

ANTÔNIO FERNANDO MOURA SANTOS

**OTIMIZAÇÃO MULTIOBJETIVO APLICADA NA IDENTIFICAÇÃO
DE PARÂMETROS PARA ANÁLISE ELETROENCEFALOGRÁFICA**



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA MECÂNICA
2019

ANTÔNIO FERNANDO MORIA SANTOS

**OTIMIZAÇÃO MULTIOBJETIVO APLICADA NA IDENTIFICAÇÃO
DE PARÂMETROS PARA ANÁLISE ELETROENCEFALOGRÁFICA**

Tese apresentada ao Programa de Pós-Graduação de Engenharia Mecânica da Faculdade de Engenharia Mecânica da Universidade Federal de Uberlândia para a obtenção do título de **Doutor em Engenharia Mecânica**.

Área de Concentração: Mecânica dos Sólidos e Vibrações

Orientador: Prof. Dr. Marcus Antonio Viana Duarte

UBERLÂNDIA

2019

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

- S237o
2019 Santos, Antônio Fernando Moura, 1952-
Otimização multiobjetivo aplicada na identificação de parâmetros para análise eletroencefalográfica [recurso eletrônico / Antônio Fernando Moura Santos. - 2019.
- Orientador: Marcus Antonio Viana Duarte.
Tese (Doutorado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Engenharia Mecânica.
Modo de acesso: Internet.
Disponível em: <http://dx.doi.org/10.14393/ufu.te.2019.324>
Inclui bibliografia.
Inclui ilustrações.
1. Engenharia mecânica. 2. Eletroencefalografia. 3. Entropia. I. Duarte, Marcus Antonio Viana, 1959- (Orient.) II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Engenharia Mecânica. III. Título.

CDU:621

Maria Salete de Freitas Pinheiro - CRB6/1262



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Engenharia Mecânica
Av. João Naves de Ávila, nº 2121, Bloco 1M, Sala 212 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4282 - www.posgrad.mecanica.ufu.br - secposmec@mecanica.ufu.br



ATA

Ata da defesa de Tese junto ao Programa de Pós-graduação em Engenharia Mecânica da Faculdade de Engenharia Mecânica da Universidade Federal de Uberlândia.

Defesa de Tese, nº 262, COPEM

Data: 22 de fevereiro de 2019

Hora início: 14:00

Hora encerramento: 17:00

Discente: Antônio Fernando Moura Santos

Título do Trabalho: Otimização Multiobjetivo Aplicada na Identificação de Parâmetros para Análise Eletroencefalográfica

Área de concentração: Mecânica dos Sólidos e Vibrações.

Linha de pesquisa: Engenharia Biomecânica

Reuniu-se na sala 313, bloco 1M - Campus Santa Mônica da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Mecânica, assim composta: Prof. Dr. Cleudmar Amaral Araújo - FEMEC/UFU; Prof. Dr. Francisco Paulo Léopore Neto - FEMEC/UFU; Prof. Dr. Israel Jorge Cárdenas Nuñez - UFTM, Jorge Nei Brito - UFSJ e Prof. Dr. Marcus Antonio Viana Duarte, orientador do candidato.

Iniciando os trabalhos o(a) presidente da mesa Dr. Marcus Antonio Viana Duarte apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu os conceitos finais.

Em face do resultado obtido, a Banca Examinadora considerou o candidato:

() Aprovado

(X) Aprovado com modificações a serem submetidas para a aprovação do orientador.

() Aprovado com modificações a serem submetidas para a aprovação da banca.

() Reprovado

Esta defesa de Tese de Doutorado é parte dos requisitos necessários à obtenção do título de Doutor. O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Assinaturas dos membros da banca via SEI



Documento assinado eletronicamente por **Marcus Antonio Viana Duarte, Professor(a) do Magistério Superior**, em 22/02/2019, às 17:01, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **JORGE NEI BRITO, Usuário Externo**, em 22/02/2019, às 17:05, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Cleudmar Amaral de Araujo, Professor(a) do Magistério Superior**, em 22/02/2019, às 17:07, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Francisco Paulo Lepore Neto, Professor(a) do Magistério Superior**, em 22/02/2019, às 17:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Israel Jorge Cardenas Nunez, Usuário Externo**, em 22/02/2019, às 17:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1034323** e o código CRC **A4695AD9**.

Aos meus pais, José Santos e Maria das Dores Moura Santos; que partiram, gratidão e saudades.

À Mara Sueli (meu primeiro amor).

Aos meus filhos:

Lívia Mara (minha princesa), Fernando Antonio, Larissa, Isabel e Samuel, Gustavo e Viviane.

Aos meus irmãos Lúcia Maria e José Ilídio.

AGRADECIMENTOS

A Deus que equilibra o Universo e concede coisas boas para todos nós.

A toda minha grande família e aos amigos (as), tudo que tenho, sempre crescendo no domínio do tempo e da frequência, sempre festeiros nas alegrias e apoiando fortemente nas tristezas.

Ao orientador Prof. D. Marcus Antonio Viana Duarte pela orientação deste trabalho.

Ao Prof. Elias Bitencourt Teodoro, Ph.D pela orientação e incentivos.

A Profa. Júlia Bertelli Duarte pelos incentivos.

Ao Programa de Pós-Graduação da Faculdade de Engenharia Mecânica, aos professores e técnicos da Faculdade de Engenharia Mecânica e à Universidade Federal de Uberlândia pelo apoio recebido.

Santos, A. F. M. Otimização Multiobjetivo Aplicada na Identificação de Parâmetros para Análise Eletroencefalográfica [tese]. Uberlândia: Faculdade de Engenharia Mecânica, Universidade Federal de Uberlândia, 2018.

RESUMO

A medicina atualmente gasta um longo tempo na análise de exames que em sua maioria não apresentam alguma anomalia. Para o caso do neurologista, mais de 80% dos exames analisados são considerados bons. Outro problema para o neurologista é que o erro diagnóstico é muito comum. Em função dos tempos despendidos na análise de eletroencefalogramas EEG normais, a percentagem elevada de erros de análise e, principalmente, a consequência que um diagnóstico falso positivo pode afetar um ser humano, toda pesquisa visando a minimização de erros de diagnóstico deve ser bem-vinda e apoiada, principalmente aquelas envolvendo Inteligência Artificial. Neste trabalho, para auxiliar o especialista, é apresentada uma metodologia, baseada em inteligência artificial, capaz de separar os sinais cerebrais em sinais com e sem anomalias. Para isto, utilizou-se de um banco de dados disponível publicamente de Bern-Barcelona, que consiste em sinais com e sem a presença de ictal. Para avaliação dos procedimentos desenvolvidos, utilizou-se das métricas de sensibilidade, especificidade, precisão, valores positivos preditos, valores negativos preditos e coeficiente de correlação de Mathew. Utilizou-se dois procedimentos de classificação (QDA e SVM) e quatro algoritmos de otimização multiobjetivo (MOPSO, PESA2, SPEA2 e NSGA-II) para análise discriminante de 1180 parâmetros candidatos a sintomas para classificar os sinais de EEG em sinais focais e não focais. Já para o cálculo dos sintomas, utilizou-se de técnicas de análise de sinais, como a análise estatística clássica (valor RMS e curtose, por exemplo), Transformadas Wavelet e de Hilbert-Huang, análise envelope e cálculos de entropia e k-NEO. A utilização do NSGA-II e funções objetivo com penalização, propostas neste trabalho, resultaram em valores de sensibilidade iguais a 94%, especificidade iguais a 82% e uma acurácia de 87% para os bancos de dados de validação e de teste.

Palavras-chave: Eletroencefalografia. Análise Discriminante Múltipla, Máquinas de Vetores de Suporte, Otimização Multiobjetivo. Entropia.

Santos, A. F. M. Multiobjective Optimization Applied in the Identification of Parameters for Electroencephalographic Analysis [thesis].. Uberlândia: Faculdade de Engenharia Mecânica, Universidade Federal de Uberlândia, 2018.

ABSTRACT

In medicine experts currently spends a long time in the analysis of tests that do not present any anomalies. For the case of the neurologist, more than 80% of the exams are good. Another problem is that the diagnostic error is very common in the neurology. Due to the time spent in analysis of normal EEG, the high percentage of errors and the consequence that to false positive diagnosis can affect to human health and life quality, all research aimed at minimizing errors in EEG's diagnosis should be welcomed, especially those involving Artificial Intelligence. In this work, to aid the EEG expert, is presented a methodology based on artificial intelligence capable of separating EEG signals into signs with and without anomalies with a high probability of success. For this was used a public available database of Bern-Barcelona University consisting of signs with ictal presence and without the presence of ictal. For the evaluation of the developed procedures was used the metrics: sensitivity, specificity, accuracy, predicted positive values, predicted negative values and Mathew correlation coefficient. Two classification procedures (QDA and SVM) and four multi-objective optimization algorithms (MOPSO, PESA2, SPEA2 and NSGA-II) were used for discriminant analysis of 1180 symptom-seeking parameters to classify EEG signals into focal and non-focal ones. Signal analysis techniques, such as classical statistical analysis (RMS value and kurtosis, for example), Wavelet and Hilbert-Huang transforms, envelope analysis and entropy and k-NEO calculations were used to choice the best parameters to classification purpose. The use of the NSGA-II and objective functions with penalization, proposed in this work, resulted in sensitivity values of 94%, specificity of 82% and an accuracy of 87% for the validation and test databases.

Keywords: Electroencephalography. Multiple Discriminant Analysis, Supporting Vector Machines, Multi-objective Optimization. Entropy.

LISTA DE FIGURAS

Fig. 2.1 - “Tempestade Elétrica”	9
Fig. 2.2 - Síntese das atividades elétricas do cérebro	10
Fig. 2.3 - Exemplo de Eletroencefalograma.....	11
Fig. 2.4 - Mapeamento de posições [10-20]	13
Fig. 2.5 - Montagem dos eletrodos	14
Fig 2.6 - Touca para 128 eletrodos	14
Fig. 2.7 - Ondas cerebrais e frequências	15
Fig 2.8 Ilustração dos conceitos de dominância de Pareto em um problema de minimização com dois objetivos.....	20
Fig. 2.9. Gráficos das funções $f(x)$ e $g(x)$ do exemplo.	21
Fig. 2.10. Fronteira de Pareto do exemplo.	21
Fig. 2.11. Exemplo ilustrativo das etapas do NSGA II com duas funções objetivos	24
Fig. 2.12. Ilustração de probabilidades a priori.	28
Fig. 2.13. Exemplo de elevação para uma dimensão linearmente separável: $\mathbf{R}^1 \rightarrow \mathbf{R}^2$	31
Fig. 2.14. Exemplo de maximização de margens	31
Fig. 2.15. Exemplo de uma representação simples de um sinal modulado em fase nos domínios do tempo e da frequência.	33
Fig. 2.16. Fase instantânea do sinal usado como exemplo de modulação de fase.	34
Fig. 2.17. Exemplo de dualidade Tempo-Frequência.	36
Fig. 2.18. Efeito das escalas na Wavelet: escalas menores resultam em janelas temporais maiores e escalas maiores resultam em janelas temporais menores.	37
Fig. 2.19. Função wavelet base db4	38
Fig. 2.20. Diagrama para extração das características e pontos definidos pelo modelo	41
Fig. 3.1. Fluxograma da Metodologia Proposta	55
Fig. 4.1. Diagrama de caixa para a curtose dos sinais filtrados na banda otimizada.	61
Fig. 4.2. Valores de sensibilidade em função do tipo de normalização para o otimizador NSGA-II.	61
Fig. 4.3. Valores de sensibilidade em função do tipo de normalização para o otimizador SPEA2 ...	62
Fig. 4.4. Valores de sensibilidade em função do tipo de normalização para o otimizador PESA2. ..	62
Fig. 4.5. Valores de sensibilidade em função do tipo de normalização para o otimizador MOPSO. ..	63

Fig. 4.6. Tempos relativos para convergência dos otimizadores PESA2, SPEA2 e NSGA-II com relação ao otimizador MOPSO	64
Fig. 4.7. Exemplo de cálculo das distâncias das aglomerações.	65
Fig. 4.8. Exemplo de gráficos das soluções não dominadas	67
Fig. 4.9. Gráficos SEN vs. NCD para os conjuntos de validação (azul) e teste (vermelho). Otimizador NSGA-II	68
Fig. 4.10. Gráficos SEN vs. NCD para os conjuntos de validação (azul) e teste (vermelho). Otimizador SPEA2	69
Fig. 4.11. Gráficos SEN vs. NCD para os conjuntos de validação (azul) e teste (vermelho). Otimizador PESA2.....	69
Fig. 4.12. Gráficos SEN vs. NPAR para os conjuntos de validação (azul) e teste (vermelho). Otimizador NSGA-II	70
Fig. 4.13. Gráficos SEN vs. NPAR para os conjuntos de validação (azul) e teste (vermelho). Otimizador SPEA2	71
Fig. 4.14. Gráficos SEN vs. NVALID para os conjuntos de validação (azul) e teste (vermelho), utilizando o otimizador NSGA-II com normalização 3S6D	72
Fig. 4.15. Gráficos SEN vs. NVALID para os conjuntos de validação (azul) e teste (vermelho), utilizando o otimizador NSGA-II sem normalização	72
Fig. 4.16. Maiores valores de SEN para as cinco funções kernel estudadas para o Otimizador SPEA2	74
Fig. 4.17. Maiores valores de SEN para a otimização multiobjetivo envolvendo SEN, SPF e ACC com o procedimento de otimização NSGA-II	76
Fig. 4.18. Frente de Pareto para $VAL_{pen} = 0,3$ para o otimizador NSGA-II	77
Fig. 4.19. Valores de SPF para os maiores valores de SEN para a otimização multiobjetivo com penalidade $VAL_{pen} = 0,3$ para o procedimento de otimização NSGA-II	77
Fig. 4.20. Maiores valores de SEN para a otimização multiobjetivo com penalidade $VAL_{pen} = 0,3$ para o procedimento de otimização NSGA-II	78
Fig. 4.21. Valores de ACC para os maiores valores de SEN para a otimização multiobjetivo com penalidade $VAL_{pen} = 0,3$ para o procedimento de otimização NSGA-II	79
Fig. 4.22. Gráficos SEN vs. NPAR para os conjuntos de validação (azul) e teste (vermelho) para o otimizador NSGA-II	80

Fig. 4.23. Gráficos ACC vs. NPAR para os conjuntos de validação (azul) e teste (vermelho) para o otimizador NSGA-II	80
Fig. 4.24. Gráfico dos tempos de processamento para todos os sintomas calculados	82
Fig. 4.25. Gráfico dos tempos de processamento para os sintomas calculados, excluindo os relacionados a EMD.	82

LISTA DE TABELAS

Tab. 2.1. Ondas cerebrais e frequências	15
Tab. 2.1. Comparação entre as bandas encontradas em EEG	16
Tab. 4.1. Valores de DMA e NSND, e respectivos desvios padrão amostrais, para o quatro otimizadores estudados	66
Tab. 4.2. Melhores resultados de sensibilidade SEN limitando o SPF a 60%	83
Tab. 4.3. Melhores resultados de sensibilidade SEN limitando o SPF a 70%	85

LISTA DE SÍMBOLOS

Letras Arábicas

A	Amplitude
A_t	População pertencente ao arquivo do otimizador SPEA 2
FC	Fator de Crista
H	Transformada de Hilbert
$H $	Entropia
J	Número de restrições de igualdade.
K	Número de restrições de desigualdade
Kernel	Núcleo
M	Número de objetivos.
N	Número de variáveis de decisão.
N_a	Tamanho do conjunto de dados com os parâmetros para treinamento
N_t	Tamanho do conjunto de dados com os parâmetros para validação
N_{arq}	Tamanho do Arquivo do otimizador SPEA 2
N_{pop}	Tamanho da população
N_{pto}	Número de pontos do sinal discretizado
Obj	Função Objetivo
P_t	População de trabalho do otimizador SPEA 2
$R(j)$	A j-ésima função de aptidão
S	Matriz de covariância amostral
$S(i)$	A i-ésima ponderação da solução na frente de Pareto
S_k	Assimetria
T	Número máximo de gerações
V_p	Valor de pico
\bar{X}	Média amostral
X_{focal}	Conjuntos de dados com sinais focal
X_{nfocal}	Conjunto de dados com sinais não focal
W	Transformada de Wavelet
a	Fator de dilatação
alfa	Ondas cerebrais na faixa de frequência de 8 a 13 Hz.
b	Translação da origem
beta	Ondas cerebrais na faixa de frequência de 13 a 30 Hz.
db4	Wavelet base Daubechies 4
delta	Ondas cerebrais na faixa de frequência de 0,5 a 4 Hz.
$f_m(x)$	m-ésima função objetivo.
gama	Ondas cerebrais na faixa de frequência acima de 30 Hz.
$g_j(x)$	A j-ésima restrição de igualdade
$h_k(x)$	A k-ésima restrição de desigualdade
k4	Produto do valor eficaz e a curtose
k6	Momento central de sexta ordem
$p(x)$	Densidade de probabilidade
teta	Ondas cerebrais na faixa de frequência de 4 a 8 Hz.

x	Variáveis de decisão do problema (parâmetros).
x	Um dos canais do banco de dados do Bern-Barcelona
x_{\inf}	Valores mínimos das variáveis de decisão.
x_{\sup}	Valores máximos das variáveis de decisão.
y	Um dos canais do banco de dados do Bern-Barcelona
w	Vetor de pesos do SVM

Letras Gregas

Σ	Matriz de covariância
Θ	Ângulo de fase da portadora
Ω	Frequência angular da portadora
β	Amplitude da moduladora
$\lambda(\mathbf{X})$	Função Discriminante
μ	Média populacional
φ	Ângulo de fase da moduladora
ψ	Função de Modos Intrínsecos
ψ_{RBF}	Núcleo função radial base
ψ_s	Núcleo função sigmoideal
ψ_{Pd}	Núcleo função polinomial
ω	Frequência angular moduladora

Lista de Abreviaturas

1S0D	Normalização por um desvio padrão amostral
1S6D	Normalização por um desvio padrão amostral limitado a 6
2S0D	Normalização por dois desvios padrões amostral
2S6D	Normalização por dois desvios padrões amostral limitado a 6
3S0D	Normalização por três desvios padrões amostral
3S6D	Normalização por três desvios padrões amostral limitado a 6
6S0D	Normalização por seis desvios padrões amostral
6S6D	Normalização por seis desvios padrões amostral limitado a 6
ACC	Acurácia
AD	Análise Discriminante Simples
ApEn	Entropia aproximada média
AR	Filtros Alto Regressivos
AVC	Acidente Vascular Cerebral
DMA	Distância Média das Aglomerações
DWT	Transformada Wavelet Discreta
ECM	Custo Médio de Classificação Incorreta
EEG	Eletroencefalograma
EMD	Decomposição de Modos Empíricos
ENF	Eletroencefalograma Não Focal
GMM	Modelo de Mistura de Gaussianas
HHT	Transformada de Hilbert-Huang

IA	Inteligência Artificial
IMF	Função de Modos Intrínsecos
k-NEO	k-point Nonlinear Energy Operator
KNN	k-Nearest Neighbor
LDA	Análise Discriminante Linear
MCC	Coeficiente de Correlação de Mathew
MDA	Análise Discriminante Múltipla
MOPSO	Multi-objective Particle Swarm Optimization II
MADM	Multiple Attribute Decision Making
MLP-BP	Rede Perceptron multicamadas com retropropagação do erro
NCD	Número de sinais dos conjuntos de dados de treinamento
NEO	Nonlinear Energy Operator
NPAR	Número de parâmetros utilizados no classificador
NPV	Valores negativos preditos
NSGA-II	Non-dominated Sorting Genetic Algorithm II
NSDN	Número Médio de Soluções não Dominadas
NVALID	Número de sinais dos conjuntos de dados de validação
POM	Problema de Otimização Multiobjetivo
QDA	Análise Discriminante Quadrática
PESA II	Pareto Envelope-based Selection Algorithm II
PPV	Valores Positivos Preditos
PSO	Otimização por Enxame de Partículas
QDA	Análise Discriminante Quadrática
RenEn	Entropia espectral média de Renyi
ROC	Curvas Característica de Operação do Receptor
RMS	Valor Médio Quadrático
ShEn	Entropia média de Shannon
SPEA 2	Strength Pareto Evolutionary Algorithm 2
SpEn	Antropia média da amostra
SVM	Máquina de Vetores de Suporte
PNN	Rede Neural Propabilística
SEN	Sensibilidade
SPF	Especificidade
TPM	Erro de Classificação Total
TAE	Teoria do Aprendizado Estatístico
TR	Tempo de processamento relativo ao tempo médio do MOPSO
VAL _{pen}	Valor do coeficiente de penalização das funções objetivo
WkNN	kNN ponderado pela distância
WT	Transformada de Wavelet

SUMÁRIO

1. INTRODUÇÃO.....	1
1.1. OBJETIVO.....	7
1.2. ORGANIZAÇÃO DA TESE.....	7
2. FUNDAMENTAÇÃO TEÓRICA.....	9
2.1. Funcionamento do cérebro humano.....	9
2.2. As atividades elétricas do cérebro.....	10
2.3. Conceito de eletroencefalograma (EEG).....	11
2.4. Aplicação do EEG.....	12
2.5. Doenças diagnosticadas pelo EEG.....	12
2.6. Mapeamento de posições e montagem dos eletrodos.....	12
2.7. Tipos de ondas cerebrais e frequências.....	14
2.8. OTIMIZAÇÃO MULTIOBJETIVO.....	17
2.8.1 Definição do Problema de Otimização Multiobjetivo.....	17
2.8.2 Algoritmos Genéticos Multiobjetivos e outras Meta-heurísticas.....	22
2.8.3. Non-dominated Sorting Genetic Algorithm II (NSGA-II).....	23
2.8.4. Strength Pareto Evolutionary Algorithm 2 (SPEA 2).....	25
2.8.5. Pareto Envelope-based Selection Algorithm II (PESA II).....	26
2.8.6. Multi-objective Particle Swarm Optimization II (MOPSO).....	26
2.9 ANÁLISE DISCRIMINANTE.....	27
2.9.1. Qadratic Discriminante Analysis (QDA).....	27
2.10 MÁQUINA DE VETORES DE SUPORTE (SUPPORT VECTOR MACHINE).....	30
2.11 FERRAMENTAS DE ANÁLISE DE SINAIS APLICADOS EM CLASSIFICAÇÃO DE EEG.....	32
2.11.1. Análise de Envelope.....	32

2.11.2. Decomposição em Modos Empíricos (EMD).....	34
2.11.3. Tranformada de Hilbert-Huang (HHT).....	35
2.11.4. Tranformada Wavelet.....	36
2.11.5. Métricas Clássicas: Valor RMS, Valor Pico, Valor Pico a Pico, Assimetria, Curtose, Fator de Crista, k4 e k6.....	38
2.11.6. Métrica k-point nonlinear energy operator (k-NEO).....	40
2.11.7. Métricas de Entropia.....	43
3. METODOLOGIA.....	46
3.1. ESCOLHA DO BANCO DE DADOS.....	46
3.2. AVALIAÇÃO DOS PROCEDIMENTOS.....	47
3.3. PARÂMETROS UTILIZADOS.....	48
3.4. PROCEDIMENTOS DE CLASSIFICAÇÃO UTILIZADOS.....	53
3.5. ANÁLISES DISCRIMINANTES.....	54
3.6. CUSTO COMPUTACIONAL.....	54
3.7. FLUXOGRAMA DA METODOLOGIA UTILIZADA.....	54
3.8. ROTINAS DE PROGRAMAÇÃO E PARÂMETROS UTILIZADOS.....	56
4. APRESENTAÇÃO E ANÁLISE DOS RESULTADOS OBTIDOS.....	58
4.1. PERFORMANCE DOS OTIMIZADORES.....	58
4.1. INFLUÊNCIA DA NORMALIZAÇÃO DOS DADOS PARA O SVM.....	59
4.2. TEMPO PARA CONVERGÊNCIA DO OTIMIZADOR.....	63
4.3. FRENTE de PARETO.....	64
4.4. QUANTIDADE DE DADOS NO CONJUNTO DE TREINAMENTO.....	67
4.5. NÚMERO DE PARÂMETROS.....	70
4.7. TAMANHO DO CONJUNTO DE DADOS DE VALIDAÇÃO.....	71
4.8. NÚCLEO (KERNEL).....	73
4.9. QDA.....	74

4.10. NÚMERO MAIOR DE FUNÇÕES OBJETIVO.....	75
4.11. PENALIZAÇÃO NAS FUNÇÕES OBJETIVO.....	76
4.12. FUNÇÃO OBJETIVO BASEADA NA ACURÁCIA.....	79
4.13. MELHOR CONJUNTO DE PARÂMETROS.....	81
5. CONCLUSÕES.....	87
6.1. TRABALHOS FUTUROS.....	89
REFERÊNCIAS BIBLIOGRÁFICAS.....	91

CAPÍTULO I

1. INTRODUÇÃO

O organismo humano é considerado a mais “perfeita” máquina do universo, sendo composto por cerca de 100 trilhões de células que nascem e morrem diariamente. A manutenção dessa máquina é feita pelos sistemas nervoso, circulatório e linfático (Oda, 2013). Da mesma maneira que acontece com as máquinas, quando algo não está ocorrendo bem, o corpo humano emite sinais diferentes dos normalmente gerados por um organismo são.

Com relação aos sinais associados aos sintomas neurológicos, o renomado cientista e filósofo Galvani foi a primeira pessoa a reportar a existência de atividade elétrica em organismos vivos no século 18. Em 1924 o eletrofisiologista Hans Berger conseguiu gravar com sucesso a atividades elétricas do cérebro humano usando o processo atualmente conhecido como eletroencefalografia (EEG), o qual mede as oscilações de voltagem devido ao fluxo de íons nos neurônios. Atualmente, EEG é uma das técnicas não invasivas mais populares para registrar a atividade cerebral em ambientes clínicos e de pesquisa.

Atualmente, já se reconhece cinco tipos de ondas cerebrais em faixas de frequência diferentes: delta (0,5 a 4 Hz), teta (4 a 8 Hz), alfa (8 a 13 Hz), beta (13 a 30 Hz) e gama (acima de 30 Hz). As ondas do tipo delta normalmente ocorrem quando se está em sono profundo, com uma variação de 75 a 200 mV, já a teta raramente é presente em seres humanos. As ondas alfa são comuns quando a pessoa está em vigília, aparecendo as ondas beta com aumento de vigília e atenção. A onda gama aparece com movimentos voluntários do organismo. (Blinowska e Durka, 2006).

A medicina atualmente gasta um longo tempo na análise de exames que em sua maioria não apresentam alguma anomalia. Para o caso do neurologista, mais de 80% dos exames analisados são considerados bons. Num estudo utilizando 5880 EEG padrões, e com privação de sono, Brogger, Eichele, Aanestad, Olberg, Hjelland e Aurlen H (2018) reportam que os especialistas levam em média 8,3 e 16,1 minutos para analisar EEG normais padrões e como privação de sono, respectivamente. Caso seja encontrado alguma anomalia estes tempos aumentam consideravelmente, por exemplo para um ataque epilético tem-se 38,6 e 41 minutos para EEG

padrões e com privação do sono. Entre o exame e a entrega dos resultados os especialistas levam em média 24 horas em função de reanálises e redação do laudo.

De acordo com Wilner (2018) os erros diagnósticos são uma grande preocupação tanto para os pacientes quanto para os médicos e o erro diagnóstico é muito comum na epilepsia. Em um estudo retrospectivo realizado com pacientes encaminhados a um centro de tratamento da epilepsia, quase 25% tinham recebido um diagnóstico equivocado (Benbadis e Tatum, 2010).

Atualmente 47 em 10000 pessoas/ano são diagnosticadas como tendo epilepsia. Um estudo com 5000 pessoas com epilepsia em 15 países europeus relatou que 96% delas receberam prescrição de drogas antiepilépticas. Destas pessoas, 88% relataram pelo menos um efeito colateral e 31% mudaram de medicamento pelo menos uma vez no último ano por causa dos efeitos colaterais. Além disso, um diagnóstico de epilepsia pode impactar em muitos aspectos da vida das pessoas tais como: afetar a capacidade de obter ou manter o emprego e perder a licença para dirigir, entre outros efeitos negativos do ponto de vista da saúde física e mental (Xu, Nguyen, Mohamed, Carcel, Li, Kutlubayev, Anderson e Hackett, 2016). Numa revisão sistemática com o objetivo de determinar a frequência de diagnósticos falso positivos de epilepsia e explorar as suas consequências, Xu et al (2016) concluíram que a frequência do diagnóstico falso positivo varia de 2% a 71%.

Em função dos tempos dispendidos na análise de EEG normais, a porcentagem elevada de erros de análise e, principalmente, a consequência que um diagnóstico falso positivo pode afetar um ser humano, toda pesquisa visando a minimização de erros de diagnóstico deve ser bem-vinda e apoiada, principalmente aquelas envolvendo Inteligência Artificial (IA).

As formas que podem ser observadas no mundo à nossa volta representam apenas uma pequena parcela de todas as formas possíveis. Logo, dado um número finito de formas, o primeiro passo da inteligência é a clusterização destas formas, ou seja: realizar um agrupamento de dados (divisão em classes distintas) de tal modo que seja minimizada a distância intragrupos e seja maximizada a distância intergrupos. O próximo passo é buscar associar padrões de entrada e saída.

Haikin (2011) toma emprestada a definição de IA dada por Sage nos anos 1990, que diz: “o objetivo da IA é o desenvolvimento de paradigmas ou algoritmos que requeiram máquinas para realizar tarefas cognitivas, para as quais os humanos são altamente melhores”. Dito isso, vale ressaltar que um sistema de IA deve ser capaz de armazenar conhecimento, aplicar este conhecimento na resolução de problemas e adquirir novos conhecimentos através da experiência,

tendo, portanto, três componentes fundamentais: representação, raciocínio e aprendizagem (Sage, 1990 *apud* Haikin, 2011).

Atualmente, para aplicar IA para a solução de um problema de decisão, primeiramente deve-se procurar padrões (parâmetros) que permitam agrupamentos com distâncias intragrupo mínimas e intergrupos máximas.

Os sinais de EEG são caracterizados como não estacionários (Pachori e Sircar, 2008). Consequentemente parâmetros associados à transformada de Fourier não são adequados para agrupamento de sinais de EEG. Portanto, os parâmetros de análise de sinais normalmente utilizados como padrões em EEG são: transformadas de wavelet, envelope, transformada de Hilbert-Huang, modos intrínsecos, métricas de entropia, forma dos picos e séries não lineares, entre outros.

No caso de aprendizado supervisionado, o próximo passo consiste na escolha de um banco de dados confiável com sinais que representam os grupos que se deseja discriminar. Em EEG o banco de dados Bern-Barcelona (Andrzejak, Schindler e Rummel, 2012), de domínio público, é muito usado em pesquisas envolvendo detecção de anomalias em sinais de EEG. Uma pesquisa no Google Acadêmico (2018) com a frase ‘Bern-Barcelona EEG database’ resultou em 51 citações envolvendo pesquisas para classificação automática de EEG.

Para a escolha dos melhores parâmetros num procedimento de classificação se utiliza:

- Diagramas de caixas e bigodes, também conhecidos como gráficos boxplot que é uma ferramenta que condensa um conjunto de informações estatísticas a respeito da distribuição de um conjunto de dados, em apenas um gráfico. Analisando os gráficos boxplot de vários agrupamentos, o analista visualiza facilmente as distâncias intra e intergrupos (Montgomery, 2000).
- O teste t de Student: É um teste de hipóteses paramétrico que usa as médias e os desvios padrões amostrais para determinar a discriminação entre dois grupos (Montgomery, 2000). Quanto maior for o valor de t, maior a capacidade discriminativa daquele parâmetro.
- O teste de Wilcoxon-Mann-Whitney (Salkind, 2010): Teste não paramétrico aplicado para duas amostras independentes, que pode ser considerado uma versão não paramétrica do teste t para comparar duas populações. O teste usa a magnitude da estatística U, por isto também é chamado de teste U, para avaliar a diferença entre duas populações.
- O algoritmo de Bhattacharyya: A distância de Bhattacharyya é uma medida da separabilidade entre duas populações (Kailath, 1967).

- Otimização (Duarte, 2018): Dados um grande número de parâmetros candidatos à descriminalização, um algoritmo de classificação, um banco de dados de treinamento e um banco de dados de validação, procedimentos de otimização são utilizados para identificar o melhor conjunto de parâmetros para diferenciar as populações.

Quanto aos procedimentos de classificação têm-se:

- k-nearest e wk-nearest: Dado um banco dados com os parâmetros das classes de saída, uma amostra desconhecida é classificada usando a relação de proximidade de seus parâmetros com os parâmetros mais próximos das amostras conhecidas (Hasan, Ahamed A, Ahmad M. e Rashid, 2017). A proximidade da amostra k mais próxima é determinada usando uma métrica de distância ou similaridade (a métrica mais utilizada para o cálculo das distâncias é a Euclidiana). No caso do wk-nearest as amostras mais próximas têm um peso maior do que as amostras distantes. Finalmente, a amostra desconhecida é considerada pertencente à classe que é comum entre os k vizinhos mais próximos.
- Análise Discriminante Linear (LDA): A LDA utiliza da função discriminante de Fisher, que é uma combinação linear das características das populações, para produzir a máxima separação entre duas populações homocedásticas (Bhattacharyya S., Khasnobish A., Chatterjee S., Konar A. e Tibarewala, 2010).
- Análise Discriminante Quadrática (QDA): É a generalização do método linear, desde que existam apenas duas classes e que suas medidas sejam normalmente distribuídas, para populações heterocedásticas, ou seja: quando as matrizes de covariância dos dois grupos não são iguais (Duarte, 2018).
- Classificador Fuzzy: Em vez de considerar que a amostra pertence a uma determinada classe, a amostra é considerada um membro de cada classe com uma função de associação diferente. A classe a qual pertence uma dada amostra é determinada com base nas regras fuzzy se-então (Acharya, Sree, Chattopadhyay, Yu e Ang, 2011).
- O Classificador da Máquina de Vetores de Suporte (SVM) pode ser usado para classificação linear e não linear selecionando-se o kernel adequado (Li, Chen, Yan, Wei e Wang, 2014). Originalmente um classificador linear dicotômico, o SVM não linear executa o mapeamento dos dados num espaço dimensional maior, onde as classes podem ser separadas por um hiperplano.
- Índice de Discriminação Único (Integrated Discrimination Index): Utilizando os parâmetros mais significativos para a discriminação das classes, deriva-se uma expressão de modo que

para cada classe resulte um intervalo exclusivo de valores (Sharma R., Pachori, R. B. e Acharya, U. R, 2015). Normalmente a expressão é derivado empiricamente.

- Redes Neurais Recorrentes (Deep Learning): São um tipo especial de rede neural artificial que contém laços na arquitetura neural para permitir que a informação persista ao longo do tempo. Para detecção de convulsões em EEG, as unidades de Memória de Longo Prazo são muito eficazes dado que as convulsões tipicamente se estendem por várias janelas consecutivas de 1 segundo (Thodoroff P., Pineau, J. e Lim, A, 2016).

Deve-se salientar que se procurou citar referências na área de detecção automática de EEG.

Existe uma infinidade de trabalhos científicos com o objetivo de automatizar a análise de EEG e uma excelente revisão bibliográfica pode ser encontrada na tese de Duarte (2018, domínio público). Nesta introdução, para fins de contextualização, apresenta-se uma breve revisão de artigos que utilizaram o banco de dados Bern-Barcelona (Andrzejak, Schindler e Rummel, 2012) que será utilizado neste trabalho.

Chua, Chandran, Acharya e Lim (2008) utilizaram parâmetros extraídos do espectro de energia e do bi-espectro para alimentar um classificador de modelo de mistura de gaussianas (GMM) para identificar sinais focal e não focal. Os resultados mostram que, com os parâmetros baseados em bi-espectros conseguiram uma precisão de 93,11% em comparação com a precisão de 88,78% da classificação de parâmetros derivados de espectros de potência.

Sharma R, Pachori R. e Acharya (2015) utilizaram como parâmetros várias métricas de entropia aplicadas a seis níveis de transformada de wavelet discreta. Para a escolha dos melhores parâmetros testaram algoritmo espacial de Bhattacharyya, teste t, teste de Wilcoxon, característica de operação do receptor (curvas ROC) e entropia. Para a classificação automatizada de sinais EEG focais e não focais, usando um número mínimo de parâmetros, utilizaram KNN, rede neural probabilística PNN, fuzzy e SVM. Concluíram que os melhores discriminadores de populações foram o teste t de Student e as curvas ROC. O melhor classificador foi o SVM que resultou numa precisão de 84%.

Das e Bhuiyan (2016) propuseram uma metodologia baseada na decomposição em modo empírico (EMD) para classificação de sinais de eletroencefalograma focal e não focal (ENF). A metodologia proposta usa o EMD e a entropia de Sharma-Mittal. Os parâmetros são os valores de distância euclidiana de K-vizinhos mais próximos (KNN) de sinais focal e não focal. Utilizando SVM a metodologia proposta resultou numa precisão de classificação de 83,18% quando aplicada em todos os sinais do banco de dados Bern-Barcelona.

Chen, Wan e Bao FS (2017), utilizaram máquinas de vetores de suporte e sete famílias de wavelets discretas para detectar ocorrência de picos focal em EEG. As bandas de frequência ótimas e os parâmetros baseados nos coeficientes de wavelet foram obtidos via otimização. Utilizando SVM para classificação os autores conseguiram uma precisão de 83,07%.

Sharma, Dhere, Pachori e Acharya, desenvolveram classes de wavelets mães específicas para detecção de sinais focal e não focal. Testando várias métricas de entropia, como parâmetros discriminantes para a classificação dos sinais e SVM, conseguiram uma porcentagem de detecção de sinais focal (sensibilidade) de 94.25%.

Utilizando transformadas tempo-frequência com filtros alto regressivos (AR) para detectar anomalias no sinal e um teste estatístico não paramétrico baseado em Martingale (Randomized Power Martingale - RPM) para classificação, Gao, Lu, Yan, Lyu, Li, Xie e Zhang (2018) conseguiram um índice de acertos nos diagnósticos de sinais focal com uma sensibilidade de 97,66%. O Martingale foi uma estratégia de apostas muito popular no século XVIII usado nas apostas de jogos com moedas em que o apostador apostava em cara ou coroa, sempre no mesmo lado e dobrando a aposta caso perdesse. Acreditava-se que apostando sempre no mesmo lado, haveria de ganhar, mais cedo ou mais tarde.

Deivasigamani, Senthilpari C. e Yong W. H (2016) utilizaram transformadas de wavelet complexa dual-tree com os parâmetros calculados a partir dos coeficientes decompostos. Na classificação foi utilizado um classificador adaptativo neuro fuzzy com uma camada de entrada, três camadas ocultas e uma camada de saída. O número de neurônios das camadas ocultas não foram reportados. Segundo os autores, uma sensibilidade de 98% e uma precisão de 99% foram encontradas quando o procedimento foi testado com 50 sinais focais e 50 não focais selecionados aleatoriamente do conjunto de dados. Detalhes sobre o conjunto de dados utilizados no treinamento não está presente no artigo.

Duarte (2018), utilizou seis procedimentos de classificação (kNN, WkNN, LDA, QDA, PNN e MLP-BP) e dois algoritmos de otimização (algoritmo genético e evolução diferencial) para achar os melhores parâmetros a partir de 1180 métricas comumente utilizadas para classificação automática de EEG. O melhor resultado observado foi a combinação evolução diferencial, para escolha dos parâmetros, e classificador WkNN, que resultou numa precisão de aproximadamente 84%. Ao contrário de vários artigos em seu trabalho Duarte descreve claramente os conjuntos de dados utilizados no treinamento, validação e teste dos procedimentos desenvolvidos.

1.1. OBJETIVO

Esta tese, que foi desenvolvida em paralelo com a de Duarte (2018), tem por objetivo desenvolver um procedimento baseado no classificador de máquinas de vetores de suporte que minimize a probabilidade de diagnose de uma EEG com indicações de possíveis patologias como um EEG normal.

Para atingir este objetivo, procedimentos de otimizações multiobjetivas serão usados para escolher os parâmetros de análise de sinais, comumente aplicados em procedimentos automáticos de classificação de EEG com funções custo que minimizem a probabilidade de ocorrência de um falso negativo, ou seja: o procedimento considerar normal um EEG com problemas.

Com relação à tese de Duarte, esta tese tem em comum:

1. O mesmo banco de dados.
2. Os mesmos parâmetros candidatos à discriminação dos grupos focal e não focal.

Como diferenças:

1. Duarte uso como ferramentas para classificação kNN, WkNN, LDA, QDA, PNN e MLP-BP. Em função do grande número de trabalhos existentes na literatura esta tese teve como foco a utilização do SVM.
2. O objetivo de Duarte era maximizar os acertos. O objetivo desta tese é minimizar a ocorrência de falsos negativos, ou seja: minimizar erros.
3. Duarte utilizou otimização mono objetivo. Esta tese estuda aplicação de otimização multi-objetivo como ferramenta para a escolha do melhor conjunto de parâmetros para a classificação de sinais focal e não focal.

1.2. ORGANIZAÇÃO DA TESE

Este trabalho é dividido em cinco capítulos. No Capítulo I é apresentado o estado da arte, uma breve introdução teórica para situar o tema da tese e seus objetivos. No Capítulo II é apresentada uma breve revisão bibliográfica, bem como a fundamentação teórica utilizada para a realização

deste trabalho. O Capítulo III apresenta a metodologia utilizada. No Capítulo IV são apresentados e discutidos os resultados obtidos através das diferentes abordagens apresentadas. Por fim, no Capítulo V, são mostradas as principais conclusões com possíveis trabalhos futuros.

CAPÍTULO II

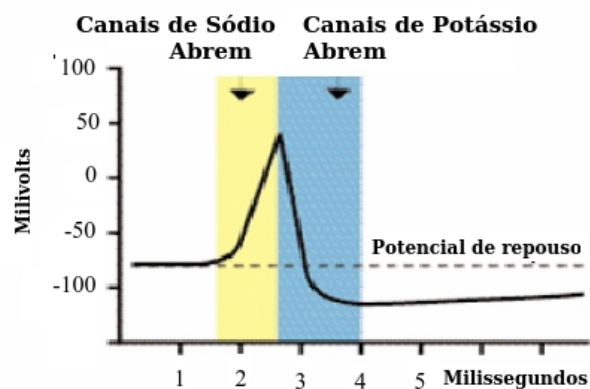
2. FUNDAMENTAÇÃO TEÓRICA

Apresenta-se neste capítulo uma breve revisão dos fundamentos teóricos necessários para a realização/entendimento da tese. Para efeitos de contextualização, inicia-se com funcionamento do cérebro humano, as atividades elétricas do cérebro, conceito de eletroencefalograma (EEG), aplicações, indicações e doenças diagnosticadas pelo EEG, montagem dos eletrodos e tipos de ondas cerebrais e frequências. A seguir apresenta-se os fundamentos e alguns procedimentos de otimização multiobjetivo, os procedimentos de classificação Análise Discriminante Quadrática e Máquinas de Vetores de Suporte e os parâmetros de análise de sinais usualmente usados como variáveis de projeto em procedimentos automatizados para classificação de EEG.

2.1. Funcionamento do cérebro humano

No cérebro humano os neurônios transmitem e recebem informações uns dos outros através de impulsos elétricos. Os impulsos são produzidos pela movimentação de átomos de sódio, potássio e cálcio para dentro e para fora da célula (Schomer e Silva, 2011). Esta “*tempestade elétrica*” pode ser registrada e analisada pelo eletroencefalograma [EEG]; como ilustra a Figura 2.1

Fig. 2.1 - “Tempestade Elétrica”



Fonte: Schomer e Silva, 2011.

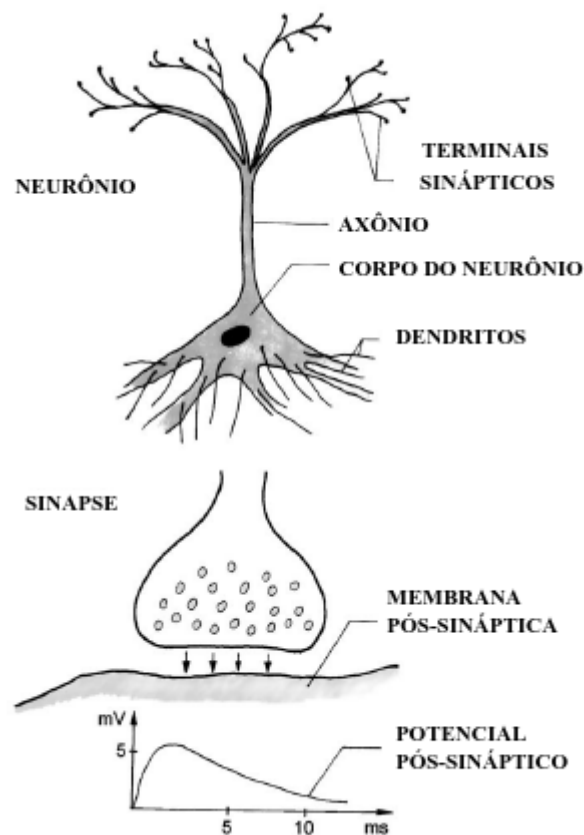
2.2. As atividades elétricas do cérebro

No cérebro existem duas classes principais de células: os neurônios e as células de glia. O corpo humano possui em média 10^{11} neurônios. As células da glia, também chamadas células da neuróglia, cobrem os axônios dos neurônios e agem como um isolante de células eletricamente condutoras. E são da ordem de 10^{12} no organismo humano. Tensão em repouso = -80 [mV].

A diferença de tensão vem da variação de concentração de cations: potássio⁺ (K^+), sódio⁺ (Na^+). Ânions: cloro⁻ (Cl^-) e íons de cálcio⁺⁺ (Ca^{++}) são menos abundantes, mas eles têm um importante papel regulador, como ilustra a Figura 2.2.

A diferença de potencial é mantida pelo transporte ativo de potássio para o interior da célula e sódio para o lado de fora, usando a energia fornecida por meio de processos metabólicos.

Fig. 2.2 – Síntese das atividades elétricas do cérebro



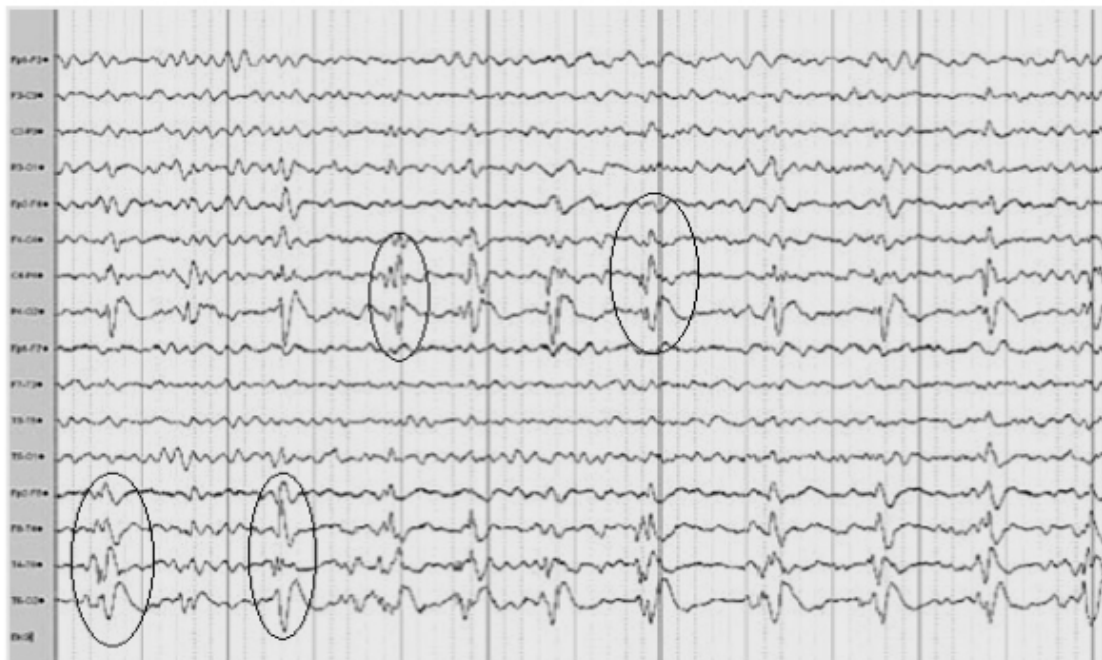
Fonte: Schomer et al (2011), modificado pelo autor.

2.3. Conceito de eletroencefalograma (EEG)

O Eletroencefalograma (EEG) é o estudo do registro gráfico das correntes elétricas desenvolvidas no encéfalo, realizado através de eletrodos aplicados no couro cabeludo, na superfície encefálica, ou até mesmo dentro da substância encefálica (www.pt.wikipedia.org, acessado em 05/08/2014). Em nível de exemplo, a Figura 2.3 mostra o EEG dum paciente de 51 anos com um tumor de gliomas onde pode-se observar as descargas ictais rítmicas características deste diagnóstico.

O encéfalo é parte do sistema nervoso central contida na cavidade do crânio, e que abrange o cérebro, o cerebelo, pedúnculos, a protuberância e o bulbo raquiano (www.pt.wikipedia.org, acessado em 05/08/2014).

Fig. 2.3 – Exemplo de Eletroencefalograma.



Fonte: Domínio Público

2.4. Aplicação do EEG

O EEG é indicado quando há suspeitas de alterações das atividades elétricas cerebral e frequências cerebrais fisiológicas, tais como:

- Epilepsia ou suspeita clínica dessa doença.
- Pacientes com alteração da consciência.
- Morte cerebral.

Avaliação diagnóstica de pacientes com outras doenças neurológicas como, por exemplo, infecciosas, degenerativas e psiquiátricas.

2.5. Doenças diagnosticadas pelo EEG

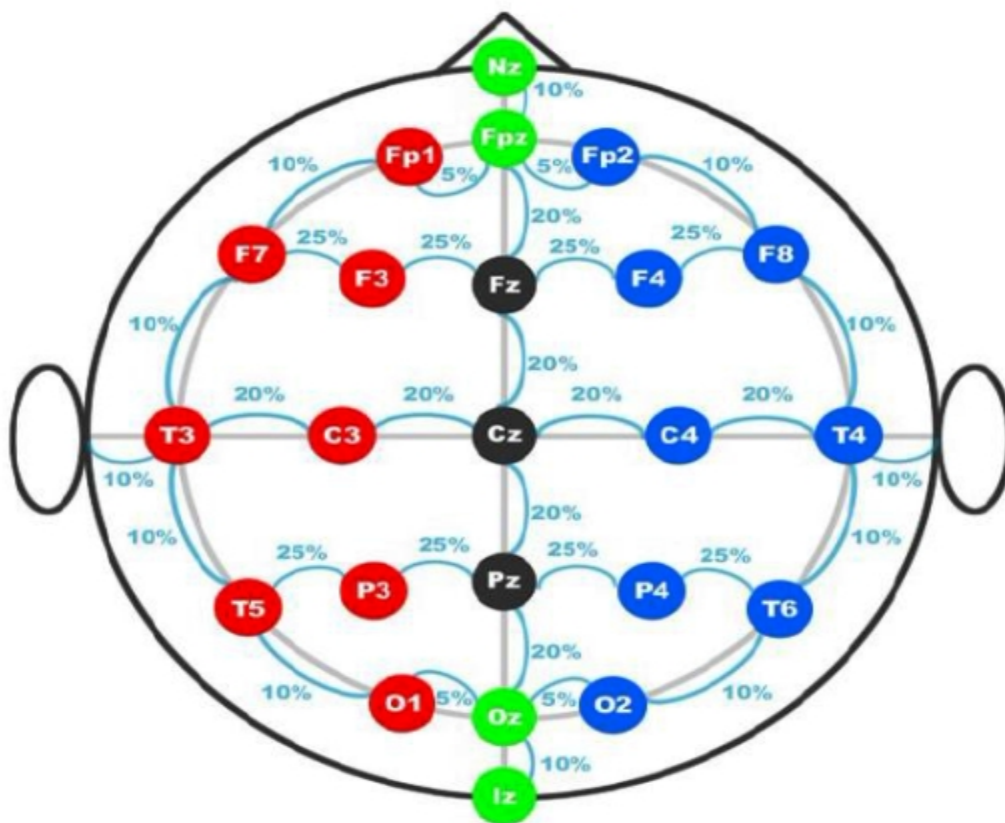
As principais doenças que podem ser detectadas analisando um Eletroencefalograma são:

- Epilepsias generalizadas.
- Epilepsias focais.
- Tumores e distúrbios vasculares.
- Processos inflamatórios tais como meningite e encefalite.
- Demências (Alzheimer) e doenças degenerativas.
- Trauma cranioencefálico.
- Morte encefálica.
- Aneurisma
- AVC
- Neurocisticercose. Cisticercose é a infecção causada pela forma cística da *tênia do porco*, a *Taenia Solium*.

2.6. Mapeamento de posições e montagem dos eletrodos

O eletrodo mais usado é o Ag-AgCl (Prata - Cloreto de Prata). Após medir a circunferência da cabeça do paciente, os eletrodos são colocados como ilustra a Figura 2.4.

Fig. 2.4 - Mapeamento de posições [10-20]

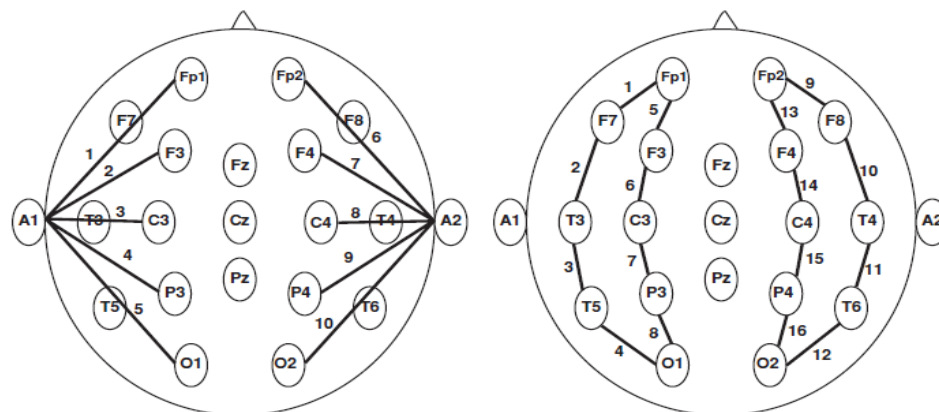


Fonte: Domínio Público

A montagem dos eletrodos podem ser de duas maneiras, como ilustrado na Figura 2.5, onde A1 e A2 são os lobos das orelhas, que apresenta tensão muito baixa, ou seja: neutro. A figura da esquerda mostra uma montagem unipolar e a da direita uma montagem bipolar.

É importante ressaltar que existe outro mapeamento, bem como outra montagem de eletrodos. Trata-se de uma touca de material plástico, para 128 eletrodos, como ilustra a Figura 2.6.

Fig. 2.5 - Montagem dos eletrodos



Fonte: Domínio Público

Fig. 2.6 - Touca para 128 eletrodos



Fonte: Domínio Público

2.7. Tipos de ondas cerebrais e frequências

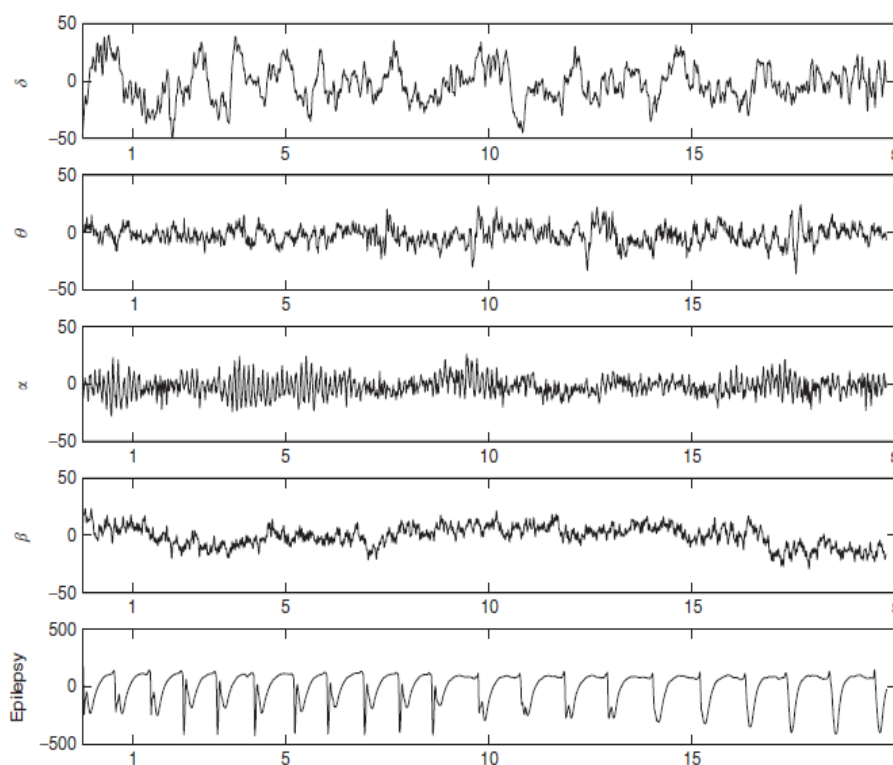
Atualmente se reconhece cinco tipos de ondas cerebrais denominadas com: delta, teta, alfa, beta e gama. A faixa de frequência destas ondas estão apresentadas na Tabela 2.1. Na Figura 2.7 tem-se a forma destas ondas num EEG.

Tab. 2.1 - Ondas cerebrais e frequências

Frequências [Hz]		
Delta (δ)	0,5 – 4,0 Hz	Sono profundo (75 – 200 mV)
Teta (θ)	4,0 – 8,0 Hz	Raramente presente em ser humanos
Alfa (α)	8,0 – 13,0 Hz	Vigília – região posterior da cabeça
Beta (β)	13,0 – 30,0 Hz	Aumento da vigilância e atenção
Gama (γ)	> 30,0 Hz	Estímulos sensoriais e movimentos voluntários

Fonte: Domínio Público

Fig. 2.7 - Ondas cerebrais e frequências



Fonte: Tatum, Husain, Benbadis e Kaplan, 2008.

Tatum et al (2014) apresentam em seu trabalho uma comparação entre as bandas existentes em EEG, bem como as patologias dentro de cada banda, conforme mostrado na Tabela 2.1.

Tab. 2.1. Comparação entre as bandas encontradas em EEG.

Banda	Freq. (Hz)	Funções Normais	Patologias
Sub-Delta	0,0 – 0,5	<ul style="list-style-type: none"> • Artefatos 	<ul style="list-style-type: none"> • ritmo interictal e ictal com crises focais
Delta	0,5 – 3,5	<ul style="list-style-type: none"> • Sono • Hiperventilação • Slow waves na juventude e Idosos 	<ul style="list-style-type: none"> • encefalopatia • lesões estruturais envolvendo a substância branca
Teta	3,5 - < 8	<ul style="list-style-type: none"> • sonolência • crianças • Idosos 	<ul style="list-style-type: none"> • Encefalopatia metabólica • Lesões (menor correlação que a Delta)
Alfa	8 – 13	<ul style="list-style-type: none"> • Ritmo posterior dominante • Ritmo Mu • “Terceiro” Ritmo 	<ul style="list-style-type: none"> • ritmo ictal com convulsões • coma
Beta	13 – 30	<ul style="list-style-type: none"> • efeito de medicação • sonolência 	<ul style="list-style-type: none"> • Ritmo “Breach” • overdose por drogas (contínuo e com alta amplitude) • ritmo ictal associadas a crises
Gamma	30 – 80	<ul style="list-style-type: none"> • Movimento motor voluntário • Aprendizagem • Memória 	<ul style="list-style-type: none"> • Convulsões
Mu	8 – 12	<ul style="list-style-type: none"> • Neurônios motores em estado de descanso 	<ul style="list-style-type: none"> • Supressão Mu pode indicar que os neurônios espelhos estão trabalhando. Sua deficiência pode desempenhar um papel no autismo.
	80 – 250	<ul style="list-style-type: none"> • Processamento Cognitivo • Consolidação da memória episódica 	<ul style="list-style-type: none"> • Frequência de crises interictal e ictal • Epileptogenesis
	250 – 500		<ul style="list-style-type: none"> • Convulsões

Fonte: Tatum et al, 2014.

2.8. OTIMIZAÇÃO MULTIOBJETIVO.

O Problema de Otimização Multiobjetivo POM está presente em diversas áreas do conhecimento, tais como as engenharias (Guerriero, Pezzella, Pisacane e Trollini, 2014; Llopis-Albert, Rubio e Valero, 2015; Lobato, Steffen Jr, Cavalini e Moreira, 2016; Xu, Liang, Yao, Chen e Li, 2017; Li e Yuan. 2017), energia e meio ambiente (Qu, Liang, Zhu, Wang e Suganthan, 2016; Williams e Kendall. 2017), agricultura (Amini, 2015; Pastori, Udias, Bouraoui e Bidoglio, 2017; Galan-Martin, Vaskan, Anton, e Jimenez, 2017), medicina (Timothy, Craig, Lee e Sharpe. 2014; Shufu Lin, Fan Lin, Chen e Zeng, 2017; Alaimo, Auricchio, Contiani e Zingales. 2017; MadhuSudana, Kannan, Manisha e Diptendu S. R., 2017) e EEG (Goh, Abbass, Tan e Mamun, 2014; Kimovski, Ortega, Ortiz e Baños, 2015; Cîmpanu, Ferariu, Dumitriu e Ungureanu, 2017), entre outras.

Neste capítulo será apresentado o problema da otimização multiobjetivo, bem como os principais métodos utilizados para a sua análise, no intuito de encontrar um conjunto de soluções não dominadas e com a maior diversidade possível, tanto no espaço de parâmetros quanto no de objetivos.

2.8.1 Definição do Problema de Otimização Multiobjetivo

Dado um conjunto x de variáveis de projeto, otimizar significa encontrar a melhor combinação de valores das variáveis de maneira a obter as melhores soluções possíveis, respeitando determinadas restrições impostas pelo problema. Quando se tem apenas um objetivo a ser otimizado a otimização é denominada mono objetivo, sendo possível de se chegar a uma única solução, conhecida como solução ótima ou ótimo global. Porém, afora problemas acadêmicos, o que se tem é uma série de objetivos com custos concorrentes entre si, tais como: menor custo e máxima resistência. Quando é necessário levar em conta diversos aspectos de um determinado problema simultaneamente, os modelos de otimização mais indicados vão apresentar mais de um objetivo, caracterizando assim uma otimização multiobjetivo.

Num problema de otimização multiobjetivo estão envolvidas várias funções objetivo, sendo que algumas devem ser maximizadas e/ou outras minimizadas. O problema pode estar sujeito a um conjunto de restrições, ao qual todas as soluções devem satisfazer. A forma geral de um problema de otimização multiobjetivo é dada pela Equação 2.1, onde x são as N variáveis de decisão do

problema que pertencem ao conjunto X^* , $f_m(x)$ a m -ésima função objetivo, M o número de objetivos, x_{\inf} e x_{\sup} os valores mínimos e máximos aceitáveis para x (restrições laterais), $g_j(x)$ a j -ésima restrição de igualdade, J o número de restrições de igualdade, $h_k(x)$ a k -ésima restrição de desigualdade e K o número de restrições de desigualdade.

$$\text{minimizar ou maximizar } f_m(x) \quad m=1,2,\dots,M \quad (2.1)$$

sujeito a:

$$\begin{aligned} x_{\inf} &< x < x_{\sup} \\ g_j(x) &= 0 \quad j=1,\dots,J \\ h_k(x) &\leq 0 \quad k=1,\dots,K \end{aligned}$$

De acordo com Silva (2012), a principal diferença entre a otimização de objetivo único e POM é que a primeira retorna a melhor solução ótima para o objetivo, enquanto POM encontra várias soluções ótimas considerando os diversos objetivos. Portanto, é necessário um método de avaliação e seleção, o qual é chamado de “Tomador de Decisão” apropriado para o POM. O “Tomador de Decisão” deve implementar uma regra para classificar as soluções, pois todos os objetivos não podem simultaneamente chegar ao seu nível ótimo. Diferentes funções têm sido utilizadas, e podem ser divididas em quatro classes principais conforme apresentadas a seguir:

a) Abordagem de Ponderação (Weighted Sum): Essa abordagem converte o POM em um único objetivo para alcançar uma solução ótima. É eficiente e fácil de implementar, mas apropriada para lidar apenas com poucos objetivos e é incapaz de produzir certas porções da Fronteira de Pareto quando há espaços não-convexos na região de busca (Denysiuk, 2013).

b) Abordagem Lexicográfica (Lexicography): Essa abordagem classifica os objetivos em ordem de importância. A solução ótima é, então, obtida minimizando as funções objetivo, começando com a mais importante e procedendo de acordo com a ordem assinalada de importância dos objetivos (Coello, 1993).

c) Abordagem ε -restrito (ε -Constraint Method): Nesta abordagem, uma das funções objetivo é escolhida para ser otimizada e as demais são convertidas em funções de restrição de desigualdades, onde ε_i representa o valor máximo de f_i . Apesar desta abordagem ser capaz de obter soluções em regiões convexas e não convexas, se os valores de ε_i não forem escolhidos de forma criteriosa o problema pode ficar sem solução (Denysiuk, 2013).

d) Abordagem de Pareto: A Abordagem de Pareto é uma técnica para encontrar as soluções na Fronteira de Pareto utilizando o conceito de dominância.

Na abordagem de Pareto, parte-se da premissa que dificilmente a solução de problemas multiobjetivo resultam na maximização/minimização de todos objetivos simultaneamente. Assim, para detectar soluções pode-se tentar melhorar um determinado objetivo evitando que outro objetivo seja influenciado de maneira negativa. Uma solução do problema que não influencie negativamente nenhum objetivo é denominada como não dominada, Pareto ótimo ou Pareto eficiente. Sendo que a finalidade de um problema multiobjetivo é detectar Paretos ideais e quantificar conflitos de escolha (Demir, Bektas e Gilbert, 2014). O conceito de Pareto ótimo foi elaborado por Vilfredo Pareto, em 1896, que enunciou a relação que diz: “não é possível melhorar uma característica do problema sem piorar outra”.

Dados dois vetores de decisão quaisquer $x, y \in X^*$, há três possibilidades para os vetores objetivos:

- $\forall i, f_i(x) \leq f_i(y)$: Diz-se que o vetor objetivo $z^1 = f(x)$ domina $z^2 = f(y)$ e que o vetor de decisão x domina y ;
- $\forall i, f_i(y) \leq f_i(x)$: Diz-se que o vetor objetivo $z^2 = f(y)$ é dominado por $z^1 = f(x)$ e que o vetor de decisão y é dominado por x ;
- dado $f_i(x) \leq f_i(y)$, se $\exists j, f_j(y) \leq f_j(x)$: Diz-se que o vetor objetivo $z^1 = f(x)$ é indiferente a $z^2 = f(y)$ e vice-versa, e que o vetor de decisão x é indiferente com y e vice-versa.

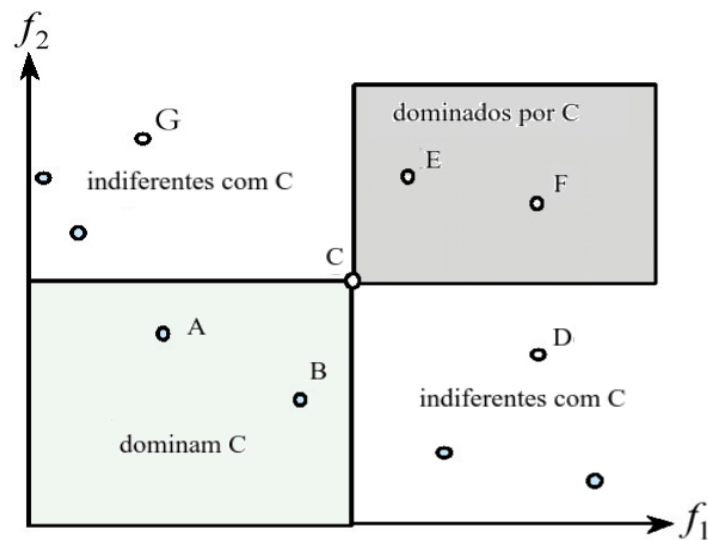
Utilizando um exemplo com dois objetivos, a Figura 2.8 ilustra os vetores objetivos onde os pontos A e B dominam C, os pontos E e F são dominados por C e os pontos D e G são indiferentes com C.

Uma solução é dita Pareto ótima, ou não-dominada ou eficiente, se não for dominada por nenhuma outra solução viável no espaço de busca X^* . Uma solução Pareto ótima não pode ser melhorada com relação a qualquer objetivo sem que exista piora para, pelo menos, algum outro objetivo. Então, o problema é encontrar uma solução eficiente, ou seja, um vetor $x \in X^*$ tal que:

$$\nexists y \in X^* \text{ tal que } f_i(y) \leq f_i(x) \text{ para todo } i = 1, \dots, M \\ \text{e } f_j(y) < f_j(x) \text{ para algum } j = 1, \dots, M$$

A imagem de um determinado conjunto Pareto ótimo, no espaço dos valores dos objetivos, é chamada de fronteira de Pareto.

Fig. 2.8 Ilustração dos conceitos de dominância de Pareto em um problema de minimização com dois objetivos.



Fonte: Arroyo (2002), *apud* Hashimoto, 2004.

Um exemplo simples, e ilustrativo, de problema de otimização multiobjetivo irrestrita foi apresentada por Silva (2009):

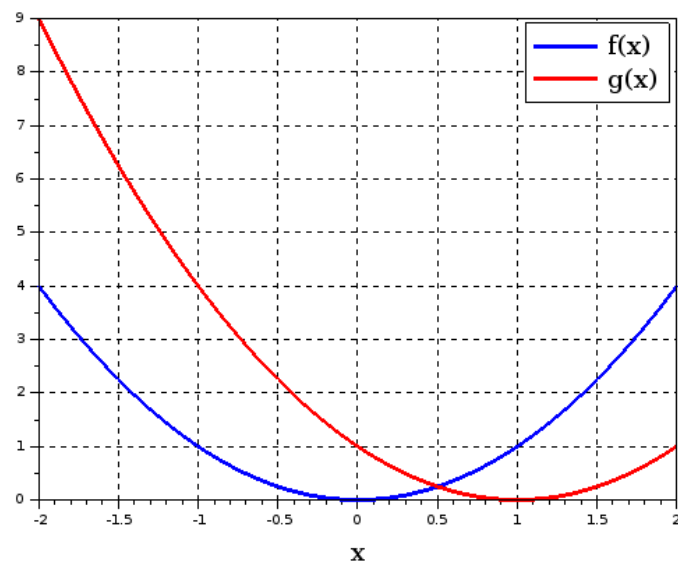
$$\text{Min } f(x) = x^2$$

$$\text{Min } g(x) = (x-1)^2$$

$$x \in \mathbb{R}$$

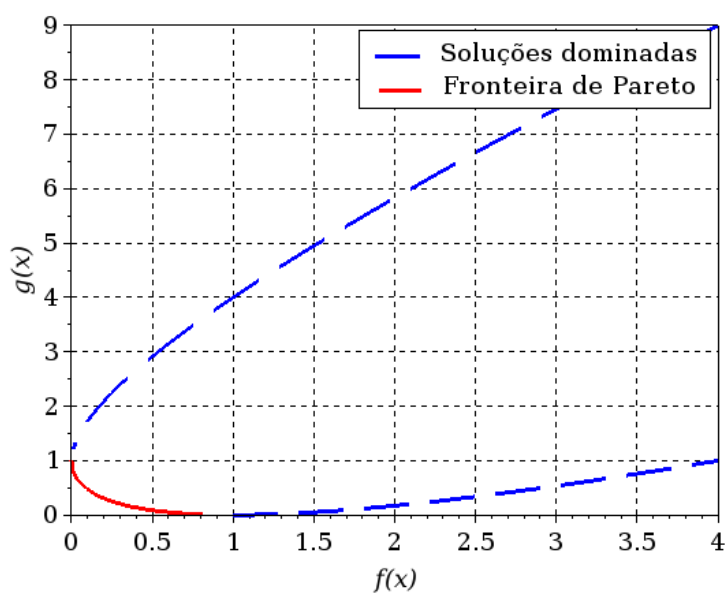
A Figura 2.9 mostra o gráfico das funções f e g , onde é fácil identificar que as soluções Pareto ótimas estão no intervalo $[0,1]$, já que fora dele ambas as funções crescem. Note que neste intervalo acontece uma compensação entre as funções objetivo, isto é, enquanto uma função cresce a outra diminui de valor. A fronteira de Pareto para o exemplo é mostrada na Figura 2.10.

Fig. 2.9. Gráficos das funções $f(x)$ e $g(x)$ do exemplo.



Fonte: Silva, 2009.

Fig. 2.10. Fronteira de Pareto do exemplo.



Fonte: Silva, 2009.

Após a construção da Fronteira de Pareto, na escolha da melhor função objetivo, ocorre um problema estudado na literatura como Multiple Attribute Decision Making (MADM), e as

principais dificuldades são: escolher a solução mais apropriada entre um conjunto de soluções e analisar a qualidade do objetivo entre vários objetivos.

O “Tomador de Decisão” pode ser classificado de acordo com os diferentes momentos do processo de otimização (Silva, 2012):

a) A priori: A decisão é tomada antes da busca. Normalmente, a otimização de único objetivo visa a encontrar a melhor solução de acordo com o conjunto de preferências no “Tomador de Decisão”.

b) A posteriori: Do conjunto de soluções, o “Tomador de Decisão” decide a melhor solução ao final da busca.

c) Interativa: o “Tomador de Decisão” interfere durante o processo de otimização guiando a busca para regiões com soluções de seu interesse.

2.8.2 Algoritmos Genéticos Multiobjetivos e outras Meta-heurísticas

Através da variação dos valores de peso ϵ pode-se construir uma frente de Pareto ótima utilizando as abordagens de ponderação ϵ -restrito, entretanto de acordo Delinsk e Colmonero (2017), os algoritmos genéticos foram desenvolvidos por autores que perceberam algumas limitações nas técnicas de programação matemática multiobjetivo, tais como:

1. Frente de Pareto é côncava ou desconectada o problema pode não ter uma solução ótima, ou seja, o problema depende da forma da frente de Pareto para conseguir detectar uma solução;
2. Os métodos heurísticos geram apenas uma solução, tendo que ser executado várias vezes para gerar um conjunto ótimo de Pareto;
3. Exigência de diferenciação das funções objetivo e restrições.

Coello (2006) destaca que os algoritmos genéticos são menos suscetíveis a forma da frente de Pareto e são capazes de gerar um conjunto de soluções ótimas em apenas uma execução do algoritmo. De acordo com Deb (2000) os algoritmos genéticos desenvolvidos para fins de POM podem ser classificados em algoritmos genéticos multiobjetivo não elitistas (primeira geração) e algoritmos genéticos elitistas multi-objetivo (segunda geração). Sendo que o elitismo acontece através da preservação das soluções não-dominadas entre uma população e outra. Assim, é possível detectar soluções não-dominadas com maior confiabilidade, pois as soluções não são deterioradas.

Dentre as várias otimizadores meta-heurísticos existentes para POM, neste trabalho far-se-á uma breve explanação sobre os procedimentos utilizados na tese: MOPSO, NSGA-II, PESA-II e SPEA2.

2.8.3. Non-dominated Sorting Genetic Algorithm II (NSGA-II)

As primeiras versões com múltiplos objetivos (multiobjetivo) dos algoritmos genéticos (Ags) foram propostas na década dos anos 1990 e, posteriormente, foram aprimoradas por vários pesquisadores. O algoritmo NSGA-II, proposto por Deb et al. (2002), é um algoritmo baseado em uma ordenação elitista por dominância que pode ser resumido em n passos:

1. Inicializa aleatoriamente N_{pop} indivíduos do conjunto x das variáveis de decisão.
2. Calcula as funções objetivo.
3. Classifica as variáveis de decisão por fronteiras F , onde na fronteira F_1 estariam os melhores indivíduos por critérios de dominância de todo o conjunto. Na fronteira F_2 estariam todos os indivíduos não dominados de $N_{pop} - F_1$ e assim sucessivamente, conforme ilustrado na Figura 2.11a.
4. Aplicação dos operadores genéticos de cruzamento (Kumar, Husian, Upreti e Gupta, 2010, ou seja, as características dos indivíduos escolhidos são recombinadas, gerando N_{pop} novos indivíduos (Figura 2.11b).
5. Elitismo: as duas populações serão concatenadas em uma só e se faz um processo de seleção de todos os indivíduos para formar uma nova população com N_{pop} indivíduos. Em caso de empate no ranqueamento, a função distância de multidão (Crowding Distance – CD), Equação 2.2, é utilizada visando manter a diversidade da população.
6. O ranqueamento atual é incrementado (Figura 2.11c) e retorna ao Passo 3 até a convergência do procedimento ou atingir o número máximo de gerações.

Ao final do procedimento espera-se que vários, ou todos, indivíduos estejam na primeira fronteira com variabilidade elevada. A primeira fronteira tornar-se a fronteira de Pareto, ou seja, todas as soluções ótimas possíveis foram encontradas.

$$cd_i = \sum_{m=1}^M \frac{d_m^i}{(f_m^{max} - f_{min}^m)} \quad (2.1)$$

Onde:

$$d_m^i = f_{i+1} - f_{i-1} \quad ;$$

f_{i+1} : função objetivo correspondente a um critério de otimização do próximo indivíduo;

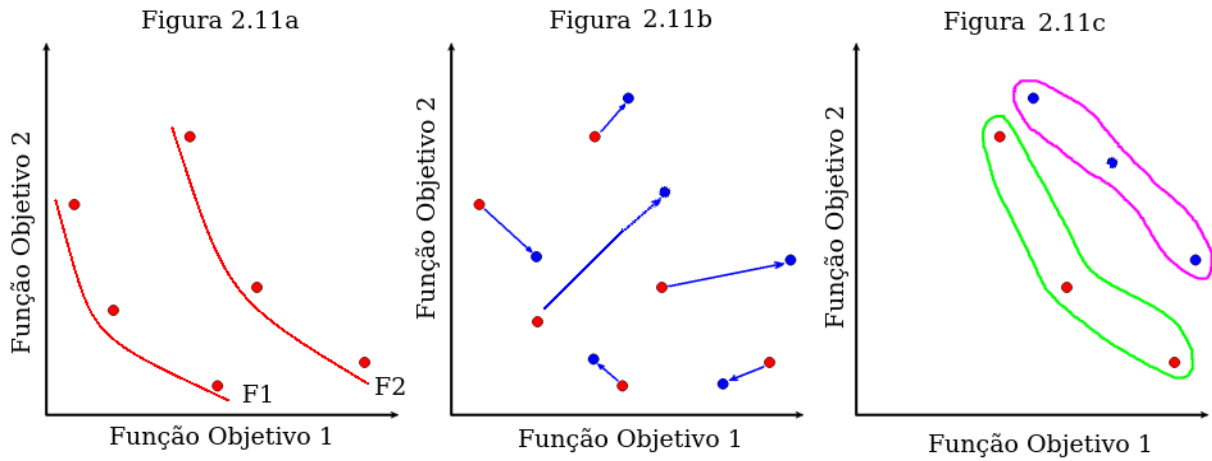
f_{i-1} : função correspondente a um critério de otimização do indivíduo anterior;

f_m^{max} : maior valor da função objetivo correspondente a um critério de otimização aplicado na fronteira m;

f_m^{min} : menor valor da função objetivo correspondente a um critério de otimização aplicado na fronteira m.

Como um problema de otimização possui N funções de otimização, o valor final da função distância de multidão será o somatório dos N cd_i .

Fig. 2.11. Exemplo ilustrativo das etapas do NSGA II com duas funções objetivo.



Fonte: o autor, 2018.

2.8.4. Strength Pareto Evolutionary Algorithm 2 (SPEA 2)

O SPEA 2 (Zitzler, Laumanns e Thiele, 2001), assim como o NSGA-II, trabalha com duas populações P_t e A_t , sendo que em P_t são armazenadas as populações geradas e em A_t (ARQUIVO) armazenadas apenas as soluções não dominadas encontradas pelo algoritmo.

Dados o tamanho da população N_{pop} , o tamanho do ARQUIVO N_{arq} e o número máximo de gerações T , o procedimento consiste de oito passos:

1. Gerar aleatoriamente uma população inicial P_0 com N_{pop} indivíduos.
2. Gerar um arquivo A_0 vazio.
3. Calcular os valores da função de aptidão dos indivíduos em P_t e A_t .
4. Copiar os indivíduos não dominados de P_t para A_t . Se o número de indivíduos não dominados for maior do que N_{arq} aplicar um operador de truncamento. Caso contrário, os melhores indivíduos dominados em P_t são copiados para A_t .
5. Se algum critério de parada for alcançado retorna A_t .
6. Seleção por torneio com reposição de indivíduos em A_t .
7. Gerar P_{t+1} aplicando Crossover e Mutação nos indivíduos selecionados no passo anterior;
8. Atualizar t e ir para o passo 2.

Para avaliar a função de aptidão para os membros das duas populações, para cada indivíduo i nos arquivos é atribuído um valor $S(i)$ (strength) $\in [0,1]$ igual ao número de indivíduos j que são dominados por i , ou iguais, com respeito aos valores das funções objetivo, dividido pelo tamanho da população mais um.

A função de aptidão $R(j)$ de um indivíduo j na população é calculado somando os $S(i)$ valores que dominam, ou são iguais, a j adicionando um no final.

A função de aptidão F_{SPEA2} é definida pela Equação 2.3, onde $R(i)$ está associado com a proximidade da fronteira de Pareto e $D(i)$ à diversidade no espaço dos objetivos.

$$F_{SPEA2}(i) = R(i) + D(i) \quad (2.2)$$

onde:

$$D(i) = \frac{1}{\sigma_k^i + 2}$$

σ_k^i : é a distância de um indivíduo para o k-ésimo vizinho mais próximo;

$$k = \sqrt[2]{N_{pop} + N_{arq}}$$

2.8.5. Pareto Envelope-based Selection Algorithm II (PESA II)

O algoritmo PESA-II (Knowles e Corne, 2000) segue o procedimento padrão de manter duas populações: uma interna de tamanho fixo e uma externa de tamanho flutuante, porém de tamanho limitado. A população interna armazena as novas soluções geradas pelo processo evolutivo e a externa contém as soluções não dominadas encontradas durante o processo de otimização. Para manter a diversidade populacional, durante o processo de otimização o algoritmo utiliza uma malha de hipercubos, nas quais os indivíduos são agrupados. O diferencial do PESA é escolher inicialmente um hipercubo e depois um indivíduo deste hipercubo é escolhido de forma aleatória para os procedimentos de operações genéticas. Este processo garante a diversidade da frente de Pareto, uma vez que a influência dos hipercubos densamente povoados é minimizada durante o processo evolutivo.

2.8.6. Multi-objective Particle Swarm Optimization II (MOPSO)

O algoritmo de Otimização por Enxame de Partículas PSO, proposto por Kennedy e Eberhart (1995), é uma meta-heurística inspirada no comportamento de bandos de aves a procura de alimento. Os bons resultados obtidos por esta técnica na otimização de problemas mono objetivo incentivaram o estudo de variações para problemas multiobjetivos (MOPSO).

A versão MOPSO-CD (crowding distance) utiliza o PSO para atualizar as velocidades das partículas de tal maneira a privilegiar, dentro do espaço de busca, as soluções não dominadas para compor o conjunto de Pareto. Como nos algoritmos de POM, apresentados anteriormente, um arquivo externo armazena as soluções não dominadas ordenadas conforme um atributo denominado distância de multidão, o qual representa quão densa populacionalmente é a região em que a solução se encontra num processo similar ao PESA-II.

2.9 ANÁLISE DISCRIMINANTE

A Análise Discriminante Simples AD, ou Múltipla MDA, é uma técnica estatística apropriada quando se tem variáveis independentes métricas e variáveis dependentes categóricas. Por exemplo, no caso sob estudo acredita-se que a variável métrica independente nível RMS do sinal está relacionada com a variável dependente classificação do EEG que possui dois valores categóricos: EEG normal e EEG anormal.

O objetivo principal da AD é identificar as variáveis que discriminam os grupos e, assim, elaborar previsões a respeito de uma nova observação, identificando o grupo mais adequado a que ela deverá pertencer, em função de suas características. Para alcançar esse objetivo, a AD gera funções discriminantes (combinações lineares das variáveis) que maximizam a discriminação dos grupos descritos pelas variáveis dependentes (Fávero, Belfiore, Silva, e Chan, 2009). Para atingir estes objetivos é necessário contemplar os três itens apresentados a seguir:

1. Identificar as variáveis que melhor discriminam dois ou mais grupos;
1. Utilizar estas variáveis para desenvolver funções discriminantes que representem as diferenças entre os grupos;
2. Fazer uso das funções discriminantes para classificar novas observações.

As hipóteses básicas para o sucesso de uma AD são:

- Normalidade multivariada das variáveis independentes.
- Homogeneidade das Matrizes de variância e covariância dos grupos.
- Multicolinearidade.

Duarte (2017) em seu trabalho sobre utilização de inteligência artificial aplicada na análise de eletroencefalografia utilizou quatro técnicas de AD para fins de classificação de EEGs: Análise Discriminante Fisher LDF, Análise Discriminante Quadrática QDA, k-Nearest Neighbor kNN e wighted k-Nearest Neighbor w-KNN. Em função dos resultados alcançados, e custo computacional, Duarte conclui que a QDA se mostrou bastante interessante.

2.9.1. Qadratic Discriminante Analysis (QDA).

A grande vantagem da análise discriminante quadrática é que não existe a necessidade de que os grupos possuam matrizes de covariância iguais.

Definindo $f_1(x)$ e $f_2(x)$ como as funções densidade de probabilidade associada aos vetores das variáveis aleatórias \mathbf{X} das populações π_1 e π_2 ; R_1 e R_2 os conjuntos de valores de \mathbf{X} que classificam os indivíduos pertencentes às populações π_1 e π_2 ; p_1 como a probabilidade a priori de π_1 e p_2 a de π_2 ; de acordo com a Figura 2.12 as probabilidades globais de classificação correta e incorreta dos indivíduos podem ser derivadas como o produto das probabilidades a priori e as probabilidades de classificação condicional conforme teorema de Byes:

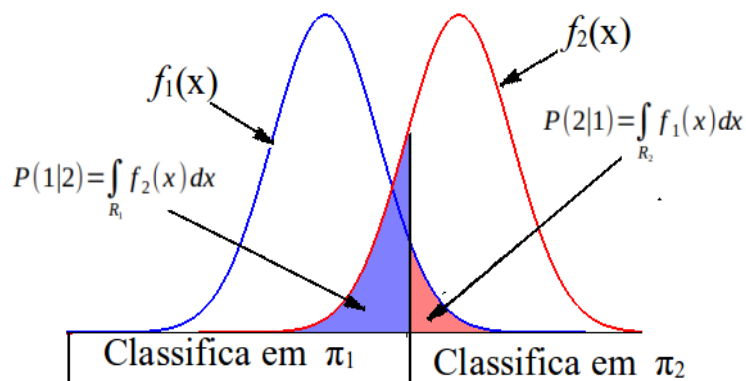
Corretamente classificado como π_1 : $P(\mathbf{X} \in R_1|\pi_1)P(\pi_1) = P(1|1)p_1$

Incorretamente classificado como π_1 : $P(\mathbf{X} \in R_1|\pi_2)P(\pi_2) = P(1|2)p_2$

Corretamente classificado como π_2 : $P(\mathbf{X} \in R_2|\pi_2)P(\pi_2) = P(2|2)p_2$

Incorretamente classificado como π_2 : $P(\mathbf{X} \in R_2|\pi_1)P(\pi_1) = P(2|1)p_1$

Fig. 2.12. Ilustração de probabilidades a priori.



Fonte: o autor, 2018.

Os custos de classificações incorretas são definidos por uma matriz de custos, conforme mostrada a seguir, onde $C(1|2)$ e $C(2|1)$ são os custos de um indivíduo ter sido classificado incorretamente.

Classificada como:

		π_1	π_2
População	π_1	0	$C(2 1)$
verdadeira	π_2	$C(1 2)$	0

Definindo-se o custo médio de classificação incorreta ECM como o produto dos custos por suas probabilidades de ocorrência, tem-se a Equação 2.4.

$$ECM = C(1|2)P(1|2)p_2 + C(2|1)P(2|1)p_1 \quad (2.4)$$

Os valores de R_1 e R_2 que minimizam ECM são os valores de x que mantêm as desigualdades, das equações 2.3 e 2.4, respectivamente.

$$R_1: \frac{f_1(x)}{f_2(x)} \geq \left(\frac{C(1|2)}{C(2|1)} \right) \left(\frac{p_2}{p_1} \right) \quad (2.5)$$

$$R_2: \frac{f_1(x)}{f_2(x)} < \left(\frac{C(1|2)}{C(2|1)} \right) \left(\frac{p_2}{p_1} \right) \quad (2.6)$$

Para minimizar ECM é necessário conhecer a razão das funções densidade de probabilidade, a razão dos custos e a razão das probabilidades a priori, o que na maioria das aplicações são difíceis de se obter. De acordo com Varela (2007) quando os custos são indeterminados assume-se que a razão entre eles é unitária e por serem normalmente desconhecidas assume-se o mesmo para a razão das probabilidades a priori. Com estas simplificações, o problema consiste em escolher os valores de R_1 e R_2 que minimizam a probabilidade de erro de classificação total TPM, de acordo com a Equação 2.7.

$$TPM = p_1 \int_{R_2} f_1(x) dx + p_2 \int_{R_1} f_2(x) dx \quad (2.7)$$

Com a hipótese de populações normais multivariadas tem-se que $f_1(x) \sim N_p(\mu_1, \Sigma_1)$ e $f_2(x) \sim N_p(\mu_2, \Sigma_2)$. Dado um vetor de variáveis aleatórias $\mathbf{X} = [X_1, X_2, \dots, X_p]^T$ pertencente à população π_1 ou π_2 , tem-se que:

$$f_i(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} e^{-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)} \quad i=1,2. \quad (2.8)$$

Utilizando a Equação 2.8 e tomando o logaritmo Neperiano da função discriminante $\lambda(\mathbf{X})$, definida coma a razão entre $f_1(\mathbf{x})$ e $f_2(\mathbf{x})$, e multiplicando por -2 (para manter a ordem das desigualdades 2.4 e 2.5), obtêm-se a Equação 2.9.

$$-2 \log_e[\lambda(\mathbf{X})] = [\log_e|\Sigma_1| - \log_e|\Sigma_2|] + [(\mathbf{X} - \mu_1)^T \Sigma_1^{-1} (\mathbf{X} - \mu_1)] - [(\mathbf{X} - \mu_2)^T \Sigma_2^{-1} (\mathbf{X} - \mu_2)] \quad (2.9)$$

Considerando as equações 2.5 e 2.6, quando se substitui um vetor \mathbf{X} na Equação 2.9, têm-se duas possibilidades de resultados:

$$\log_e(\lambda(\mathbf{X})) \geq 0 = \log_e\left[\left(\frac{C(1|2)}{C(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right], \quad \mathbf{X} \text{ está mais próximo de } \mu_1 \text{ do que de } \mu_2.$$

$$\log_e(\lambda(\mathbf{X})) < \log_e\left[\left(\frac{C(1|2)}{C(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right], \quad \mathbf{X} \text{ está mais próximo de } \mu_2 \text{ do que de } \mu_1.$$

Na prática utiliza-se as médias e covariâncias amostrais (\bar{X}_i, S_i) de amostras das populações π_1 e π_2 (conjuntos de aprendizado supervisionado).

Para constar, no caso em que as populações são homocedásticas ($\Sigma_1 = \Sigma_2$), o procedimento resulta no Discriminante Linear de Fischer.

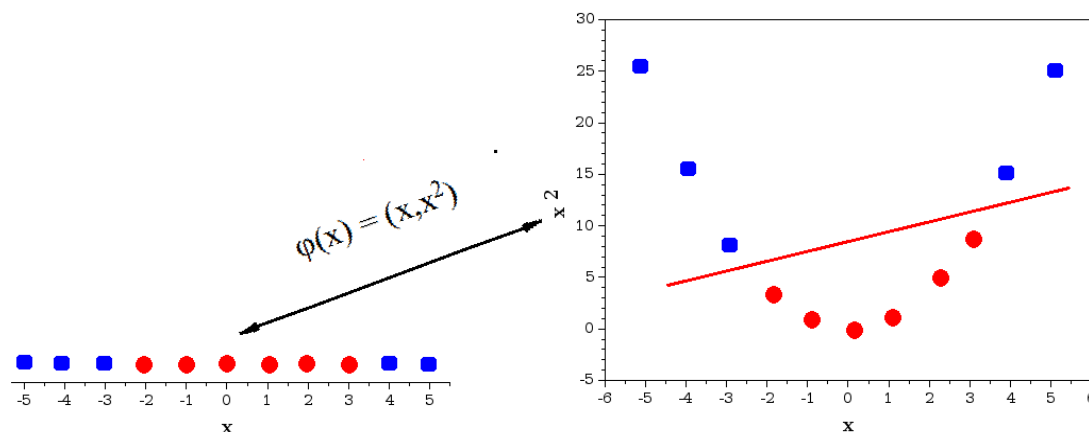
2.10 MÁQUINA DE VETORES DE SUPORTE (SUPPORT VECTOR MACHINE)

As Máquinas de Vetores de Suporte (SVM), cuja fundamentação teórica é muito bem-apresentada por Lorean e Carvalho (2007), é um classificador baseado na Teoria de Aprendizado Estatístico (TAE) que estabelece as condições matemáticas para auxiliar na escolha de um classificador particular \bar{f} a partir de um conjunto de dados de treinamento. Essas condições levam em conta o desempenho do classificador no conjunto de treinamento e a sua complexidade, com o objetivo de obter um bom desempenho também para novos dados do mesmo domínio. As máquinas de vetores de suporte são muito utilizadas na classificação de EEG, sendo que uma pesquisa no google “Artigos Acadêmicos sobre EEG vector machine” resultou em

aproximadamente 53800 resultados no dia 05 de junho de 2018. Só do ano 2018 foram encontrados cinco trabalhos publicados. As premissas do procedimento são:

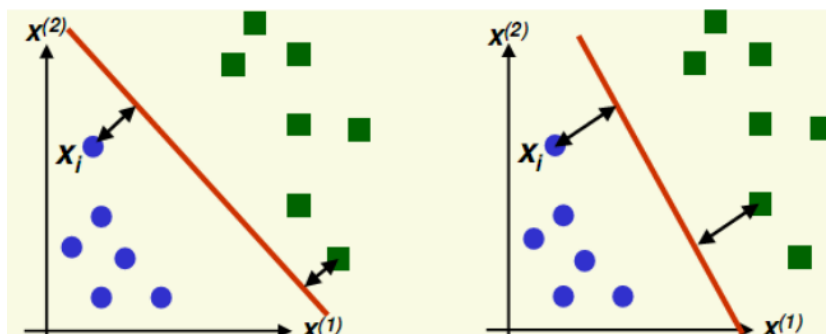
- Aprendizagem supervisionada.
- Classificador binário.
- Classificação linear. Para problemas de classificação não lineares se utiliza o Teorema de Cover: “os dados são mais propensos a serem separados linearmente em altas dimensões”. As funções **kernel** são utilizadas para projetar os dados de um espaço não linearmente separáveis em um que seja possível separá-los linearmente (Figura 2.13).
- Trabalha com o conceito de maximização da margem dos hiperplanos (Figura 2.14).

Fig. 2.13. Exemplo de elevação para uma dimensão linearmente separável: $\mathbf{R}^1 \rightarrow \mathbf{R}^2$.



Fonte: o autor, 2018.

Fig. 2.14. Exemplo de maximização de margens



Fonte: Zavaschi, 2018.

2.11 FERRAMENTAS DE ANÁLISE DE SINAIS APLICADOS EM CLASSIFICAÇÃO DE EEG.

Neste tópico faz-se uma breve revisão das ferramentas de análise de sinais usadas normalmente em procedimentos de classificação de EEGs.

2.11.1. Análise de Envelope.

Um sinal modulado em fase é um sinal que contém informações de fase variante no tempo. Uma representação simples de um sinal com modulação de fase é dado pela Eq. 2.10, onde A é a amplitude do sinal, β é a amplitude da moduladora, Ω é a frequência angular da portadora, $\bar{\omega}$ é a frequência angular da moduladora e θ e ϕ são os ângulos de fase da portadora e moduladora, respectivamente.

$$x(t) = A \sin(\Omega t + \theta + \beta \sin(\bar{\omega} t + \phi)) \quad (2.10)$$

Expressando a Eq. 2.10 na forma complexa, tem-se a Equação 2.11.

$$x(t) = A e^{j[\Omega t + \theta + \beta \sin(\bar{\omega} t + \phi)]} \quad (2.11)$$

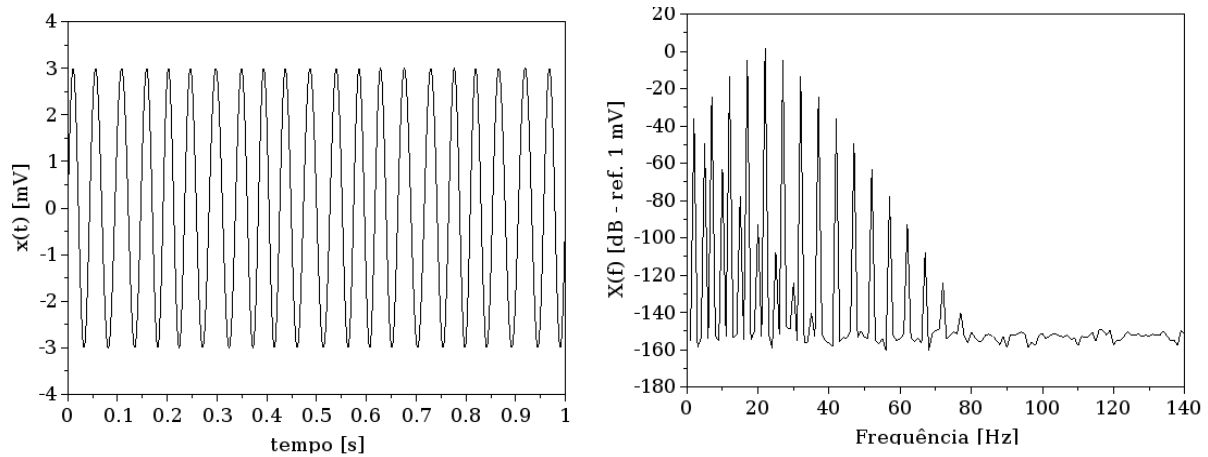
Utilizando a relação de Euler, tem-se:

$$x_a(t) = A \cos(\Omega t + \theta + \beta \sin(\bar{\omega} t + \phi)) + j A \sin(\Omega t + \theta + \beta \sin(\bar{\omega} t + \phi)) \quad (2.12)$$

Na Equação 2.12, $x_a(t)$ é conhecido como sinal analítico.

Em nível de exemplo, a Figura 2.15 mostra os gráficos no tempo e na frequência para $A=3$, $\Omega=329,86$ rad/s, $\beta = 0,5$, $\bar{\omega} = 34,42$ rad/s e $\theta = \phi = 0$. Apesar de a modulação ser quase imperceptível no domínio do tempo, o surgimento das bandas laterais no domínio da frequência é um forte indicativo de modulação em amplitude ou fase. Deve-se ressaltar que o espaçamento das bandas laterais é igual à frequência moduladora.

Fig. 2.15. Exemplo de uma representação simples de um sinal modulado em fase nos domínios do tempo e da frequência.



Fonte: O autor, 2018.

Para construir um sinal analítico basta usar uma função que transforme um cosseno num seno e somar o resultado multiplicado pelo imaginário j ao sinal original. Uma função que faz esta operação é a transformada de Hilbert dada pela Eq. 2.13.

$$Hx(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(u)}{u-t} du \quad (2.13)$$

As propriedades

$$H[\sin(t)] = -\cos(t) \text{ e } H[\cos(t)] = \sin(t)$$

permitem escrever um sinal analítico como apresentado na Equação 2.14.

$$xa(t) = x(t) + jH[x(t)] \quad (2.14)$$

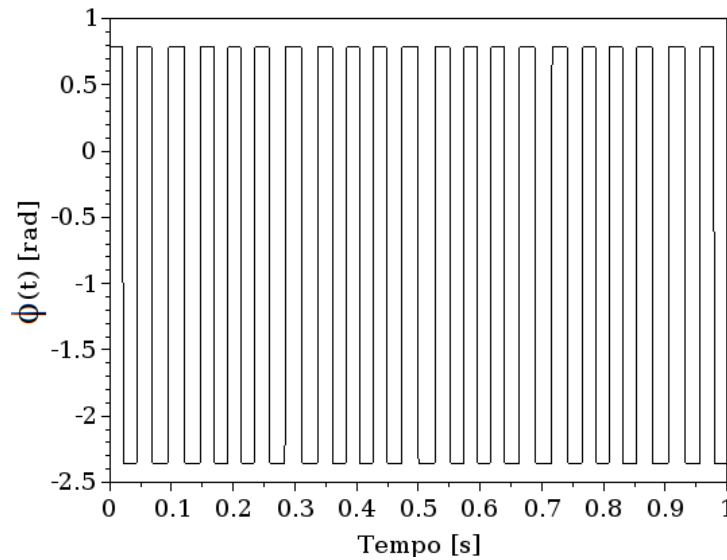
Tendo-se um sinal analítico, o envelope $x_{\text{env}}(t)$ e a fase instantânea $\varphi(t)$ podem ser obtidos utilizando as equações Eq. 2.15 e 2.16, respectivamente.

$$x_{env}(t) = |xa(t)| \quad (2.15)$$

$$\phi(t) = \text{tg}^{-1}\left(\frac{H[x(t)]}{x(t)}\right) \quad (2.16)$$

Para exemplificar, na Figura 2.16 tem-se a fase instantânea calculada para o sinal utilizado como exemplo de modulação de fase.

Fig. 2.16. Fase instantânea do sinal usado como exemplo de modulação de fase.



Fonte: o autor, 2018.

A análise de envelope é muito utilizada para fins de classificação em EEG (Şeker e Özerdem, 2016).

2.11.2. Decomposição em Modos Empíricos (EMD)

Huang, Shen, Long, Wu, Shih, Zheng, Yen, Tung e Liu (1998) propuseram uma classe de funções denominadas como Funções de Modos Intrínsecos (IMF) definidas pela Equação 2.17, onde a amplitude e a fase são física e matematicamente significativas. Para isto, deve-se considerar:

1. Funções simétricas em relação à média local igual a zero;

2. O número de extremos deve ser igual ou diferir de um em relação ao número de cruzamentos em zero.

$$\psi(t) = r(t) \sin[\theta(t)] \quad (2.17)$$

A Decomposição Empírica de Modos (EMD) (Huang *et al.* 1998) é um método que se destaca pela adaptabilidade, não dependendo de características como linearidade ou estacionariedade dos sinais. Este método busca dividir o sinal analisado $x(t)$ em uma série de Funções de Modo Intrínseco (IMFs), onde cada uma deve satisfazer a condições 1 e 2 das IMF.

O algoritmo proposto em (Huang *et al.* 1998), para a obtenção destas IMFs, é composto por cinco passos:

- Passo 1: Inicializa $r_0(t) = h(t) = x(t)$.
- Passo 2: Gerar os envelopes e_{\max} e e_{\min} por meio de interpolação de spline cúbica, passando pelos máximos e mínimos locais de $h(t)$, respectivamente.
- Passo 3: Para o i -ésimo modo, calcular a função $m_i(t)$ como a média das duas splines.
- Passo 4: Atualizar $h(t) = h(t) - m_i(t)$. Se $h(t)$ satisfazer as condições apresentadas anteriormente para uma IMF, a IMF $\psi_i(t)$ é encontrada. Caso contrário voltar para o passo 2.
- Passo 5 Um novo resíduo $r_i(t)$ é gerado: $r_i(t) = r_{i-1}(t) - \psi_i(t)$. Iniciar $h(t) = r_i(t)$ e repetir os passos de 2 a 4 de modo a encontrar as IMFs restantes. O processo para quando não é mais possível computar uma IMF de um resíduo, que então é definido como resíduo final r_N .

Após a decomposição, o sinal $x(t)$ é decomposto em N IMFs mais o resíduo r_N , conforme mostrado na Equação 2.18.:

$$x(t) = r_N + \sum_{k=1}^N \psi_k(t) \quad (2.18)$$

2.11.3. Transformada de Hilbert-Huang (HHT)

A transformada de Hilbert-Huang é uma técnica de decomposição de sinais em tempo-

frequência que rompe com os pressupostos clássicos de projeção do sinal em funções de base ortogonais definidas à priori. Esta técnica tem especial importância quando os sinais não têm características de linearidade e de estacionaridade, que é o caso dos sinais de EEG (Zhu e Lin, 2014).

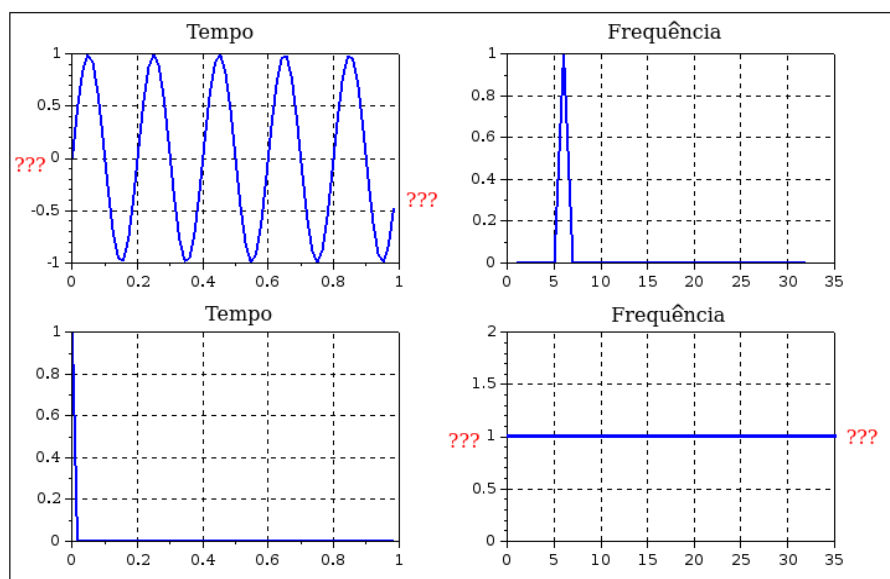
As HHTs de um sinal nada mais é do que o sinal analítico (Eq. 2.11) de uma IMF (Eq. 2.17) obtida por decomposição em modos empíricos.

2.11.4. Transformada Wavelet

A Transformada Wavelet (WT), outra transformada tempo-frequência, é muito utilizada para identificação de epilepsia em sinais de EEG (aproximadamente 57800 resultados de busca “EEG Wavelet” no google Artigos Acadêmicos em 05 de junho de 2018).

O teorema da dualidade tempo frequência demonstra que utilizando uma Transformada Rápida de Fourier é impossível representar bem um sinal simultaneamente nos domínios do tempo e da frequência, como pode ser observado na Figura 2.17.

Fig. 2.17. Exemplo de dualidade Tempo-Frequência.

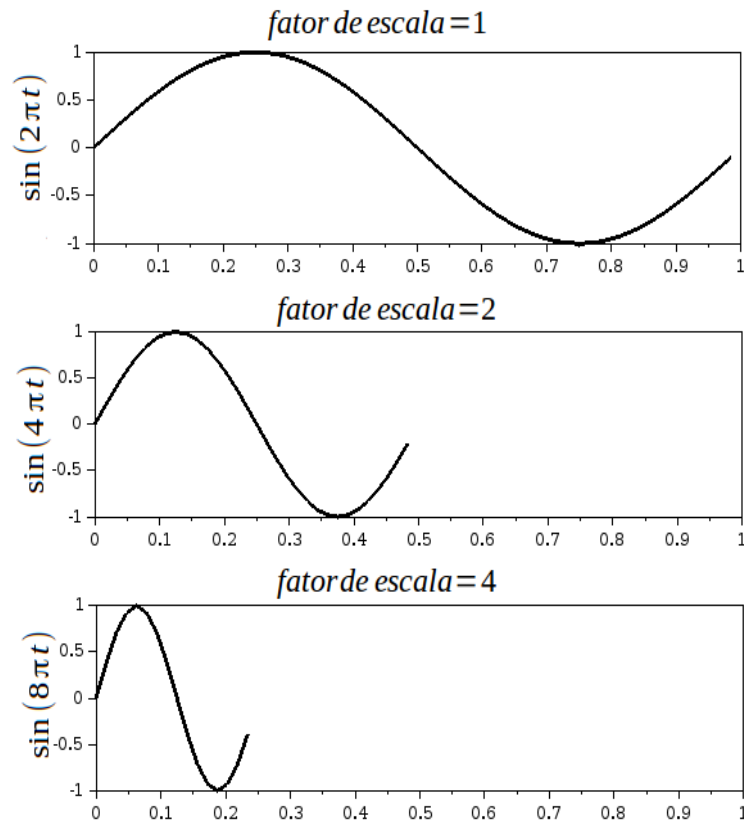


Fonte: o autor, 2018.

A Transformada de Wavelet tenta contornar este problema utilizando funções em diferentes escalas e diferentes resoluções (Figura 2.18), sendo esta uma das suas principais características (Daubechies, 1992; apud Silva, 2008). Simplificando, a WT é uma convolução do tempo do sinal original $x(t)$ com uma função wavelete base ψ^* , onde o eixo do tempo é dilatado por um fator de dilatação a e uma translação da origem b , conforme Equação 2.19.

$$W(a, b, x, \psi) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (2.19)$$

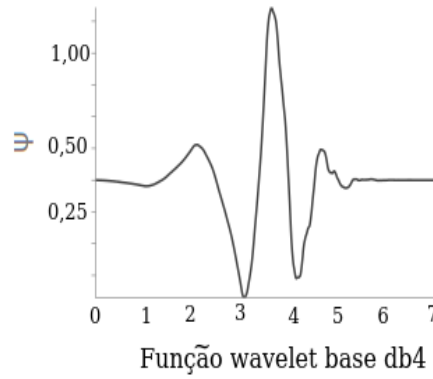
Fig. 2.18. Efeito das escalas na Wavelet: escalas menores resultam em janelas temporais maiores e escalas maiores resultam em janelas temporais menores.



Fonte: adaptado pelo autor, 2018.

A função de sua forma, a wavelet base mais usada em EEG para classificação de eventos focal e no focal é a Daubechies 4 (db4) mostrada na Figura 2.19.

Fig. 2.19. Função wavelet base db4.



Fonte: adaptado pelo autor, 2018.

2.11.5. Métricas Clássicas: Valor RMS, Valor Pico, Valor Pico a Pico, Assimetria, Curtose, Fator de Crista, k4 e k6.

Para reduzir as informações contidas em um sinal no domínio do tempo, frequência e, principalmente, no domínio tempo-frequência é necessário utilizar métricas que condensem a informação em apenas um parâmetro. Uma vez que as perturbações que ocorrem num EEG são do tipo pulso, deu-se especial atenção neste trabalho às métricas que mesuram excursões fora da média (curtose, fator de crista, k4 e k6).

O valor médio quadrático (RMS) de um sinal discretizado $x_i(t)$, também chamado de valor eficaz, é uma das métricas mais utilizadas em análise de sinais por quantificar a energia contida na onda. A equação para calcular o valor RMS de um sinal discretizado com N_{pto} pontos é calculado através da Eq. 2.20.

$$RMS = \sqrt{\frac{1}{N_{pto}} \sum_{i=1}^{N_{pto}} x_i^2} \quad (2.20)$$

O Valor Pico (V_p) de um sinal é o valor máximo (pico positivo) ou mínimo (pico negativo) de amplitude encontrado no sinal. Já o Valor Pico a Pico (V_{pp}) é a diferença entre o valor máximo de

amplitude e o valor mínimo de amplitude encontrado no sinal.

A Assimetria (*skewness*) S_k é um parâmetro estatístico que mede a simetria de uma distribuição em torno da média. O coeficiente de assimetria amostral é dado pela Eq. 2.21.

$$S_k = \frac{\frac{1}{N_{pto}} \sum_{i=1}^{N_{pto}} (x_i - \bar{x})^3}{S^3} \quad (2.21)$$

onde:

$$S = \sqrt{\frac{1}{N_{pto}-1} \sum_{i=1}^{N_{pto}} (x_i - \bar{x})^2} \quad \text{é o desvio padrão amostral;}$$

$$\bar{x} = \frac{1}{N_{pto}} \sum_{i=1}^{N_{pto}} x_i \quad \text{é a média amostral}$$

Estatisticamente a curtose mede o grau de achatamento de uma distribuição, considerado usualmente em relação à distribuição normal, Eq. 2.22. Em um sinal temporal, um valor elevado da curtose indica a presença de excursões longe da média por um curto período de tempo. A bibliografia tem mostrado que esta é uma boa métrica para classificação de EEG (Wang, Shepherd, Beggs, Rao e Zhang, 2015, por exemplo).

$$S_k = \frac{\frac{1}{N_{pto}} \sum_{i=1}^{N_{pto}} (x_i - \bar{x})^4}{S^4} \quad (2.22)$$

O Fator de Crista (FC), definido pela Eq. 2.23, assim como a curtose, é um detector de excursões fora da média, muito utilizado para detecção de falhas em rolamentos.

$$FC = 20 \log_{10} \frac{V_p}{RMS} \quad (2.23)$$

Sinais com grandes excursões fora da média por períodos maiores resultam em valores de curtose menores. Uma métrica para este tipo de comportamento é estimador k4, também muito

utilizado para análise de vibrações em manutenção preditiva, definido pela Eq. 2.24. Grandes excursões fora da média em espaços de tempo curtíssimos são melhores detectadas pelo estimador k_6 é dado pela Eq. 2.25.

$$k_4 = RMS * S_k \quad (2.24)$$

$$S_k = \frac{\frac{1}{N_{pto}} \sum_{i=1}^{N_{pto}} (x_i - \bar{x})^6}{S^6} \quad (2.25)$$

2.11.6. Métrica k-point nonlinear energy operator (k-NEO)

Este estimador foi utilizado por Liu, Lin, Tsai e Sun (2013) com o intuito de detectar picos seguidos de uma onda de baixa frequência em sinais de EEG de pessoas com epilepsia. Ele é uma variação do estimador NEO (*Nonlinear Energy Operator*), comumente utilizado na detecção de ocorrências de picos de alta nos sinais, o que torna um ótimo estimador com aplicação em processamento de sinais, processamento de imagem e demodulação AM/FM (Maragos e Kaiser, 1993). Uma vez que NEO é muito sensível a altos picos de ruído, Choi, em 2002, propôs a versão estendida k-NEO, apresentada na Eq. 2.26.

$$\psi_k[x(n)] = x^2(n) - x(n-k)x(n+k) \quad (2.26)$$

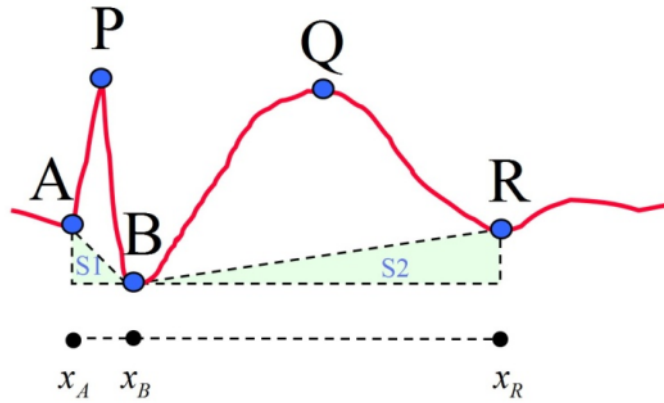
onde k é um parâmetro de resolução relacionado à largura do pico dos candidatos a serem detectados e $x(n)$ é o n -ésimo valor do sinal digitalizado.

Para uma região do sinal como candidato a pico P , detectado pela Eq. 2.23, Liu *et al* definem outros quatro pontos A , B , Q e R , mostrados na Figura 2.20. O grupo de pontos A , P e B representa o pico, enquanto o grupo dos pontos B , Q e R representa a porção da onda lenta. Um único pico possui apenas o primeiro grupo de pontos, enquanto um pico seguido por uma onda lenta tem os dois grupos de pontos. Os dois primeiros pontos característicos são designados por A e B , que correspondem ao início e fim do pico, respectivamente. Ao rastrear uma inclinação positiva, anteriormente ao ponto P , define-se o ponto A . O Ponto B é definido de maneira análoga, porém

posteriormente ao ponto P. Já os pontos Q e R correspondem ao máximo local e ao ponto final da onda lenta.

Identificados os pontos A, P, B, Q e R, calcula-se as 13 métricas, divididas em quatro categorias (duração, amplitude, inclinação e área), utilizando as equações de Eq. 2.27 a 2.39.

Fig. 2.20. Diagrama para extração das características e pontos definidos pelo modelo.



Fonte: Liu *et al*, 2013.

$$Dur_{AP} = A'P_x \quad (2.27)$$

$$Dur_{PB} = P'B_x \quad (2.28)$$

$$Dur_{spike} = A'P_x + P'B_x \quad (2.29)$$

$$Dur_{slowwave} = B'Q_x + Q'R_x \quad (2.30)$$

$$Amp_{AP} = A'P_y \quad (2.31)$$

$$Amp_{PB} = P'B_y \quad (2.32)$$

$$Amp_{spike} = (A'P_y + P'B_y)/2 \quad (2.33)$$

$$Amp_{slowwave} = (B'Q_y + Q'R_y)/2 \quad (2.34)$$

$$Slope_{AP} = A'P_y / A'P_x \quad (2.35)$$

$$Slope_{PB} = -P'B_y / P'B_x \quad (2.36)$$

$$Slope_{sharpness} = Slope_{AP} - Slope_{PB} \quad (2.37)$$

$$Area_{spike} = \int_{x_A}^{x_B} f(x) dx - S_1 \quad (2.38)$$

$$Area_{slowwave} = \int_{x_B}^{x_R} f(x) dx - S_2 \quad (2.39)$$

onde:

- Dur_{AP} é a duração da primeira meia onda do pico;
- $A'P_x$ é a projeção do vetor AP no eixo x;
- Dur_{PB} é a duração da segunda meia onda do pico;
- $P'B_x$ é a projeção do vetor PB no eixo x;
- Dur_{spike} é a soma das durações para o pico;
- $Dur_{slowwave}$ é a soma das durações para a onda lenta;
- $B'Q_x$ é a projeção do vetor BQ no eixo x;
- $Q'R_x$ é a projeção do vetor QR no eixo x;
- Amp_{AP} é a amplitude da primeira meia onda do pico;
- $A'P_y$ é a projeção do vetor AP no eixo y;
- Amp_{PB} é a amplitude da segunda meia onda do pico;
- $P'B_y$ é a projeção do vetor PB no eixo y;
- Amp_{spike} é a média das amplitudes para o pico;
- $Amp_{slowwave}$ é a média das amplitudes para a onda lenta;
- $B'Q_y$ é a projeção do vetor BQ no eixo y;
- $Q'R_y$ é a projeção do vetor QR no eixo y;
- $Slope_{AP}$ é a inclinação da primeira meia onda do pico;
- $Slope_{PB}$ é a inclinação da segunda meia onda do pico;
- $Slope_{sharpness}$ é a nitidez do pico;
- $Area_{spike}$ é a área do pico;
- x_A é a posição do ponto A;
- x_B é a posição do ponto B;
- S_1 é a área mostrada na Figura 2.20;
- $Area_{slowwave}$ é a área da onda lenta;
- x_R é a posição do ponto R;
- S_2 é a área mostrada na Figura 2.20.

2.11.7. Métricas de Entropia

A entropia, é um conceito da termodinâmica que mede a desordem das partículas de um sistema físico. Já contextualizando na teoria de informação, a entropia é um indicador da quantidade de informação armazenada em uma distribuição de probabilidade (Kannathal, Choo, Acharaya e Sadasivan, 2005), Eq. 2.37.

$$H = - \int p(x) \log_e[p(x)] dx \quad \text{Eq. 2.37}$$

Para deixar o conceito estatístico de entropia mais intuitivo, Subramaniyam (2018) considera um experimento aleatório simples com espaço amostral $S=\{1,2,3\}$ com probabilidades $P(1)$, $P(2)$ e $P(3)$, onde se avalia três situações:

$P(1) = P(2) = P(3) = 1/3$, o que resulta em $H=1,099$ (Eq. 2.37).

$P(1)=1$, $P(2)=P(3)=0$, o que resulta em $H=0,00$.

$P(1)=P(2)=1/2$ e $P(3)=0$, resultando em $H=0,69$.

Observa-se que a entropia é máxima quando o número de possibilidades é máxima e mínima quando se tem apenas uma possibilidade. Portanto, quando aplicada em análise de sinais, a entropia é uma métrica da variabilidade (desordem) dentro do mesmo.

Para Sharma, Pachori e Acharya (2015), “a entropia é uma medida da complexidade da série temporal. Os parâmetros não-lineares podem ser úteis para descrever a dinâmica dos sinais EEG considerando a natureza não-linear e não estacionária dos mesmos”.

Neste trabalho, foram utilizados quatro parâmetros de entropia, aplicados nas IMF, descritos em Sharma, Pachori e Acharya (2015), os quais são:

- Entropia média de Shannon (ShEn);
- Entropia espectral média de Renyi (RenEn);
- Entropia aproximada média (ApEn);
- Entropia média da amostra (SpEn);

A partir de um espectro de potência discreto $P(f)$, a Entropia Média de Shannon é calculada pela Eq. 2.41.

$$ShEn = \sum_f p_f \log\left(\frac{1}{p_f}\right) \quad (2.41)$$

Onde:

$$p_f = \frac{P_f}{\sum P_f}$$

Outra medida de entropia que mede a complexidade espectral de uma série temporal é a Entropia Quadrática de Renyi definida pela Eq. 2.42.

$$RenEn(2) = -\log_e \sum_f p_f^2 \quad (2.42)$$

Dada um sinal discretizado em N valores igualmente espaçados, para calcular a Entropia Média Aproximada *EpEn*, inicialmente é necessário definir o tamanho dos subgrupos d e, a seguir, criar sequências de dados

$$X(i) = [x(i), x(i+1), \dots, x(i+d-1)] \quad , \quad 1 \leq i \leq N-d+1$$

e calcular o valor de *ApEn* pela expressão mostrada em Eq. 2.43.

$$ApEn(d, r, N) = \phi^d(r) - \phi^{r+1}(r) \quad (2.43)$$

Onde:

$$\psi^d(r) = \frac{1}{N-d+1} \sum_i \log_e(C_i^d(r))$$

$$C_i^d(r) = (\text{número de ocorrências } x(j) \text{ tal que } D[x(i), x^*(j)] \leq r)$$

$$D[x, x^*] \text{ é a diferença escalar máxima entre os elementos de } x \text{ e } x^*$$

r é a tolerância que depende do tipo de aplicação.

A entropia média da amostra *SpEn* (Eq. 2.44) é uma versão simplificada da *ApEn* que consome um tempo computacional para sua avaliação em torno de 50% do tempo de cálculo da

$SpEn$, entre outras vantagens tais como consistência e ser independente do tamanho da amostra (Sharma et al, 201).

$$SpEn(d, r, N) = -\log_e \left(\frac{A^d(r)}{A^d(r)} \right) \quad (2.44)$$

Onde:

$$B^d(r) = \frac{1}{N-d} \sum_{i=1}^{N-d} C_i^d(r) \quad A^d(r) = \frac{1}{N-d} \sum_{i=1}^{N-d} C_i^{d+1}(r)$$

CAPÍTULO III

3. METODOLOGIA

Apresenta-se neste capítulo a metodologia utilizada para desenvolver um procedimento de anomalias nos sinais de EEG. Inicialmente, escolheu-se os parâmetros de análise sinais candidatos a serem utilizados nos procedimentos de classificação e um banco de dados muito usado em pesquisas envolvendo detecção automática de anomalias em EEG. A seguir, utilizando sinais de EEG normais (não focal) e com anomalias (focal), foram construídos três conjuntos de dados utilizados para treinamento do procedimento de classificação, para a escolha dos melhores parâmetros a serem utilizados na classificação (validação) e o último reservado para testar os procedimentos desenvolvidos.

3.1. ESCOLHA DO BANCO DE DADOS

Para o desenvolvimento deste trabalho, utilizou-se o banco de dados Bern-Barcelona (Andrzejak, Schindler e Rummel, 2012) de domínio público e que é muito usado em pesquisas envolvendo detecção de anomalias em sinais de EEG. Uma pesquisa no Google Acadêmico (2018) com a frase ‘Bern-Barcelona EEG database’ resultou em 51 citações. Os principais motivos para a escolha deste banco de dados foram o fato de que o banco de dados é muito bem documentado e que as anomalias não são tão evidentes, como pode ser observado em outros bancos de dados tais como o do departamento de Epileptologia da universidade de Bonn (Andrzejak, Lehnertz, Rieke, Mormann, David, Elger, 2001).

Os sinais utilizados pelos autores são gravações de EEG intracraniais de cinco pacientes com epilepsia. É importante ressaltar que todos os pacientes apresentaram epilepsia de lóbulo temporal fármaco-resistente de longa data e foram candidatos à cirurgia de epilepsia. Os sinais foram adquiridos com uma taxa de amostragem de 512 e 1024 Hz. Como pré-processamento, todos os sinais foram filtrados por um filtro Butterworth de quarta ordem do tipo passa banda entre 0,5 e

150 Hz. Todos os sinais adquiridos com 1024 Hz foram re-amostrados para a frequência de 512 Hz.

O banco é composto por 7510 arquivos ASCII com duas colunas – canais - de dados (x e y). Nos sinais do primeiro canal (x) de 3750 arquivos, denominados canais focal, pelo menos dois neurologistas certificados detectaram uma mudança ictal. Os canais y são utilizados para a montagem bipolar (x-y). Os outros 3570 arquivos são constituídos de sinais x e y de regiões do eletroencefalograma em que não foram detectadas mudanças ictais (canais não focal). As gravações correspondem a janelas de 20 segundos, o que corresponde a um sinal com 10240 pontos. Gravações das crises convulsivas e de três horas após a última crise foram excluídas. Antes de incluir o par de sinais no banco de dados, todos foram novamente inspecionados visualmente.

Uma vez que os pares de canais focal e não focal foram selecionados aleatoriamente para a construção do banco de dados, para este trabalho foram utilizados os primeiros 350 arquivos focal e não focal agrupados da seguinte maneira:

1. 50 conjuntos para fins de treinamento.
2. 200 conjuntos para validação.
3. 100 conjuntos para teste.

3.2. AVALIAÇÃO DOS PROCEDIMENTOS

Para a avaliação dos procedimentos desenvolvidos, utilizou-se as métricas propostas por Sharma, Pachori e Acharya (2015), que também utilizaram o banco de dados de Bern-Barcelona: Sensibilidade (SEN), dada pela Eq.3.1; Especificidade (SPF), dada pela Eq. 3.2; Precisão (ACC); dada pela Eq. 3.3, Valores Positivos Preditos (PPV), dados pela Eq. 3.4; Valores Negativos Preditos (NPV), dados pela Eq. 3.5; e Coeficiente de Correlação de Mathew (MCC), dado pela Eq. 3.6.

$$SEN = \frac{TP}{TP + FN} \times 100(\%) \quad (3.1)$$

$$SPF = \frac{TN}{TN + FP} \times 100 (\%) \quad (3.2)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 (\%) \quad (3.3)$$

$$PPV = \frac{TP}{TP + FP} \times 100 (\%) \quad (3.4)$$

$$NPV = \frac{TN}{TN + FN} \times 100 (\%) \quad (3.5)$$

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN)(TP + FP)} \times \sqrt{(TN + FN)(TN + FP)}} \times 100 (\%) \quad (3.6)$$

onde:

TP (verdadeiro positivo) é a quantidade de pares focais identificados de maneira correta,

TN (verdadeiro negativo) a quantidade de pares não focais identificados de maneira correta,

FN (falso negativo) é o número de pares focais identificados erroneamente,

FP (falso negativo) é o número de pares não focais identificados erroneamente.

3.3. PARÂMETROS UTILIZADOS

As métricas parâmetros utilizados como dados de entrada para os procedimentos de classificação foram os mesmos utilizadas por Duarte (2018): Nível RMS, Valor Pico, Valor Pico a Pico, Assimetria, Curtose, Fator de Crista, estimador k4 e estimador k6, descritos no item 2.3.8.

Várias ferramentas de análise de sinais foram aplicadas aos dados dos 300 conjuntos utilizados neste trabalho sendo elas: filtragem digital, análise envelope, Transformada Wavelet Contínua, Funções de Modos Intrínsecos, métricas de avaliação de picos de baixa frequência K-NEO,

Transformada de Hilbert-Huang e cálculo de entropia (28 parâmetros aplicados às IMF retirados de Sharma, Pachori e Acharya, 2015). Estas ferramentas foram aplicadas para sinais sem e com normalização (pela média e desvio padrão) para arranjo unipolar e bipolar. A combinação das oito métricas básicas com os diversos tratamentos resultaram em 1180 parâmetros candidatos a servirem de candidatos a sintomas para a classificação dos sinais em focal ou não focal.

É importante ressaltar que nos processos de filtragem foram utilizados filtros Butterworth com seis polos, e as bandas dos filtros foram escolhidas de acordo com os tipos de onda encontradas nos EEG. A seguir tem-se a denominação dos filtros utilizados:

- Total: sinal filtrado com filtro passa baixa em 256 Hz;
- Delta: sinal filtrado com filtro passa baixa em 4 Hz;
- Teta: sinal filtrado com filtro passa banda de 3 a 7 Hz;
- Alfa: sinal filtrado com filtro passa banda de 7 a 13 Hz;
- Beta: sinal filtrado com filtro passa banda de 13 a 30 Hz;
- Gama: sinal filtrado com filtro passa alta de 30 Hz;
- Usuário: sinal filtrado com filtro passa banda de 19,68 a 188 Hz (Duarte, 2018).

A aplicação das oito métricas e os filtros (Total, Delta, Teta, Alfa, Beta, Gama e Usuário) com os pares de sinais unipolares e bipolares normalizados e não normalizados resultaram nos sintomas de 3.1 a 3.224:

- 3.1. Valor eficaz RMS do sinal monopolar não normalizado e sem filtragem
- 3.2. Valor eficaz do sinal monopolar não normalizado e filtro Delta
- 3.3. Valor eficaz do sinal monopolar não normalizado e filtro Teta
- 3.4. Valor eficaz do sinal monopolar não normalizado e filtro Alfa
- 3.5. Valor eficaz do sinal monopolar não normalizado e filtro Beta
- 3.6. Valor eficaz do sinal monopolar não normalizado e filtro Gama
- 3.7-3.12: 3.1 a 3.6 com a métrica Valor Pico.
- 3.13-3.18: 3.1 a 3.6 com a métrica Valor Pico a Pico.
- 3.19-3.24: 3.1 a 3.6 com a métrica Assimetria.
- 3.25-3.30: 3.1 a 3.6 com a métrica Curtose.
- 3.31-3.36: 3.1 a 3.6 com a métrica K4.
- 3.37-3.42: 3.1 a 3.6 com a métrica Fator de Crista.
- 3.43-3.48: 3.1 a 3.6 com a métrica K6.

- 3.49-3.56: métricas RMS, Pico, Pico a Pico, Assimetria, Curtose, K4, Fator de Crista e K6 aplicadas aos sinais monopulares não normalizados e filtro na banda 19,68 a 188 Hz.
- 3.57-3.112: 3.1 a 3.56 para sinais monopulares normalizados.
- 3.113-3.168: 3.1 a 3.56 para arranjos bipolares não normalizados.
- 3.169-3.224: 3.1 a 3.56 para arranjos bipolares normalizados.

A aplicação da técnica de envelope (módulo) nos sinais normalizados unipolares e bipolares, na sequência 3.1-3.56, resultaram nos sintomas de 3.225 a 3.336.

Os parâmetros de 3.337 a 3.416 foram obtidos utilizando a transformada de wavelet WT contínua e as oito métricas aplicadas nos sinais unipolares e bipolares normalizados. As escalas foram escolhidas para contemplar as faixas de frequência delta, teta, alfa, beta e gama. Para cada faixa de frequências, um procedimento de algoritmos genéticos foi utilizado para determinar os valores de escala que maximizasse a separação entre as médias das métricas utilizadas para dois conjuntos de dados focal X_{focal} e não focal X_{nfocal} com 50 sinais cada. A função objetivo a maximizar é apresentada na Eq. 3.7 (teste t para populações com mesma variância), onde: \bar{X}_{Focal} é a média amostral do conjunto focal, \bar{X}_{nFocal} é a média amostral do conjunto não focal, S^2_{Focal} é a variância amostral do conjunto focal e S^2_{nFocal} é a variância amostral do conjunto não focal.

$$Obj = \max\left(\frac{\|\bar{X}_{Focal} - \bar{X}_{nFocal}\|}{S^2_{Focal} + S^2_{nFocal}}\right) \quad (3.7)$$

Na Equação 3.7, assume-se a hipótese de que as populações possuem variância constante, ou seja, são homocedásticas.

A seguir tem-se os parâmetros de 3.337 a 3.416 para a faixa de frequência gama:

- 3.337: Valor eficaz RMS da db4 do sinal monopolar normalizado escala 160,84473.
- 3.338: Valor Pico da db4 do sinal monopolar normalizado escala 91,451092.
- 3.339: Valor Pico a Pico da db4 do sinal monopolar normalizado escala 90,500382.
- 3.340: Valor da Assimetria da db4 do sinal monopolar normalizado escala 120,62891.
- 3.341: Valor da Curtose da db4 do sinal monopolar normalizado escala 329,74694.
- 3.342: Valor do K4 da db4 do sinal monopolar normalizado escala 91,427468.
- 3.343: Valor do Fator de Crista da db4 do sinal monopolar normalizado escala 348,27371.

- 3.344: Valor do K6 da db4 do sinal monopolar normalizado escala 327,82598.
- 3.345-3.352: 3.337 a 3.444 para arranjo bipolar normalizado com escalas: 115,30142, 159,5283, 61,81469, 108,15915, 231,15775, 155,41359, 241,0147 e 315,09591.
- 3.353-3.360: 3.337 a 3.444 (faixa teta) com escalas: 118,76358, 87,35949, 83,970987, 73,144005, 56,209172, 76,690788, 83,70937 e 55,188296.
- 3.361-3.368: 3.345 a 3.352 (faixa teta) com escalas: 115,06844, 112,94824, 119,34503, 57,949938, 119,40504, 55,956421, 82,078132 e 120,000000.
- 3.369-3.376: 3.337 a 3.444 (monopolar faixa alfa) com escalas: 28,183701, 30,723399, 29,183817, 31,617307, 31,717204, 28,111775 e 28,706688.
- 3.377-3.384: 3.345 a 3.352 (bipolar faixa alfa) com escalas: 28,706688, 31,849115, 31,017556, 54,067467, 28,528216, 35,501333, 28,809325 e 32,563611.
- 3.385-3.392: 3.337 a 3.444 (monopolar faixa beta) com escalas: 16,260607, 12,476955, 12,705316, 17,472496, 27,596968, 12,512035, 12,891702 e 27,686198.
- 3.393-3.400: 3.345 a 3.352 (bipolar faixa beta) com escalas: 17,608476, 12,418916, 16,270828, 25,986877, 12,377078, 16,307959 e 27,207219.
- 3.401-3.408: 3.337 a 3.444 (monopolar faixa gama) com escalas: 12,000972, 2,1928478, 2,3170607, 4,4101201, 12,052171, 4,6424592, 11,425589 e 1,5025539.
- 3.409-3.416: 3.345 a 3.352 (bipolar faixa gama) com escalas: 12,032615, 1,6339708, 1,643376, 2,4351534, 1,7466053, 3,3060348, 12,081792 e 1,6621105.

O módulo dos envelopes das WT na sequência 3.337 a 3.416 resultaram nos sintomas de 3.417 a 3.496.

Os sintomas de 3.497 a 3.700 são as oito métricas aplicadas aos modos intrínsecos IMF, e englobam o índice de ortogonalidade, os seis primeiros modos intrínsecos e respectivos resíduos, tanto para os sinais unipolares quanto para os sinais bipolares. A sequência consiste em:

- 3.497. Valor do índice de ortogonalidade da EMD do sinal monopolar normalizado.
- 3.498. Valor RMS do modo 1 da EMD do sinal monopolar normalizado.
- 3.499. Valor Pico do modo 1 da EMD do sinal monopolar normalizado.
- 3.500. Valor Pico-Pico do modo 1 da EMD do sinal monopolar normalizado.
- 3.501. Valor de assimetria do modo 1 da EMD do sinal monopolar normalizado.
- 3.502. Valor da curtose do modo 1 da EMD do sinal monopolar normalizado.
- 3.503. Valor do K4 do modo 1 da EMD do sinal monopolar normalizado.
- 3.504. Valor do Fator de Crista do modo 1 da EMD do sinal monopolar normalizado.

- 3.505. Valor K6 do modo 1 da EMD do sinal monopolar normalizado.
- 3.506-3.513. 3.499 a 3.505 aplicados ao resíduo da EMD do primeiro modo do sinal monopolar normalizado.
- 3.514-3.530. 3.497 a 3.513 aplicados aos sinais bipolares normalizados.
- 3.531-3.564. 3.497 a 3.530 para o segundo modo intrínseco.
- 3.565-3.598. 3.497 a 3.530 para o terceiro modo intrínseco.
- 3.499 a 3.632. 3.497 a 3.530 para o quarto modo intrínseco.
- 3.633 a 3.666. 3.497 a 3.530 para o quinto modo intrínseco.
- 3.667 a 3.700. 3.497 a 3.530 para o sexto modo intrínseco.

Os sintomas de 3.701 a 3.892 são resultantes da aplicação da Transformada de Hilbert Huang HHT nas seis primeiras IMF, e respectivos resíduos, para os sinais unipolares e bipolares, na mesma sequência de 3.497 a 3.700.

Os 13 parâmetros do K-NEO foram calculados para um limite de 2 segundos e larguras de pico variando entre 20 a 65 milissegundos, com um espaçamento de 5 milissegundos. Para cada par de sinal, foram identificados no máximo dez picos. A utilização dos valores das métricas máximas e médias de cada par resulta em 260 sintomas:

- 3.893-3.905 – valores máximos de Dur_{AP} , Dur_{PB} , Dur_{Spike} , $Dur_{slowwave}$, Apm_{AP} , Amp_{PB} , Amp_{Spike} , $Amp_{Slowwave}$, $Slope_{AP}$, $Slope_{PB}$, $Slope_{sharpness}$, $Area_{spike}$ e $Area_{slowwave}$ de no máximo dez pico identificados para o k correspondente a uma largura de pico de 20 ms.
- 3.906-3.918 – valores médio de Dur_{AP} , Dur_{PB} , Dur_{Spike} , $Dur_{slowwave}$, Apm_{AP} , Amp_{PB} , Amp_{Spike} , $Amp_{Slowwave}$, $Slope_{AP}$, $Slope_{PB}$, $Slope_{sharpness}$, $Area_{spike}$ e $Area_{slowwave}$ de no máximo dez pico identificados para o k correspondente a uma largura de pico de 20 ms.
- 3.919-3.944. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 30 ms.
- 3.945-3.970. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 40 ms.
- 3.971-3.996. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 50 ms.
- 3.997-3.1022. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 25 ms.

- 3.1023-3.1048. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 35 ms.
- 3.1049-3.1074. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 45 ms.
- 3.1075-3.1100. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 55 ms.
- 3.1101-3.1126. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 60 ms.
- 3.1127-3.1152. métricas de 3.893 a 3.918 para o k correspondente a uma largura de pico de 65 ms.

Por fim, os valores de entropia aplicados às seis primeiras IMF e seus respectivos resíduos, resultaram nos sintomas de 3.1153 a 3.1180:

- 3.1153-3.1156. valores de entropia RenEn, ShEn, ApEn e SampEn calculadas para o primeiro modo (EMD) unipolar normalizado.
- 3.1157-3.1160, métricas de 3.1153 a 3.1156 para o segundo modo intrínseco.
- 3.1161-3.1164, métricas de 3.1153 a 3.1156 para o terceiro modo intrínseco.
- 3.1165-3.1168, métricas de 3.1153 a 3.1156 para o quarto modo intrínseco.
- 3.1169-3.1172, métricas de 3.1153 a 3.1156 para o quinto modo intrínseco.
- 3.1173-3.1176, métricas de 3.1153 a 3.1156 para o sexto modo intrínseco.
- 3.1177-3.1180, métricas de 3.1153 a 3.1156 para o resíduo do sexto modo intrínseco.

3.4. PROCEDIMENTOS DE CLASSIFICAÇÃO UTILIZADOS

Os procedimentos de classificação utilizados são o QDA, em função dos resultados obtidos por Duarte (2018) e o SVM que é amplamente utilizado em trabalhos envolvendo detecção de anomalias em EEG (Li et al,2014). Serão analisadas as performances dos SVM lineares e com kernel polinomial, gaussiano e sigmoidal. A possibilidade de se utilizar SVM não supervisionada também será estudada. Os dois procedimentos foram descritos no Capítulo II.

3.5. ANÁLISES DISCRIMINANTES

Para a definição dos melhores parâmetros para serem utilizados nos procedimentos de classificação serão utilizados os otimizadores multiobjetivo MOPSO, PESAI, SPEA2 e NSGA-II. Dado como entrada o número de parâmetros utilizados pelo procedimento de classificação, o otimizador escolherá entre os 1180 candidatos as combinações que pertencem à curva de Pareto. o “Tomador de Decisão” é feito a posteriori.

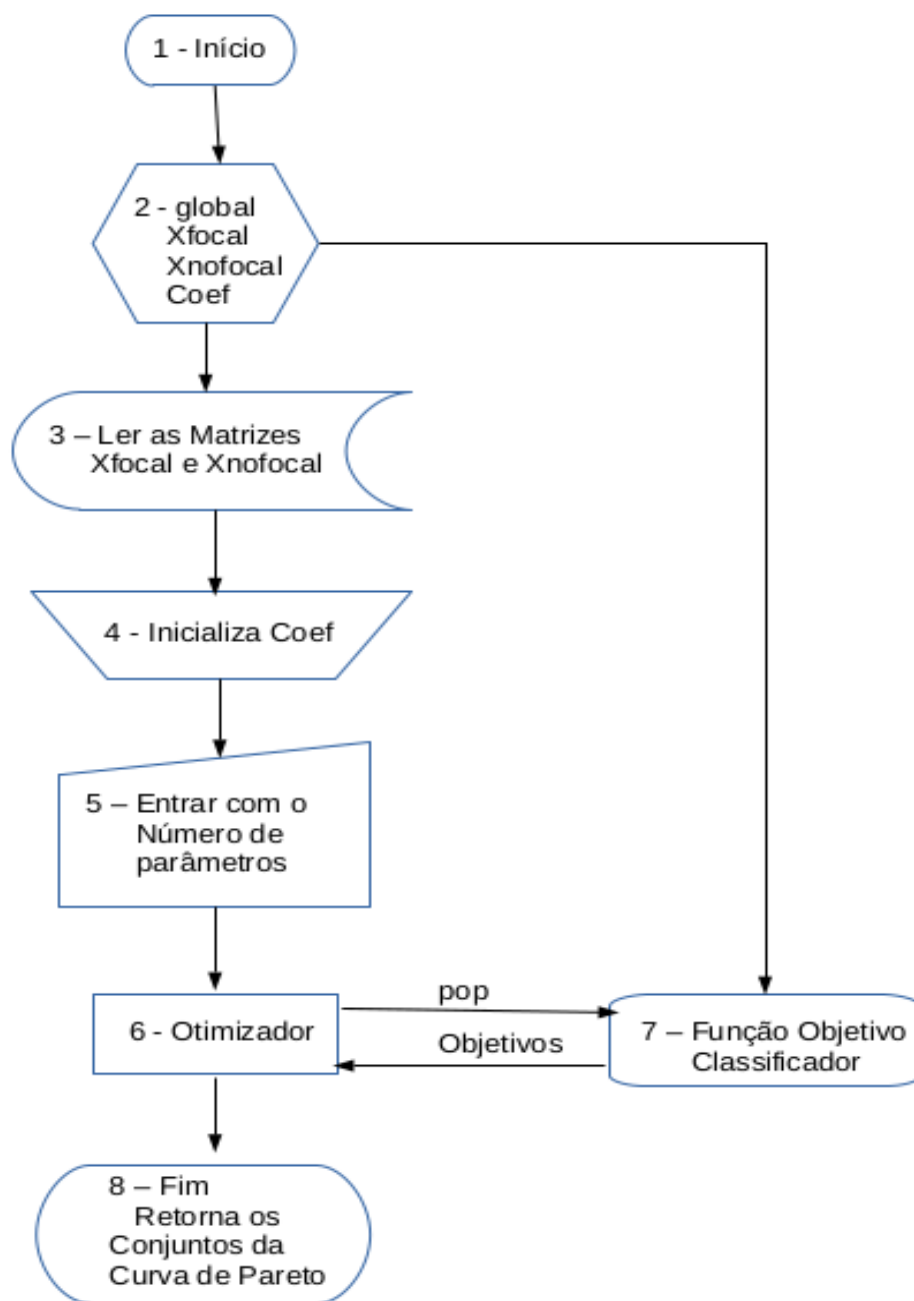
3.6. CUSTO COMPUTACIONAL

Utilizou-se uma métrica para estimar o custo computacional dos procedimentos adotados neste trabalho. Para isto, o tempo médio total para avaliar dez segundos de sinal é calculado. Apesar de variar de máquina para máquina, optou-se por utilizar o tempo médio em segundos por ser uma medida mais realista. Define-se tempo médio total como sendo o tempo médio gasto para o cálculo do sintoma de maior custo computacional somado ao tempo médio gasto para classificação. Utilizou-se apenas o tempo médio do sintoma com maior custo computacional porque o processo de cálculo dos sintomas é facilmente paralelizado.

3.7. FLUXOGRAMA DA METODOLOGIA UTILIZADA

Escolhido o procedimento de classificação, o algoritmo de otimização e as funções objetivo para fins de descriminalização, o fluxograma utilizado no desenvolvimento da metodologia de trabalho proposta, a fim de obter o melhor conjunto de sintomas capaz de classificar os pares de sinais em sinais focais e não focais é apresentado na Figura 3.1.

Fig. 3.1. Fluxograma da Metodologia Proposta



Fonte: O autor (2018)

O fluxograma descreve a seguinte sequência de ações:

1. Início: inicialização das variáveis e parâmetros do processo de otimização.
2. Criação das matrizes globais X_{focal} , X_{nfocal} e Coef. As matrizes X_{focal} e X_{nfocal} contêm os valores de todos os 1180 parâmetros para os bancos de dados de treinamento, validação

e teste. Coef é um vetor de inteiros que apontam para a localização dos parâmetros nas matrizes X_{focal} e X_{nfocal} . Em função do tamanho, e dos índices, de Coef pode-se otimizar utilizando um conjunto selecionado de parâmetros dentre as 1180 possibilidades.

3. Leitura das matrizes X_{focal} , X_{nfocal} .
4. Entra-se com o vetor Coef contendo as posições dos sintomas selecionados.
5. Entra-se com o número de parâmetros N_p que serão utilizados na classificação.
6. O algoritmo de otimização passa para a função objetivo um vetor pop contendo as N_p posições.
7. Uma vez que os parâmetros da otimização são números inteiros (posição de colunas numa matriz), a função objetivo arredonda os valores de pop para o inteiro mais próximo, confere se há parâmetros se repetindo e aciona o procedimento de classificação que vai utilizar N_a conjuntos de sintomas para treinamento e N_t conjuntos de sintomas utilizados para avaliação, através do cálculo de SEN. No caso de sintomas repetidos, a penalização se dá pela diminuição do número de sintomas utilizados no processo de classificação. Deste modo, é possível que se tenha uma população com sintomas repetidos. A rotina que calcula os objetivos utiliza um conjunto de dados focal e não focal para treinar o classificador e um conjunto de dados de validação para calcular os valores dos Objetivos.
8. O processo de otimização termina quando se atinge o número máximo de iterações definidos pelo usuário.

Finalizado o processo de otimização, utiliza-se um conjunto de dados para o teste do procedimento de descriminalização.

3.8. ROTINAS DE PROGRAMAÇÃO E PARÂMETROS UTILIZADOS

Todas as rotinas e procedimentos foram desenvolvidos no Laboratório Matricial Matlab, que é um programa científico para programação numérica, multiplataforma e que possui uma linguagem de programação de alto nível dedicado às aplicações científicas e de engenharia.

Para o cálculo dos sintomas foram utilizadas funções nativas e funções desenvolvida pelo autor.

Os otimizadores multiobjetivo, scripts matlab, foram baixados de <http://yarpiz.com> em 04 de abril de 2017.

Em todos procedimentos de otimização utilizou-se um número máximo de iterações igual a 200 e tamanho da população inicial igual a 200. Os demais parâmetros foram os padrões dos otimizadores.

CAPÍTULO IV

4. APRESENTAÇÃO E ANÁLISE DOS RESULTADOS OBTIDOS

Neste capítulo, são apresentados os resultados obtidos utilizando as rotinas de otimização multiobjetivo apresentadas no Capítulo 3. A performance destas rotinas para diversas opções de funções objetivo serão analisadas para os classificadores QDA e SVM visando a escolha dos melhores sintomas que minimizam a função objetivo, associados aos procedimentos de classificação.

Para as análises, o banco de dados composto por parâmetros de 350 conjuntos de dados focal/não focal serão divididos em:

1. 50 conjuntos utilizados para o treinamento dos classificadores;
2. 200 conjuntos utilizados para a validação;
3. 100 conjuntos utilizados para o teste.

A seguir será apresentado um resumo dos mais de 5000 procedimentos de otimização realizados neste trabalho.

4.1. PERFORMANCE DOS OTIMIZADORES.

Para avaliar a performance dos quatro procedimentos de otimização estudados foram utilizadas duas funções objetivo com o SVM para classificar os sinais em focal e não focal.

As funções objetivo a serem maximizadas foram a Sensibilidade (SEN) e a Especificidade (SPF), que possuem objetivos excludentes: uma maximiza a porcentagem de acertos de EEGs com anomalias (SEN) e, em contrapartida, a outra maximiza a porcentagem de acertos de EEGs normais (SPF). A tomada de decisão foi feita a posteriori, considerando a melhor solução aquela que maximiza a porcentagem de acertos de sinais focal dos conjuntos de dados de validação e teste, tendo como restrição valores de SPF maiores que 60% para ambos os bancos de dados. Em

resumo, o objetivo é minimizar a ocorrência de falsos negativos na análise aceitando até 40% de falsos positivos, ou seja: é preferível que o especialista tenha que conferir um EEG normal classificado como focal do que aceitar um EEG erroneamente classificado como normal.

Na classificação foi utilizado o SVM com a opção classificador multiclasse, com regularizador **nu**, e kernel função radial base ϕ_{RBF} mostrada na Equação 4.1, onde os valores de w são ajustados pelo procedimento SVM. As constantes nu (entre 0 e 1) e γ entram como parâmetros a serem otimizados com mais seis parâmetros de análise de sinais candidatos a sintomas para detecção de EEGs focal.

$$\psi_{RBF} = e^{\gamma |w-x|^2} \quad (4.1)$$

4.1. INFLUÊNCIA DA NORMALIZAÇÃO DOS DADOS PARA O SVM.

Para esta análise será utilizado o kernel função de base radial RBF com regularização do tipo Nu [0 1] e as funções objetivo conflitantes Sensibilidade (SEN) e Especificidade (SPF), para otimização de seis parâmetros e os coeficientes Nu e gama. A escolha de seis parâmetros se justifica pelos resultados obtidos por Duarte (2018). Os valores limites para Nu e gama foram [0 1] e [0 5], respectivamente. Os parâmetros independentes são valores inteiros [1 180]. Uma vez que os otimizadores trabalham com parâmetros reais, um fator igual a $5N_{Rep}$ foi somado às funções objetivo com o intuito de penalizar as N_{Rep} repetições que possam ocorrer na conversão de real para inteiro, conforme Equação 4.2.

$$Obj_i(p_1, \dots, p_{6, Nu}, gama) = Obj_i(p_1, \dots, p_{6-N_{Rep}}, Nu, gama) + 5 * N_{Rep} \quad (4.2)$$

Uma vez que os otimizadores minimizam as funções objetivo, a maximização de SEN e SPF toma a forma:

$$Obj_1(p) = 100(1 - SEN/100) \quad \text{e} \quad Obj_2(p) = 100(1 - SPF/100)$$

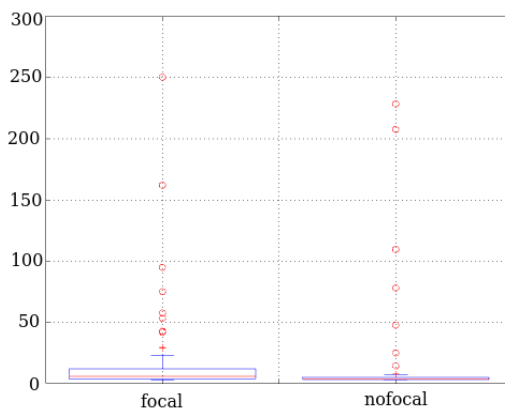
Além dos parâmetros sem nenhuma normalização SN, foram avaliados oito tipos de normalizações:

1. 1S0D: Do valor do parâmetro é subtraído o valor da média amostral e o resultado dividido por um desvio padrão amostral. As médias e desvios padrão amostrais foram calculados usando o banco de dados dos parâmetros focal.
2. 1S6D: Os valores dos parâmetros foram normalizados por um desvio padrão amostral do conjunto não focal e limitados a seis desvios padrão.
3. 2S0D: Os valores dos parâmetros foram normalizados por duas vezes o desvio padrão amostral do conjunto não focal.
4. 2S6D: Os valores dos parâmetros foram normalizados por duas vezes o desvio padrão amostral do conjunto não focal e limitados a seis desvios padrão.
5. 3S0D: Os valores dos parâmetros foram normalizados por três vezes o desvio padrão amostral do conjunto não focal.
6. 3S6D: Os valores dos parâmetros foram normalizados por três vezes o desvio padrão amostral do conjunto não focal e limitados a seis desvios padrão.
7. 6S0D: Os valores dos parâmetros foram normalizados por seis vezes o desvio padrão amostral do conjunto não focal.
8. 6S6D: Os valores dos parâmetros foram normalizados por seis vezes o desvio padrão amostral do conjunto não focal e limitados a seis desvios padrão.

A experiência com limitação em seis desvios padrão foi testada em função da grande dispersão dos parâmetros testados, principalmente para a curtose, o K4 e o K6. Como exemplo, a Figura 4.1 mostra os diagramas de caixas para os conjuntos focal e não focal do conjunto de treinamento para o parâmetro 3.53 (valores de curtose dos sinais filtrados na banda otimizada), onde observa-se valores de parâmetros bem afastados da mediana (*outliers*).

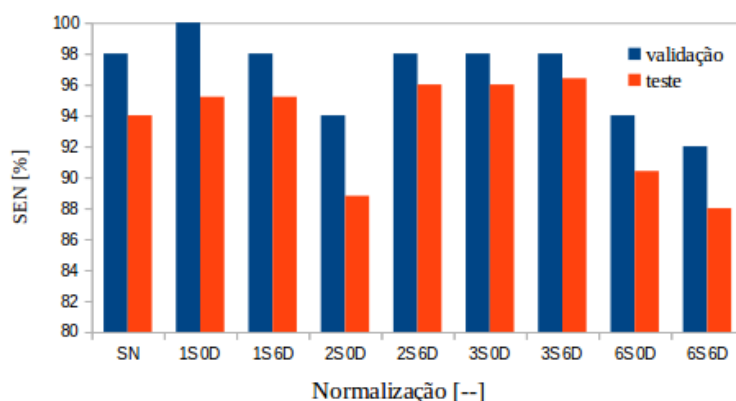
Já, a Figura 4.2 mostra, para o otimizador NSGA-II, os valores de sensibilidade SEN para o conjunto de validação (azul) e teste (vermelho), respectivamente. Observa-se na figura que a normalização 1SD0 resultou em 100% de acerto para a sensibilidade do grupo de validação e 95,2% para o grupo de teste. O maior índice de acerto para o grupo de teste (96,4%) foi para a normalização 3S6D com um índice de acerto de 98% na validação.

Fig. 4.1. Diagrama de caixa para a curtose dos sinais filtrados na banda otimizada.



Fonte: O autor, 2018.

Fig. 4.2. Valores de sensibilidade em função do tipo de normalização para o otimizador NSGA-II.



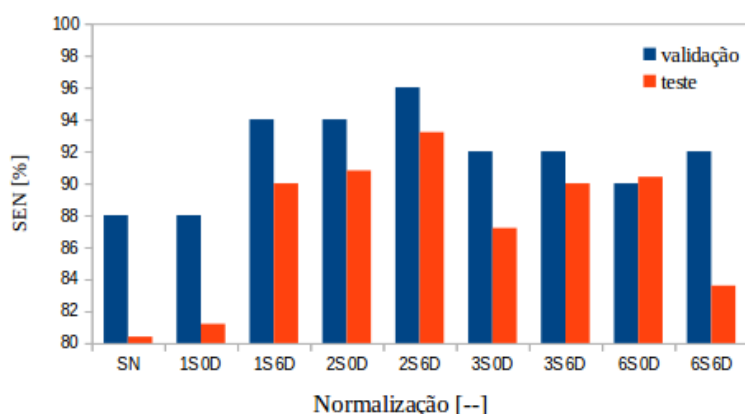
Fonte: O autor, 2018.

Utilizando o SPEA2, Figura 4.3, o melhor resultado para o conjunto de validação foi utilizando a normalização 1S0D que resultou numa porcentagem de acertos de 98% e 94,4% para o conjunto de teste. O maior índice de acerto para o grupo de teste (96,4%) foi para a normalização 3S6D que resultou num índice de acerto de 98% para o grupo de validação, ou seja: os mesmos valores obtidos utilizando o NSGA-II.

Para o PESA2, Figura 4.4, os melhores resultados foram obtidos com a normalização 3S6D: uma sensibilidade de 96% para o conjunto de validação e 94% para o conjunto de teste. O índice de

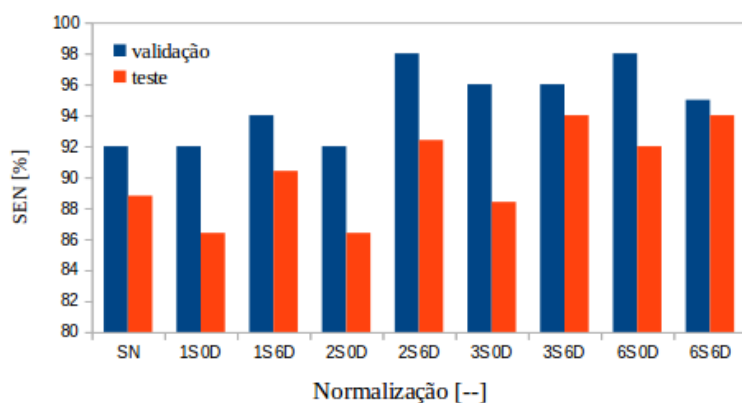
acerto de 98% alcançados com as normalizações 2S6D e 6S0D para o conjunto de validação, implicaram índices de acerto iguais a 92,4 e 92% para o conjunto de teste.

Fig. 4.3. Valores de sensibilidade em função do tipo de normalização para o otimizador SPEA2.



Fonte: O autor, 2018.

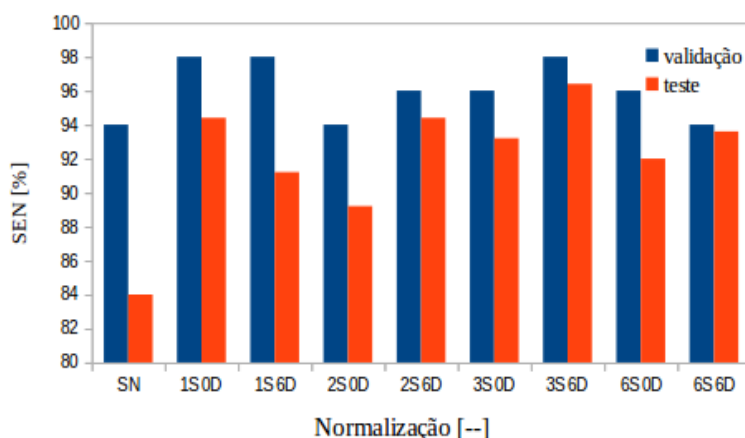
Fig. 4.4. Valores de sensibilidade em função do tipo de normalização para o otimizador PESA2.



Fonte: O autor, 2018.

Os piores resultados foram obtidos utilizando o otimizador MOPSO (Figura 4.5), sendo que os melhores resultados para este otimizador foram obtidos com a normalização 3S6D (92 e 91,6% de acerto para o conjunto de validação e de teste, respectivamente).

Fig. 4.5. Valores de sensibilidade em função do tipo de normalização para o otimizador MOPSO.



Fonte: O autor, 2018.

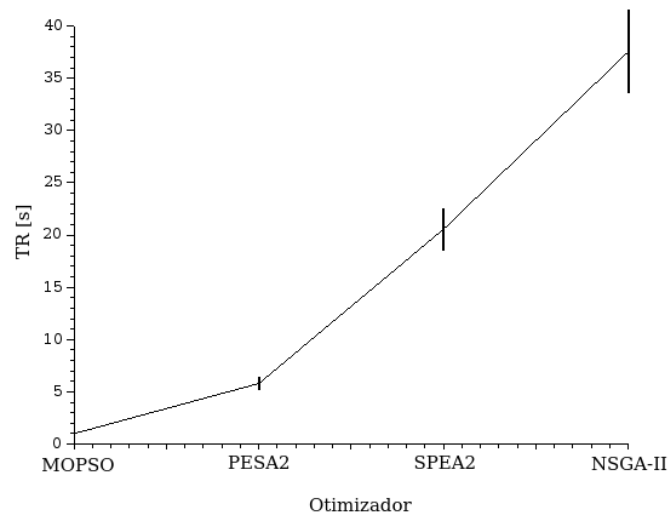
Concluindo, os melhores resultados globais foram obtidos com os otimizadores NSGA-II e SPEA2 com normalização 3S6D. Para o PESA2, os melhores resultados também foram alcançados com a normalização 3S6D. Conclui-se, assim, que a utilização da limitação em 6 desvios padrão é válida,

4.2. TEMPO PARA CONVERGÊNCIA DO OTIMIZADOR

Para avaliar o tempo médio de convergência dos quatro otimizadores foram realizados 45 procedimentos de otimização para a mesma configuração do subitem 4.1. Uma vez que foram utilizados computadores diferentes no estudo, optou-se por analisar um tempo para convergência relativo ao procedimento de MOPSO que foi o otimizador que convergiu mais rápido para todos os casos testados. Para evitar a influência da carga de processamento nos tempos de convergência, os computadores sempre trabalhavam com os quatro otimizadores operando em paralelo. O número de avaliações da função objetivo, por otimização, foi igual a 88.401, 80.000 e 80.001 para os otimizadores NSGA-II, SPEA2 e PESA2. Para o otimizador MOPSO o número de avaliações até a convergência foi variável com média igual a 87.821,23 e desvio padrão igual a 34,1.

Na Figura 4.6 é apresentado o gráfico dos tempos relativos TR com barras de erro verticais iguais a 1 desvio padrão. Deve-se constar que o tempo médio de convergência do MOPSO foi de 324,2 segundos com um desvio padrão de 36,5 segundos.

Fig. 4.6. Tempos relativos para convergência dos otimizadores PESA2, SPEA2 e NSGA-II com relação ao otimizador MOPSO.



Fonte: O autor, 2018.

Analisando a figura, uma vez que o número de avaliações das funções objetivo foi aproximadamente igual para todos os otimizadores (diferença máxima de 1%), pode-se creditar as diferenças de tempo de convergência aos procedimentos de otimização. Para o problema estudado têm-se que os procedimentos PESA2, SPEA2 e NSGA-II levam, em média, 5,79, 20,56 e 37,52 vezes mais tempo para a convergência do que o MOPSO.

4.3. FRENTE de PARETO

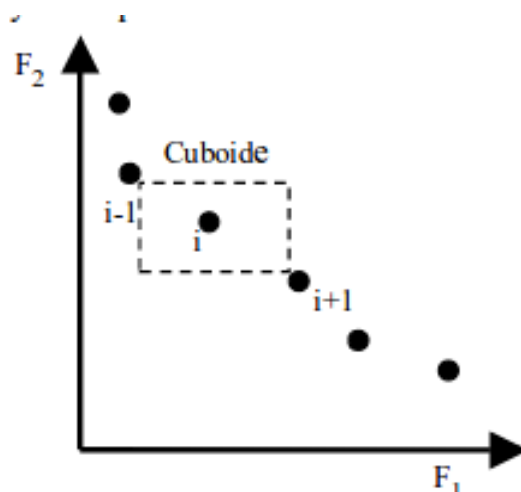
Comparar o desempenho de métodos heurísticos de busca é sempre uma tarefa difícil uma vez que o desempenho é considerado um atributo multidimensional envolvendo a qualidade da solução, esforço computacional e robustez, dentre outros (Wanner e Takahashi, 2007). É comum utilizar os gráficos das soluções não dominadas para indicar o desempenho de qualquer algoritmo multiobjetivo, o que torna o processo de comparação muito subjetivo. Segundo os autores, um parâmetro objetivo para mesurar a qualidade do conjunto Pareto ótimo deve englobar três aspectos:

1. O número de soluções Pareto ótimas no conjunto de estimativas.
2. A proximidade das soluções obtidas em relação ao verdadeiro conjunto de Pareto.
3. A distribuição e o espalhamento das soluções.

Para as funções teste, que se conhece o conjunto de Pareto real, as métricas populares para cálculo de desempenho são a Distância Geracional e Erro Máximo da Frente de Pareto (Coello, 1999), por exemplo. No caso de problemas de otimização reais, uma medida de desempenho amplamente utilizada é a *S Metric* (Wanner e Takahashi, 2007).

O número de soluções Pareto ótimas no conjunto das estimativas é bem fácil de calcular. A métrica distância das aglomerações *crowding distance* fornece uma estimativa da densidade de soluções em torno desta solução. A Figura 4.7 mostra o significado da *crowding distance* da solução i , ou seja, é a dimensão do maior cuboide ao redor da solução i que não inclua nenhuma outra solução do conjunto de soluções não dominadas. O algoritmo para o cálculo das distâncias das aglomerações pode ser encontrado no trabalho de Raquel e Prospero Jr. (2005).

Fig. 4.7. Exemplo de cálculo das distâncias das aglomerações.



Fonte – Raquel e Prospero Jr., 2005.

Na Tabela 4.1 são apresentados os valores da distância média das aglomerações DMA e o número médio de soluções não dominadas NSND, resultante da utilização dos quatro estimadores em 43 problemas de otimização multiobjetivo. Para fins de análise, os desvios padrão amostrais também são apresentados na mesma tabela.

Tab. 4.1. Valores de DMA e NSND, e respectivos desvios padrão amostrais, para o quatro otimizadores estudados.

Otimizador	DMA	σ_{DMA}	NSND	σ_{NSND}
MOPSO	6,75	1,91	28,80	6,50
PESA2	4,07	0,90	42,98	9,65
SPEA2	4,41	1,34	39,56	10,88
NSGA-II	2,39	0,80	64,23	18,51

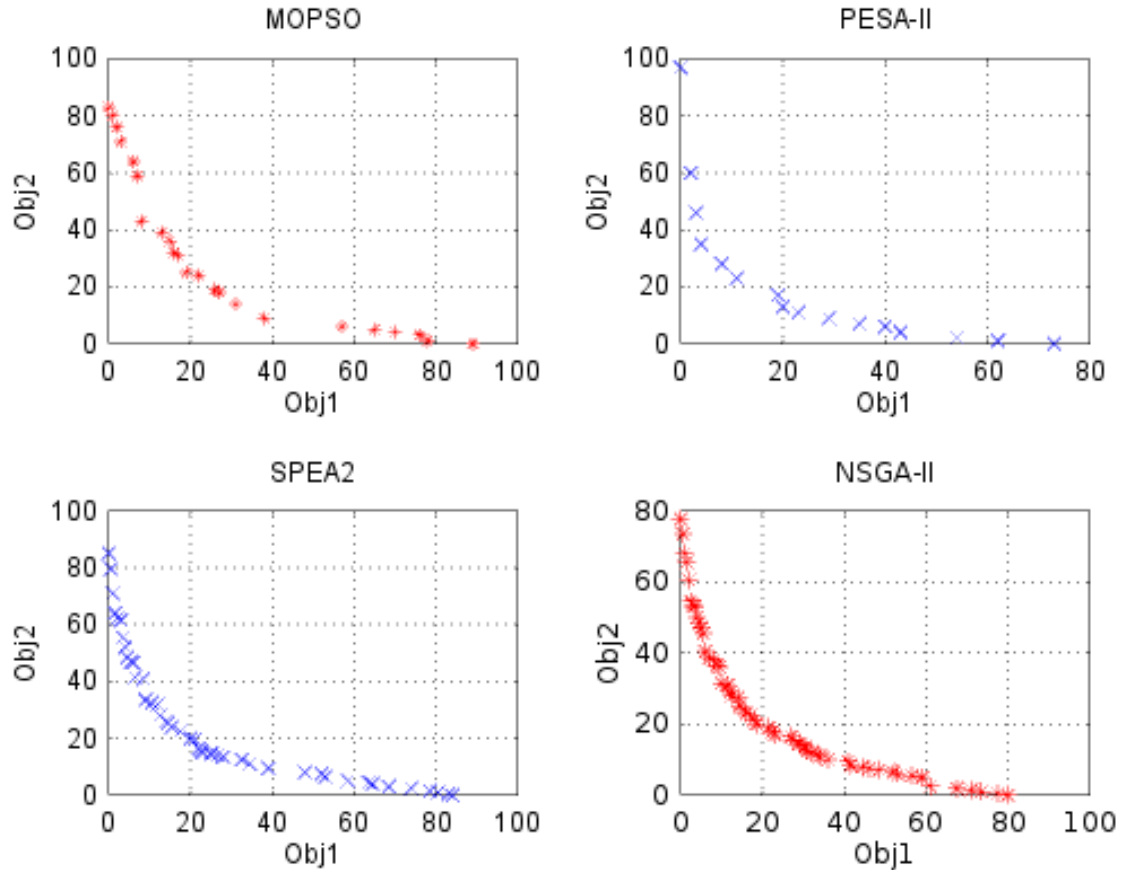
Da análise dos gráficos do subitem 4.1 e Tabela 4.1, observa-se:

- Através de um teste t, estima-se que estatisticamente, os valores de DMA e NSND dos otimizadores PESA2 e SPEA2 pertencem à mesma população, ou seja: podem ser considerados iguais.
- A correlação entre os valores médios de DMA e NSND e os resultados do subitem 4.1 é direta: quanto menor a distância média das aglomerações, e maior o número médio de solução não dominadas, melhores foram os resultados almejados neste trabalho.
- Os melhores resultados foram alcançados com o otimizador NSGA-II e os piores com o MOPSO.
- Os otimizadores SPEA2 e PESA2 resultaram em funções objetivo da mesma ordem de grandeza, porém o otimizador PESA2 converge aproximadamente quatro vezes mais rápido do que o SPEA2.

Uma vez que se trabalhou com duas funções objetivo, as conclusões supracitadas já eram esperadas a partir da observação dos gráficos das soluções não dominadas. Como exemplo, a Figura 4.8 mostra os gráficos das soluções não dominadas para os quatro otimizadores com a melhor normalização de cada um.

Observa-se na figura que a frente de Pareto do otimizador NSGA-II possui mais soluções quase que uniformemente espaçadas.

Fig. 4.8. Exemplo de gráficos das soluções não dominadas.



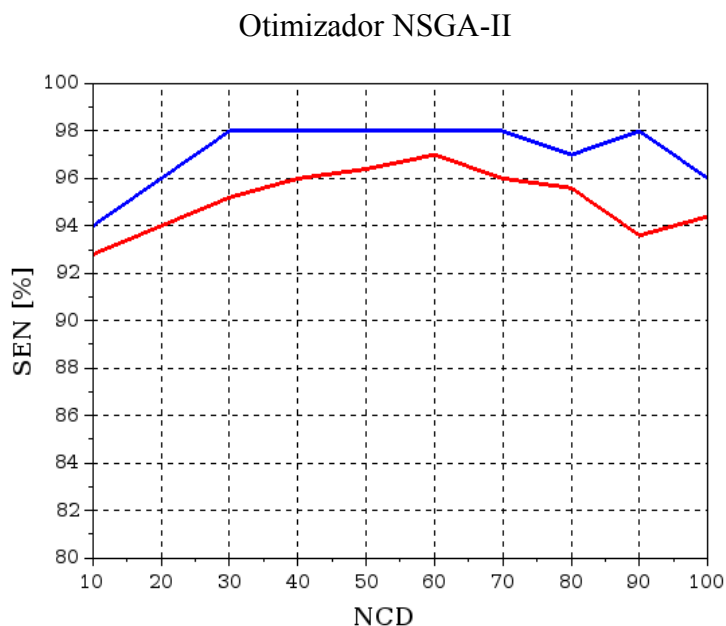
Fonte: O autor, 2018

4.4. QUANTIDADE DE DADOS NO CONJUNTO DE TREINAMENTO

É sabido que um dos fatores que influenciam na precisão do SVM supervisionado é o tamanho do conjunto de dados utilizados para o ajuste de seus parâmetros na fase de treinamento (Rodríguez-Pérez, Vogt e Bajorath, 2017). Para avaliar esta influência, neste trabalho realizou-se um experimento onde o tamanho do conjunto de dados NCD variou de 10 até 100 com um incremento de 10. Deve-se ressaltar que um conjunto de dados é composto do mesmo número de parâmetros calculados para EEGs classificados como focal e não focal, ou seja: um NCD igual a 10 implica que o conjunto de treinamento é composto por vinte conjuntos de parâmetros.

Na Figura 4.9 tem-se os valores de sensibilidade SEN, para o conjunto de validação (azul) e teste (vermelho) para o otimizador NSGA-II em função de NCD.

Fig. 4.9. Gráficos SEN vs. NCD para os conjuntos de validação (azul) e teste (vermelho).



Fonte: O autor, 2018.

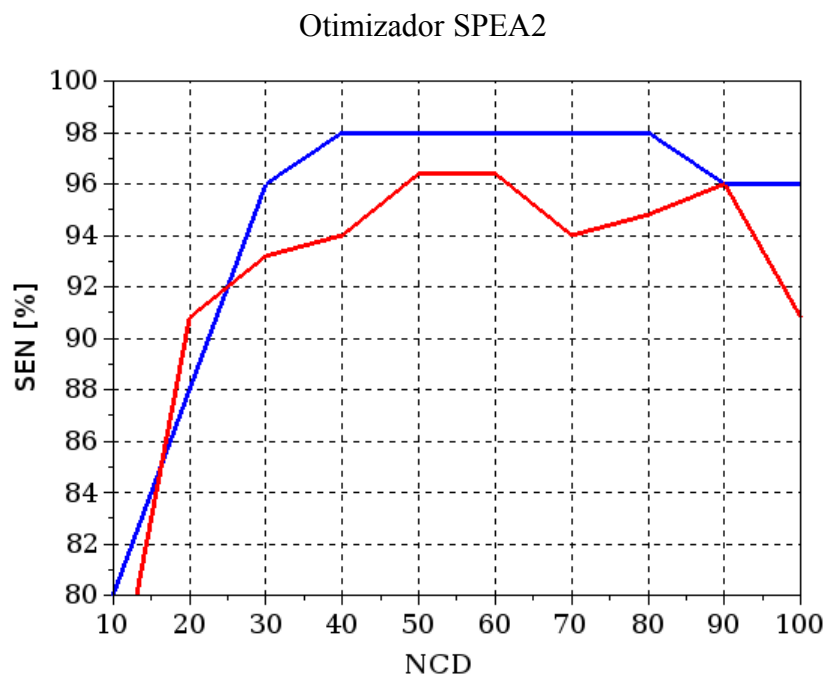
Da análise da Figura 4.9, observa-se que os melhores valores de funções objetivo, máximo de SEN para os conjuntos de validação e teste, foram alcançados para valores de NCD variando entre 40 e 70, com máximos de sensibilidade iguais a 98% para o conjunto de validação e 70% para o conjunto de teste.

As figuras 4.10 e 4.11 mostram os valores de sensibilidade SEN, para o conjunto de validação (azul) e teste (vermelho), dos otimizadores SPEA2 e PESA2 em função de NCD. Os valores resultantes do otimizador MOPSO não serão mais mostrados, uma vez que resultaram em valores menores para a função objetivo.

Para o SPEA2, os melhores valores de SEN para o conjunto de validação (98%) e teste (96,4%) foram obtidos para valores de NCD iguais a 50 e 60.

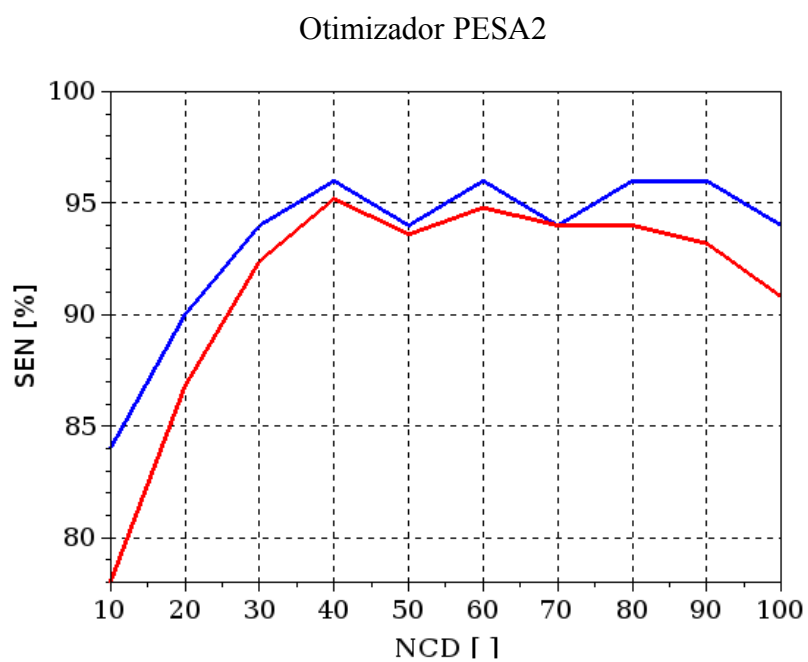
Os resultados para o PESA2 não foram tão bons quanto os observados para o NSGA-II e SPEA2.

Fig. 4.10. Gráficos SEN vs. NCD para os conjuntos de validação (azul) e teste (vermelho).



Fonte: O autor, 2018.

Figura 4.11. Gráficos SEN vs. NCD para os conjuntos de validação (azul) e teste (vermelho).



Fonte: O autor, 2018.

Na Figura 4.11 observa-se que os melhores valores de funções objetivo foram obtidos para valores de NCD variando entre 40 e 80, com máximos de sensibilidade iguais em torno de 96% para o conjunto de validação e 95% para o conjunto de teste.

Resumindo, os melhores resultados foram obtidos com valores de NCD variando entre 40 e 70.

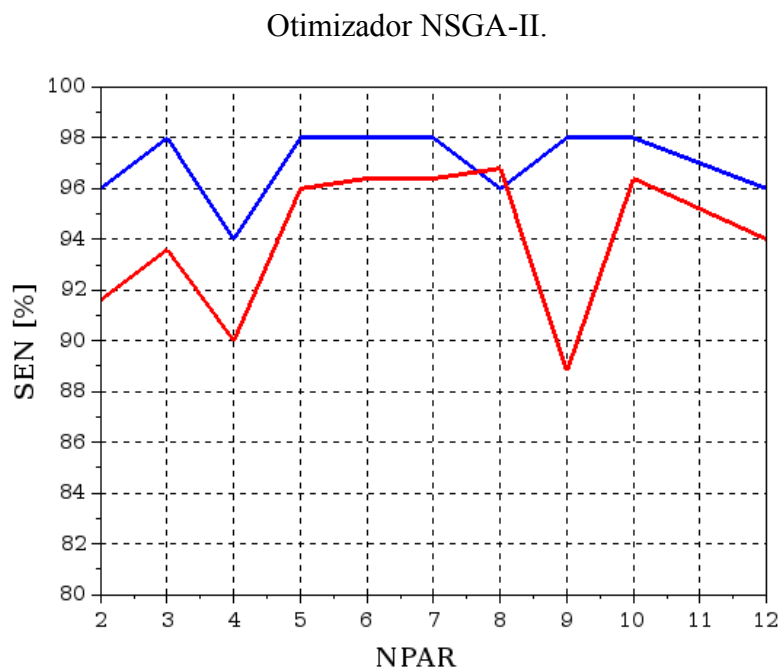
4.5. NÚMERO DE PARÂMETROS

Para avaliar a influência do número de parâmetros NPAR na qualidade da classificação, otimizações com o NSGA-II e o SPEA2 foram realizadas, com a seguinte configuração:

- Normalização 3S6D.
- NCD mantido fixo e igual a 60.
- Número de parâmetros na classificação NPAR variando entre 2 e 12.

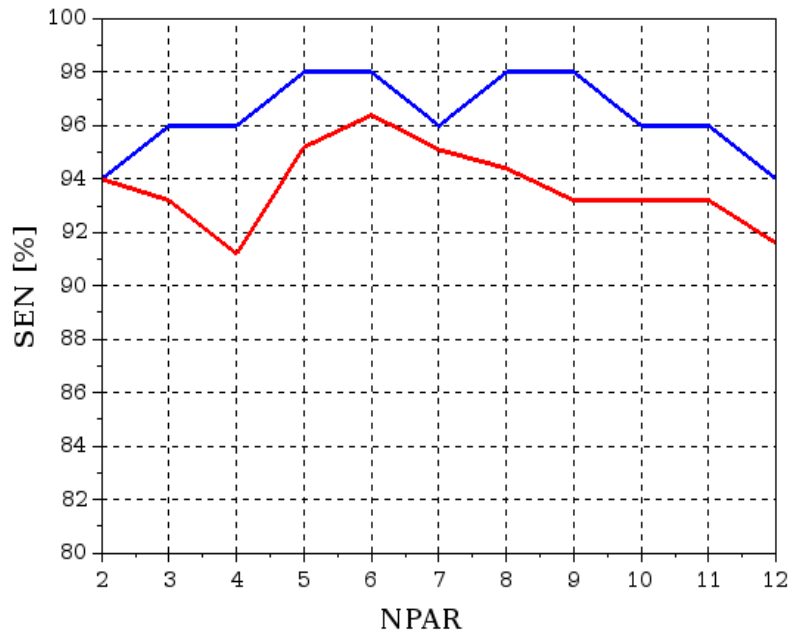
As figuras 4.12 e 4.13 mostram os valores de sensibilidade SEN, para o conjunto de validação (azul) e teste (vermelho), dos otimizadores NSGA-II e SPEA2 em função de NPAR.

Fig. 4.12. Gráficos SEN vs. NPAR para os conjuntos de validação (azul) e teste (vermelho).



Fonte: O autor, 2018.

Fig. 4.13. Gráficos SEN vs. NPAR para os conjuntos de validação (azul) e teste (vermelho).
Otimizador SPEA2.



Fonte: O autor, 2018.

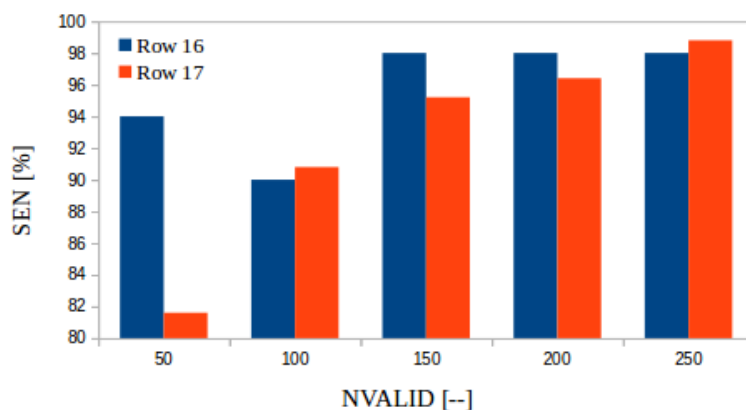
4.7. TAMANHO DO CONJUNTO DE DADOS DE VALIDAÇÃO

Na metodologia desenvolvida neste trabalho, o procedimento de otimização escolhe os melhores conjuntos de parâmetros para construir uma frente de Pareto com os valores das funções objetivo do procedimento de classificação aplicado no conjunto de dados do conjunto de validação. Para avaliar a influência do conjunto de validação na performance da metodologia, realizou-se otimizações com o NSGA-II com a seguinte configuração:

- Normalização 3S6D.
- NCD mantido fixo e igual a 60.
- Número de parâmetros na classificação NPAR igual a 6.
- Número de pares de dados focal/não focal no conjunto de validação NVALID iguais a 50, 100, 150, 200 e 250.

Na Figura 4.14 são apresentados os valores de sensibilidade SEN, para o conjunto de validação (azul) e teste (vermelho) em função de NVALID, onde [--] significa adimensional.

Fig. 4.14. Gráficos SEN vs. NVALID para os conjuntos de validação (azul) e teste (vermelho), utilizando o otimizador NSGA-II com normalização 3S6D.

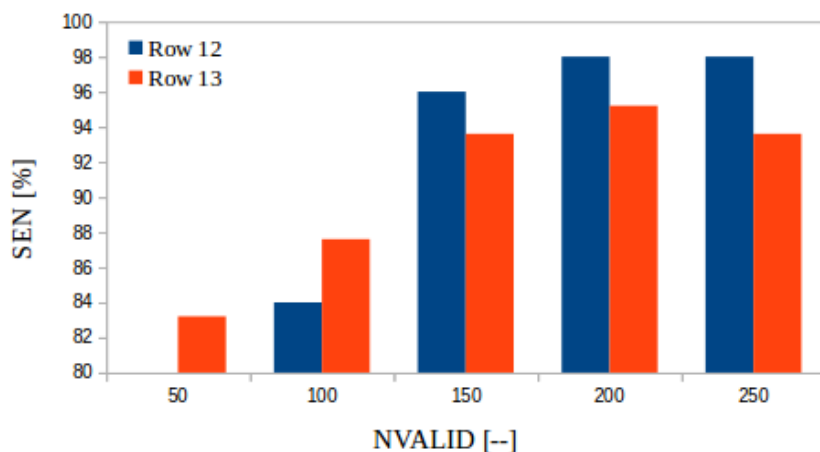


Fonte: O autor, 2018.

Observa-se, na Figura 4.14, que o aumento da sensibilidade SEN está diretamente relacionada com o tamanho do conjunto de dados da validação. Para o banco com os parâmetros de 250 sinais focal/não focal os valores de SEN foram iguais a 98 e 98,8% para os conjuntos de validação e teste, respectivamente.

Para mostrar a importância da normalização no procedimento de classificação, a Figura 4.15 mostra os valores de sensibilidade SEN para o banco de dados sem normalização SN.

Fig. 4.15. Gráficos SEN vs. NVALID para os conjuntos de validação (azul) e teste (vermelho), utilizando o otimizador NSGA-II sem normalização.



Fonte: O autor, 2018.

Ao comparar as figuras 4.14 e 4.15, a importância da normalização dos dados fica bem evidente. Para o banco de dados sem normalização, os valores máximos de sensibilidade acontecem para NVALID igual a 200 (98 e 95,2% para os conjuntos de validação e teste).

4.8. NÚCLEO (KERNEL)

Neste subitem serão apresentados os resultados relativos aos vários kernels estudados neste trabalho. Em função do grande número de simulações realizadas, utilizou-se o otimizador SPEA2 que tem uma boa relação entre o tempo de processamento e a qualidade da solução. Assim como para o kernel base radial, na média, os maiores valores das funções objetivo foram obtidos com a normalização 3S6D, NCD igual a 60 e NVALID igual a 50. Para NPAR, houve muita variação nos resultados em função do kernel utilizado. Além da base radial (Eq. 4.3), os kernels estudados foram:

1. S: sigmoíde (Eq. 4.4).
2. P2: polinomial de grau 2 (Eq. 4.5).
3. P3: polinomial de grau 3 (Eq. 4.5).
4. P4: polinomial de grau 4 (Eq. 4.5).
5. P5: polinomial de grau 5 (Eq. 5.5).

$$\psi_{RBF} = e^{(-\gamma |w-v|^2)} \quad (4.3)$$

$$\psi_S = \text{tgh}(\gamma w'x + \text{coef}_0) \quad (4.4)$$

$$\psi_{Pd} = (\gamma w'x + \text{coef}_0)^d \quad (4.5)$$

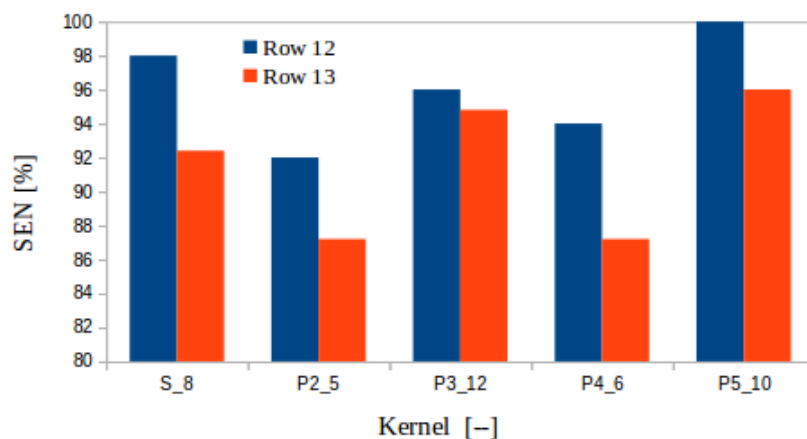
As constantes γ e coef_0 , equações 4.4 e 4.5, entram como parâmetros a serem otimizados, junto a \mathbf{nu} .

Pesquisas com o kernel linear e o SVM de uma classe foram abandonadas em função de resultados não satisfatórios, em comparação com os de base radial (valores de SEN máximos da ordem de 70%). Bastante esforço foi dispendido na SVM de uma classe (treinamento não

supervisionado) em função da perspectiva de se trabalhar apenas com dados não focal no treinamento.

Na Figura 4.16 são mostrados os valores de sensibilidade SEN, para o conjunto de validação (azul) e teste (vermelho) em função do kernel utilizado, sendo o valor de NPAR mostrado depois do sublinhado do símbolo do kernel. Em relação à base radial, o único resultado digno de atenção foi para o kernel P5 com 10 parâmetros que resultou em valores de sensibilidade iguais a 100% para o conjunto de validação e 96% para o conjunto de teste. Para efeito de comparação, é importante lembrar que o melhor resultado com kernel de base radial, otimizador NSGA-II, resultou em valores de SEN iguais a 100 e 96,4% para os conjuntos de validação e teste.

Fig. 4.16. Maiores valores de SEN para as cinco funções kernel estudadas para o Otimizador SPEA2.



Fonte: O autor – 2018

4.9. QDA

Utilizando o mesmo banco de dados, e parâmetros estudados, Duarte (2018) obteve valores da ordem de 90% para a sensibilidade utilizando algoritmo evolucionário com a função acurácia ACC como função mono objetivo a ser maximizada e QDA como classificador. Em função da boa performance do QDA, neste trabalho foram realizadas uma série de testes envolvendo normalização, número de parâmetros NPAR, tamanho do conjunto de validação NVALID e tamanho do conjunto de dados de treinamento NCD. Os melhores valores de sensibilidade (100%

para o conjunto de validação e 95,2% para o conjunto de teste) foram obtidos para normalização 1S0D, NPAR igual a 6, NCD igual a 50 e NVALID igual a 100, com o otimizador NSGA-II. Os melhores valores de sensibilidade (100% para o conjunto de validação e 95,2% para o conjunto de teste) foram obtidos para normalização 1S0D, NPAR igual a 6, NCD igual a 50 e NVALID igual a 100, com o otimizador NSGA-II. Em resumo: o QDA apresentou performance equivalente ao SVM com kernel P5.

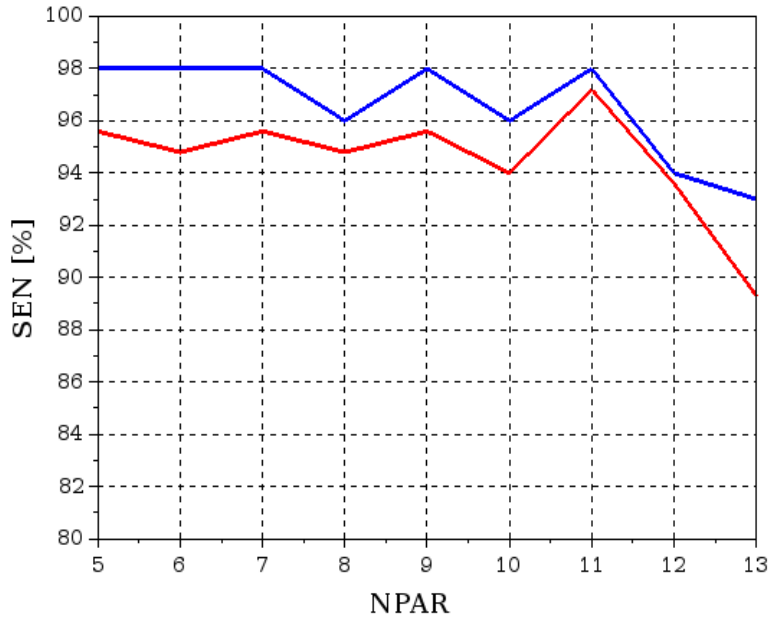
4.10. NÚMERO MAIOR DE FUNÇÕES OBJETIVO

Várias combinações de funções objetivo foram testadas envolvendo sensibilidade SEN, especificidade SPF, precisão ACC, valores positivos preditos PPV e valores negativos preditos NPV. Outras possibilidades testadas foram dividir o conjunto de validação em subconjuntos e maximizar o valor média amostral de SEN e SPF e minimizar os respectivos desvios padrões (otimização robusta). De todas as possibilidades testadas será apresentada a maximização de três funções objetivo: SEN, SPF e ACC, em função dos resultados foram obtidos.

Na Figura 4.17 são mostrados os resultados para a maximização das três funções objetivo em função do número de parâmetros NPAR. Estes resultados foram obtidos para normalização 3S3D, NCD igual a 60 e NVALID igual a 200, com o otimizador NSGA-II.

Observa-se na Figura 4.17 que os maiores valores de SEN (98% na validação e 97.2% no teste) foram para NPAR igual a onze. Observa-se uma piora sensível nos resultados para valores de NPAR maiores do que onze. Um resultado interessante ocorreu para NPAR igual a sete com valores dos conjuntos de validação e teste iguais a 98 e 95,6% para o SEN, 70 e 69,4% para o SPF e 84 e 82% para o ACC. Para os outros valores de NPAR, os valores de SPF e ACC ficaram próximos dos 60% estabelecidos como um dos objetivos do “Tomador de Decisão” a posteriori.

Fig. 4.17. Maiores valores de SEN para a otimização multiobjetivo envolvendo SEN, SPF e ACC com o procedimento de otimização NSGA-II



Fonte: O autor, 2018.

4.11. PENALIZAÇÃO NAS FUNÇÕES OBJETIVO

A ideia de introduzir algum tipo de penalização nas funções objetivo teve como intuito, tentar aumentar os valores de SPF sem diminuir os valores de SEN. As equações 4.6 e 4.7 mostram duas funções objetivo testadas, onde VAL_{pen} é o valor escolhido para a penalização.

$$Obj_1(p) = 1 - SEN/100 + (1 - SPF/100) - VAL_{pen} \quad \text{se} \quad (1 - SPF/100) > VAL_{pen} \quad (4.6)$$

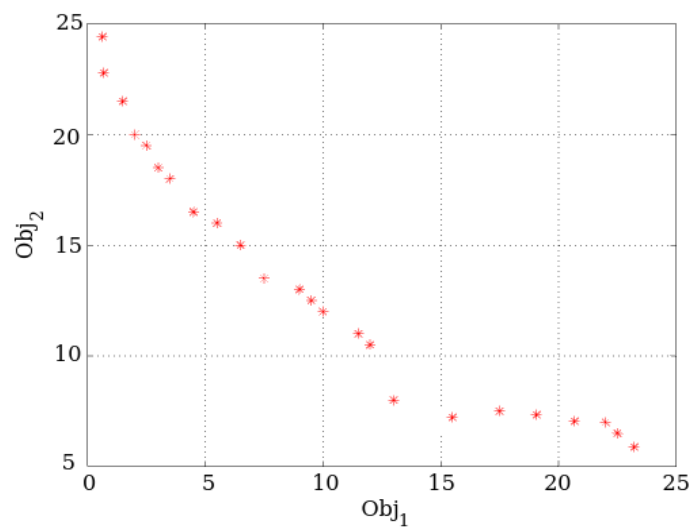
$$Obj_1(p) = 1 - SEN/100 \quad \text{se} \quad (1 - SPF/100) \leq VAL_{pen}$$

$$Obj_2(p) = 1 - SPF/100 + (1 - SEN/100) - VAL_{pen} \quad \text{se} \quad (1 - SEN/100) > VAL_{pen} \quad (4.7)$$

$$Obj_2(p) = 1 - SPF/100 \quad \text{se} \quad (1 - SEN/100) \leq VAL_{pen}$$

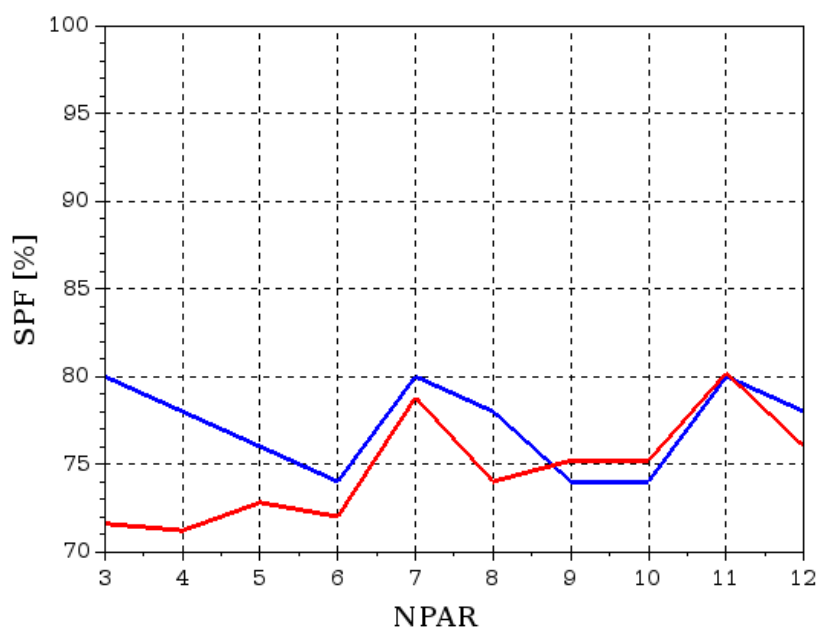
Analisando as duas funções, observa-se que o objetivo é restringir os limites da frente de Pareto, como pode ser visto na Figura 4.18 para VAL_{pen} igual a 0,3 que equivale a um valor mínimo de função objetivo de 70% (frente de Pareto entre zero e trinta).

Fig. 4.18. Frente de Pareto para $VAL_{pen} = 0,3$ para o otimizador NSGA-II



Fonte: O autor, 2018.

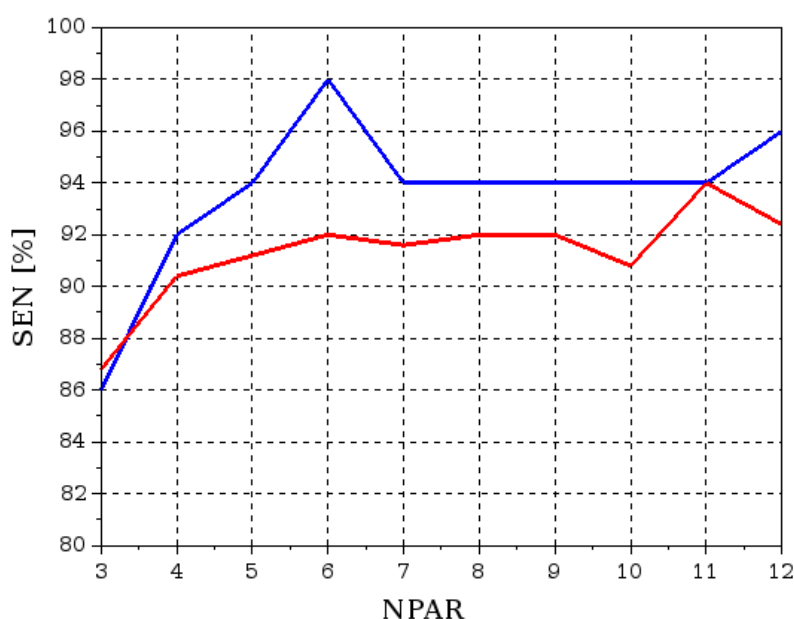
Fig. 4.19. Valores de SPF para os maiores valores de SEN para a otimização multiobjetivo com penalidade $VAL_{pen} = 0,3$ para o procedimento de otimização NSGA-II.



Fonte: O autor, 2018.

Dentre as otimizações realizadas com penalizações, os melhores resultados foram obtidos com o otimizador NSGA-II e $VAL_{pen} = 0,3$ com normalização 3S3D, NCD igual a 60 e NVALID igual a 200. A Figura 4.20 mostra os maiores valores de SEN em função de NPAR onde observa-se que o melhor resultado para o conjunto de validação (SEN igual a 98%) foi obtido para NPAR igual a 6.

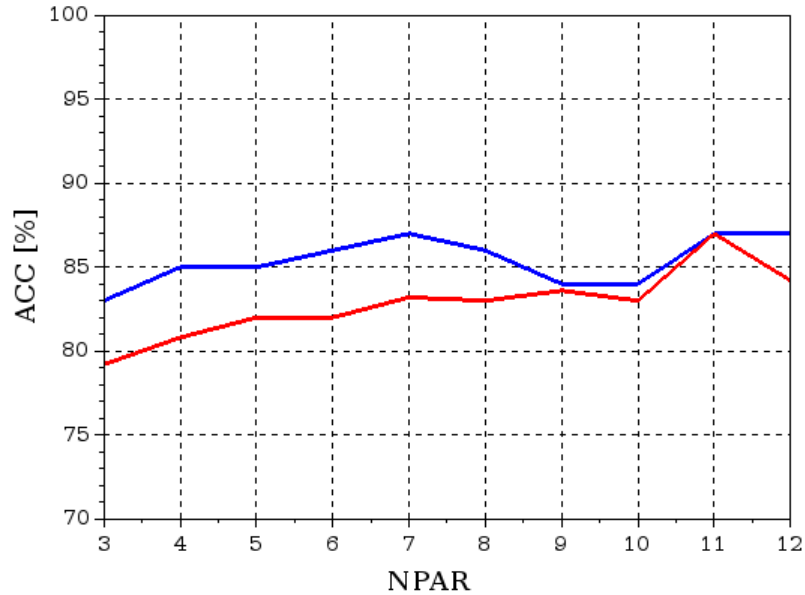
Fig. 4.20. Maiores valores de SEN para a otimização multiobjetivo com penalidade $VAL_{pen} = 0,3$ para o procedimento de otimização NSGA-II.



Fonte: O autor, 2018.

Para o conjunto de teste o melhor resultado foi alcançado com NPAR igual a 11, o que resultou em valores de SEN iguais a 94% e SPF iguais a 82%. A precisão ACC obtida foi de 87% para os dois conjuntos de dados, como pode ser visualizado na Figura 4.21. Uma melhora da precisão para valores de NPAR maiores do que seis já era esperado visto que foi fato recorrente em todas os procedimentos de otimização realizados.

Fig. 4.21. Valores de ACC para os maiores valores de SEN para a otimização multiobjetivo com penalidade $VAL_{pen} = 0,3$ para o procedimento de otimização NSGA-II



Fonte: O autor, 2018.

4.12. FUNÇÃO OBJETIVO BASEADA NA ACURÁCIA

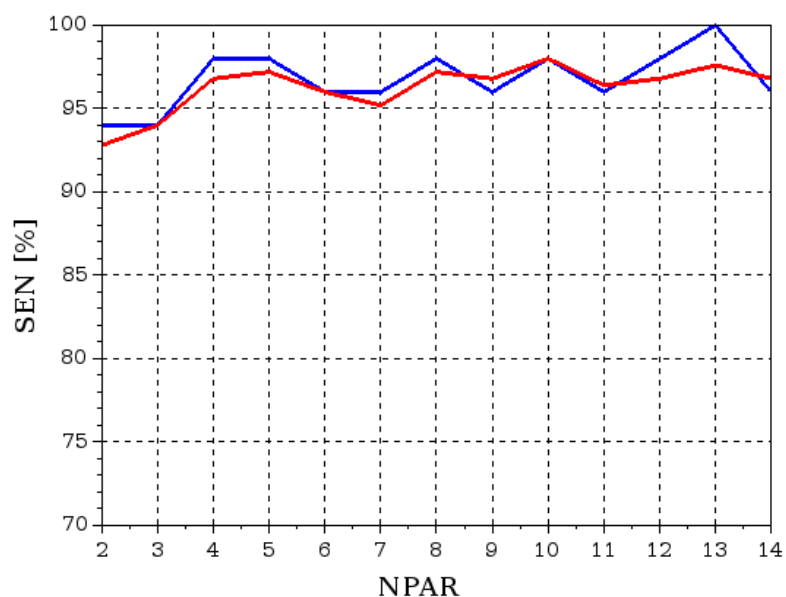
O conjunto de parâmetros ideal seria aquele que resultasse em valores de ACC iguais a 100% para os conjuntos de validação e teste. Por este motivo, foram realizados vários procedimentos de otimização envolvendo a métrica ACC nas funções objetivo e os melhores resultados foram obtidos para os objetivos mostrados na Equação 4.8 utilizando SEN e ACC.

$$Obj_1(p) = 100(1 - SEN/100) \quad E \quad Obj_2(p) = 100(1 - ACC/100) \quad (4.8)$$

Nesta fase do trabalho foi utilizado o otimizador NSGA-II com normalização 3S3D, NCD igual a 60, NVALID igual a 200 e NPAR variável.

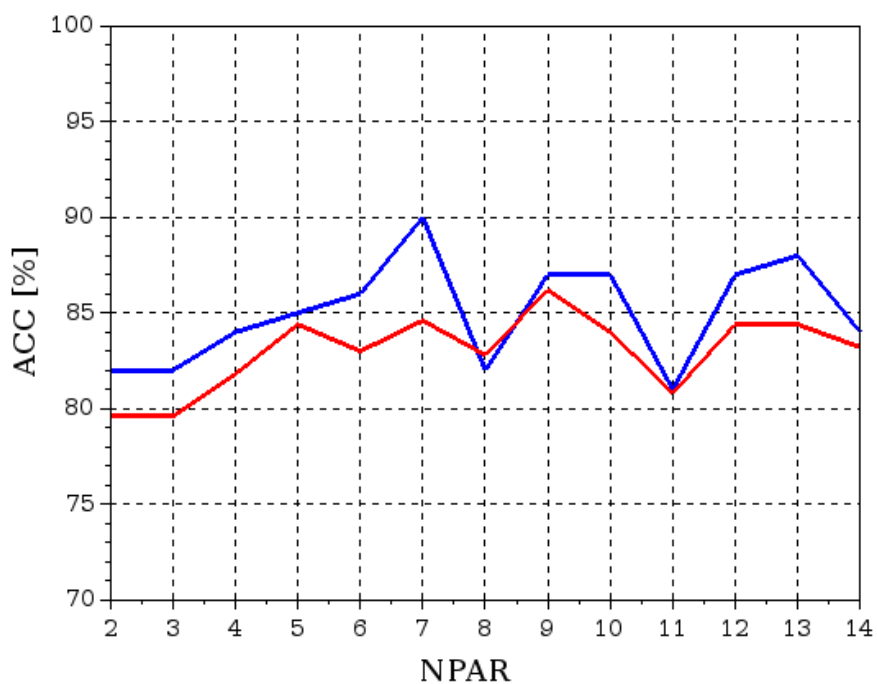
A Figura 4.22 mostra os maiores valores de SEN, em função de NPAR. Os maiores valores de ACC, em função de NPAR, são mostrados na Figura 4.23.

Fig. 4.22. Gráficos SEN vs. NPAR para os conjuntos de validação (azul) e teste (vermelho) para o otimizador NSGA-II.



Fonte: O autor, 2018.

Fig. 4.23. Gráficos ACC vs. NPAR para os conjuntos de validação (azul) e teste (vermelho) para o otimizador NSGA-II.



Fonte: O autor, 2018.

O melhor valor para SEN ocorreu para NPAR igual a 13, que resultou em valores de SEN iguais a 100 e 97,6% para os conjuntos de validação e teste, respectivamente. Os valores de SEN para NPAR igual a 10 foi igual a 98% para os dois conjuntos de dados. O melhor resultado para ACC foi para NPAR igual a 7 (90% para o conjunto de validação e 84,6 para o conjunto de teste) com valores de SEN iguais a 94 e 92%.

O melhor resultado conjunto foi para NPAR igual a 9 que resultou, para os conjuntos de validação e teste, em valores de SEN iguais a 96 e 96,8%, SPF iguais a 72 e 69,8% e ACC iguais a 84 e 83%, respectivamente.

4.13. MELHOR CONJUNTO DE PARÂMETROS

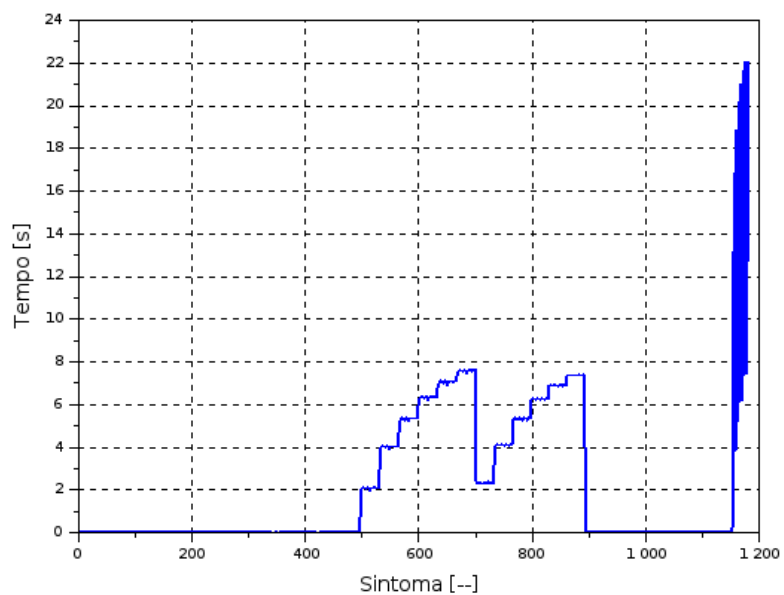
Definir o melhor conjunto dos parâmetros estudados para serem utilizados num procedimento de classificação de EEGs utilizando SVM é bastante complicado uma vez que esta é uma tarefa multiobjetivo. O ideal seria um conjunto de parâmetros que resultasse em valores de ACC iguais a 100% com um custo computacional mínimo, uma vez que é necessário classificar vários sinais (eletrodos) com até várias horas de duração.

Ao montar o Banco de Dados utilizados para testar as metodologias adotadas no trabalho de Duarte (2018), e neste trabalho, foram computados os tempos médios para o cálculo de cada parâmetro, considerando dez segundos de sinal. O parâmetro tempo médio em segundos foi escolhido para se ter uma ideia realista do tempo de computação dos sintomas para a realidade do ano de 2017. Para o cálculo do tempo médio optou-se por utilizar um notebook de médio porte com processador Intel® Core™ i5-4200M CPU@ 2.50GHz x 4 com memória de 8 GB.

Na Figura 4.24 é mostrado o gráfico com os resultados dos tempos de processamento para todos os 1180 parâmetros estudados. Nessa figura, os parâmetros de 497 até 892 envolvem Decomposição em Modos Empíricos (EMD), o que exige muito processamento. Já os parâmetros de 1153 até 1180 envolvem, além da EMD, cálculos de entropia com custos máximos encontrados para parâmetros com avaliações de Entropia Média Aproximada.

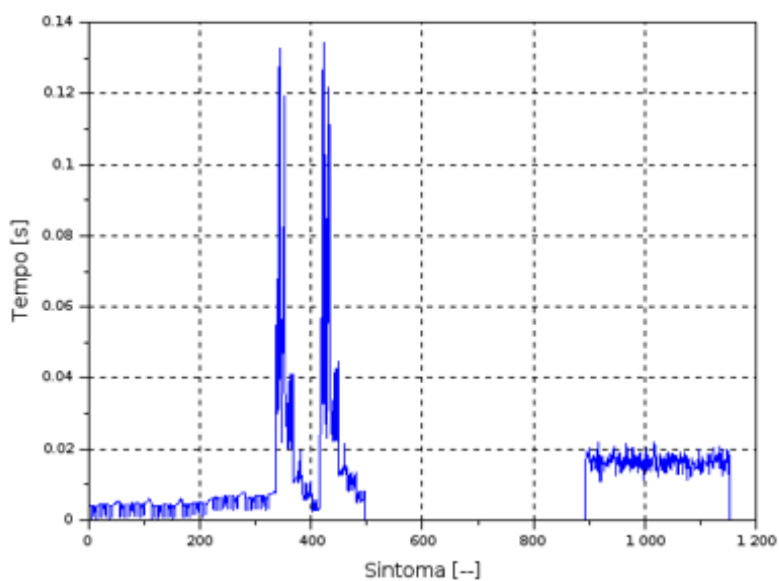
Retirando os tempos computacionais envolvendo modos empíricos, tem-se o gráfico dos tempos de processamento para os outros parâmetros mostrados na Figura 4.25.

Fig. 4.24. Gráfico dos tempos de processamento para todos os sintomas calculados.



Fonte: Duarte, 2018.

Fig. 4.25. Gráfico dos tempos de processamento para os sintomas calculados, excluindo os relacionados a EMD.



Fonte: Duarte, 2018.

Analisando a Figura 4.25 tem-se que os maiores tempos de computação ocorrem para os parâmetros envolvendo as wavelets contínuas WT. Os tempos de processamento dos parâmetros k_NEO (lado direito da figura) são da ordem de 0,016 segundos. Os outros parâmetros têm um tempo médio de processamento de 6,7 milissegundos.

Na Tabela 4.2 são apresentados os parâmetros, os valores de sensibilidade para o conjunto de validação SEN_{Valid} e teste SEN_{Teste} e os tempos médios T_{10} , para o cálculo de cada parâmetro em dez segundos de sinal, para os melhores resultados de sensibilidade obtidos neste trabalho, para um valor mínimo de SPF igual a 60%.

Tab. 4.2. Melhores resultados de sensibilidade SEN limitando o SPF a 60%

Parâmetros	SEN_{Valid} [%]	SEN_{Teste} [%]	T_{10} [s]
284 322 352 431 441 521 612 647 744 795 1044 1079 1117	100	97,6	14,23
223 361 362 466 723 877	100,0	97,2	14,68
248 373 441 476 548 564 680 723 836 1008	100,00	96,00	13,90
52 57 396 441 513 904	100,00	95,20	4,26
131 144 345 357 468 580 581 642 720 773	98,00	98,00	10,75
283 441 475 479 571 1162	98,00	97,60	10,77
210 441 521 574 930	98,00	97,20	10,57
97 106 286 332 441 521 571 773 820 948 1044	98,00	97,20	12,62

Parâmetros	SEN_{Valid} [%]	SEN_{Teste} [%]	T₁₀ [s]
332 425 521 646 805 925 983 997	98,00	97,20	14,22
198 283 441 723	98,00	96,80	4,58
284 322 425 431 495 587 859 903 1178	96,00	96,80	14,82
229 279 441 442 444 493 517 547 627 675 723 766	98,00	96,80	15,26
228 298 345 495 868 1020	98,00	96,4	14,84
55 361 463 608 865 957	98,00	96,00	14,83
96 116 172 345 495 507 621 680 765 888 949	96,00	96,40	15,26
55 361 463 608 865 957	98,00	96,00	14,83
279 409 425 516 580 1178	96,00	96,00	14,82
154 228 440 441 518 642	96,00	94,80	14,03

Analisando os dados da Tabela 4.2 e os gráficos de tempo para computação dos parâmetros conclui-se que valores de sensibilidade altos implica, necessariamente, no cálculo de modos intrínsecos e, na maioria das vezes, parâmetros envolvendo entropia.

Considerando valores de sensibilidade do conjunto de teste e tempo para cálculo dos parâmetros o melhor resultado foi obtido para o conjunto de parâmetros 198, 283, 441 e 723, que resultou em um valor de SEN_{Teste} igual a 96,8% para um tempo T₁₀ de 4,58 segundos. Os parâmetros são:

- 198: sinais bipolares normalizados e filtro Gama.

- 283: valor RMS dos sinais bipolares normalizados e filtro teta.
- 441: valor RMS da db4 do sinal bipolar normalizado escala 115,06844.
- 723: fator de crista do primeiro modo intrínseco bipolar normalizado.

É interessante observar que todos os parâmetros foram calculados utilizando sinais bipolares normalizados.

Na Tabela 4.3 são apresentados os parâmetros, os valores de sensibilidade para o conjunto de validação SEN_{Valid} e teste SEN_{Teste} e os tempos médios T_{10} para o cálculo de cada parâmetro em dez segundos de sinal, para os melhores resultados de sensibilidade obtidos neste trabalho com valores de SPF superiores a 70%.

Tab. 4.3. Melhores resultados de sensibilidade SEN limitando o SPF a 70%

Parâmetros	SEN_{Valid} [%]	SEN_{Test} e [%]	T_{10} [s]
18 229 345 391 914 1056 1072	98,00	95,60	0,04
33 210 397 425 519 669	94,00	96,80	15,06
147 228 345 395 415 773	98,00	94,40	10,51
11 52 322 361 405 451 631 731 737 773 866 1145	96,00	94,00	14,85
144 197 325 367 415 425 511 566 612 732	94,0	94,0	12,80

Avaliando as sensibilidades e tempo, o melhor resultado da Tabela 4.3 foi obtido para o conjunto de parâmetros 18, 229, 345, 391, 914, 1056 e 1072, que resultou em um valor de SEN_{Teste} igual a 95,6% para um tempo T_{10} de 0,04 segundos. Os valores de SPF foram iguais a 70% para os conjuntos de validação e teste. Os parâmetros são:

- 18: amplitude do sinal monopolar filtro gama.
- 229: módulo do envelope RMS do sinal monopolar filtro beta.
- 345: valor RMS da db4 do sinal bipolar normalizado escala 115,30142.

- 391: valor do fator de crista da db4 do sinal monopolar normalizado escala 12,891702.
- 914: valor médio da métrica Dur_{spike} do k_NEO para um valor de k correspondente a 20 ms.
- 1056: valor máximo da métrica $Slope_{PB}$ do k_NEO para um valor de k correspondente a 45 ms.
- 1072: valor médio da métrica Amp_{AP} do k_NEO para um valor de k correspondente a 45 ms.

Considerando o SPF, no lugar do tempo, o melhor resultado foi para o conjunto 144, 197, 325, 367, 415, 425, 511, 566, 612 e 732, com valores de SEN iguais a 94% para os conjuntos de validação e teste e de 80 e 80,2% para os valores de SPF. Os parâmetros são:

- 144: valor de K4 do sinal bipolar não normalizado e filtro delta.
- 197: valor da Curtose do sinal bipolar normalizado e filtro beta.
- 325: valor de K6 do sinal bipolar normalizado e filtro teta.
- 367: Valor do fator de crista da db4 do sinal bipolar normalizado escala 82.078132.
- 415: valor do fator de crista da db4 do sinal bipolar normalizado escala 12,081792.
- 425: valor RMS da db4 do sinal bipolar normalizado escala 115,30142.
- 511: valor de K4 do resíduo do primeiro modo intrínseco unipolar normalizado.
- 566: valor RMS do terceiro modo intrínseco unipolar normalizado.
- 612: valor da Curtose do resíduo do quarto modo intrínseco unipolar normalizado.
- 732: valor do K6 da HHT do resíduo do primeiro modo intrínseco bipolar normalizado.

CAPÍTULO V

5. CONCLUSÕES

Foi proposta desta tese a utilização de otimizadores multiobjetivo como procedimento de discriminante para a escolha dos melhores parâmetros, dentro de um conjunto de 1180 parâmetros, a serem utilizadas para a classificação de sinais de EEG, tendo como ferramenta de classificação a Máquinas de Vetores de Suporte SVM e Análise Discriminante Múltipla QDA. O principal objetivo do trabalho foi maximizar a sensibilidade, ou seja: minimizar o risco de classificar um sinal focal como não focal.

Como conclusão geral, a análise discriminante proposta neste trabalho mostrou-se eficaz para a classificação de sinais cerebrais obtidos via eletroencefalografia do banco de Bern-Barcelona. Os resultados, utilizando Máquinas de Vetores de Suporte SVM como classificador, foram iguais ou superiores aos obtidos por vários trabalhos existentes na literatura. Considerando a porcentagem de acertos para o banco de teste, obteve-se uma sensibilidade máxima de 98% para uma especificidade de 60%. Fixando uma especificidade mínima de 70% na fronteira de Pareto, obteve-se uma sensibilidade máxima de 95,6%. Os melhores resultados globais, dentre as métricas utilizadas para avaliação dos procedimentos de classificação de EEG, foram obtidos com a utilização de funções objetivo com penalização, propostas neste trabalho, que resultaram em valores de sensibilidade iguais a 94%, especificidade iguais a 82% e uma acurácia de 87% para os bancos de dados de validação e de teste. Para efeitos de contextualização, Numa revisão sistemática com o objetivo de determinar a frequência de diagnósticos falso positivos de epilepsia e explorar as suas consequências, Xu et al (2016) concluíram que a frequência do diagnóstico falso positivo, realizados por especialistas, varia de 2% a 71%.

A seguir apresenta-se as conclusões específicas:

Classificador SVM

Com relação à normalização dos dados, pré-processamento importante para a utilização dos SVM, os melhores resultados foram alcançados com a estratégia 3S6D: normalização pela média

amostral e três desvios padrões (3S), dos valores dos parâmetros do conjunto focal, limitados a seis desvios padrões (6D).

É sabido que um dos fatores que influenciam na precisão do SVM supervisionado é o tamanho do conjunto de dados utilizados para o ajuste de seus parâmetros na fase de treinamento. Neste trabalho, os melhores valores de funções objetivo foram obtidos com bancos de dados para treinamento variando entre 40 a 70 conjuntos de sinais focal e não focal.

Com relação ao número de parâmetros os melhores resultados foram obtidos utilizando de 5 a 8 parâmetros discriminados via otimização.

Pesquisas com o núcleo (*kernel*) linear e o SVM de uma classe foram abandonadas em função de resultados não satisfatórios, em comparação com os de base radial (valores de SEN máximos da ordem de 70%). Apesar de infrutífero, muito esforço foi dispendido na SVM de uma classe (treinamento não supervisionado) em função da perspectiva de se trabalhar apenas com dados não focal na fase de treinamento.

Os melhores núcleos para o problema estudado foram o de base radial e polinômio de grau 5 P5. O kernel P5, com 10 parâmetros, resultou em valores de sensibilidade iguais a 100% para o conjunto de validação e 96% para o conjunto de teste. Nos testes para avaliação dos núcleos, o melhor resultado com núcleo de base radial resultou em valores de SEN iguais a 100 e 96,4% para os conjuntos de validação e teste.

Classificador QDA

Utilizando o QDA como procedimento de classificação, os melhores valores de sensibilidade (100% para o conjunto de validação e 95,2% para o conjunto de teste) foram obtidos para normalização média amostral e 1 desvio padrão 1S0D, número de parâmetros iguala seis e o banco usado no treinamento com 50 conjuntos de sinais focal e não focal. Por ser fácil de implementar e não ser necessário nenhum ajuste de coeficientes, o QDA torna-se uma opção muito interessante para classificação de EEG.

Otimizadores Multiobjetivos

O MOPSO foi o otimizador mais eficiente no que diz respeito ao tempo de processamento. Para o problema estudado observou-se que os procedimentos PESA2, SPEA2 e NSGA-II levam, em média, 5,79, 20,56 e 37,52 vezes mais tempo para a convergência do que o MOPSO.

Com relação ao desempenho os melhores resultados foram alcançados com o otimizador NSGA-II e os piores com o MOPSO. Os otimizadores SPEA2 e PESA2 tiveram desempenho semelhantes, porém o otimizador PESA2 é aproximadamente quatro vezes mais rápido do que o SPEA2 em tempo de processamento.

funções objetivo

Utilizando a Sensibilidade (SEN) e a Especificidade (SPF) como funções objetivo, o melhor resultado para a Sensibilidade do conjunto de testes foi de 98,8% para uma Especificidade de 60%. Estas funções objetivo foram muito estudadas uma vez que possuem objetivos excludentes: uma maximiza a porcentagem de acertos de EEGs com anomalias (SEN) e, em contrapartida, a outra maximiza a porcentagem de acertos de EEGs normais (SPF).

Utilizando três funções objetivo SEN SPF e ACC (acurácia), o maior valor de SEN para o conjunto de teste foi de 97,2% para um SPF de 60%. Considerando a maximização dos três métricas de avaliação dos procedimentos, obteve-se uma combinação de sete parâmetros que resultou em valores dos conjuntos de validação e teste iguais a 98 e 95,6% para o SEN, 70 e 69,4% para o SPF e 84 e 82% para o ACC.

Um procedimento baseado na penalização das funções objetivo para restringir a frente de Pareto proposto neste trabalho resultou em valores de SEN iguais a 94%, SPF iguais a 82% e ACC iguais a 87%, para os conjuntos de dados de validação e teste.

Utilizando a Sensibilidade (SEN) e a Acurácia (ACC) como funções objetivo, o melhor conjunto foi para um número de parâmetros igual a 9 resultando, para os conjuntos de validação e teste, em valores de SEN iguais a 96 e 96,8%, SPF iguais a 72 e 69,8% e ACC iguais a 84 e 83%, respectivamente.

6.1. TRABALHOS FUTUROS

Para trabalhos futuros, sugere-se:

- Testar o procedimento desenvolvido com outros bancos de dados.
- Aumentar o número de parâmetros candidatos para classificação, incluindo filtros autoregressivos e transformadas discretas de wavelet, por exemplo.
- Utilizar outros procedimentos de classificação.

- Paralelizar, via GPU, os procedimentos para o cálculo dos parâmetros. Esta etapa é fundamental para a aplicação de inteligência artificial para a classificação de EEG com vários canais, a tendência é 28 canais, com horas e horas de gravação.

REFERÊNCIAS BIBLIOGRÁFICAS

Abbas Amini (2015). Application of Fuzzy Multi-Objective Programming in Optimization of Crop Production Planning. Asian Journal of Agricultural Research, 9: 208-222. <https://doi.org/10.3923/ajar.2015.208.222>

Acharya, U.R., Sree, S.V., Chattopadhyay, S., Yu, W., Ang, P.C.A (2011). Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. Int. J. Neural Syst. 2011, 21, 199–211. <https://doi.org/10.1142/S0129065711002808>

Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001). Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, *Phys. Rev. E*, 64, 061907. <https://doi.org/10.1103/PhysRevE.64.061907>

Andrzejak, R. G., Schindler, K., & Rummel, C. (2012). Nonrandomness, Nonlinear Dependence, And Nonstationarity Of Electroencephalographic Recordings From Epilepsy Patients. *Physical Review E*, 86(4), 046206. <https://doi.org/10.1103/PhysRevE.86.046206>

Benbadis SR, Tatum WO. 2003. Overinterpretation of EEGs and misdiagnosis of epilepsy. *J Clin Neurophysiol*. 2003;20:42-44. <https://doi.org/10.1097/00004691-200302000-00005>

Benbadis SR. The tragedy of over-read EEGs and wrong diagnoses of epilepsy. *Expert Rev Neurother*. 2010;10:343-346. <https://doi.org/10.1586/ern.09.157>

Bhattacharyya S., Khasnobish A., Chatterjee S., Konar A. e Tibarewala D.N (2010). "Performance Analysis of LDA, QDA and KNN Algorithms in Left-Right Limb Movement Classification from EEG Data, " *Proc. of 2010 International Conference on Systems in Medicine and Biology*, pp. 126-131, 16-18 December 2010. <https://doi.org/10.1109/ICSMB.2010.5735358>

Blinowska, K. And Durka, P (2006). Electroencephalography (EEG), John & Sons, Inc.; Wiley Encyclopeida Of Biomedical Enginneering, 15p. <https://doi.org/10.1002/9780471740360.ebs0418>

Brogger J., Eichele T., Aanestad E., Olberg H., Hjelland I. e Aurlien H (2018). Visial EEG reviewing times whith SCORE EEG. Clinical Neurophysioloy Prectice, V. 3, 2018, pg. 59-64. <https://doi.org/10.1016/j.cnp.2018.03.002>

B.Y. Qu, J.J. Liang, Y.S. Zhu, Z.Y. Wang, P.N. Suganthan (2016). Economic emission dispatch problems with stochastic wind power using summation based multi-objective evolutionary algorithm. Information Sciences 351, p. 48–66. <https://doi.org/10.1016/j.ins.2016.01.081>

Chen D., Wan S. e Bao FS (2017). Epileptic Focus Localization Using Discrete Wavelet Transform Based on Interictal Intracranial EEG. IEEE Trans Neural Syst Rehabil Eng. 2017 May;25(5):413-425. doi: 10.1109/TNSRE.2016.2604393. Epub 2016 Aug 30. <https://doi.org/10.1109/TNSRE.2016.2604393>

Choi, J. H., & Kim, T. (2002). Neural Action Potential Detector Using Multi-Resolution Teo. Electronics Letters, 38(12), 541-542. <https://doi.org/10.1049/el:20020386>

Chua K. C.,Chandran V., Acharya U.R. e Lim C.M (2008). Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: A comparative study, Proc IEEE Eng Med Biol Soc Conf (2008), 3824–3827. <https://doi.org/10.1109/IEMBS.2008.4650043>

Cîmpanu, C., Ferariu, L., Dumitriu, T. and Ungureanu, F. (2017). Multi-Objective Optimization of Feature Selection Procedure for EEG Signals Classification. The 6th IEEE International Conference on E-Health and Bioengineering - EHB 2017 Grigore T. Popa University of Medicine and Pharmacy, Sinaia, Romania, June 22-24, 2017. <https://doi.org/10.1109/EHB.2017.7995454>

Coello-Coello, C. A. A Short Tutorial on Evolutionary Multiobjective Optimization. First International Conference on Evolutionary Multi-Criterion Optimization, 2001, Springer-Verlag. Lecture Notes in Computer Science, No. 1993, p.21-40. https://doi.org/10.1007/3-540-44719-9_2

Coello-Coello, C. A (1999). A Comprehensive Survey of Evolutionary Based Multiobjective Optimization Techniques. Knowledge and Information Systems , 1999, 1(3), p. 269–308. <https://doi.org/10.1007/BF03325101>

Das A.B. e Bhuiyan M.I.H (2016). Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain, Biomed Signal Proces Control 29 (2016), 11–21. <https://doi.org/10.1016/j.bspc.2016.05.004>

Deb, K (2000). An efficient constraint handling method for genetic algorithms. Computer Methods In Applied Mechanics And Engineering, v. 186, n. 2, p.311-338, jun. 2000. [https://doi.org/10.1016/S0045-7825\(99\)00389-8](https://doi.org/10.1016/S0045-7825(99)00389-8)

Deivasigamani S., Senthilpari C. e Yong W. H (2016). Classification of Focal and Nonfocal EEG Signals Using ANFIS Classifier for Epilepsy Detection. Wiley Periodicals, Inc. Int. J. Imaging. Syst. Technol., 26, 277–283, 2016. <https://doi.org/10.1002/ima.22199>

Delinsk, L. M. M. e Colmonero, J. C (2017). Otimização Multiobjetivo: uma Abordagem Conceitual. VII Congresso Brasileiro de Engenharia de Produção. 06 a 08 de dezembro de 2017 – Ponta Grossa, PR. Brasil.

Demir, E.; Bektas, T.; Laporte, G (2014). The bi-objective Pollution-Routing Problem. European Journal of Operational Research, v. 232, n. 3, p.464-478. <https://doi.org/10.1016/j.ejor.2013.08.002>

Denysiuk, R (2012). Evolutionary Multiobjective Optimization: Review, Algorithms, and Applications. Tese de doutorado do Programa Doutoral em Engenharia Industrial e Sistemas – Universidade do Minho – Portugal.

Dujardin Y, Chadès I (2018) Solving multi-objective optimization problems in conservation with the reference point method. PLoS ONE 13(1): e0190748. Perry J. Williams and William L. Kkendall. <https://doi.org/10.1371/journal.pone.0190748>

Duarte, J. B (2018). Inteligência Artificial Aplicada a Análise de Eletroencefalografia. Tese (Doutorado em Engenharia Mecânica) – Faculdade de Engenharia Mecânica, Universidade Federal de Uberlândia.

Fávero, L. P., Belfiore, P., Silva, F. L., Chan, B. L (2009). Análise de dados: modelagem multivariada para tomada de decisões. Rio de Janeiro: Campus/Elsevier. 646 p.

Francesca Guerriero, Ferdinando Pezzella, Ornella Pisacane and Luigi Trollini (2014). Multi-objective optimization in dial-a-ride public transportation. *Transportation Research Procedia* 3, p. 299 – 308. <https://doi.org/10.1016/j.trpro.2014.10.009>

Galan-Martin, A., Vaskan, P., Anton, A., and Jimenez Esteller, L. (2017). Multi-objective optimization of rainfed and irrigated agricultural areas considering production and environmental criteria: a case study of wheat production in Spain. *Journal Of Cleaner Production*, 140, p. 816-830. <https://doi.org/10.1016/j.jclepro.2016.06.099>

G. Alaimo, F. Auricchio, M. Contand M. Zingales (2017). Multi-objective optimization of nitinol stent design. *Medical Engineering & Physics*, Vol. 47, p. 13–24. <https://doi.org/10.1016/j.medengphy.2017.06.026>

Gao Z., Lu G., Yan P., Lyu C., Li X., Xie Z. e Zhang W (2018). Automatic Change Detection for Real-Time Monitoring of EEG Signals. *Front Physiol.* 2018; 9: 325. Published online 2018 Apr 4. doi: 10.3389/fphys.2018.00325. <https://doi.org/10.3389/fphys.2018.00325>

Goh S.K., Abbass H.A., Tan K.C., Mamun A.A (2014). Artifact Removal from EEG Using a Multi-objective Independent Component Analysis Model. In: Loo C.K., Yap K.S., Wong K.W., Teoh A., Huang K. (eds) *Neural Information Processing. ICONIP 2014. Lecture Notes in Computer Science*, vol 8834. Springer, Cham. https://doi.org/10.1007/978-3-319-12637-1_71

Hasan K., Ahamed A, Ahmad M. e Rashid A (2017). Prediction of Epileptic Seizure by Analysing Time Series EEG Signal Using k-NN Classifier. Appl Bionics Biomech. Published online 2017 Aug 13. doi: 10.1155/2017/6848014. <https://doi.org/10.1155/2017/6848014>

Hashimoto, K. (2004). Técnicas de otimização combinatória multiobjetivo aplicadas na estimação do desempenho elétrico de redes de distribuição. Tese (Doutorado em Sistemas de Potência) - Escola Politécnica, Universidade de São Paulo, São Paulo. doi:10.11606/T.2.2004.tde-19112004-165342. Acesso em: 2018-06-01.

Haykin, S (2001), “Redes Neurais – Princípios E Prática.”, 2ª Ed., Ed. Bookman, Porto Alegre.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Liu, H. H (1998). The Empirical Mode Decomposition And The Hilbert Spectrum For Nonlinear And Non-Stationary Time Series Analysis. In Proceedings Of The Royal Society Of London A: Mathematical, Physical And Engineering Sciences (Vol. 454, No. 1971, Pp. 903-995). The Royal Society. <https://doi.org/10.1098/rspa.1998.0193>

Kailath, T (1967). The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Commun. Technol. 15, 52–60. <https://doi.org/10.1109/TCOM.1967.1089532>

Kennedy, J. e Eberhart, R. C (1995). Particle Swarm Optimization. Proceedings of IEEE International Conference on Neural Networks, Washington, DC, USA. <https://doi.org/10.1109/ICNN.1995.488968>

Kimovski, D., Ortega, J. Ortiz, A. and Baños R (2015). Leveraging cooperation for parallel multi-objective feature selection in high-dimensional EEG data. Concurrency Computat.: Pract. Exper. 2015; 27:5476–5499 Published online 14 August 2015 in Wiley Online Library (wileyonlinelibrary.com). <https://doi.org/10.1002/cpe.3594>

Knowles, J. and Corne D (2000). Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy. *Evolutionary Computation* 8(2). p. 149-172.
<https://doi.org/10.1162/106365600568167>

Kumar, M., Husian, M., Upreti, N., Gupta, D. (2010). Genetic Algorithm: Review And Application. *International Journal Of Information Technology And Knowledge Management*, Vol. 2, No 2, Pp. 451 – 454.

Li, X., Chen, X., Yan, Y., Wei, X. e Wang Z. J (2014). Classification of EEG Signals Using a Multiple Kernel Learning Support Vector Machine. *Sensors (Basel)*. 2014 Jul; 14(7): 12784–12802. <https://doi.org/10.3390/s140712784>

Liu, Y. C., Lin, C. C. K., Tsai, J. J., & Sun, Y. N. (2013). Model-Based Spike Detection Of Epileptic Eeg Data. *Sensors*, 13(9), 12536-12547. <https://doi.org/10.3390/s130912536>

Lobato, F. S. ; Steffen Jr, V. ; Cavalini Jr, A. A. ; Moreira, F. R (2016). Robust Multi-Objective Optimization Applied to Engineering Systems Design. *Latin American Journal of Solids and Structures* v. 13, p. 1802-1822. <https://doi.org/10.1590/1679-78252801>

Llopis-Albert, C., Rubio, F. and Valero, F. (2015). Improving productivity using a multi-objective optimization of robotic trajectory planning. *Journal of Business Research* 68, 1429 – 1431.
<https://doi.org/10.1016/j.jbusres.2015.01.027>

Lorena, A. C. E Carvalho, A. C. P. L. F (2007). Uma Introdução às Support Vector Machines. *RITA*, Vol. XIV, Nº 2. p. 44 – 67.

MadhuSudana R. N., Kannan K., Manisha M. and Diptendu S. R (2017). Hybrid Disease Diagnosis Using Multiobjective Optimization with Evolutionary Parameter Optimization. *J Healthc Eng*. Published online 2017 Jul 4. doi: 10.1155/2017/5907264. <https://doi.org/10.1155/2017/5907264>

Maragos, P., Kaiser, J. F., & Quatieri, T. F (1993). On Amplitude And Frequency Demodulation Using Energy Operators. *Ieee Transactions On Signal Processing*, 41(4), 1532-1550. <https://doi.org/10.1109/78.212729>

Ming Qing Li, Chun Yan Yuan (2017). Multi-objective optimization for the aerodynamic noise of the high-speed train in the near and far field based on the improved NSGA-II algorithm. *Journal of Vibroengineering*, Vol. 19, Issue 6, p. 4759-4782.

Montgomery, D. C., (2000), *Design and Analysis of Experiments*, Fifth Edition, John Wiley & Sons, New York, NY.

Oda, Shigueo, 2013, “O Corpo Humano Possui Um Mecanismo Automático De Manutenção” <<http://www.kotsubantousei.com.br/doencas/o-corpo-humano-possui-um-mecanismo-automatico-de-manutencao>>. Acessado em janeiro de 2018.

Pachori R.B. e Sircar P (2008). EEG signal analysis using FB expansion and second-order linear TVAR process, *Signal Process.* 88 (2) 415–420. <https://doi.org/10.1016/j.sigpro.2007.07.022>

Pastori, M., Udias, A., Bouraoui, A. and Bidoglio (2017). A Multi-Objective Approach to Evaluate the Economic and Environmental Impacts of Alternative Water and Nutrient Management Strategies in Africa. *Journal of Environmental Informatics* 29(1), p16-28.

Perry J. Williams and William L. Kkendall (2017). A guide to multi-objective optimization for ecological problems with an application to cackling goose management. *Ecological Modelling*, 343, p. 54–67. <https://doi.org/10.1016/j.ecolmodel.2016.10.010>

Raquel, C. R. e Prospero Jr., C. N (2005). An effective use of crowding distance in multiobjective particle swarm optimization. *GECCO 05 Proceedings of the 7th annual conference o Genetic and evolutionary computation*. Washington DC, USA – June 25 – 29, p. 257-264.

Rodríguez-Pérez R., Vogt M. e Bajorath J (2017). Influence of Varying Training Set Composition and Size on Support Vector Machine-Based Prediction of Active Compounds .J Chem Inf Model. Apr 24; 57(4): 710–716. Published online 2017 Apr 4. <https://doi.org/10.1021/acs.jcim.7b00088>

Salkind, N. J. (Ed.). (2010). Encyclopedia Of Research Design(Vol. 1). Sage. <https://doi.org/10.4135/9781412961288>

Scola, L. A. (2014). Otimização multiobjetivo evolutiva da operação de sistemas de reservatórios multiusos. Tese (doutorado) Universidade Federal de Minas Gerais, Escola de Engenharia.

Seker, M. and Özerdem , M. S. (2016). Classification of EEG data sets with Hilbert transform. 24th Signal Processing and Communication Application Conference (SIU). Zonguldak, Turkey, 16-19 May 2016 – p. 1949 – 1952. <https://doi.org/10.1109/SIU.2016.7496148>

Sharma R., Pachori, R. B. e Acharya, U. R (2015). An Integrated Index for the Identification of Focal Electroencephalogram Signals Using Discrete Wavelet Transform and Entropy Measures. Entropy 2015, 17, 5218-5240. <https://doi.org/10.3390/e17085218>

Sharma, R., Pachori, R. B., & Acharya, U. R (2015). Application Of Entropy Measures On Intrinsic Mode Functions For The Automated Identification Of Focal Electroencephalogram Signals. Entropy, 17(2), 669-691. <https://doi.org/10.3390/e17020669>

Sharma M., Dhare A., Pachori R. B. e Acharya U. R (2017). An automatic detection of focal EEG signals using new class of time–frequency localized orthogonal wavelet filter banks. Knowledge-Based Systems 118 (2017) 217–227. <https://doi.org/10.1016/j.knosys.2016.11.024>

Schomer, D. L. e Silva, F. H. L (2011). Basic Principles, Clinical Applications and Related Fields, Neidermeyer's Electroencephalography, sixth edition, 2011, Copyright ©2011 Lippincott Williams & Wilkins, Vol I/Vol II.

Shufu Lin, Fan Lin, Haishan Chen and Wenhua Zeng (2017). A MOEA/D-based multi-objective optimization algorithm for remote medical. Neurocomputing 220, p. 5 – 16. <https://doi.org/10.1016/j.neucom.2016.01.124>

Silva, H. A (2012). Algoritmo de otimização multinível aplicado a problemas de planejamento de Redes. Tese doutorado – Pontifícia Universidade Católica do Paraná, Curitiba.

Silva, J. F. (2008). Sistema De Armazenamento De Imagens Comprimidas Através da Transformada Wavelet. Dissertação de Mestrado – UNESP – Ilha Solteira – 99p.

Subramaniam, N. P. Measuring Entropy in the EEG. Disponível em: <http://sapienlabs.co/measuring-entropy-in-the-eeeg/> Acesso em 05 de junho de 2018.

Tatum, W.O; Husain, A.M; Benbadis, S.R; Kaplan, P.W (2008). Handbook of EEG Interpretation, Demos, 289p.

Thodoroff P., Pineau, J. e Lim, A (2016). Learning Robust Features using Deep Learning for Automatic Seizure Detection. Proceedings of Machine Learning and Healthcare Conference Held at Children's Hospital LA, Los Angeles, CA, USA on 19-20 August 2016.

Timothy C. Y. Chan, Tim Craig, Taewoo Lee and Michael B. Sharpe (2014). Generalized Inverse Multiobjective Optimization with Application to Cancer Therapy. Operations Research 201462:3, 680-695. <https://doi.org/10.1287/opre.2014.1267>

Varella, C. A. A. “Análise Discriminante”, Apostila De Aula De Análise Multivariada Aplicada As Ciências Agrárias, – Seropédica, Rj, 2007.

Wang G., Shepherd S. J., Beggs C. B., Rao N. and Zhang Y (2015). The use of kurtosis denoising for EEG analysis of patients suffering from Alzheimer’s disease. Bio-Medical Materials and Engineering (26), S1 135 – S1 148. <https://doi.org/10.3233/BME-151410>

Wanner, E. F.; Takahashi, R.H.C (2007). Uma nova medida de desempenho para otimização multiobjetivo: a integral da contagem de esferas. In: Congresso Nacional de Matemática Aplicada e Computacional, 2007, Florianópolis. Anais do CNMAC 2007.

Wilner, A. N (2016). Erros diagnósticos em pacientes com sintomas neurológicos. <https://portugues.medscape.com/features/slides/65000041#page=1>. Acessado em junho de 2018.

Xu G., Liang X., Yao S., Chen D., Li Z (2017). Multi-Objective Aerodynamic Optimization of the Streamlined Shape of High-Speed Trains Based on the Kriging Model. PloS ONE 12(1): e0170802. <https://doi.org/10.1371/journal.pone.0170803>

Xu, Y., Nguyen D., Mohamed A., Carcel C., Li Q., Kutlubaev M. A., Anderson C. S. e Hackett M. L. (2016). Frequency of a false positive diagnosis of epilepsiy: A systematic review of observational studies. Seizure 41, p. 167-174. <https://doi.org/10.1016/j.seizure.2016.08.005>

Zavaschi, T. Aprendizagem Supervisionada Através do Método de Support Vector Machine (SVM). Disponível em: <https://www.google.com/search?lient=ubuntu&channel=fs&q=Thiago+Zavaschi+%28+zavaschi%40ppgia.pucpr.br+%29+Orientador%3A+Alessandro+Koerich&ie=utf-8&oe=utf-8>> Acesso em 05 de junho de 2018.

Zhu, Jin-De; Lin, Chin-Feng (2014). Analyzing Spike I Wave Electroencephalogram Signals For Epilepsy Based On Hilbert-Huang Transformation. In: Bioelectronics And Bioinformatics (Isbb), 2014 Ieee International Symposium On. Ieee pg. 1-2. <https://doi.org/10.1109/ISBB.2014.6820937>

Zitzler, E., Laumanns, M. and Thiele, L. (2001). SPEA2: Improving the Performance of the Strength Pareto Evolutionary Algorithm. Technical Report 103, Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH) Zurich.