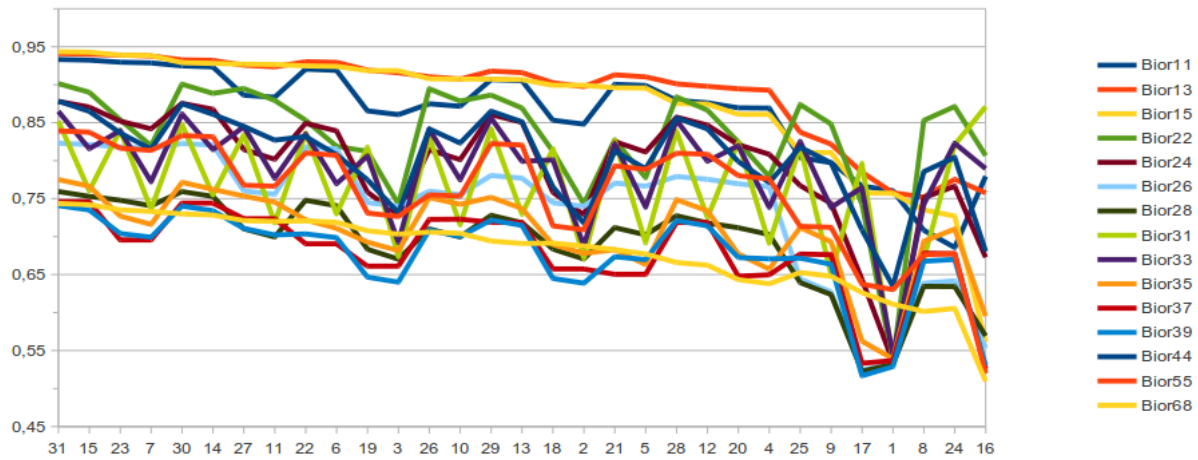
**Tabela 22:** Índice de desempenho AUC para as combinações de decomposições apresentadas na tabela 15 para as wavelets do grupo 5 apresentadas na tabela 16. Nesta tabela são apresentadas as wavelets da família Coiflet, Dmeyer, Haar e Legendre. As combinações de níveis de decomposições estão ordenadas em ordem decrescente de índice AUC de acordo com a wavelet-mãe Haar Ortogonal.

Decomposições	Combinação	Coif1	Coif2	Coif3	Coif4	Coif5	Dmey	Haar Ort.	Haar	Lege2	Lege4	Lege6
D12345	31	0,9226	0,8438	0,7413	0,7640	0,7384	0,6842	0,9332	0,9332	0,9332	0,9476	0,8852
D1234-	15	0,9152	0,8425	0,7410	0,7611	0,7381	0,6627	0,9323	0,9323	0,9323	0,9476	0,8818
D123-5	23	0,8962	0,8354	0,7331	0,7357	0,6951	0,6739	0,9296	0,9296	0,9296	0,9476	0,8811
D123--	7	0,8753	0,8334	0,7323	0,7343	0,6945	0,6546	0,9287	0,9287	0,9287	0,9476	0,8748
D-2345	30	0,9221	0,8352	0,7323	0,7575	0,7346	0,6840	0,9248	0,9248	0,9248	0,9279	0,8453
D-234-	14	0,9140	0,8337	0,7317	0,7545	0,7341	0,6616	0,9234	0,9234	0,9234	0,9276	0,8410
D-23-5	22	0,8944	0,8269	0,7213	0,7264	0,6876	0,6741	0,9205	0,9205	0,9205	0,9276	0,8374
D-23--	6	0,8724	0,8250	0,7203	0,7247	0,6870	0,6539	0,9189	0,9189	0,9189	0,9272	0,8285
D1-345	29	0,9064	0,8120	0,6790	0,7329	0,6889	0,6286	0,9067	0,9067	0,9067	0,9321	0,8722
D1-34-	13	0,8966	0,8094	0,6785	0,7291	0,6887	0,5913	0,9052	0,9052	0,9052	0,9323	0,8692
D1-3-5	21	0,8706	0,7988	0,6603	0,6934	0,6152	0,6097	0,9005	0,9005	0,9005	0,9320	0,8684
D1-3--	5	0,8392	0,7960	0,6592	0,6907	0,6155	0,5665	0,8989	0,8989	0,8989	0,9321	0,8623
D12-45	27	0,9084	0,7777	0,7291	0,7392	0,7264	0,6824	0,8861	0,8861	0,8861	0,9358	0,8710
D12-4-	11	0,8992	0,7761	0,7289	0,7359	0,7257	0,6586	0,8835	0,8835	0,8835	0,9356	0,8650
D--345	28	0,9029	0,7875	0,6500	0,7160	0,6722	0,6309	0,8794	0,8794	0,8794	0,8856	0,8035
D--34-	12	0,8913	0,7842	0,6496	0,7108	0,6718	0,5844	0,8762	0,8762	0,8762	0,8849	0,7996
D-2-45	26	0,9080	0,7633	0,7180	0,7322	0,7237	0,6818	0,8749	0,8749	0,8749	0,9030	0,8101
D-2-4-	10	0,8981	0,7618	0,7177	0,7287	0,7229	0,6581	0,8716	0,8716	0,8716	0,9019	0,8025
D--3-5	20	0,8613	0,7721	0,6187	0,6665	0,5796	0,6171	0,8695	0,8695	0,8695	0,8842	0,7942
D--3--	4	0,8261	0,7682	0,6180	0,6640	0,5790	0,5545	0,8692	0,8692	0,8692	0,8840	0,7848
D12--5	19	0,8565	0,7536	0,7194	0,6973	0,6760	0,6710	0,8656	0,8656	0,8656	0,9347	0,8615
D12---	3	0,8208	0,7520	0,7189	0,6953	0,6752	0,6506	0,8606	0,8606	0,8606	0,9342	0,8468
D-2--5	18	0,8538	0,7376	0,7049	0,6840	0,6693	0,6707	0,8532	0,8532	0,8532	0,8973	0,7845
D-2---	2	0,8145	0,7362	0,7044	0,6814	0,6682	0,6496	0,8481	0,8481	0,8481	0,8947	0,7610
D1--45	25	0,8809	0,7105	0,6470	0,6919	0,6707	0,6229	0,8041	0,8041	0,8041	0,8842	0,8443
D1--4-	9	0,8638	0,7044	0,6462	0,6835	0,6705	0,5678	0,7974	0,7974	0,7974	0,8841	0,8379
D----5	16	0,8105	0,5278	0,5343	0,6104	0,5980	0,6993	0,7794	0,7794	0,7794	0,7456	0,7297
D1----	17	0,7764	0,6486	0,6196	0,6128	0,5822	0,6010	0,7667	0,7667	0,7667	0,8774	0,8313
D----4-	1	0,6550	0,6415	0,6185	0,6085	0,5820	0,5364	0,7604	0,7604	0,7604	0,8764	0,8163
D----4-	8	0,8627	0,6575	0,5956	0,6476	0,6381	0,5146	0,7084	0,7084	0,7084	0,7118	0,6975
D---45	24	0,8741	0,6575	0,5958	0,6676	0,6378	0,6480	0,6857	0,6857	0,6857	0,6901	0,6928

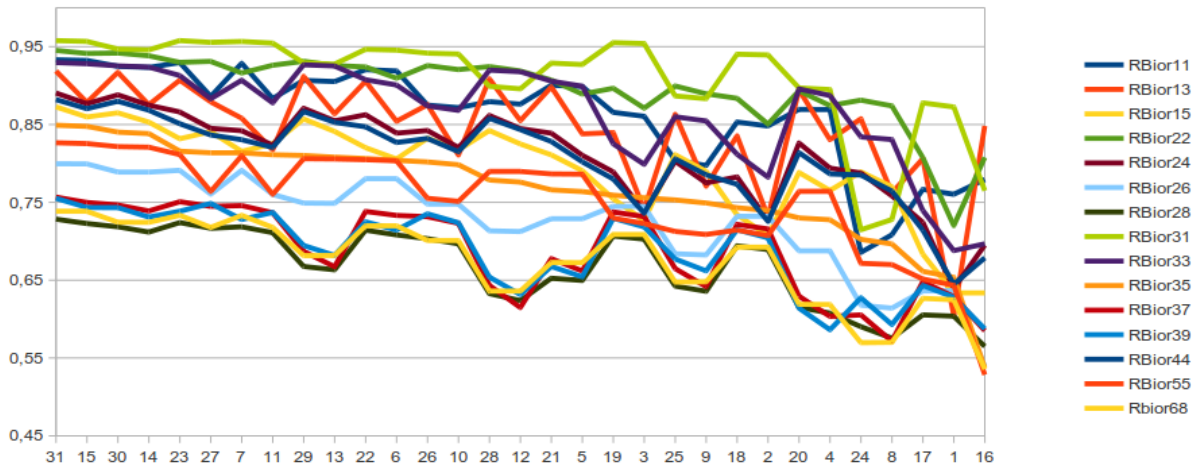
A figura 38 apresenta o índice de desempenho AUC obtido apresentado pelas tabelas 17 a 22. Os gráficos estão ordenados de maneira decrescente pelo índice de desempenho AUC, da esquerda para a direita, pelas combinações que apresentaram na média o melhor desempenho. Os resultados foram reunidos em grupos de 1 a 5 de acordo com a classificação da tabela 16.



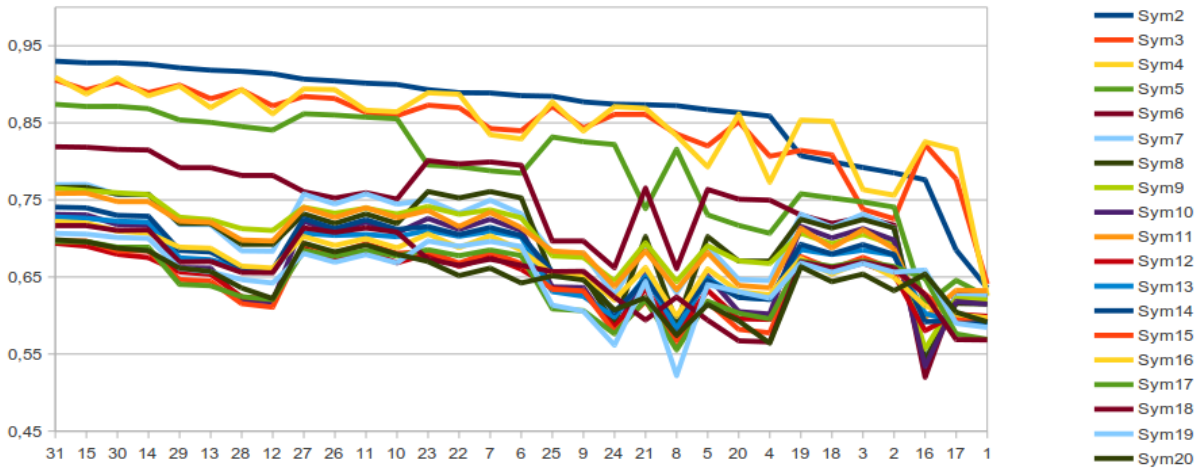
(a) Grupo 1: Família Daubechies



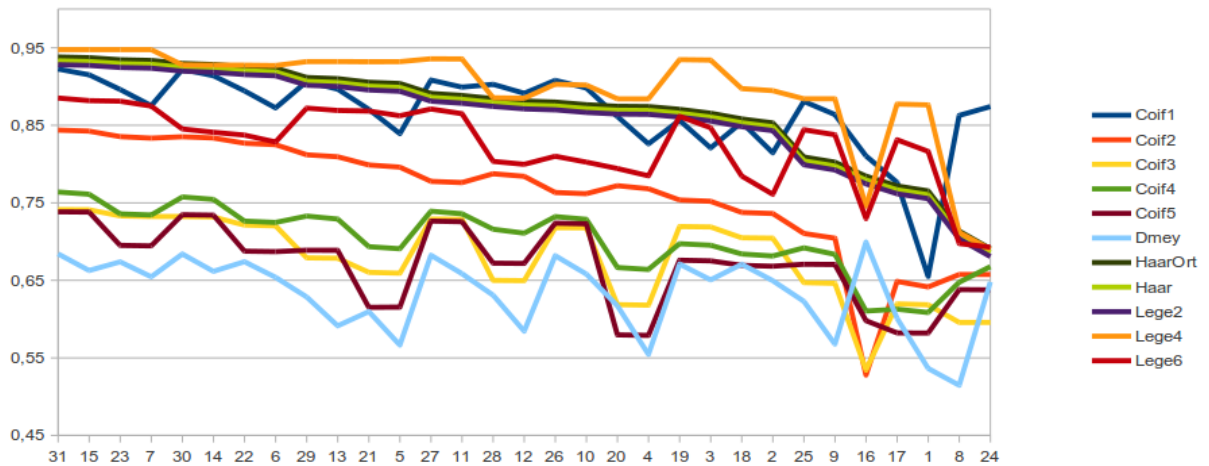
(b) Grupo 2: Família Biortogonal



(c) Grupo 3: Família Reverse Biortogonal



(d) Grupo 4: Família Symlet



(e) Grupo 5: Famílias Coiflet, Haar, Legendre e DMeyer

**Figura 38:** Resultados de índice de desempenho AUC para cada wavelet-mãe, para cada combinação de decomposição apresentada na tabela 15. Os resultados foram reunidos em grupos como apresentado pela tabela 16. Os gráficos estão ordenados de maneira decrescente pelo índice de desempenho AUC, da esquerda para a direita, pelas combinações que apresentaram na média o melhor desempenho.

É possível observar que as combinações 31 (D12345) e 15 (D1234-) apresentam o melhor resultado em todos os grupos. Também é possível observar que dentro de seus respectivos grupos, algumas wavelets-mães apresentam resultados melhores do que as demais do mesmo grupo. As cinco wavelets-mães com maiores índices de desempenho AUC de cada grupo são apresentadas na tabela 23.

**Tabela 23:** Resumo com as cinco wavelets-mães com maiores índices de desempenho AUC de cada grupo apresentado na tabela 16 de acordo com as tabelas 17 a 22.

Grupo		Classificação				
		1º	2º	3º	4º	5º
1	Wavelet	DB1	DB2	DB3	DB4	DB5
	AUC	0,9332	0,9300	0,9057	0,9055	0,8915
	Combinação	D12345	D12345	D12345	D12345	D12345
2	Wavelet	Bior15	Bior13	Bior11	Bior22	Bior24
	AUC	0,9434	0,9409	0,9332	0,9015	0,8781
	Combinação	D12345	D12345	D12345	D12345	D12345
3	Wavelet	RBior31	RBior22	RBior11	RBior33	RBior13
	AUC	0,9579	0,9452	0,9332	0,9296	0,9187
	Combinação	D12345	D12345	D12345	D12345	D12345
4	Wavelet	Sym2	Sym4	Sym3	Sym5	Sym6
	AUC	0,9300	0,9091	0,9057	0,8740	0,8189
	Combinação	D12345	D12345	D12345	D12345	D12345
5	Wavelet	Lege4	Haar Ort.	Coif1	Lege6	Coif2
	AUC	0,9476	0,9332	0,9226	0,8852	0,8438
	Combinação	D12345	D12345	D12345	D12345	D12345

A wavelet-mãe RBior31 apresentou o melhor índice de desempenho na detecção de espículas do que qualquer wavelet-mãe apresentada na tabela 16.

A wavelet-mãe Lege4, pertencente ao grupo 5, não apresenta variação de desempenho entre as combinações 31 (D12345) e 15 (D1234-). A wavelet-mãe Haar Ortogonal, do mesmo grupo, apresenta variação de 0,09% entre as mesmas combinações.

Observando os resultados obtidos entre as combinações de decomposições wavelet apresentados pelas tabelas 17, 18, 19, 20, 21, e 22, e considerando as wavelets-mães que obtiveram o melhor resultado apresentadas na tabela 23, verifica-se que a diferença percentual de desempenho entre a melhor combinação e a segunda melhor combinação de níveis de decomposição não é maior do que 0,21%.

Considerando essa diferença percentual entre as combinações 31 e 15, foi verificado também a diferença com relação à combinação 7 (D123--), contendo apenas três níveis de decomposição. A tabela 24 apresenta a diferença percentual de desempenho das wavelets-mães entre as combinações  $[(D12345) - (D1234-)]\%$  e  $[(D12345) - (D123--)]\%$ .

**Comentário:** Considerando a dimensionalidade de características extraídas, quanto maior o número de características, maior será a quantidade de informações a serem processadas pelo algoritmo do LDA. Assim, será considerado também para fins de comparação a combinação 7 (D123--), por utilizar três decomposições.

**Tabela 24:** Diferença percentual entre o desempenho obtido pelo classificador utilizado as combinações 31 e 15, e as combinações 31 e 7. A maior diferença obtida entre as combinações 31 e 15, para as wavelets-mães com melhor desempenho não excede 0,21% de diferença.

Grupo	Wavelet	AUC			Diferença %	
		31 (D12345)	15 (D1234-)	7 (D123--)	(31)-(15)	(31)-(7)
1	DB2	0,9300	0,9279	0,8888	0,21%	4,12%
2	Bior15	0,9434	0,9428	0,9384	0,06%	0,50%
3	RBior31	0,9575	0,9568	0,9567	0,11%	0,12%
4	Sym2	0,9300	0,9279	0,8888	0,21%	4,12%
5	Lege4	0,9476	0,9476	0,9476	0,00%	0,00%
5	Haar Ort.	0,9332	0,9323	0,9287	0,09%	0,45%
5	Coif1	0,9226	0,9152	0,8753	0,74%	4,73%

Realizadas as verificações de desempenho dos níveis de decomposição wavelet na detecção de espículas, é verificado a seguir o uso combinado de wavelets-mães de uma mesma família.

### 5.3.2 ANÁLISE DE AGRUPAMENTOS DE WAVELETS-MÃES

A seguir, da mesma maneira em que foram analisadas as combinações de níveis de decomposição, foram feitas combinações de wavelets-mães dos grupos apresentados pela tabela 16. Para esse processamento, foram utilizadas as decomposições 31, 15, e 7 (conforme apresentado pela tabela 15), ou seja, as cinco (D12345), quatro (D1234-), e três (D123--) primeiras decomposições para obtenção dos índices de desempenho. A extensão de borda SP0 foi utilizada para as transformações wavelet desse experimento.

Os resultados foram ordenados em ordem decrescente de índice de desempenho AUC para cada grupo, sendo estes apresentados nas tabelas 25, 26, 27, e 28.

**Tabela 25:** Índice de desempenho AUC para agrupamentos de wavelets-mães da família Daubechies considerando níveis de decomposição (D123--), (D1234-), e (D12345).

Daubechies			
Agrupamento	(D123--)	(D1234-)	(D12345)
1-38	0,9653	0,9801	0,9843
1-37	0,9653	0,9801	0,9842
1-36	0,9652	0,9800	0,9842
1-35	0,9649	0,9797	0,9837
1-34	0,9649	0,9796	0,9836
1-33	0,9646	0,9792	0,9832
1-32	0,9647	0,9791	0,9827
1-31	0,9645	0,9791	0,9826
1-30	0,9643	0,9787	0,9821
1-29	0,9624	0,9776	0,9811
1-28	0,9620	0,9771	0,9809
1-27	0,9620	0,9771	0,9809
1-26	0,9615	0,9768	0,9807
1-25	0,9614	0,9766	0,9805
1-24	0,9614	0,9767	0,9805
1-23	0,9610	0,9763	0,9803
1-22	0,9607	0,9764	0,9799
1-21	0,9606	0,9753	0,9786
1-20	0,9601	0,9748	0,9776
1-19	0,9586	0,9736	0,9767
1-18	0,9583	0,9735	0,9767
1-17	0,9578	0,9733	0,9761
1-16	0,9575	0,9729	0,9755
1-15	0,9566	0,9722	0,9751
1-14	0,9566	0,9721	0,9750
1-13	0,9566	0,9721	0,9751
1-12	0,9565	0,9719	0,9747
1-11	0,9563	0,9718	0,9746
1-10	0,9561	0,9712	0,9741
1-9	0,9560	0,9705	0,9738
1-8	0,9557	0,9705	0,9729
1-7	0,9549	0,9685	0,9708
1-6	0,9546	0,9685	0,9703
1-5	0,9544	0,9682	0,9702
1-4	0,9543	0,9662	0,9679
1-3	0,9512	0,9642	0,9660
1-2	0,9477	0,9601	0,9615

**Tabela 26:** Índice de desempenho AUC para agrupamentos de wavelets-mães da família Biortogonal e Reverse Biortogonal considerando níveis de decomposição (D123--), (D1234-), e (D12345).

Agrupamento	Biortogonal			Reverse Biortogonal		
	(D123--)	(D1234-)	(D12345)	(D123--)	(D1234-)	(D12345)
11-68	0,9673	0,9712	0,9756	0,9754	0,9799	0,9812
11-55	0,9672	0,9708	0,9749	0,9753	0,9797	0,9810
11-44	0,9658	0,9694	0,9737	0,9752	0,9794	0,9808
11-39	0,9655	0,9692	0,9735	0,9750	0,9793	0,9806
11-37	0,9653	0,9690	0,9731	0,9743	0,9778	0,9793
11-35	0,9651	0,9688	0,9729	0,9742	0,9776	0,9790
11-33	0,9648	0,9686	0,9722	0,9744	0,9772	0,9786
11-31	0,9642	0,9679	0,9710	0,9724	0,9743	0,9761
11-28	0,9639	0,9673	0,9693	0,9588	0,9681	0,9707
11-26	0,9634	0,9671	0,9688	0,9579	0,9679	0,9708
11-24	0,9629	0,9669	0,9681	0,9575	0,9674	0,9703
11-22	0,9624	0,9664	0,9674	0,9570	0,9671	0,9694
11-15	0,9613	0,9622	0,9631	0,9471	0,9532	0,9614
11-13	0,9554	0,9566	0,9573	0,9435	0,9483	0,9569

**Tabela 27:** Índice de desempenho AUC para agrupamentos de wavelets-mães da família Symlet considerando níveis de decomposição (D123--), (D1234-), e (D12345).

Agrupamento	Symlet		
	(D123--)	(D1234-)	(D12345)
1-20	0,9203	0,9527	0,9611
1-19	0,9203	0,9530	0,9616
1-18	0,9201	0,9531	0,9612
1-17	0,9201	0,9532	0,9604
1-16	0,9203	0,9532	0,9602
1-15	0,9195	0,9519	0,9594
1-14	0,9194	0,9517	0,9588
1-13	0,9192	0,9516	0,9588
1-12	0,9188	0,9515	0,9589
1-11	0,9175	0,9513	0,9582
1-10	0,9173	0,9505	0,9576
1-9	0,9170	0,9504	0,9573
1-8	0,9166	0,9502	0,9570
1-7	0,9164	0,9501	0,9564
1-6	0,9165	0,9499	0,9560
1-5	0,9142	0,9484	0,9543
1-4	0,9133	0,9458	0,9509
1-3	0,9079	0,9426	0,9473
1-2	0,8888	0,9279	0,9300

**Tabela 28:** Índice de desempenho AUC para agrupamentos de wavelets-mães da família Legendre e Coiflet considerando níveis de decomposição (D123--), (D1234-), e (D12345).

Legendre			
Agrupamento	(D123--)	(D1234-)	(D12345)
2-6	0,9639	0,9639	0,9643
2-4	0,9626	0,9627	0,9627

Coiflet			
Agrupamento	(D123--)	(D1234-)	(D12345)
1-5	0,8926	0,9314	0,9393
1-4	0,8917	0,9307	0,9383
1-3	0,8913	0,9290	0,9365
1-2	0,8916	0,9293	0,9361

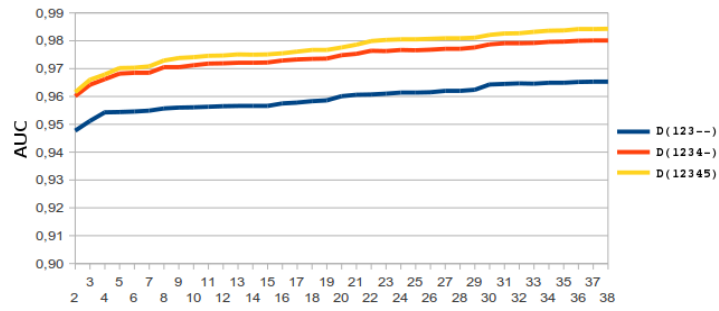
O resultado apresentado nas tabelas 25, 26, 27, e 28 são apresentados na figura 39.

Observa-se pelos resultados apresentados na tabela 25 e figura 39(a) que quanto maior o número de wavelets-mães da família Daubechies, melhores são os resultados obtidos pelo classificador. A quantidade de decomposições utilizadas se manteve estável conforme o estudo de combinações de decomposições wavelets onde os agrupamentos que utilizam cinco decomposições apresentam melhores resultados do que agrupamentos que utilizam quatro ou três níveis de decomposição, respectivamente. Esse comportamento se manteve para as wavelets-mães dos grupos 2, 3, 4, e 5 da tabela 16.

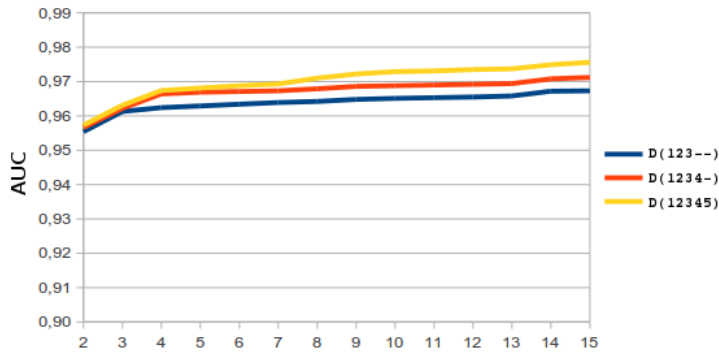
O melhor resultado foi obtido pelo agrupamento 1-38 da família Daubechies (DB1-DB38) utilizando os cinco níveis de decomposição (D12345) com índice de desempenho AUC igual a 0,9843. O segundo melhor resultado foi obtido pelo agrupamento 11-68 da família Reverse Biortogonal, utilizando os cinco níveis de decomposição (D12345) com índice de desempenho AUC igual a 0,9812.

É interessante observar que o agrupamento 1-5 da família Daubechies (DB1-DB5) utilizando os cinco níveis de decomposição (D12345), apresentado pela tabela 25, obteve índice de desempenho AUC igual a 0,9702, superior a wavelets-mães dos grupos 4 e 5, e 0,54% inferior ao melhor resultado do grupo 2.

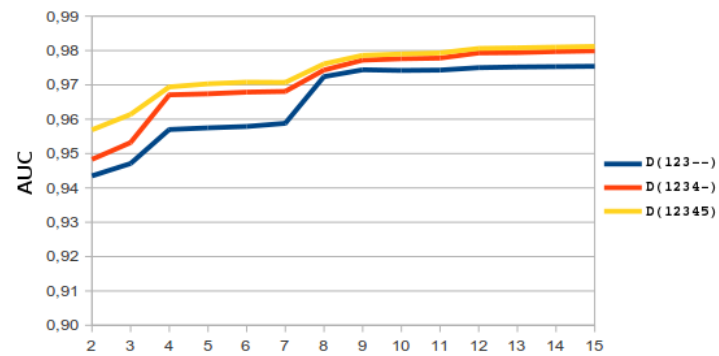




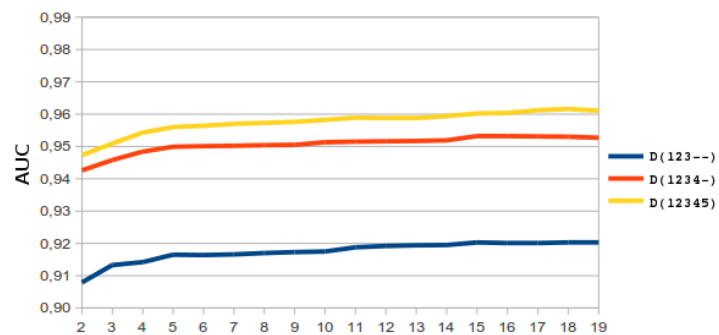
(a) Grupo 1: Família Daubechies



(b) Grupo 2: Família Biortogonal



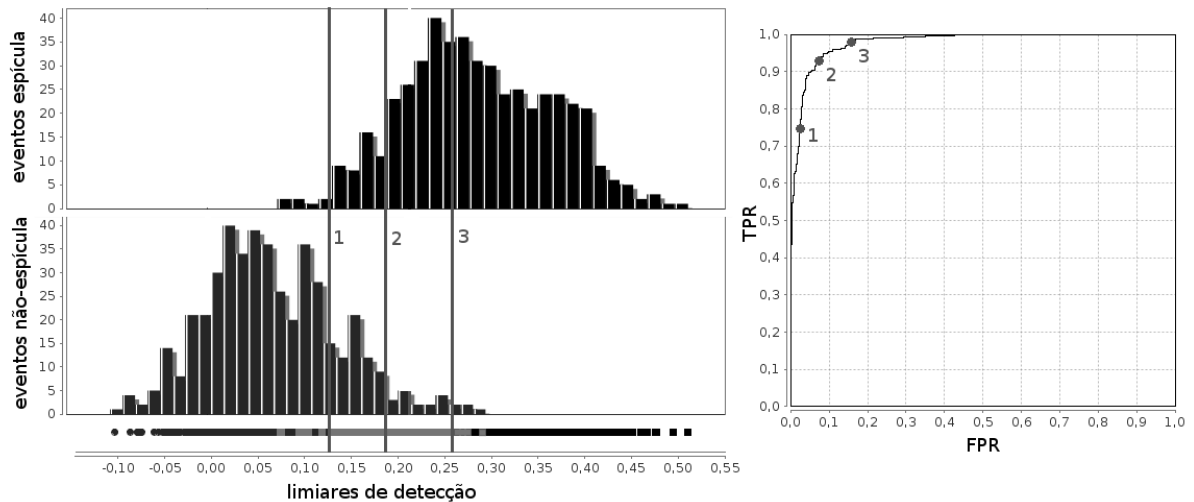
(c) Grupo 3: Família Reverse Biortogonal



(d) Grupo 4: Família Symlet

**Figura 39:** Resultados em termos de índice de desempenho AUC para combinações de wavelets-mães de uma mesma família. (a) resultado respectivo à tabela 25 (b) e (c) resultado respectivo à tabela 26 (d) resultado respectivo à tabela 27, utilizando as três, quatro e cinco primeiras decomposições da transformada wavelet.

Como exemplo é apresentado na figura 40 a distribuição dos 494 eventos espícula e 1500 eventos não-espícula após o processamento LDA usando a configuração DB1-DB5, com quatro decomposições (D1234-) e entropia aproximada como descritor.



**Figura 40:** Distribuição dos eventos espícula e não-espícula após o algoritmo LDA. O histograma de distribuição dos eventos espícula e não-espícula são apresentados do lado esquerdo superior e inferior, respectivamente. As amostras sobrepostas são apresentadas em cinza claro abaixo dessas distribuições. Do lado direito é apresentada a respectiva curva ROC dessas distribuições. Os pontos 1, 2, e 3 são limiares de detecção diferentes e seus valores respectivos de limiar, sensibilidade e especificidade são apresentados na tabela 29.

Do lado esquerdo da figura 40, tem-se a projeção unidimensional determinada pelo LDA juntamente com dois histogramas sobrepostos. O histograma com a distribuição dos eventos espícula são apresentados no lado superior esquerdo da figura, enquanto que o histograma de distribuição dos eventos não-espícula são apresentados no lado inferior da figura. A linha horizontal abaixo dos histogramas (saída unidimensional do LDA), apresenta as amostras sobrepostas em cinza claro. É possível verificar que as médias das distribuições estão bem separadas. A figura 40 apresenta do lado direito a respectiva curva ROC do classificador linear sobre essas distribuições, com índice AUC igual a 0,9682. Na curva ROC apresentada na figura 40 existem três pontos especificados como 1, 2, e 3. Estes três pontos se referem às três linhas representando limiares de detecção diferentes sobre os histogramas de distribuição dos eventos espícula e não-espícula no lado esquerdo do gráfico. A sensibilidade e especificidade desses três pontos são apresentados na tabela 29.

Com os agrupamentos e níveis de decomposição obtidos nesta seção, serão analisadas agora composições entre as famílias de wavelets-mães, comparando os resultados com os obtidos até então.

**Tabela 29:** Sensibilidade e especificidade para níveis diferentes de limiares de detecção do classificador configurado para DB1-DB5, (D1234-) e ApEn para os pontos apresentados na figura 40.

Ponto	Limiar de Detecção	Sensibilidade (TPR)	Especificidade	1-Especificidade (FPR)
1	0,1237	0,7393	0,9780	0,0220
2	0,1931	0,9246	0,9231	0,0769
3	0,2572	0,9919	0,6659	0,3341

### 5.3.3 ANÁLISE ENTRE FAMÍLIAS DE WAVELETS-MÃES

Para esse estudo, cada espícula da base de dados foi decomposta em cinco níveis (D1 a D5) para cada uma das 98 wavelets propostas na seção 2.5.3. Foi calculada a energia de cada nível de decomposição de cada transformada wavelet para encontrar a wavelet que apresentasse os maiores níveis de energia em suas decomposições. Os resultados são apresentados na tabela 30.

**Tabela 30:** Energia dos coeficientes dos níveis de decomposição D1, D2, D3, D4, e D5 de 98 wavelets-mães aplicadas na base de espículas. (obs: os resultados foram divididos por  $10^3$ .)

Posição	Wavelet	D5	Wavelet	D4	Wavelet	D3	Wavelet	D2	Wavelet	D1
01	Haar Ort.	2168,3	Haar Ort.	443,8	Haar Ort.	106,2	RBior31	30,3	RBior31	13,7
02	RBior31	235,3	RBior31	147,0	RBior31	86,8	Haar Ort.	19,0	Haar Ort.	6,2
03	Bior31	200,9	RBior37	54,3	Lege4	14,9	Lege6	7,6	RBior33	4,9
04	Bior33	172,2	RBior39	52,2	RBior33	14,8	Lege4	6,5	RBior35	4,5
05	Bior35	132,0	RBior22	51,3	Bior13	14,6	Bior15	4,9	RBior37	4,3
06	Bior28	131,3	RBior35	49,8	Bior15	14,3	Bior13	4,8	RBior39	4,3
07	RBior55	128,9	RBior33	49,5	Bior11	13,2	Bior11	4,7	Lege6	3,6
08	Bior26	123,2	Bior15	49,0	DB1	13,2	DB1	4,7	RBior22	3,3
09	Bior24	101,2	DB2	36,6	RBior11	13,2	RBior11	4,7	Lege4	3,1
10	Bior37	101,0	Sym2	36,6	Haar	13,2	Haar	4,7	Bior15	3,0
11	Bior39	97,2	Bior13	36,6	Lege2	13,2	Lege2	4,7	Bior13	3,0
12	RBior44	88,6	Bior37	36,2	Lege6	12,6	Bior31	2,9	Bior11	3,0
13	Bior68	83,4	RBior44	35,4	RBior22	10,8	Bior39	2,9	DB1	3,0
14	Coif2	81,5	Bior39	34,8	RBior35	6,4	Bior37	2,8	RBior11	3,0
15	RBior68	80,3	Sym5	34,6	DB2	5,3	Bior35	2,8	Haar	3,0
16	Bior13	79,1	RBior55	34,6	Sym2	5,3	Bior33	2,8	Lege2	3,0
17	Sym8	78,1	Coif1	34,0	RBior55	5,1	RBior22	2,3	RBior24	3,0
18	Sym10	78,0	DB5	33,9	Coif1	5,1	Bior28	2,0	RBior26	3,0
19	RBior22	77,7	Bior35	32,9	RBior44	5,0	RBior33	2,0	RBior28	3,0
20	Coif3	77,2	DB8	32,1	RBior24	4,8	Bior26	2,0	Bior55	2,6
21	RBior33	76,7	DB11	31,2	RBior37	4,5	Bior24	1,9	RBior68	2,2
22	Sym6	75,8	Coif4	30,8	RBior39	4,3	Bior22	1,9	RBior44	2,2
23	DB4	75,6	Sym11	30,5	DB3	3,7	DB2	1,7	RBior15	2,2
24	DB7	75,4	Bior22	30,3	Sym3	3,7	Sym2	1,7	RBior13	2,2
25	Bior15	75,1	DB14	29,7	Bior31	3,6	Coif1	1,7	Coif1	2,2
26	Sym4	73,9	Sym18	29,6	Bior26	3,5	RBior55	1,5	DB2	2,1
27	Bior22	73,0	RBior24	29,3	Bior24	3,5	Bior44	1,5	Sym2	2,1
28	Sym12	72,3	Coif5	29,2	RBior13	3,4	Bior68	1,5	Sym4	2,1
29	DB21	71,5	DB17	28,9	RBior26	3,3	Dmey	1,5	Coif2	2,1
30	DB10	70,6	Sym9	28,7	Bior22	3,1	DB34	1,4	DB3	2,1
31	DB18	69,9	DB20	28,4	Bior28	3,1	DB35	1,4	Sym3	2,1

*Continua na próxima página*

Tabela 30 – Continuação da página anterior

Posição	Wavelet	D5	Wavelet	D4	Wavelet	D3	Wavelet	D2	Wavelet	D1
32	DB24	69,2	Sym20	28,3	Coif2	2,9	DB33	1,4	DB5	2,1
33	Sym14	68,8	Sym3	28,2	DB4	2,8	DB36	1,4	DB4	2,1
34	Sym17	68,6	DB3	28,2	Sym4	2,8	DB27	1,4	DB6	2,1
35	DB3	68,5	Sym16	28,1	Bior33	2,6	DB28	1,4	Sym6	2,1
36	Sym3	68,5	Bior11	27,7	Sym6	2,4	DB26	1,4	Coif3	2,1
37	Bior11	67,7	DB1	27,7	DB5	2,3	DB25	1,4	Sym5	2,1
38	DB1	67,7	RBior11	27,7	Bior39	2,3	DB37	1,4	DB7	2,1
39	RBior11	67,7	Haar	27,7	DB6	2,2	Sym19	1,4	Sym8	2,1
40	Haar	67,7	Lege2	27,7	Bior35	2,2	DB32	1,4	Coif4	2,1
41	Lege2	67,7	DB23	27,4	Bior37	2,1	DB31	1,4	Dmey	2,1
42	DB36	67,7	DB26	27,1	RBior28	2,1	Sym20	1,4	Sym10	2,1
43	Coif1	67,4	Dmey	26,9	RBior68	2,1	DB24	1,4	Coif5	2,0
44	RBior26	66,5	DB29	26,7	Sym5	2,1	DB18	1,4	DB8	2,0
45	RBior28	66,3	DB32	26,0	DB8	2,1	DB3	1,4	Sym9	2,0
46	DB28	65,8	DB35	25,9	Sym9	2,0	Sym3	1,4	Sym7	2,0
47	RBior24	65,8	Bior33	25,8	Bior44	2,0	DB29	1,4	DB13	2,0
48	DB25	65,3	DB38	25,6	RBior15	2,0	DB38	1,4	DB14	2,0
49	DB27	65,2	Bior24	25,6	Coif4	2,0	DB30	1,4	DB12	2,0
50	DB15	65,0	Sym13	25,3	Sym8	2,0	DB19	1,4	Sym18	2,0
51	DB33	64,4	Sym4	25,1	DB10	1,9	Sym15	1,4	DB9	2,0
52	Sym7	64,2	DB37	24,9	Sym11	1,9	Sym16	1,4	DB11	2,0
53	DB13	63,9	DB31	24,8	Coif3	1,9	DB21	1,4	Sym13	2,0
54	DB35	63,7	DB34	24,7	DB12	1,8	DB22	1,4	DB15	2,0
55	Sym16	63,6	DB28	24,4	DB7	1,8	DB20	1,4	DB10	2,0
56	DB32	63,4	DB25	24,2	Sym14	1,8	DB17	1,4	Sym17	2,0
57	DB6	63,2	DB22	24,1	DB14	1,7	DB23	1,4	Sym11	2,0
58	DB30	63,0	Sym14	24,1	Sym7	1,7	Sym14	1,4	DB22	2,0
59	Sym19	63,0	Sym7	23,9	Sym19	1,7	DB16	1,4	Sym12	2,0
60	DB31	62,9	DB19	23,4	DB21	1,6	DB15	1,4	DB21	2,0
61	DB14	62,3	Lege4	23,3	Sym16	1,6	Sym12	1,4	DB16	2,0
62	DB22	61,9	DB6	23,1	DB19	1,6	Sym11	1,4	DB23	2,0
63	DB38	61,1	Sym15	23,1	Sym17	1,6	Sym17	1,4	DB30	2,0
64	DB17	61,0	DB16	22,9	DB23	1,6	Sym18	1,4	DB38	2,0
65	Bior44	60,7	DB13	22,4	Sym12	1,6	DB12	1,4	DB31	2,0
66	DB12	60,5	Bior31	22,0	Bior55	1,6	Sym13	1,4	DB20	2,0
67	Coif4	60,4	Bior28	21,5	DB17	1,6	DB11	1,4	Sym14	2,0
68	DB29	60,1	DB09	21,5	DB32	1,5	DB10	1,4	DB29	2,0
69	Dmey	60,0	Sym12	21,4	DB9	1,5	DB9	1,4	DB17	2,0
70	Sym20	59,7	Bior26	21,2	DB30	1,5	DB14	1,4	DB24	2,0
71	DB9	59,7	DB10	20,9	DB16	1,5	DB13	1,4	Sym16	2,0
72	DB34	59,4	Sym6	20,3	DB25	1,5	RBior13	1,4	DB19	2,0
73	DB16	59,1	DB12	20,1	Coif5	1,5	Sym7	1,4	DB37	2,0
74	DB11	58,9	RBior15	20,0	Bior68	1,5	Sym9	1,4	Sym15	2,0
75	DB19	58,6	Coif3	20,0	DB34	1,5	DB8	1,4	DB18	2,0
76	DB37	58,1	DB15	19,9	Sym10	1,5	Sym10	1,4	Bior44	2,0
77	Sym15	57,3	DB33	19,8	Sym13	1,5	Coif05	1,4	DB32	2,0
78	DB20	57,2	DB7	19,8	DB15	1,5	DB4	1,4	Sym20	2,0
79	Sym18	56,6	DB36	19,7	DB28	1,5	Coif4	1,4	DB28	2,0
80	RBior13	56,3	DB18	19,7	DB11	1,4	Sym8	1,4	DB25	2,0
81	DB26	56,1	DB24	19,6	DB13	1,4	DB7	1,4	Sym19	2,0
82	Coif5	56,1	DB27	19,6	DB36	1,4	Sym5	1,4	DB36	2,0
83	Sym13	54,9	DB30	19,5	DB27	1,4	Sym4	1,4	DB27	2,0
84	DB23	54,8	DB21	19,3	DB26	1,4	Coif3	1,3	DB26	2,0
85	RBior35	53,4	Bior44	18,6	DB18	1,4	Coif2	1,3	DB33	2,0
86	Sym9	52,5	RBior26	18,4	DB37	1,4	Sym6	1,3	DB35	2,0
87	DB8	51,8	Lege6	18,4	Sym20	1,4	DB5	1,3	DB34	2,0
88	Sym11	48,7	Sym19	18,0	DB38	1,3	RBior15	1,3	Bior68	1,9
89	RBior15	48,2	DB04	18,0	Sym15	1,3	DB6	1,3	RBior55	1,7

Continua na próxima página

Tabela 30 – Continuação da página anterior

Posição	Wavelet	D5	Wavelet	D4	Wavelet	D3	Wavelet	D2	Wavelet	D1
90	DB2	45,6	Sym17	17,7	Dmey	1,3	RBior44	1,3	Bior28	1,6
91	Sym2	45,6	Coif2	17,5	DB24	1,3	Bior55	1,3	Bior26	1,6
92	DB5	45,2	RBior68	17,5	DB35	1,3	RBior68	1,3	Bior24	1,6
93	Bior55	45,0	Sym8	17,2	DB20	1,3	RBior35	1,2	Bior22	1,6
94	RBior37	43,7	Sym10	17,1	DB29	1,3	RBior24	1,2	Bior31	1,3
95	Sym5	43,5	Bior68	16,1	Sym18	1,3	RBior37	1,1	Bior39	1,3
96	RBior39	40,2	RBior13	15,7	DB22	1,3	RBior28	1,1	Bior37	1,3
97	Lege4	34,1	RBior28	15,2	DB33	1,2	RBior39	1,1	Bior35	1,3
98	Lege6	27,0	Bior55	13,2	DB31	1,2	RBior26	1,1	Bior33	1,3

Da tabela 30, a wavelet Rbior31 produziu os maiores níveis de energia para o primeiro e segundo nível de decomposição (D1,D2). Para o terceiro, quarto e quinto nível de decomposição (D3,D4,D5), a wavelet Haar Ortogonal apresentou os maiores resultados.

Sabendo que as wavelets Haar Ortogonal e RBior31 são as wavelets que apresentaram maior energia em suas decomposições, estas foram utilizadas para a detecção de espículas individualmente e agrupadas. O resultado é apresentado comparativamente com os agrupamentos DB1-DB5 e DB1-DB38 na tabela 31.

**Tabela 31:** Resultados obtidos usando as wavelets-mães com os maiores valores de energia na decomposição de espículas. Em negrito os melhores resultados por nível de decomposição. Sublinhados o segundo melhor resultado por nível de decomposição.

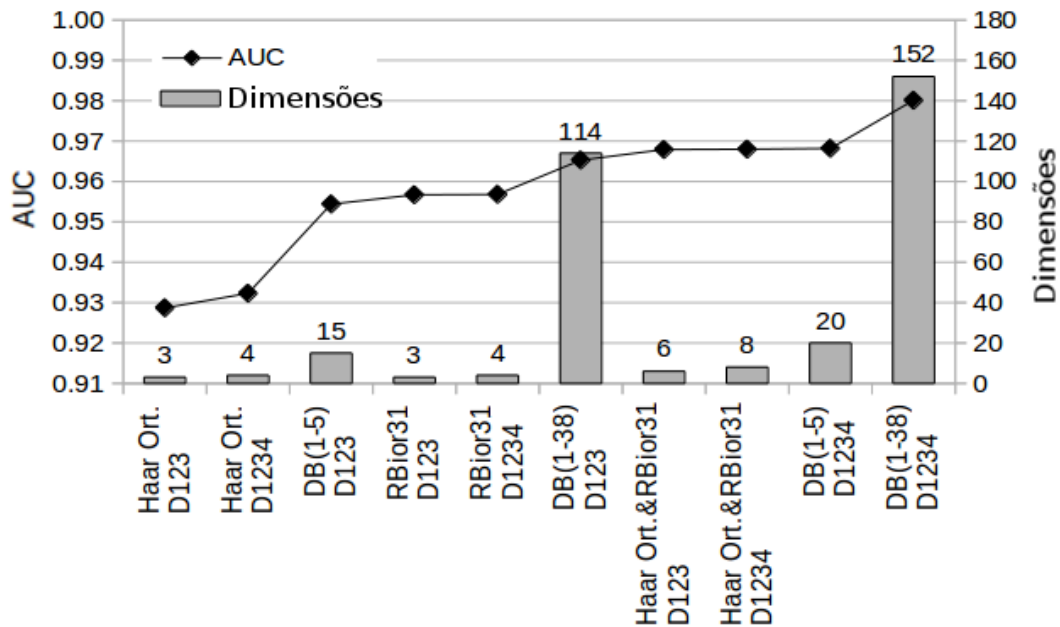
Wavelets	Decomposições	Dimensões	AUC
Haar Ortogonal	(D123--)	3	0,9287
RBior31	(D123--)	3	0,9567
Haar Ort. + RBior31	(D123--)	6	<u>0,9679</u>
DB1-DB5	(D123--)	15	<u>0,9544</u>
DB1-DB38	(D123--)	114	<b>0,9653</b>
Haar Ortogonal	(D1234-)	4	0,9323
RBior31	(D1234-)	4	0,9568
Haar Ort. + RBior31	(D1234-)	8	<u>0,9680</u>
DB1-DB5	(D1234-)	20	<u>0,9682</u>
DB1-DB38	(D1234-)	152	<b>0,9801</b>
Haar Ortogonal	(D12345)	5	0,9332
RBior31	(D12345)	5	0,9579
Haar Ort. + RBior31	(D12345)	10	0,9685
DB1-DB5	(D12345)	25	<u>0,9702</u>
DB1-DB38	(D12345)	190	<b>0,9843</b>

Observa-se pelos resultados apresentados na tabela 31 que o grupo DB1-DB38 gerou os melhores índices de detecção para os níveis de decomposição (D123--), (D1234-), e (D12345). O agrupamento Haar Ort. + Rbior31 gerou os segundos melhores índices de detecção para as decomposições (D123--) e (D1234-). Como comentado no estudo de decomposições wavelet (seção 5.3.1), a wavelet-mãe RBior31 apresentou sozinha melhores

índices de detecção do que qualquer wavelet-mãe apresentada nas tabelas 17, 18, 19, 20, 21, e 22.

**Comentário:** Na tabela 31 é apresentada a dimensão de características extraídas de cada configuração. Quanto maior a dimensão, maior o tempo de processamento necessário para a obtenção de características e processamento pelo algoritmo LDA. Em caso de implementação do classificador proposto em sistemas embarcados com restrição de desempenho computacional, é interessante que seja utilizado o menor número de características possíveis.

Os dados da tabela 31 referente às decomposições (D123--) e (D1234-), assim como as dimensões respectivas a cada configuração podem ser vistos na figura 41.



**Figura 41:** Wavelets-mães selecionadas ordenadas por ordem crescente de índice de desempenho AUC. Observe que não necessariamente o número de dimensões (número acima de cada coluna) está associado a um desempenho melhor do classificador.

É possível observar pela figura 41 que não necessariamente o aumento do número de dimensões acarreta no aumento do índice de desempenho AUC. O uso das wavelets-mães Haar Ort. + Rbior31 com decomposições (D123--) apresenta melhor índice de desempenho com seis dimensões do que o grupo DB1-DB38 e decomposições (D123--) com 114 dimensões.

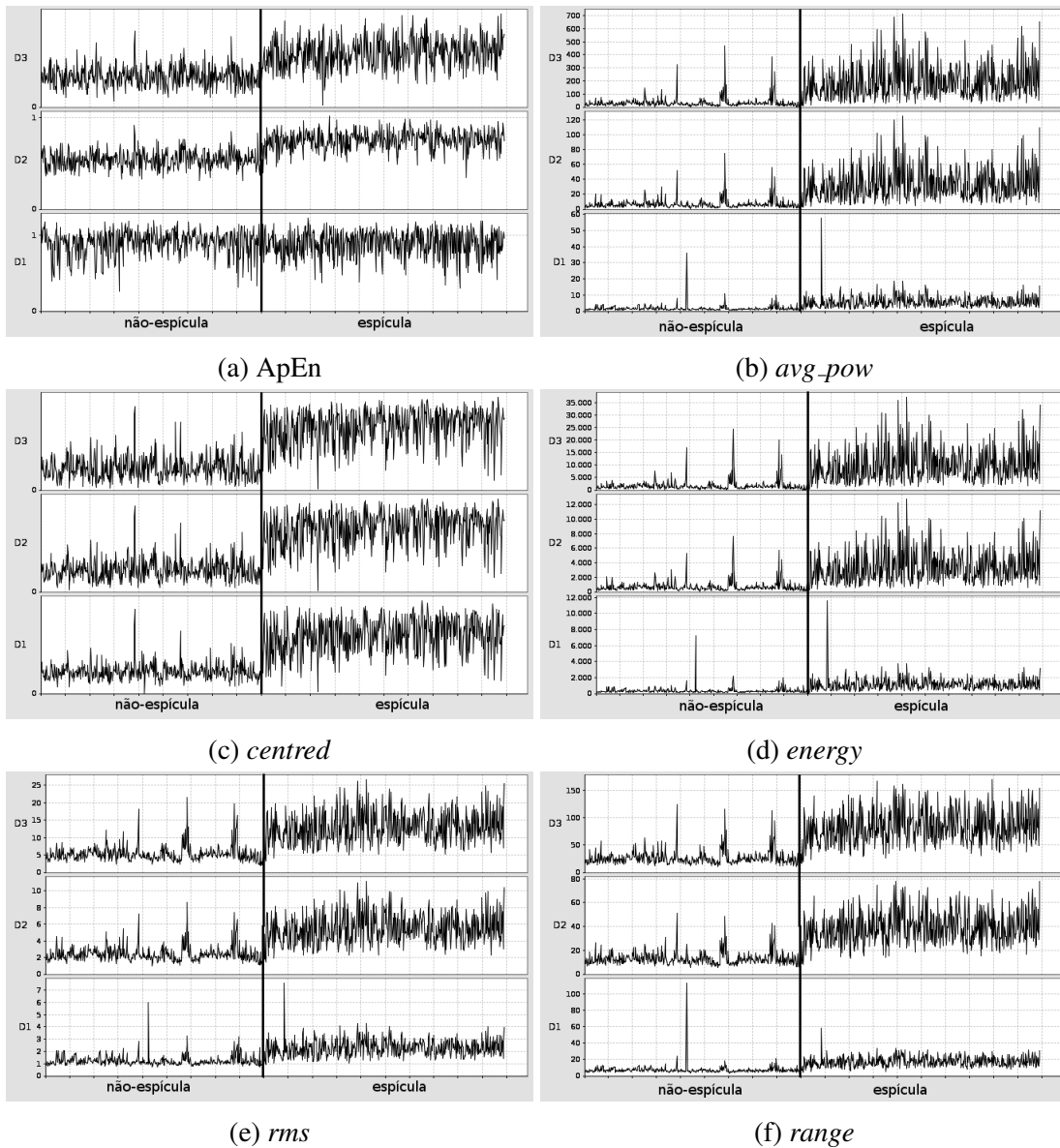
A seguir, será analisada a contribuição de diferentes descritores em conjunto com as wavelets-mães selecionadas.

### 5.3.4 ANÁLISE DE DESCRITORES

Um descritor adequado para extração de características é aquele que consegue produzir maior informação discriminante entre as classes que se deseja classificar. Para essa análise foram utilizados os descritores comentados na literatura, vistos no Capítulo 3. Adicionalmente, foi analisado o descritor *centred*, desenvolvido para esse estudo e apresentado na seção 4.3.2. Os 14 descritores utilizados para esse estudo foram apresentados na tabela 11, na seção 4.3.

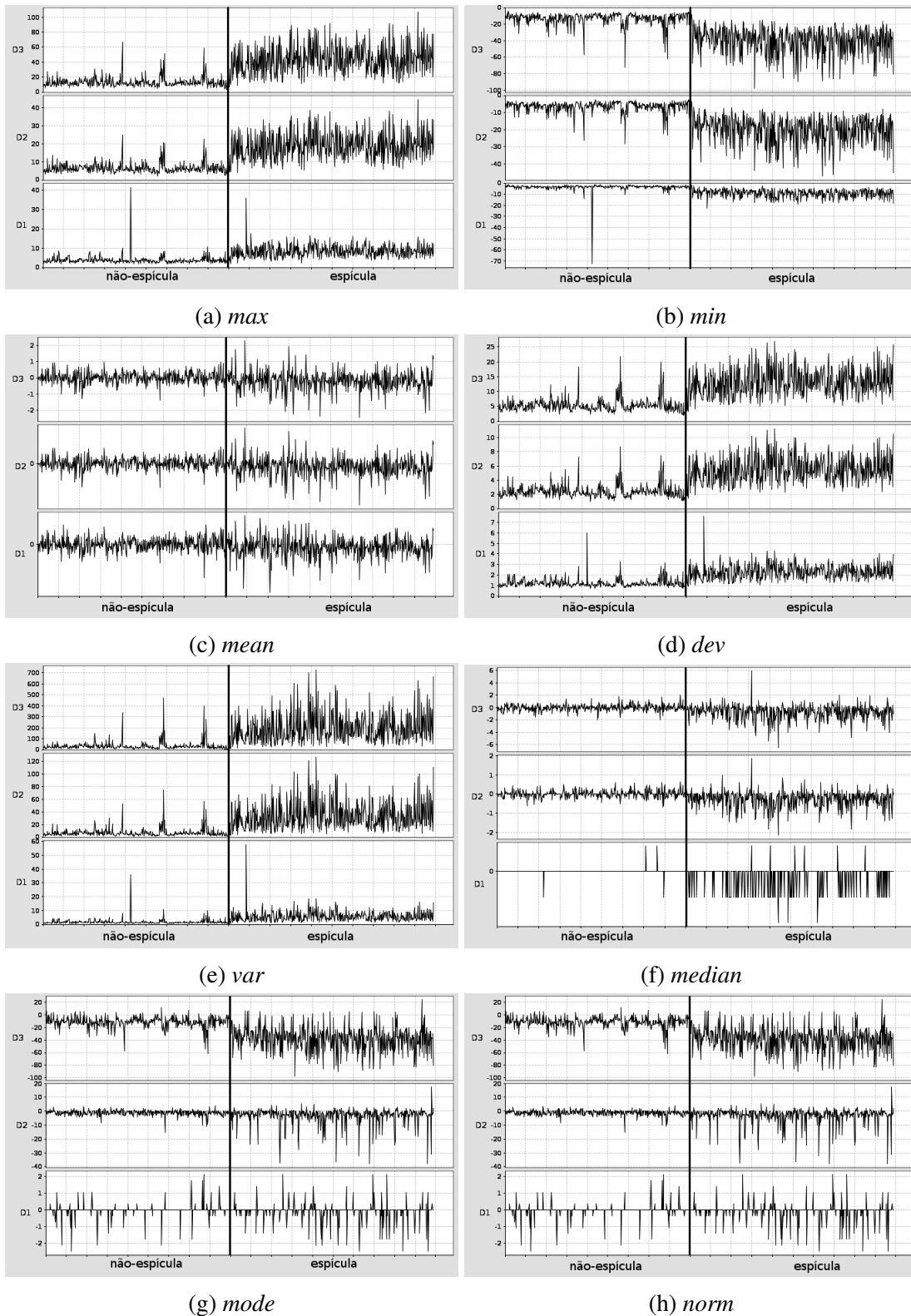
Primeiramente, foi feita uma análise visual da capacidade discriminatória de cada descritor. Foi aplicado a wavelet-mãe RBior31 com método de extensão de borda SP0, com três níveis de decomposição, para 494 eventos não-espícula e 494 eventos espícula. De cada evento, para cada nível de decomposição, foram calculados os 14 descritores da tabela 11. As figuras 42 e 43 apresentam o resultado para cada descritor individualmente. Para cada nível de decomposição, na lado esquerdo da figura são exibidos o valor do descritor para os 494 eventos não-espícula, e do lado direito da figura são exibidos o valor do descritor para os 494 eventos espícula.

É possível observar na figura 42(c) que, visivelmente, há informação discriminatória entre eventos espícula e não-espícula, assim como nas figuras 42(b) e 43(b). Porém, nas figuras 43(c), 43(f) e 43(h) há pouca informação discriminante.



**Figura 42:** Resultado dos 6 primeiros descritores da tabela 11 para 3 decomposições da transformada wavelet RBior31. Os valores dos gráficos estão ordenados de modo que, no lado esquerdo, são vistos resultados de 494 realizações de fundo de EEG. Ao lado direito são vistos resultados de 494 realizações de espículas. As realizações de espículas estão ordenadas em ordem crescente da esquerda para a direita de 60 ms a 200 ms.





**Figura 43:** Resultado dos 8 últimos descritores da tabela 11 para 3 decomposições da transformada wavelet RBior31. Os valores dos gráficos estão ordenados de modo que, no lado esquerdo, são vistos resultados de 494 realizações de fundo de EEG. Ao lado direito são vistos resultados de 494 realizações de espículas. As realizações de espículas estão ordenadas em ordem crescente da esquerda para a direita de 60 ms a 200 ms.

A seguir foram realizadas utilizadas wavelets-mães RBior31 e Haar Ort. para aferir numericamente a informação discriminatória apresentada pelas figuras 42 e 43. O resultado é apresentado na tabela 32 somente para a decomposição (D1234-).

**Tabela 32:** Valores do índice AUC para os descritores apresentados na tabela 11 e as wavelets RBior31 e Haar Ort. para as decomposições (D1234-). Foram destacadas as linhas e colunas com os melhores resultados.

	RBior31	Haar Ortogonal	<b>RBior31 + Haar Ort.</b>
ApEn	0,9568	0,9323	0,9680
<i>avg_pow</i>	0,9613	0,9501	0,9627
<b><i>centred</i></b>	0,9770	0,9723	0,9782
<i>energy</i>	0,9613	0,9501	0,9627
<i>rms</i>	0,9694	0,9521	0,9706
<b><i>range</i></b>	0,9777	0,9710	0,9781
<b><i>max</i></b>	0,9637	0,9592	0,9633
<b><i>min</i></b>	0,9725	0,9598	0,9727
<i>mean</i>	0,6462	0,6466	0,6906
<i>dev</i>	0,9695	0,9517	0,9706
<i>var</i>	0,9614	0,9499	0,9629
<i>median</i>	0,7518	0,7286	0,7587
<i>mode</i>	0,9449	0,6927	0,9459
<i>norm</i>	0,9694	0,9521	0,9706

É possível observar que os descritores que apresentaram os melhores índices AUC foram os descritores *centred*, *max*, *min*, assim como o *range*. Os descritores *mean*, *median*, e *mode* apresentaram os piores resultados como era esperado pela informação discriminante entre eventos espícula e não-espícula apresentados nas figuras 43(c), 43(f) e 43(g).

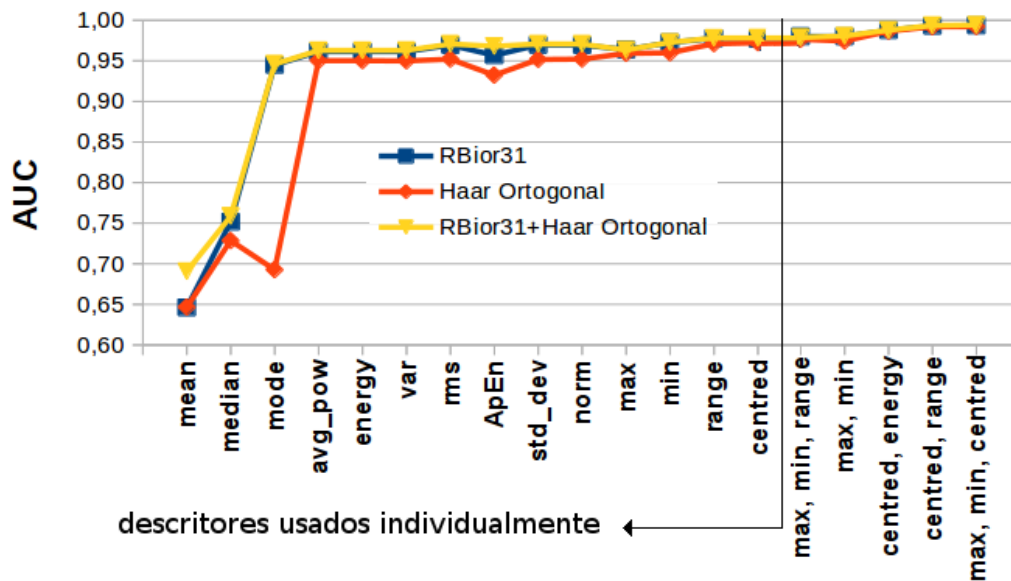
Sabendo quais descritores apresentaram melhor desempenho na discriminação de eventos, foram combinados os descritores *centred*, *max*, *min* e *range*. O resultado é apresentado para as combinações que resultaram maiores índices nas decomposições (D1234-).

**Tabela 33:** Valores do índice AUC para agrupamentos de descritores apresentados na tabela 32 processadas pelo classificador proposto no item 4.2 utilizando as wavelets selecionadas no item 5.3.3 com decomposições (D1234-). Foram destacadas as linhas e colunas com os melhores resultados.

Composição	RBior31	Haar Ortogonal	<b>RBior31 + Haar Ort.</b>
<i>max, min</i>	0,9800	0,9744	0,9805
<b><i>max, min, centred</i></b>	0,9934	0,9921	<b>0,9941</b>
<i>centred, max</i>	0,9919	0,9908	0,9928
<b><i>centred, range</i></b>	<u>0,9930</u>	0,9920	0,9935

A figura 44 apresenta os resultados apresentados nas tabelas 32 e 33.

É possível observar pelos resultados da tabela 33 e pela figura 44 que a combinação



**Figura 44:** Índice de desempenho AUC para os descritores apresentados na tabela 11 para wavelets-mães selecionadas assim como para a combinação de descritores apresentada na tabela 33.

de descritores [*max*, *min*, *centred*], juntamente com o agrupamento RBior31 + Haar Orthogonal apresentou o melhor índice AUC, igual a 0,9941. Considerando o número de dimensões que foi reduzido, o melhor resultado foi obtido pelo wavelet RBior31 com os descritores [*centred*, *range*]. O uso combinado de descritores resultou melhores índices AUC do que os descritores usados individualmente.

A seguir o conjunto de descritores selecionados são processados em conjunto com os melhores agrupamentos de wavelets obtidos nos experimentos anteriores. A tabela 34 apresenta os resultados obtidos para os melhores conjuntos de wavelets-mães, descritores e número de decomposições.

**Tabela 34:** Comparação entre classificadores com diferentes configurações.

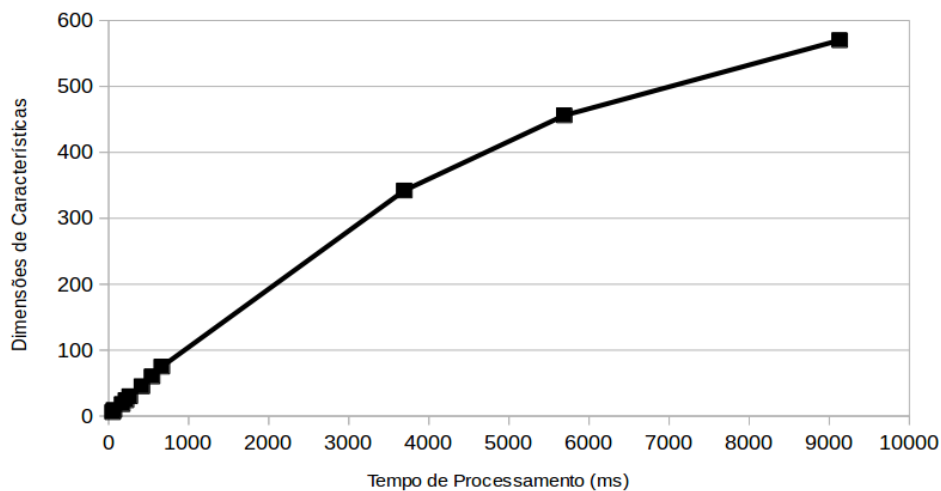
Wavelets	Decomposições	Descritor	Dimensões	AUC	tempo (ms)
RBior31	(D123--)	<i>centred, range</i>	6	0,9936	51
RBior31	(D123--)	<i>max, min, centred</i>	9	0,9940	67
Haar Ort.+RBior31	(D123--)	<i>max, min, centred</i>	18	0,9941	170
Haar Ort.+RBior31	(D1234-)	<i>max, min, centred</i>	24	0,9944	218
Haar Ort.+RBior31	(D12345)	<i>max, min, centred</i>	30	0,9946	265
DB1-DB5	(D123--)	<i>max, min, centred</i>	45	0,9952	416
DB1-DB5	(D1234-)	<i>max, min, centred</i>	60	0,9958	543
DB1-DB5	(D12345)	<i>max, min, centred</i>	75	0,9966	666
DB1-DB38	(D123--)	<i>max, min, centred</i>	342	0,9983	3694
DB1-DB38	(D1234-)	<i>max, min, centred</i>	456	0,9993	5691
DB1-DB38	(D12345)	<i>max, min, centred</i>	570	0,9999	9130

Observa-se pelos dados da tabela 34 que o índice AUC melhora conforme aumenta a

dimensionalidade das características. Por outro lado, a diferença percentual entre o pior e o melhor resultado apresentado na tabela 34 é de 0,63%.

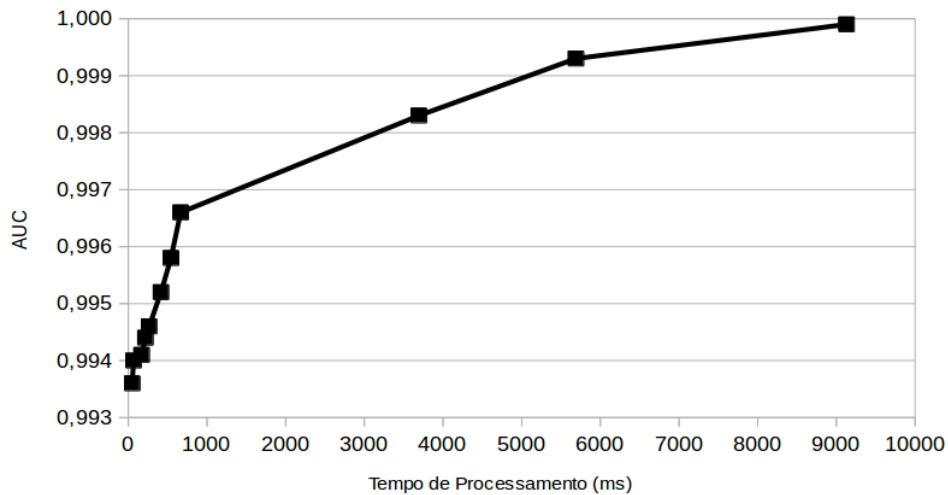
O computador utilizado para esse estudo é um Intel Core i5-2500K CPU @ 3.30GHz x 4 com 8 GB de memória. O classificador de espículas proposto foi programado em Java com recurso multi-tarefa, onde para cada cálculo é alocado um novo processo. O tempo necessário para realizar a transformada wavelet dos 1994 eventos da base de sinais, calcular os descritores selecionados, reduzir a dimensionalidade das características com o algoritmo LDA e aplicar o classificador linear varia bastante. Para o classificador com seis dimensões, o tempo de processamento foi de 51 milissegundos enquanto que para o classificador com 570 dimensões o tempo de processamento foi de 9,13 segundos.

Da tabela 34, é possível verificar dois comportamentos diferentes apresentados pelas figuras 45 e 46. Na figura 45, há um comportamento quase linear entre o tempo de processamento e as dimensões de características utilizadas pelo classificador, enquanto que a curva apresentada pela figura 46 apresenta um crescimento rápido do índice de desempenho AUC até o número de dimensões ser igual a 75, após esse valor, o índice de desempenho AUC continua aumentando, porém, suave e assintoticamente até 0,9999.



**Figura 45:** Tempo de processamento em milissegundos em relação às dimensões de características utilizadas pelo classificador para os resultados da tabela 34.

**Contribuição:** Considerando os resultados obtidos nos experimentos das seções 5.3.1, 5.3.2 e 5.3.4, a conclusão é a de que é mais adequado utilizar um conjunto de wavelets com vários níveis de decomposição e vários descritores do que uma wavelet-mãe ou descritor específico para a detecção de espículas. A escolha desses conjuntos permite obter níveis de detecção elevados conforme a carga computacional que se deseje ou a plataforma computacional disponível para a implementação do sistema de detecção.



**Figura 46:** Tempo de processamento em milissegundos em relação ao índice de desempenho AUC obtido pelo classificador para os resultados da tabela 34.

**Comentário:** Um estudo preliminar abordando a arquitetura do classificador proposto na seção 4.2, figura 19, juntamente com resultados preliminares de classificação foram publicados em Pacola *et al.* (2012a).

Foi apresentado nesta seção a análise de níveis de decomposição wavelet, famílias de wavelets-mães que resultam melhores índices de detecção de espícula assim como um estudo detalhado de descritores utilizados na extração de características dos eventos processados pela transformada wavelet.

#### 5.4 ANÁLISE DA TAXA DE AMOSTRAGEM EM SINAIS DE EEG

Como apresentado no Capítulo 3, a maioria dos estudos encontrados na literatura abrangem sinais amostrados até 200 Hz, com poucos estudos abrangendo a detecção de espículas com sinais amostrados acima de 200 Hz. Isso se deve ao fato das espículas e ondas agudas possuírem uma duração de 20 a 200 ms, o que corresponde a frequências de 5 a 50 Hz. Isso leva à conclusão que amostragens de sinal realizadas a 100 Hz contemplariam esse espectro de frequências.

**Comentário:** Apesar de equipamentos comerciais de eletroencefalografia conseguirem atualmente coletar sinais com frequências de amostragem superiores a 200 Hz, os neurologistas ainda utilizam e armazenam os sinais amostrados nessa frequência. Em parte pelo sinal na tela do equipamento não apresentar diferença visual. Além disso, sinais amostrados com frequências superiores ocupam mais espaço em sistemas de armazenamento, tornando-se desinteressante manter arquivos volumosos sem melhora no aspecto visual do

exame.

Estudos na literatura ((GANDHI; PANIGRAHI; ANAND, 2011; BIROT *et al.*, 2013)) com frequências de amostragem elevadas como 400 Hz e 600 Hz (BIROT *et al.*, 2013) comentam que há características no sinal de EEG, denominadas como oscilações de alta frequência, que não são exploradas e que podem ser relevantes para a detecção de espículas.

Dessa maneira, o objetivo desse estudo é verificar a influência do espectro adicional de frequências, acima de 200 Hz, na detecção de espículas.

Para esse estudo foi utilizado o classificador proposto na seção 4.2, figura 19. Foram selecionadas wavelets-mães normalmente utilizadas em estudos encontrados na literatura: wavelet-mãe DB2, DB3, DB4, DB5, e Haar assim como a wavelet RBior31 selecionada na seção de resultados 5.3. Para essas seis wavelets-mães foi utilizada a extensão de borda SP0 durante a transformada wavelet como definido na seção de resultados 5.2. Os descritores *max*, *min*, e a *centred* foram utilizados para extrair as características do sinal.

A base de eventos original foi amostrada a 1000 Hz, onde foram aplicados filtros notch em 60 Hz e suas harmônicas, juntamente com um filtro passa-baixas com frequência de corte em 350 Hz. O filtro utilizado é do tipo FIR, com 21 coeficientes e janelamento por Blackman, equivalente a um filtro passa-baixas Butterworth de segunda ordem.

A seguir, a base original de sinais foi reamostrada a 500, 333, 250, 200, 167, 143, 125, 111 e 100 Hz. A reamostragem foi feita removendo-se amostras do sinal original de modo que o espaçamento temporal entre as amostras fosse de 2, 3, 4, 5, 6, 7, 8, 9, e 10 milissegundos respectivamente. Para cada frequência, antes da reamostragem, foi aplicado um filtro passa-baixas com frequência de corte na mesma proporção do filtro aplicado na base original igual a  $1000/350 = 2,8571$ . As informações de reamostragem são apresentadas na tabela 35.

A transformada wavelet decompõe os sinais em sub-bandas de frequências como apresentado na seção 2.5.2. As frequências contidas em cada sub-banda dependem da frequência de amostragem do sinal em transformação. Considerando a base de eventos amostrada a 200 Hz, o primeiro nível de decomposição wavelet D1 abrange as frequências de 50 a 100 Hz. O segundo nível de decomposição wavelet D2 abrange as frequências de 25 a 50 Hz e o terceiro nível de decomposição wavelet D3 abrange as frequências de 12,5 a 25 Hz. Tendo como referência esse três níveis de decomposição para a base de eventos amostrada a 200 Hz, tem-se uma quantidade de informação delimitada pelo espectro de frequências de 12,5 a 100 Hz. Conforme é alterada a frequência de amostragem da base de eventos, a informação contida no espectro de frequências de 12,5 a 100 Hz desloca-se entre as novas sub-bandas respectivas

**Tabela 35:** Frequências de reamostragem da base original, intervalo de tempo entre amostras e frequência de corte do filtro passa-baixas aplicado. As informações da base original estão em negrito.

Frequência de amostragem (Hz)	Intervalo de tempo entre amostras (ms)	Frequência de corte do filtro passa-baixas (Hz)
<b>1000</b>	<b>1</b>	<b>350,0</b>
500	2	175,0
333	3	116,5
250	4	87,5
200	5	70,0
167	6	58,5
143	7	50,0
125	8	43,8
111	9	38,9
100	10	35,0

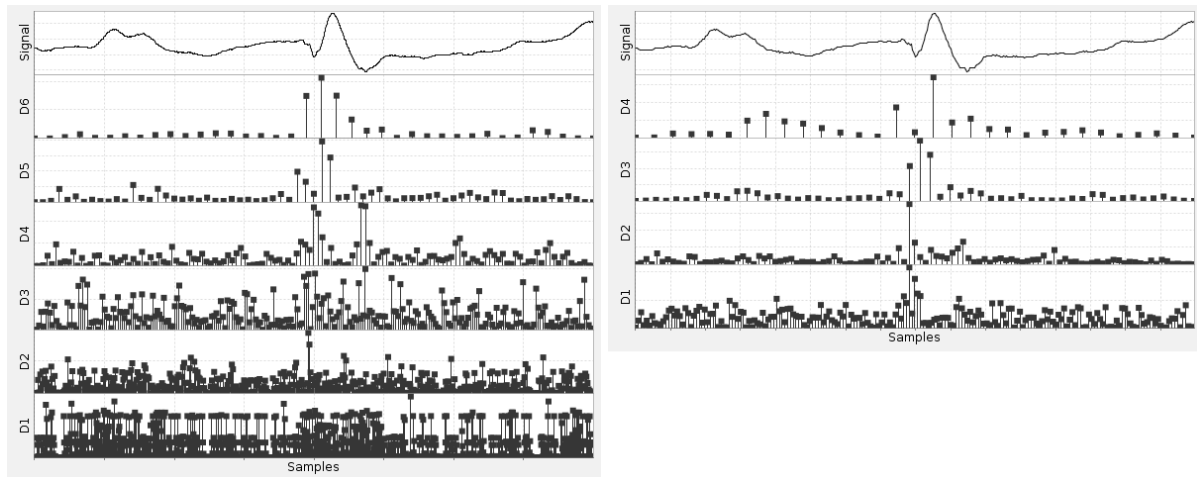
às novas frequências. Desse modo, considerando a informação contida no espectro de 12,5 a 100 Hz, e que para conseguir esse espectro a partir de uma base com frequência de amostragem superior a 200 Hz é necessário realizar mais decomposições, obtém-se informações adicionais não contempladas por sinais amostrados em frequências inferiores a 1000 Hz.

A tabela 36 apresenta as frequências das sub-bandas inerentes a sinais de EEG amostrados nas frequências apresentadas na tabela 35. Em negrito são apresentadas as sub-bandas de decomposição wavelet contendo o espectro de frequências de 12,5 a 100 Hz. Observa-se que para a base amostrada a 100 Hz, o grupo de frequências está contido nos níveis D1 e D2. Para a base amostrada a 250 e 333 Hz, o mesmo grupo de informação está contido entre os níveis D1 a D4. Em 500 Hz, o mesmo grupo de informação está contido nas decomposições D2, D3, D4, e D5, sendo que a decomposição D1, contendo as frequências de 125 a 250 Hz, deve ser calculada para se obter as decomposições seguintes. Da mesma forma temos as decomposições da base amostrada a 1000 Hz, onde o grupo de informação de 12,5 a 100 Hz está contido nas decomposições D3, D4, D5, e D6, com espectro de frequência adicional de 125 a 500 Hz contido nas sub-bandas D1 e D2.

A figura 47 mostra um exemplo das decomposições wavelet calculadas a partir de um mesmo sinal amostrado em frequências diferentes. A figura 47(a) apresenta uma espícula amostrada a 1000 Hz e suas seis primeiras decomposições. A figura 47(b) apresenta a mesma espícula, reamostrada a 200 Hz com quatro níveis de decomposição. Em ambos os casos foi utilizada a wavelet-mãe DB4. Observa-se que as quatro primeiras decomposições da figura 47(b) são equivalentes às quatro últimas decomposições da figura 47(a).

**Tabela 36:** Frequências das sub-bandas de decomposições wavelet para sinais amostrados nas frequências apresentadas na tabela 35. Em negrito estão as sub-bandas que abrangem as frequências de 12,5 Hz a 100 Hz.

Frequência de Amostragem (Hz)	Decomposições Wavelet											
	D6		D5		D4		D3		D2		D1	
1000	<b>7,81</b>	<b>15,63</b>	<b>15,63</b>	<b>31,25</b>	<b>31,25</b>	<b>62,50</b>	<b>62,50</b>	<b>125,00</b>	125,00	250,00	250,00	500,00
500	3,91	7,81	<b>7,81</b>	<b>15,63</b>	<b>15,63</b>	<b>31,25</b>	<b>62,50</b>	<b>62,50</b>	<b>125,00</b>	<b>125,00</b>	125,00	250,00
333	2,60	5,2	5,20	10,41	<b>10,41</b>	<b>20,81</b>	<b>20,81</b>	<b>41,63</b>	<b>41,63</b>	<b>83,25</b>	<b>83,25</b>	<b>166,50</b>
250	1,95	3,91	3,91	7,81	<b>7,81</b>	<b>15,63</b>	<b>15,63</b>	<b>31,25</b>	<b>31,25</b>	<b>62,50</b>	<b>62,50</b>	<b>125,00</b>
200	1,56	3,13	3,13	6,25	6,25	12,50	<b>12,50</b>	<b>25,00</b>	<b>25,00</b>	<b>50,00</b>	<b>50,00</b>	<b>100,00</b>
167	1,30	2,61	2,61	5,22	5,22	10,44	<b>10,44</b>	<b>20,88</b>	<b>20,88</b>	<b>41,75</b>	<b>41,75</b>	<b>83,50</b>
143	1,12	2,23	2,23	4,47	4,47	8,94	<b>8,94</b>	<b>17,88</b>	<b>17,88</b>	<b>35,75</b>	<b>35,75</b>	<b>71,50</b>
125	0,98	1,95	1,95	3,91	3,91	7,81	<b>7,81</b>	<b>15,63</b>	<b>15,63</b>	<b>31,25</b>	<b>31,25</b>	<b>62,50</b>
111	0,87	1,73	1,73	3,47	3,47	6,94	<b>6,94</b>	<b>13,88</b>	<b>13,88</b>	<b>27,75</b>	<b>27,75</b>	<b>55,50</b>
100	0,78	1,56	1,56	3,13	3,13	6,25	6,25	12,50	<b>12,50</b>	<b>25,00</b>	<b>25,00</b>	<b>50,00</b>



(a) 1000 Hz

(b) 200 Hz

**Figura 47:** WT DB4 com (a) seis primeiros níveis de decomposição de uma espícula amostrada a 1000 Hz e (b) quatro primeiras decomposições da mesma espícula de (a) reamostrada a 200 Hz. As frequências respectivas a cada sub-banda, D1, D2, D3, D4, D5, e D6 de cada amostragem são apresentadas na tabela 36.

Considerando as informações contidas no espectro de frequências entre 12,5 e 100 Hz, o classificador proposto na seção 4.2 opera com níveis diferentes de decomposições, conforme a frequência de amostragem da base de eventos. Isso é apresentado pela tabela 37, onde os níveis de decomposição D1 e D2 são utilizados para sinais amostrados a 100 Hz, D1, D2, e D3 para sinais amostrados a 200 Hz, e os níveis de decomposição D1, D2, D3, D4, D5, e D6 para sinais amostrados a 1000 Hz.

Os índices de desempenho AUC obtidos são apresentados na tabela 38. Para fins de comparação, os cálculos foram repetidos para o descritor entropia aproximada (ApEn). O



**Tabela 37:** Níveis de decomposição utilizado pelo classificador proposto na seção 4.2, figura 19 para sinais amostrados em 100, 111, 125, 143, 167, 200, 250, 333, 500 e 1000 Hz.

Frequência de amostragem	Níveis de decomposição
1000	D1, D2, D3, D4, D5, D6
500	D1, D2, D3, D4, D5
333	D1, D2, D3, D4
250	D1, D2, D3, D4
200	D1, D2, D3
167	D1, D2, D3
143	D1, D2, D3
125	D1, D2, D3
111	D1, D2, D3
100	D1, D2

resultado é apresentado na tabela 39.

**Tabela 38:** Índice de Desempenho AUC para a base de eventos amostrada nas frequências apresentadas na tabela 35 utilizando os descritores *centred*, *max*, e *min*.

Decomposições	Freq. (Hz)	DB2	DB3	DB4	DB5	Haar	RBior31
D123456	1000	0,9911	0,9894	0,9877	0,9705	0,9942	0,9949
D12345-	500	0,9870	0,9891	0,9842	0,9565	0,9937	0,9931
D12345--	333	0,9773	0,9829	0,9796	0,9327	0,9971	0,9943
D1234---	250	0,9873	0,9892	0,9634	0,9663	0,9930	0,9938
D123----	200	0,9863	0,9693	0,9631	0,9335	0,9912	0,9935
D123----	167	0,9742	0,9835	0,9326	0,9620	0,9955	0,9940
D123----	143	0,9789	0,9848	0,9380	0,9708	0,9948	0,9930
D123----	125	0,9831	0,9794	0,9603	0,9814	0,9883	0,9927
D123----	111	0,9828	0,9885	0,9550	0,9822	0,9905	0,9924
D12----	100	0,9728	0,9693	0,9469	0,9698	0,9901	0,9926

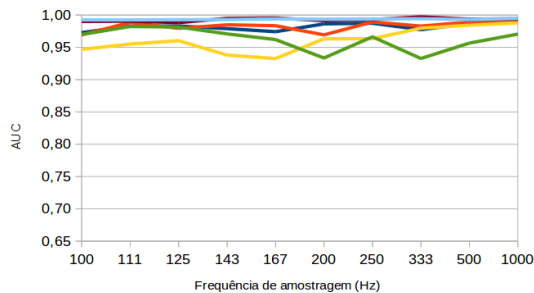
É possível verificar pela tabelas 38 e 39 que o desempenho do classificador utilizando eventos de EEG amostrados a 1000 Hz são superiores aos índices obtidos utilizando a mesma base de eventos reamostrada em frequências inferiores. Observa-se que o índice de desempenho AUC do classificador não aumenta constantemente conforme aumenta a frequência de amostragem. Porém, na média, o crescimento é progressivo conforme os coeficientes referentes às frequências de 100 Hz a 500 Hz são adicionados.

Os dados das tabelas 38 e 39 são apresentados pela figura 48.

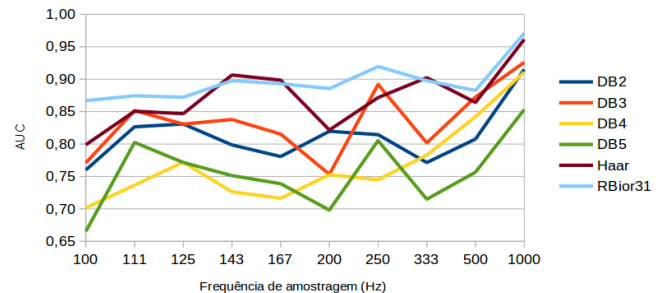
Observa-se na figura 48 que a utilização de descritores diferentes acarreta uma diferença significativa no comportamento do desempenho do classificador em função da

**Tabela 39:** Índice de desempenho AUC para a base de eventos amostrada nas frequências apresentadas na tabela 35 utilizando o descritor entropia aproximada (ApEn).

Decomposições	Freq. (Hz)	DB2	DB3	DB4	DB5	Haar	RBior31
D123456	1000	0,9150	0,9259	0,9108	0,8532	0,9613	0,9708
D12345-	500	0,8077	0,8726	0,8421	0,7566	0,8641	0,8826
D12345-	333	0,7715	0,8017	0,7829	0,7149	0,9023	0,8977
D1234--	250	0,8144	0,8919	0,7450	0,8053	0,8716	0,9195
D123---	200	0,8197	0,7533	0,7528	0,6982	0,8216	0,8855
D123---	167	0,7807	0,8153	0,7163	0,7388	0,8985	0,8930
D123---	143	0,7986	0,8378	0,7264	0,7513	0,9064	0,8976
D123---	125	0,8309	0,8307	0,7719	0,7716	0,8468	0,8721
D123---	111	0,8266	0,8517	0,7366	0,8025	0,8506	0,8745
D12----	100	0,7600	0,7710	0,7018	0,6653	0,7988	0,8668



(a) *centred, max, min.*



(b) ApEn

**Figura 48:** Índice de desempenho AUC conforme frequência de amostragem da tabela 35 utilizando os descritores (a) *centred, max, min* e (b) entropia aproximada.

frequência de amostragem do sinal. Utilizando os descritores *centred, max, e min* (tabela 38 e figura 48(a)), há uma variação de 4,8% entre o melhor resultado obtido com a wavelet-mãe RBior31 com amostragem a 1000 Hz e índice de desempenho AUC igual a 0,9949 e o pior resultado, obtido pela wavelet-mãe DB4 com amostragem a 100 Hz e índice de desempenho AUC igual a 0,9469.

Utilizando o descritor entropia aproximada (tabela 39 e figura 48(b)), o índice de desempenho AUC varia 30,55% entre o melhor resultado obtido com a wavelet-mãe RBior31 com amostragem a 1000 Hz e índice de desempenho AUC igual a 0,9708 e o pior resultado, obtido pela wavelet-mãe DB4 com amostragem a 100 Hz e índice de desempenho AUC igual a 0,6653.

Amostrando sinais a 1000 Hz ao invés de 100 Hz, possibilitou uma melhora no desempenho do classificador de 4,08% utilizando DB4, enquanto as wavelets-mães DB2, DB3, e DB5 obtiveram melhora de 1,83%, 2,01%, e 0,07%, respectivamente, utilizando os descritores

selecionados *centred*, *max*, e *min*.

As wavelets-mães Haar e RBior31 apresentaram melhora em seus resultados com o aumento da frequência de amostragem. Porém, não de maneira significativa, com acréscimo de 0,23% e 0,41%, respectivamente.

**Contribuição:** É apresentado neste trabalho que a informação espectral contida acima de 200 Hz influencia significativamente a detecção de espículas, melhorando o índice de detecção de espículas em até 4,08% considerando wavelets-mães comumente encontradas na literatura.

**Comentário:** Somente dois trabalhos foram encontrados na literatura envolvendo taxas de amostragem de sinais de EEG superiores a 200 Hz e detecção de espículas através da transformada wavelet (GANDHI; PANIGRAHI; ANAND, 2011; BIROT *et al.*, 2013). Mesmo assim, é apenas informado que existe a possibilidade de melhora na detecção, não sendo apresentados resultados quantitativos.

## 5.5 COMENTÁRIOS

Neste capítulo foi avaliada a arquitetura baseada em LDA proposta na seção 4.2 para a detecção de espícula. Foram avaliados os métodos de detecção de borda na transformada wavelet, quais wavelets-mães, número de decomposições e descritores que apresentam os melhores resultados para a detecção de espícula assim como a frequência de amostragem do sinal influencia a detecção de eventos espícula e não-espícula.

No próximo capítulo são feitos comentários sobre os resultados obtidos, conclusões e sugestões para trabalhos futuros.

## 6 CONCLUSÃO E COMENTÁRIOS FINAIS

É apresentado neste capítulo um resumo das análises, comentários e contribuições realizadas no capítulo 5. É feito ainda uma discussão geral sobre o trabalho realizado e sugestões de trabalhos futuros.

### 6.1 AVALIAÇÃO DOS MÉTODOS DE EXTENSÃO DE BORDA

A avaliação dos métodos de extensão de borda pode ser considerado um conhecimento básico necessário no uso da transformada wavelet discreta. Isso é embasado no fato de que os efeitos de seu uso são conhecidos em outras áreas da engenharia como apresentado por Katunin (2012) e Montanari *et al.* (2015). Os efeitos da extensão de borda são inclusive comentados no estudo inicial da DWT feito por Mallat (1989).

Entretanto, considerando a aplicação da transformada wavelet para a detecção de espículas em sinais de EEG, não foi encontrado nenhum trabalho na literatura que aborde esse tema, nem menção sobre métodos de extensão de borda utilizados em experimentos realizados.

Dessa maneira, a contribuição desse experimento é justamente o estudo das distorções causadas pela interpretação do método de extensão de borda na detecção de espículas.

Foi verificado que os métodos de extensão de borda são dependentes da quantidade de coeficientes que a wavelet-mãe possui.

Além disso, as extensões de borda PER, PPD, ZPD e ASYMH sempre apresentaram os piores resultados para a classificação de espículas, não sendo indicadas para uso nessa aplicação. Isso é importante pois a extensão de borda ZPD é a que exige menor esforço de implementação, onde os valores adicionais de extensão de borda são iguais a zero.

E concluindo esse experimento, a extensão de borda SP0 se mostrou mais adequada para uso de wavelets com até oito coeficientes (Haar, DB2, Sym2, DB3, Sym3, RBior13, DB4, e Sym4). Isso é importante pois wavelets com até oito coeficientes são as mais utilizadas em estudos encontrados na literatura.

Considerando wavelets com mais de oito coeficientes, a extensão de borda SYMH se mostrou melhor para a detecção de espículas para wavelets DB6, Coif2, Coif3, Coif4, Coif5, e DB20.

A extensão de borda SYMW apresentou os melhores resultados para wavelets DB5, Sym5, e DB8, enquanto que a extensão de borda ASYMW se mostrou adequada para wavelets DB10 e Dmey.

A extensão de borda ASYMH apresentou o pior resultado, independente do número de coeficientes da wavelet, degradando os resultados em até 44,11%.

A wavelet-mãe Haar, apresentada nas figuras 35 e 37 não sofre influência da extensão de borda usada, uma vez que o pior resultado obtido foi com a extensão de borda ZPD, a qual degradou o desempenho do classificador em 0,09%.

## 6.2 ANÁLISE DE FAMÍLIAS DE WAVELETS-MÃES E DESCRITORES

Como verificado no estado da arte, no capítulo 3, em mais de 20 anos de pesquisas, nos 42 estudos avaliados, não há consenso entre os autores sobre qual seria a melhor wavelet-mãe, número de decomposições wavelet ou descritor que seja mais adequado para a detecção de espículas.

Nesse estudo, primeiramente, foi feita uma avaliação detalhada dos níveis de decomposição que resultam em melhores resultados de classificação. Porém, o que se constatou é que cada nível de decomposição tem a sua parcela de influência na classificação dos eventos. A união dos níveis de decomposição apresentam índices melhores do que níveis de decomposição analisados individualmente.

A partir do quarto nível de decomposição ainda ocorre melhora no desempenho do classificador, porém menos significativa, com aumento da dimensionalidade das características extraídas do sinal.

Em seguida, foram analisados cinco grupos de famílias de wavelets-mães da seguinte maneira: análise individual de cada wavelet-mãe; agrupadas aditivamente com outras wavelets-mães da mesma família; e por fim, entre famílias. O teste envolveu 98 wavelets-mães.

A análise individual revelou que a wavelet-mãe Reverse Biortogonal 31 (RBior31) apresentou o melhor desempenho.

A análise entre wavelets-mães da mesma família revelou que o uso das 38 wavelets-mães da família Daubechies (DB1-DB38) apresenta o melhor desempenho.

A análise de wavelets-mães entre famílias revelou que, seguido do agrupamento (DB1-DB38) o melhor desempenho vem do grupo (RBior31+Haar Ortogonal) com a vantagem de apresentar um número menor de dimensões de características.

Por fim, a análise de descritores validou, entre os descritores analisados, que os descritores valor máximo, valor mínimo e a energia centrada do vetor usados de modo conjunto, apresentam o melhor resultado de classificação. Esses resultados levam às seguintes contribuições:

1. É mais adequado utilizar um conjunto de wavelets com vários níveis de decomposição e vários descritores do que uma wavelet-mãe ou descritor específico para a detecção de espículas. A escolha desses conjuntos permitem obter níveis de detecção elevados conforme a carga computacional que se deseje ou a plataforma computacional disponível para implementação do sistema de detecção.
2. O descritor energia centrada do vetor (*centred*), (seção 4.3.2), em conjunto com o uso da transformada wavelet e de acordo com os resultados apresentados, obteve melhores índices de desempenho na classificação de espículas. Esse descritor foi desenvolvido ao longo desse estudo como método para identificar coeficientes com mais alta energia no centro da janela resultante da transformada wavelet. Esse descritor foi concebido baseado na possibilidade de utilizar um descritor para a detecção de espículas de modo contínuo, ou seja, sem a prévia seleção ou janelamento das amostras de exames de EEG, onde a leitura dos dados dos canais do exame de EEG seriam feitos continuamente e o processamento realizado através de janelas deslizantes.
3. A arquitetura do classificador proposto na seção 4.2 (figura 19), é inédito na classificação de espículas em utilizar o LDA na redução de características a ponto de possibilitar um classificador linear de separar eventos espícula de eventos não-espícula com os índices de desempenho obtidos nesse estudo. Além disso, se mostrou versátil nos experimentos aceitando mudanças de configuração na arquitetura, seja aumentando o número de wavelets-mães, o número de decomposições wavelet ou a quantidade de descritores, sem a necessidade de treinamento intensivo.

Os resultados destes estudos foram publicados por Pacola *et al.* (2012a).

### 6.3 ANÁLISE DA TAXA DE AMOSTRAGEM EM SINAIS DE EEG

Foram encontrados na literatura apenas dois estudos que contemplam sinais de EEG amostrados acima de 200 Hz, como descrito no capítulo 3. Outro fato que reforça essa constatação é que as bases públicas de sinais de EEG foram amostradas a 256, 200 e 173,5 Hz. Isso motivou a obtenção de sinais amostrados acima de 200 Hz para verificar se há influência do espectro adicional de frequências na detecção de espículas.

O estudo apresentou diferenças na capacidade de detecção de espículas pelo classificador atribuídas à frequência de amostragem do sinal. Considerando o resultado da tabela 39 referente a wavelet-mãe DB4, que é uma wavelet-mãe utilizada frequentemente pela literatura, tem-se um ganho de 15,8% pelo índice de desempenho AUC. Considerando a base de sinais utilizada nesse estudo, o uso de uma base amostrada a 1000 Hz ao invés de 200 Hz permitiu ao classificador, dentre os 1500 eventos não-espícula, reconhecer corretamente 237 eventos que antes eram tidos como falso positivos. Considerando os eventos espícula, somente 45 espículas foram classificadas como falso negativo, contra 124 espículas classificadas como falso negativo com o sinal sendo amostrado a 200 Hz. Isso represente um ganho significativo.

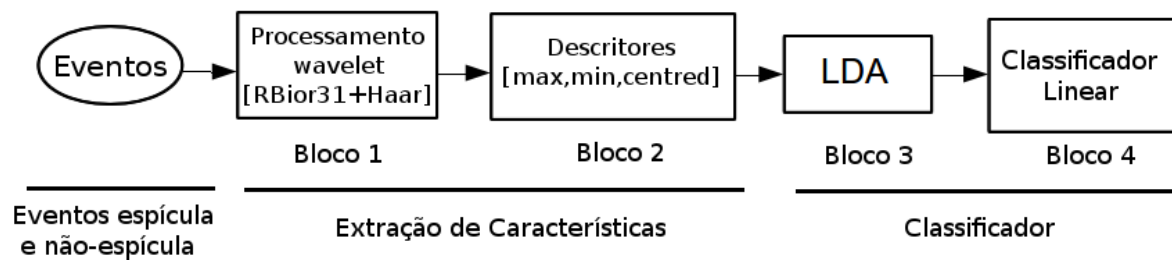
Utilizando descritores e wavelets-mães selecionadas, a contribuição apresentada neste trabalho, é que a informação espectral contida acima de 200 Hz também influencia a detecção de espículas, melhorando o índice de detecção de espículas em até 4,08%.

Aparelhos atuais de EEG permitem a amostragem do sinal em frequências superiores a 200 Hz, porém não é utilizada essa característica pelo fato da análise visual do especialista não perceber diferença entre registros feitos a 200 Hz ou em frequências superiores. Assim, seria possível um ganho imediato no desempenho de detectores de espícula caso os exames fossem registrados em taxas de amostragem elevadas (1000 Hz), e novas descobertas poderiam ser feitas por pesquisadores caso esses registros amostrados a essas taxas estivessem disponibilizados publicamente.

### 6.4 CLASSIFICADOR DE EVENTOS EPILEPTIFORMES BASEADO EM LDA

Com base nos resultados apresentados no Capítulo 5 juntamente com as informações apresentadas na tabela 34, o classificador proposto na seção 4.2, figura 19, tem como configuração sugerida para a detecção de espículas em exames de EEG o uso combinado das wavelets-mães RBior31 e Haar Ortogonal, ambas com quatro níveis de decomposição. Após a aplicação da transformada wavelet, são calculados os descritores max, min, e centred. A

configuração sugerida para a detecção de espículas é apresentada na figura 49.



**Figura 49:** Diagrama em blocos do classificador proposto para o estudo dos métodos de extensão de bordas. O classificador proposto na seção 4.2, figura 19, tem como configuração sugerida o uso em conjunto das wavelets-mães RBior31 e Haar Ortogonal com quatro níveis de decomposição, assim como o uso conjunto dos descritores max, min, e centred.

A configuração sugerida, para a base de dados apresentada na seção 4.1, resulta em um índice de desempenho AUC igual a 0,9944.

A arquitetura do classificador proposto baseado no LDA é a contribuição desse trabalho. Nos estudos realizados verificou-se que o classificador identifica os eventos com tempo computacional aceitável, abrindo a possibilidade para implementação em um sistema embarcado.

## 6.5 LIMITAÇÕES DO ESTUDO

Conforme descrito nas seções 3.2.1, as bases de espícula públicas foram registradas a uma frequência de amostragem inferior a 200 Hz. Bases contendo espículas amostradas acima dessa frequência não estão atualmente disponibilizadas.

Um objetivo específico desse trabalho é justamente a análise da influência de componentes espectrais acima de 200 Hz no sinal de EEG na detecção de espículas. Para que isso fosse possível, era necessário obter uma base de sinais amostrada a 1000 Hz, permitindo reamostrar o sinal em frequências inferiores e analisar o desempenho do classificador na detecção de espículas.

Encontrar um neurologista com disponibilidade e equipamento condizente com o objetivo do estudo não é uma tarefa fácil. Além disso, o desenvolvimento do software de seleção de espículas, seleção do paciente e liberação do comitê de ética exigiram 6 meses.

A identificação e seleção das espículas pelo neurologista, assim como o processamento digital das espículas para obter a base descrita na seção 4.1 exigiram outros 6 meses de trabalho.

A base de espículas comentada na seção 4.1 foi obtida de um único paciente



com espículas variando entre 60 a 200 milissegundos como pode ser visto na figura 18. Eventualmente essa distribuição pode variar de paciente para paciente. Para cobrir a variabilidade das espículas (20 a 200 milissegundos) assim como obter espículas de um maior número de pacientes, não haveria tempo hábil para a conclusão do trabalho de doutorado.

Além da limitação do número de pacientes, há também a limitação do número de neurologistas. Como apresentado no capítulo 3 pelos autores Halford *et al.* (2013), não há concordância entre os próprios especialistas no reconhecimento de padrões epileptiformes. Isso é apresentado na tabela 10 onde cada especialista apresenta índices de sensibilidade e especificidade diferentes uns dos outros.

Esse é o limitante dessa pesquisa, pois os resultados apresentados no capítulo 5 podem ter mapeado a detecção de espículas desse paciente específico sob o ponto de vista desse neurologista específico.

Essas limitações todavia não invalidam a pesquisa. Os métodos de pesquisa apresentados, os resultados e contribuições preliminares, assim como o classificador proposto, continuam válidos. O que pode sofrer alteração será a configuração final do classificador, onde o agrupamento de wavelets-mães, número de decomposições e descritores utilizados podem não estar otimizados para um grupo maior de pacientes e neurologistas.

Além disso, a arquitetura do classificador proposto permite determinar a melhor wavelet-mãe e descritor de maneira versátil, adaptando-se facilmente à outros pacientes e/ou neurologistas.

## 6.6 COMENTÁRIOS FINAIS

O classificador de espículas proposto na seção 4.2, figura 19, é a base dos resultados apresentados nesse trabalho. A arquitetura proposta se mostrou versátil e ágil na mudança de sua configuração. Isso foi possível devido ao uso da análise de discriminante linear.

A programação dos métodos e ferramentas em Java também propiciou vantagens que não foram contempladas no início desse estudo. A primeira vantagem, detectada na validação de ferramentas, foi a constatação das distorções causadas pelo método de extensão de bordas. Isso gerou uma contribuição na área de detecção de espículas e os resultados foram publicados por Pacola *et al.* (2012b).

Outra vantagem foi a velocidade de obtenção dos resultados. Após o processamento wavelet e cálculo dos descritores, foram geradas mais de uma mil planilhas. O carregamento dessas planilhas no software implementado em java durava menos de 30 segundos. Uma tentativa preliminar de carregamento de uma dessas planilhas no Matlab levou mais de um minuto. A implementação da arquitetura de modo *multi-thread* permitiu, além de tudo, testar várias combinações e obter resultados significativos em um tempo razoável.

## 6.7 TRABALHOS FUTUROS

A detecção de eventos epileptiformes não é uma tarefa fácil. Percebe-se pela quantidade de trabalhos existentes aplicando as mais diversas técnicas e que, de modo geral, não conseguem implementar um algoritmo de fato que permita o avanço tanto tecnológico quanto o bem estar de milhares de pessoas que sofrem com a epilepsia.

Esse trabalho contempla um número significativo de eventos espícula, atingindo índices de desempenho muito próximos da detecção completa, como observado no experimento da seção 5.3.4, tabela 34, onde o índice de desempenho AUC atinge 0,9999. Porém, a base utilizada para esses experimentos contempla espículas de 60 a 200 ms, sendo necessário colher mais eventos que cubram todo o espectro de duração das espículas.

Apesar do número significativo de eventos espícula utilizados nesse trabalho, esses eventos foram retirados de um único paciente, e selecionados por um único neurologista. Isso não invalida o trabalho, porém, abre a possibilidade de trabalhos futuros para a coleta de eventos epileptiformes de mais pacientes e repetição das técnicas apresentadas nesse trabalho.

Considerando a base de dados obtida, se for devidamente trabalhada e obtendo a liberação do comitê de ética do Hospital Pequeno Príncipe, a base de dados atual seria uma contribuição prévia caso se tornasse acessível a pesquisadores, e se tornaria mais abrangente com o incremento de novos eventos epileptiformes, pacientes e especialistas avaliadores.

Considerando a arquitetura do classificador proposto, um trabalho futuro seria a implementação do classificador LDA de modo embarcado. Isso poderia viabilizar um equipamento versátil na detecção de espículas que operaria de acordo com o paciente ou conforme configuração do especialista.

Outro trabalho futuro necessário, constatado ao longo do desenvolvimento desse estudo, foi a resposta em frequência da transformada wavelet em conjunto com o tempo de duração da espícula. Sabendo-se que a duração da espícula é de 20 ms a 200 ms, e que os coeficientes das wavelets-mães são filtros passa-altas e passa-baixas, é coerente afirmar

que a detecção de espículas com determinada duração deve estar atrelada ao uso de uma determinada gama de wavelets-mãe. Esse trabalho não foi encontrado na literatura e traria novos conhecimentos à área da engenharia biomédica.

## REFERÊNCIAS

- ACHARYA, R.; SREE, V.; SWAPNAD, G.; MARTIS, R. J.; SURI, J. Automated EEG analysis of epilepsy: A review. **Knowledge-Based Systems**, v. 45, p. 147–165, 2013.
- ADELI, H.; GHOSH-DASTIDAR, S.; DADMEHR, N. A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. **IEEE Transactions On Biomedical Engineering**, v. 54, n. 2, p. 205–211, February 2007.
- ADELI, H.; ZHOU, Z.; DADMEHR, N. Analysis of EEG records in an epileptic patient using wavelet transform. **Journal of Neuroscience Methods**, v. 123, n. 1, p. 69–87, February 2003.
- ANDRZEJAK, R. G.; SCHINDLER, K.; RUMMEL, C. **The Bern-Barcelona EEG database**. October 2012. Disponível em: <<http://ntsa.upf.edu/downloads/andrzejak-rg-schindler-k-rummel-c-2012-nonrandomness-nonlinear-dependence-and>>.
- ANDRZEJAK, R. G.; SCHINDLER, K.; RUMMEL, C. Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients. **PHYSICAL REVIEW E**, v. 86, p. 046206, October 2012.
- ANDRZEJAK, R. G.; WIDMAN, G.; LEHNERTZ, K.; RIEKE, C.; DAVID, P.; ELGER, C. The epileptic process as nonlinear deterministic dynamics in a stochastic environment: an evaluation on mesial temporal lobe epilepsy. **Epilepsy**, v. 44, n. 2-3, p. 129–140, 2001.
- ANGHINAH, R.; BASILE, L.; SCHMIDT, M.; SAMESHIMA, K.; GATTAZ, W. Artefatos biológicos no EEG quantitativo. **Arquivos de Neuro-Psiquiatria**, v. 64, n. 2, p. 264–268, June 2006.
- APOSTOL, T. M. **Calculus, multi variable calculus and linear algebra, with applications to difdifferent equations and probability**. 2nd. ed. USA: Indiana University, 1969.
- BALASUBRAMANIAN, K.; OBEID, I. Fuzzy logic-based spike sorting system. **Journal of Neuroscience Methods**, v. 198, n. 1, p. 125–134, May 2011.
- BIROT, G.; KACHENOURA, A.; ALBERA, L.; BENAR, C.; WENDLING, F. Automatic detection of fast ripples. **Journal of Neuroscience Methods**, v. 213, n. 2, p. 236–249, March 2013.
- BLANCO, S.; FIGLIOLA, A.; QUIROGA, R. Q.; ROSSO, O. A.; SERRANO, E. Time-frequency analysis of electroencephalogram series. iii. wavelet packet and information cost function. **Physical Review E: Statistical, Nonlinear, and Soft Matter Physics**, v. 57, n. 1, p. 932–940, January 1998.
- BLUME, W. T.; LUDERS, H. O.; MIZRAHI, E.; TASSINARI, C.; BOAS, W. v. E.; ENGEL, J. Glossary of descriptive terminology for ictal semiology: Report of the ILAE task force on classification and terminology. **Epilepsia**, v. 42, n. 9, p. 1212–1218, 2001.

- BRECHET, L.; LUCAS, M.-F.; DONCARLI, C.; FARINA, D. Compression of biomedical signals with mother wavelet optimization and best-basis wavelet packet selection. **IEEE Transactions on Biomedical Engineering**, v. 54, n. 12, p. 2186–2192, December 2007.
- CALAIS, G. S. P.; LIMA, S. O. Epilepsia. **Saúde e Economia**, Brasília/DF, v. 3, n. 5, p. 1–4, Abril 2011.
- CHATRIAN, G.; BERGAMINI, L.; DONDEY, M.; KLASS, D.; M.L., B.; PETERSÉN, I. A glossary of terms most commonly used by clinical electroencephalographers. **Electroencephalography and Clinical Neurophysiology**, v. 37, n. 5, p. 538–548, 1974.
- CLARK, I.; BISCAY, R.; ECHEVERRÍA, M.; VIRUÉS, T. Multiresolution decomposition of non-stationary EEG signals: a preliminary study. **Computers in Biology and Medicine**, v. 25, n. 4, p. 373–382, July 1995.
- CORREA, A. G.; OROSCO, L.; DIEZ, P.; LACIAR, E. Automatic detection of epileptic seizures in long-term EEG records. **Computers in Biology and Medicine**, v. 57, p. 66–73, February 2015.
- DAUBECHIES, I. Orthonormal bases of compactly supported wavelets. **Communications Pure and Applied Mathematics**, v. 41, n. 7, p. 909–996, October 1988.
- DAUBECHIES, I. The wavelet transform, time-frequency localization and signal analysis. **IEEE Transactions on Information Theory**, v. 36, n. 5, p. 961–1005, September 1990.
- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2nd. ed. 605 Third Avenue, New York, NY: Wiley Interscience, 2001.
- EHRENTREICH, F.; SUEMMCHEN, L. Spike removal and denoising of raman spectra by wavelet transform methods. **Analytical Chemistry**, v. 73, n. 17, p. 4364–4373, September 2001.
- ERCELEBI, E.; SUBASI, A. Classification of EEG for epilepsy diagnosis in wavelet domain using artificial neural network and multi linear regression. In: **Signal Processing and Communications Applications, 2006 IEEE 14th**. Antalya: IEEE, 2006. p. 1–4.
- ERKEL, A. R. v.; PATTYNAMA, P. M. T. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. **European Journal of Radiology**, v. 27, n. 2, p. 88–94, May 1998.
- FATOURECHI, M.; BIRCH, G. E.; WARD, R. K. Application of a hybrid wavelet feature selection method in the design of a self-paced brain interface system. **Journal of Neuro Engineering and Rehabilitation**, v. 4, n. 11, p. 1–13, April 2007.
- FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, June 2006.
- FISHER, R.; BOAS, W.; BLUME, W.; ELGER, C.; GENTON, P.; LEE, P.; JR., J. E. Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE). **Epilepsia**, v. 46, n. 4, p. 470–472, 2005.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v. 7, n. 2, p. 179–188, September 1936.

- GAJIC, D.; DJUROVIC, Z.; GENNARO, S.; GUSTAFSSON, F. Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. **Biomedical Engineering: Applications, Basis and Communications**, v. 26, n. 2, p. 1–13, April 2014.
- GANDHI, T.; PANIGRAHI, B. K.; ANAND, S. A comparative study of wavelet families for EEG signal classification. **Neurocomputing**, v. 74, n. 17, p. 3051–3057, October 2011.
- GEVA, A. B.; KEREM, D. H. Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervised fuzzy clustering. **IEEE Transactions on Biomedical Engineering**, v. 45, n. 10, p. 1205–1216, October 1998.
- HALFORD, J. J.; SCHALKOFF, R. J.; ZHOU, J.; BENBADIS, S. R.; TATUM, W. O.; TURNER, R. P.; SINHA, S. R.; FOUNTAIN, N. B.; ARAIN, A.; PRITCHARD, P. B.; KUTLUAY, E.; MARTZ, G.; EDWARDS, J. C.; WATERS, C.; DEAN, B. C. Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis. **Journal of Neuroscience Methods**, v. 212, n. 2, p. 308–316, January 2013.
- HSU, K. C.; YU, S. N. Classification of seizures in EEG using wavelet-chaos methodology and genetic algorithm. In: DOESSEL, O.; SCHLEGEL, W. (Ed.). **World Congress on Medical Physics and Biomedical Engineering**. Munich: IFMBE Proceedings, 2009. v. 25, n. 4, p. 564–567.
- INDIRADEVI, K. P.; ELIAS, E.; SATHIDEVI, P. S.; NAYAK, S. D.; RADHAKRISHNAN, K. A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. **Computers in Biology and Medicine**, IEEE, v. 38, n. 7, p. 805–816, July 2008.
- INUSO, G.; FORESTA, F. L.; MAMMONE, N.; MORABITO, F. C. Brain activity investigation by EEG processing: Wavelet analysis, kurtosis and renyi's entropy for artifact detection. In: **International Conference on Information Acquisition, 2007. ICIA '07**. Jeju City, Korea: IEEE, 2007. p. 195–200.
- KALAYCI, T.; OZDAMAR, O. Wavelet preprocessing for automated neural network detection of EEG spikes. **IEEE Engineering in Medicine and Biology**, v. 14, n. 2, p. 160–166, March/April 1995.
- KALAYCI, T.; OZDAMAR, O.; ERDOL, N. The use of wavelet transform as a preprocessor for the neural network detection of EEG spikes. In: **Proceedings of the 1994 IEEE Southeastcon '94. Creative Technology Transfer - A Global Affair**. Miami, FL: IEEE, 1994. p. 1–3.
- KATUNIN, A. Reduction of boundary effect during structural damage identification using wavelet transform. **Selected Engineering Problems**, v. 1, n. 3, p. 97–102, 2012.
- KUMAR, Y.; DEWAL, M. L.; ANAND, R. S. Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network. **Signal, Video and Image Processing**, v. 8, n. 8, p. 1323–1334, 2014.
- LAGET, P.; SALBREUX, R. **Atlas D'electroencephalographie infantile**. Paris: Masson et Cie, 1967.

LATHI, B. P. **Sinais e Sistemas Lineares**. 2. ed. Porto Alegre: Artmed, 2007.

LEONEL, L. D. Abril 2015. Disponível em: <<http://www.epilepsiabrasil.org.br/noticias/caminhada-purple-day-reune-mais-de-500-pessoas-no-parque-villa-lobos-2>>.

MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 7, p. 674–693, July 1989.

MALLAT, S. G. **A Wavelet Tour of Signal Processing**. 2nd. ed. London: Academic Press, 2009.

MALMIVUO, J.; PLONSEY, R. **Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields**. 1. ed. New York: Oxford University Press, 1995.

MARTINEZ, A. M.; KAK, A. C. PCA versus LDA. **IEEE Transactions on Pattern Analysis and Machine Intelligence** **23**, v. 23, n. 2, p. 228–233, February 2001.

MIRZAEI, A.; AYATOLLAHI, A.; VAVADI, H. Statistical analysis of epileptic activities based on histogram and wavelet-spectral entropy. **Biomedical Science and Engineering**, v. 4, n. 3, p. 207–213, March 2011.

MISITI, M.; MISITI, Y.; OPPENHEIM, G.; POGGI, J.-M. **Wavelet Toolbox User's Guide**. 3 Apple Hill Drive, Natick: The MathWorks, Inc., 1996.

MISITI, M.; MISITI, Y.; OPPENHEIM, G.; POGGI, J.-M. **Wavelet Toolbox Reference**. 3 Apple Hill Drive, Natick: The MathWorks, Inc., 2013.

MONTANARI, L.; BASU, B.; SPAGNOLI, A.; BRODERICK, B. M. A padding method to reduce edge effects for enhanced damage identification using wavelet analysis. **Mechanical Systems and Signal Processing**, v. 52–53, p. 264 – 277, February 2015.

NENADIC, Z.; BURDICK, J. W. Spike detection using the continuous wavelet transform. **IEEE Transactions on Biomedical Engineering**, v. 52, n. 1, p. 74–87, January 2005.

NOACHTAR, S.; BINNIE, C.; EBERSOLE, J.; MAUGUIÈRE, F.; SAKAMOTO, A.; WESTMORELAND, B. **A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form the EEG Findings**. Munique: Elsevier Sciences B.V., 1999.

NUNES, T. M.; COELHO, A. L.; LIMA, C. A. M.; PAPA, J. P.; ALBUQUERQUE, V. H. C. EEG signal classification for epilepsy diagnosis via optimum path forest – a systematic assessment. **Neurocom**, v. 136, p. 103–123, 2014.

OCAK, H. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. **Expert Systems with Applications**, v. 36, n. 2, p. 2027–2036, March 2009.

OLIVEIRA, H. M. d. **Análise de Sinais para Engenheiros: Uma abordagem via Wavelets**. 1. ed. Rio de Janeiro: Brasport, 2007.

PACOLA, E. R.; QUANDT, V. I.; ARRUDA, L. V. R.; SCHNEIDER, F. K.; SOVIERZOSKI, M. A. Reconhecimento de espículas em sinais de EEG utilizando wavelets e LDA. In: **XXIII Congresso Brasileiro em Engenharia Biomédica**. Porto de Galinhas: ANAIS - CBEB2012, 2012a. v. 1, p. 1899–1903.

PACOLA, E. R.; QUANDT, V. I.; SCHNEIDER, F. K.; SOVIERZOSKI, M. A. The wavelet transform border effect in EEG spike signals. In: **World Congress on Medical Physics and Biomedical Engineering**. Beijing, China: Springer Berlin Heidelberg, 2012b. v. 39, p. 593–596.

PINCUS, S. M. Approximate entropy as a measure of system complexity. **Proceedings of the National Academy of Sciences of the United States of America**, v. 88, n. 6, p. 2297–2301, March 1990.

POLIMENO, A. Epilepsia – referência mundial. **Pesquisa Médica**, v. 5, p. 26–33, 2008.

RAMANAN, S. V.; KALPAKAM, N. V.; SAHAMBI, J. S. A novel wavelet based technique for detection and de-noising of ocular artifact in normal and epileptic electroencephalogram. In: **International Conference on Communications, Circuits and Systems**. Chengdu, China: IEEE, 2004. v. 2, p. 1027–1031.

SADATI, N.; MOHSENI, H.; MAGHSOUDI, A. Epileptic seizure detection using neural fuzzy networks. In: IEEE (Ed.). **International Conference on Fuzzy Systems**. Vancouver, Canada: IEEE, 2006. p. 596–600.

SHNEKER, B.; FOUNTAIN, N. Epilepsy. **Disease-a-Month**, v. 49, n. 7, p. 426–478, July 2003.

SHUREN, Q.; ZHONG, J. Extraction of feature information in EEG signal by virtual EEG instrument with the functions of time-frequency analysis. In: **Proceedings of 6th International Image and Signal Processing and Analysis**. Salzburg: IEEE, 2009. v. 25, n. 4, p. 7–11.

SONG, Y.; LIO, P. Epileptic EEG detection via a novel pattern recognition framework. In: **5th International Conference on Bioinformatics and Biomedical Engineering**. Wuhan: IEEE, 2011. p. 1–6.

SOVIERZOSKI, M. A. **Avaliação de descritores morfológicos na identificação de eventos epileptiformes**. Tese (Doutorado) — Universidade Federal de Santa Catarina - UFSC, Abril 2009.

SU, H.; LIU, Q.; LI, J. Boundary effects reduction in wavelet transform for time-frequency analysis. **WSEAS Transactions on Signal Processing**, v. 8, n. 4, p. 169–179, October 2012.

SUBASI, A.; GURSOY, M. I. EEG signal classification using PCA, ICA, LDA and support vector machines. **Expert Systems with Applications**, n. 37, p. 8659–8666, 2010.

TAMBOURGI, P. **Ministério da Saúde redobra atenção para Epilepsia em 2013**. Dezembro 2012. Website. Disponível em: <<http://www.cinapce.org.br/index.php/br/noticia/80-slides/409-ministerio-da-saude-redobra-atencao-para-epilepsia-em-2013>>.

TEIXEIRA, L. M. **O Uso da Transformada Wavelet Contínua na Detecção do Complexo Ponta-e-Onda - Estudo de Desempenho**. Dissertação (Mestrado) — UTFPR, 2001.



- UBEYLI, E. D. Combined neural network model employing wavelet coefficients for EEG signals classification. **Digital Signal Processing**, v. 19, n. 2, p. 297–308, March 2009.
- UNSER, M. Wavelets, statistics, and biomedical applications. In: **8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing**. Corfu: IEEE, 1996. p. 244–249.
- VAVADI, H.; AYATOLLAHI, A.; MIRZAEI, A. A wavelet-approximate entropy method for epileptic activity detection from EEG and its sub-bands. **Biomedical Science and Engineering**, v. 3, n. 12, p. 1182–1189, December 2010.
- VENERI, G.; FEDERIGHI, P.; ROSINI, F.; FEDERICO, A.; RUFA, A. Spike removal through multiscale wavelet and entropy analysis of ocular motor noise: a case study in patients with cerebellar disease. **Journal of Neuroscience Methods**, v. 196, n. 2, p. 318–326, March 2011.
- WANG, C.; ZOU, J.; ZHANG, J.; ZHANG, Z.; ZHANG, C. Classifying detection of epileptic EEG based on approximate entropy in wavelet domain. In: **2nd International Conference on Biomedical Engineering and Informatics**. Tianjin: IEEE, 2009. p. 1–5.
- WASAY, M.; GRISOLD, W. **Epilepsy: Theme for World Brain Day 2015**. June 2015. Disponível em: <<http://www.worldneurologyonline.com/article/epilepsy-theme-for-world-brain-day-2015/>>.
- WATERS, C. G.; DEAN, B. C.; HAL. **EEGnet: A Web platform for collaborative EEG research**. 2013. Graduate Research and Discovery Symposium (GRADS). Disponível em: <<http://eegnet.clemson.edu>>.
- WILSON, S. B.; EMERSON, R. Spike detection: a review and comparison of algorithms. **Clinical Neurophysiology**, v. 113, n. 12, p. 1873–1881, December 2002.
- XU, B.; SONG, A. Pattern recognition of motor imagery EEG using wavelet transform. **Journal of Biomedical Science and Engineering Research Publishing**, v. 1, n. 1, p. 64–67, May 2008.
- YAMAGUCHI, C. Fourier and wavelet analyses of normal and epileptic electroencephalogram (EEG). In: **Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering**. Capri Island: IEEE, 2003. p. 406–409.
- YONG, L.; SHENGXUN, Z. Apply wavelet transform to analyse EEG signal. In: **18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society**. Amsterdam: IEEE, 1996. v. 3, p. 1007–1008.
- YU, L. EEG de-noising based on wavelet transformation. In: **3rd International Conference on Bioinformatics and Biomedical Engineering ICBBE**. Beijing, China: IEEE, 2009. p. 1–4.
- ZANDI, A. S.; JAVIDAN, M.; DUMONT, G. A.; TAFRESHI, R. Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform. **IEEE Transactions on Biomedical Engineering**, v. 57, n. 7, p. 1631–1651, July 2010.