



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Paulo Henrique Gubert

**CONTRIBUIÇÃO À MATRIZ DE INFORMAÇÕES EM SISTEMAS DE
IMAGÉTICA MOTORA BASEADAS NO MÉTODO DE CSP**

Florianópolis

2019

Paulo Henrique Gubert

**CONTRIBUIÇÃO À MATRIZ DE INFORMAÇÕES EM SISTEMAS DE
IMAGÉTICA MOTORA BASEADAS NO MÉTODO DE CSP**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Engenharia Elétrica

Orientador: Prof. Dr. Márcio Holsbach Costa

Coorientador: Prof. Dr. Alexandre Trofino Neto

Florianópolis

2019

Ficha de identificação da obra

Gubert, Paulo Henrique

Contribuição à matriz de informações em sistemas de
imagética motora baseadas no método de CSP /

Paulo Henrique Gubert; orientador, Márcio Holsbach
Costa, coorientador, Alexandre Trofino Neto, 2019.

115 p.

Dissertação (mestrado) - Universidade Federal de
Santa Catarina, Centro Tecnológico, Programa de
Pós-Graduação em Engenharia Elétrica,
Florianópolis, 2019.

Inclui referências.

1. Engenharia Elétrica. 2. CSP. 3. FLDA. 4. Imagética motora. 5.
BCI. I. Holsbach Costa, Márcio. II. Trofino Neto, Alexandre. III.
Universidade Federal de Santa Catarina. Programa de Pós-
Graduação em Engenharia Elétrica. IV. Título.

Paulo Henrique Gubert

**CONTRIBUIÇÃO À MATRIZ DE INFORMAÇÕES EM SISTEMAS DE
IMAGÉTICA MOTORA BASEADAS NO MÉTODO DE CSP**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Alejandro Rafael Garcia Ramirez, Dr.
Universidade do Vale do Itajaí

Prof. José Carlos Moreira Bermudez, Dr.
Universidade Federal de Santa Catarina

Prof. Ranniery da Silva Maia, Dr.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Engenharia Elétrica.

Prof. Dr. Bartolomeu Ferreira Uchôa Filho
Coordenador do Programa

Prof. Dr. Márcio Holsbach Costa
Orientador

Florianópolis, 12 de julho de 2019.

Este trabalho é dedicado aos meus colegas do LPDS e à minha querida família.

AGRADECIMENTOS

Agradeço primeiramente à minha família, que me propiciou a segurança, conforto e afeto para que eu me tornasse uma pessoa de boa índole, digna, inteligente, com senso crítico e ético.

À minha namorada Júlia Fernandes Holvorcem, pela sua atenção, carinho e conselhos.

À UFSC por ter cedido o espaço de trabalho para a elaboração da dissertação, além disso, agradeço-a pelos excelentes professores que foram fundamentais para o meu crescimento acadêmico e intelectual.

Ao meu orientador Márcio Holsbach Costa, pelos ensinamentos e instruções para a elaboração desta dissertação.

Ao Cleison Daniel Silva por ter fornecido o material inicial para a composição desta dissertação.

Ao professor Alexandre Trofino Neto pela sua disponibilidade e sugestões apresentadas nas reuniões.

Ao professor Adriano Borgatto pela sua ajuda na parte estatística da dissertação.

Aos meus colegas de LPDS: Augusto Boff Pizzetta, Diego Marques do Carmo e Ricardo Augusto Borsoi, que contribuíram com sua amizade e conselhos para a confecção desta dissertação.

À CAPES pelo apoio financeiro.

E a todas as outras pessoas, que de alguma forma me ajudaram a tornar esse sonho possível.

“Tudo o que temos de decidir é o que fazer com o tempo que nos é dado (TOLKIEN, 1994)”.

RESUMO

Este trabalho propõe uma nova contribuição à formação da matriz de informações em um classificador de imagética motora para interfaces cérebro-máquina. A técnica consiste na inclusão da correlação temporal inter e intra-eletrodos na matriz de informações utilizada pelo método de padrões espaciais comuns (CSP). Simulações com sinais eletroencefalográficos obtidos a partir de dois bancos públicos de imagética motora, em um contexto de diferenciação entre movimentos binários (mão esquerda e direita) e quaternários (mão esquerda, mão direita, ambos os pés e língua) indicaram que o método proposto aumenta os índices de classificação quando comparado ao CSP utilizando a matriz de correlação convencional. Os resultados obtidos, corroborados por meio de inferência estatística e da curva ROC, indicam que a correlação inter e intra-eletrodos possui informação relevante, até então descartada pela literatura em aplicações de imagética motora. Utilizando uma abordagem conjunta do CSP e um classificador por Análise de Discriminantes Lineares de Fisher (FLDA) em problemas de imagética motora binários, para um conjunto de 12 participantes, houve um aumento médio de acurácia de classificação de 3,3% utilizando o método proposto em relação ao convencional. Em outra simulação, houve um aumento de acurácia média de até 5% em relação ao método convencional, quando utilizados pelo menos 7 canais de EEG. Utilizando métodos regularizados de CSP em conjunto com o classificador FLDA em problemas binários, o aumento de acurácia de 2 a 5,4% obtido, varia de acordo com a regularização empregada e do banco de dados utilizado. Por fim, em problemas quaternários usando a abordagem CSP-FLDA, houve um aumento de 1,8% de acurácia média usando a abordagem proposta.

Palavras-chave: BCI. LDA. FLDA. CSP. ERD. ERS. ROC. Imagética motora. Estimação. Covariância. Atrasos.

ABSTRACT

This work proposes a new contribution to the information matrix in a classifier of motor imagery for brain-computer interfaces. The technique consists in the inclusion of intra and inter electrode temporal correlation in the information matrix used by the common spatial patterns method (CSP). Simulations performed with electroencephalographic signals obtained from two public motor imagery datasets, in a context of differentiation between binary (left and right hand) and quaternary movements (left hand, right hand, both feet and tongue), indicated that the proposed method increases the classification indices when compared to the CSP method using the conventional correlation matrix. The results obtained, corroborated through statistical inference and the ROC curve, indicate that the inter- and intra-electrode correlation have relevant information previously discarded by the literature. Using a joint approach based on the CSP and the Fisher Linear Discriminant Analysis (FLDA) classifier to binary motor imagery problems for a set of 12 subjects, there was a mean increase in classification accuracy of 3,3% using the proposed method over the conventional one. By optimizing the conventional and proposed methods, the average accuracy increase is 5%, either using all available electrodes or only 7 of them. Using regularized CSP methods in conjunction with FLDA in binary problems, the accuracy increase is between 2 and 5,4%, according to the regularization technique and database employed. Finally, in quaternary problems using the CSP-FLDA approach, there was a 1,8% increase in mean accuracy using the proposed approach and 7 subjects.

Keywords: BCI. LDA. FLDA. CSP. ERD. ERS. ROC. Motor imagery. Estimation. Covariance. Lags.

LISTA DE FIGURAS

Figura 1 – Áreas funcionais do córtex cerebral. O córtex cerebral contém áreas sensoriais para percepção, áreas motoras e áreas de associação que integram informações.	31
Figura 2 – Hemisférios e lateralização cerebral.	32
Figura 3 – Categorias morfológicas e funcionais dos neurônios.	34
Figura 4 – Geração do sinal elétrico em um neurônio.	35
Figura 5 – Geração de um potencial de ação a partir de um potencial graduado.	36
Figura 6 – Os neurotransmissores formam a mensagem em uma sinapse química. O terminal axônico contém mitocôndrias e vesículas sinápticas preenchidas com neurotransmissores. A membrana pós-sináptica possui receptores para o neurotransmissor que se difunde através da fenda sináptica.	37
Figura 7 – Inibição pós-sináptica.	38
Figura 8 – Inibição pré-sináptica.	38
Figura 9 – A curva em azul representa a redução de energia (ERD) no espectro do EEG em torno de 10 Hz. Em vermelho, o espectro de EEG sem dessincronização.	42
Figura 10 – Posição dos eletrodos no escalpo de acordo com o sistema 10-20.	43
Figura 11 – Diagrama em blocos de um sistema genérico de BCI.	46
Figura 12 – Abordagem para classificação de quatro movimentos motores.	47
Figura 13 – Disposição de eletrodos: (a) banco de sinais IIA com 22 eletrodos; e (b) banco de sinais IIIA com 60 eletrodos.	48
Figura 14 – Geração dos sinais de imagética motora.	49
Figura 15 – Exemplos de autovalores associados aos autovetores que compõem a matriz \mathbf{W} de filtragem espacial CSP. Os autovalores foram obtidos usando os sinais do sujeito B_1 do banco de dados IIIA.	52
Figura 16 – FLDA, hiperplano ótimo (linha sólida) que separa duas classes. O hiperplano é descrito pelo vetor \mathbf{a} e o termo de <i>offset</i> b	60
Figura 17 – SVM de margens suaves.	62
Figura 18 – SVM não linear. A primeira imagem não é linearmente separável por um hiperplano. A segunda imagem representa o espaço aumentado por meio de um Kernel, no qual é possível separar os dados.	63
Figura 19 – Curva ROC.	69
Figura 20 – Testes de hipótese unilateral e bilateral.	73
Figura 21 – Diagrama em caixa (<i>boxplot</i>).	78

Figura 22 – Diagrama em caixa e sua distribuição de probabilidade equivalente.	79
Figura 23 – Teste pareado para dois espaços amostrais.	84
Figura 24 - Teste para mais de 2 espaços amostrais.	85
Figura 25 – Histograma para diferentes canais de eletrodos em conjunto com a distribuição normal sobreposta.	88
Figura 26 – Protocolo para geração da curva ROC. A variável l representa o atraso adicionado para a construção da matriz de informação. A variável i representa o número da realização. Score (l, i) é o cálculo da pontuação da época em análise.	90
Figura 27 – Curva ROC para o sujeito S_9	91
Figura 28 – Curvas ROCs para os sujeitos (a) S_5 e (b) S_6	92
Figura 29 - Protocolo para otimização de parâmetros e geração das maiores acurácias médias. O parâmetro os depende do banco de dados, sendo definido como 3,5s para o IIIA e 0,5s para o IIA. A variável l representa o atraso adicionado para a construção da matriz de informação. A variável h representa a quantidade de pares de filtros. A variável i representa o número de realizações da etapa de classificação. Acurácia (O, l, h, i) é o cálculo da acurácia de classificação das épocas de execução disponíveis de acordo com os parâmetros internos da função.	93
Figura 30 – Comparação de acurácias médias entre o CSP proposto e o clássico, usando todos os eletrodos disponíveis. Em conjunto com a acurácia está o respectivo desvio padrão para cada participante.	95
Figura 31 – Comparação de acurácias médias entre o CSP proposto e o clássico, usando 7 eletrodos posicionados sobre o córtex motor. Em conjunto com a acurácia, está o respectivo desvio padrão para cada participante.	96
Figura 32 – Protocolo para avaliação de métodos regularizados. A variável l representa o atraso adicionado para a construção da matriz de informação. A variável i representa o número de realizações da etapa de classificação. Acurácia (l, i) é o cálculo da acurácia de classificação das épocas de execução disponíveis de acordo com os parâmetros internos da função.	98
Figura 33 – Comparação entre os métodos, com e sem atrasos. (a) B_1 , (b) B_2 e (c) B_3	100
Figura 34 – Protocolo para avaliação de problemas quaternários. A variável l representa o atraso adicionado para a construção da matriz de informação. A variável i representa o número de realizações da etapa de classificação. Acurácia (l, i) é o cálculo da acurácia de classificação das épocas de execução disponíveis de acordo com os parâmetros internos da função.	102
Figura 35 – Comparação do método proposto com o clássico para a classificação de 4 movimentos motores.	104

LISTA DE TABELAS

Tabela 1 – Funções Kernel mais comuns.	64
Tabela 2 – Matriz de confusão para um problema envolvendo duas classes.	66
Tabela 3 – Assimetria e curtose média para os canais C_3 , C_z e C_4	87
Tabela 4 – Acurácia do classificador CSP-FLDA para imagética motora de mão direita e esquerda. A primeira coluna apresenta o método clássico ($L = 0$) e as demais o método proposto para $L = \{1, 2, \dots, 5\}$. A última linha apresenta a média de cada coluna em conjunto com o desvio padrão. A coluna $L = 0$ é a de referência para comparações múltiplas. Em negrito e azul são destacadas as distribuições de maior média amostral.	89
Tabela 5 – Acurácias médias para cada sujeito de acordo com os seus parâmetros ótimos dentro do espaço de busca. A última linha representa a média e a variância global. As duas últimas colunas apresentam os parâmetros ótimos encontrados dentro do espaço de busca. Em negrito e azul, as maiores acurácias. Na tabela, O representa o <i>offset</i> , L o atraso e W é referente à quantidade de pares de filtros espaciais selecionados.	94
Tabela 6 – Acurácias médias para cada sujeito de acordo com os seus parâmetros ótimos dentro do espaço de busca utilizando apenas 7 canais. A última linha representa a média e o desvio padrão para todos os sujeitos. As duas últimas colunas, representam os parâmetros ótimos obtidos dentro do espaço de busca. Em negrito e azul, as maiores acurácias. Na tabela, O representa o <i>offset</i> , L o atraso e W é referente à quantidade de pares de filtros espaciais selecionados.	96
Tabela 7 – Comparação entre os métodos com e sem atrasos para os sujeitos do banco de dados IIIA. Em azul e negrito os melhores resultados. A penúltima coluna, representa a média e o desvio padrão, considerando todos os participantes da respectiva linha. A última coluna indica o atraso utilizado na matriz de informação respectiva a cada sujeito: $L = \{B_1, B_2, B_3\}$	99
Tabela 8 – Comparação entre os métodos com e sem atraso para estimação da matriz de informação. Banco de dados IIA. Em negrito e azul são destacados os melhores resultados. A última coluna, apresenta a média e o desvio padrão do conjunto de sujeitos.	100
Tabela 9 – Atrasos ótimos para os métodos propostos da Tabela 8: $L = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9\}$	101
Tabela 10 – Experimento de classificação quaternária para os métodos CSP clássico ($L = 0$) e CSP proposto ($L = 1$). Em negrito e azul são apresentadas as maiores acurácias. Em caso de duplicidade, significa que não há diferença estatística entre os métodos.	103

LISTA DE ABREVIATURAS E SIGLAS

α_r	Ritmo alfa
β_r	Ritmo beta
δ_r	Ritmo delta
θ_r	Ritmo teta
γ_r	Ritmo gama
μ_r	Ritmo mu
AUC	Área abaixo da curva
B ₁ , B ₂ e B ₃	Participantes do banco de dados IIIA
BCI	Interface cérebro-máquina
Ch ₁	Eletrodo principal
Ch ₂	Eletrodo de referência
CCSP	CSP composto
CSP	Padrões espaciais comuns
DLCSP	CSP regularizado com carregamento
DLCSPauto	CSP regularizado com carregamento diagonal automático
DLCSPcv	CSP regularizado com carregamento diagonal e validação-cruzada
DLCSPcvd	CSP regularizado com carregamento diagonal e validação-cruzada diferente
EEG	Eletroencefalografia
ERP	Potenciais relacionados a eventos
ERD	Dessincronização relacionada a eventos
ERS	Sincronização relacionada a eventos
FLDA	Análise de discriminantes lineares de Fisher
FN	Falsos negativos
FP	Falsos positivos
fMRI	Imagem por ressonância magnética funcional
fNIRS	Espectroscopia funcional de infravermelho próximo
GLRCSP	CSP regularizado com aprendizado genérico

HCI	Interface homem-computador
Hz	Hertz
IBGE	Instituto Brasileiro de Geografia e Estatística
K ⁺	Cátion de potássio
KL	Kullback-Leibler
KNN	Classificador de K vizinhos mais próximos
LDA	Análise de discriminantes lineares
MEG	Magnetoencefalografia
Na ⁺	Cátion de sódio
PET	Tomografia por emissão de pósitrons
PPSE	Potencial pós-sináptico excitatório
PPSI	Potencial pós-sináptico inibitório
ROC	<i>Receiving operating characteristics</i>
S ₁ , S ₂ ... S ₉	Participantes do banco de dados IIA
SCP	Potenciais corticais lentos
SMR	Ritmos sensores-motor
SNC	Sistema nervoso central
SNR	Razão sinal-ruído
SRCSP	CSP regularizado espacialmente
SSVEP	Potenciais evocados visuais em regime permanente
STFT	Transformada de Fourier de tempo curto
SVM	Máquina de vetores de suporte
TFN	Taxa de falsos negativos
TFP	Taxa de falsos positivos
TVP	Taxa de verdadeiros positivos
VN	Verdadeiros negativos
VP	Verdadeiros positivos
WTRCSP	CSP regularizado por pesos de Tikhonov

LISTA DE SÍMBOLOS

A	Parâmetro de regularização
α	Nível de significância estatística
α_c	Parâmetro de regularização
β_3	Assimetria
β_4	Curtose
ε_g	Esfericidade de Greenhouse-Geisser
ζ	Variável de folga
ζ, φ	Parâmetros de regularização
μ_3	Momento central de terceira ordem
μ_4	Momento central de quarta ordem
μ_d	Média do vetor \mathbf{d}
μ_p	Média do método proposto
$\tilde{u}_p\{.\}$	Mediana do argumento para o método proposto
μ_r	Média do método de referência
$\tilde{u}_r\{.\}$	Mediana do argumento para o método de referência
λ	Multiplicador de Lagrange
γ_n	Vetor de componentes para a época n
ρ	Classe a ser analisada
σ	Desvio padrão
δ_s	Parâmetro de kernel
κ_s	Parâmetro de kernel
$\phi(\mathbf{t})$	Função que mapeia T em F
Ω	Conjunto de sujeitos
$\lfloor . \rfloor$	Arredondamento para baixo
\mathbf{A}	Vetor para definição do hiperplano
a_1, a_2	Eletrodos
$ac(f)$	Acurácia do classificador
A	Classe A - movimento da mão esquerda
B	Constante para definição do hiperplano
b_1, b_2	Atrasos
B	Classe B – movimento da mão direita

C	Coluna
C	Classe C - movimento da língua
C_A	Matriz de covariância média para a classe A
\hat{C}_A	Matriz de covariância regularizada para a classe A
C_B	Matriz de covariância média para a classe B
\hat{C}_B	Matriz de covariância regularizada para a classe B
C_{ρ}^i	Matriz de covariância espacial para a classe ρ e sujeito i
C_{ρ}	Matriz de covariância da classe ρ
\hat{C}_{ρ}	Matriz de covariância regularizada
\tilde{C}_{ρ}	Matriz de covariância usando atrasos
C_s	Termo de regularização SVM
D	Parâmetro do Kernel polinomial
\mathbf{d}	Vetor contendo amostras de um canal de EEG
d_q	Intervalo interquartilico
D	Classe D – movimento dos pés
\mathbf{D}_g	Matriz diagonal gerada por uma função específica
\mathbf{D}_w	Matriz diagonal
$\det\{.\}$	Determinante do argumento
$\text{diag}\{.\}$	Diagonal do argumento
$\text{err}(f)$	Taxa de erro total
$\text{err}-(f)$	Taxa de erro na classe negativa
$\text{err}+(f)$	Taxa de erro na classe positiva
$\text{esp}(f)$	Especificidade do classificador
$F_m(f)$	Medida F
$F_1(f)$	Medida F_1
$f(\cdot)$	Classificador binário de imagética motora
$f(\gamma_n)$	Classificador binário em função da componente
F	Espaço de características
G	Número total de grupos ou espaços amostrais
\mathbf{G}_{ρ}	Matriz de covariância genérica
H	Número de vetores (características extraídas)
H_o	Hipótese nula

H_I	Hipótese alternativa
\mathbf{I}	Matriz identidade
$J(\mathbf{w})$	Função de Rayleigh
$J_p(\mathbf{w})$	Função de Rayleigh com regularização
\mathbf{K}	Matriz de penalidade
$KL(i,t)$	Divergência de KL entre o sujeito desejado t e sujeito i
L	Linha
L	Número de atrasos
$L(\lambda, \mathbf{w})$	Função Lagrangiana
$\log\{\cdot\}$	Logaritmo do argumento
$\text{mod}\{\cdot\}$	Resto da divisão do argumento
m_a	Média para a classe A
\mathbf{M}_A	Autovetores para a classe A
m_B	Média para a classe B
\mathbf{M}_B	Autovetores para a classe B
M_C	Número de épocas usadas para calcular a matriz de covariância
m_l	Média da l -ésima linha
$MCC(f)$	Coefficiente de Matthews do classificador
Min	Minimizar
max	Maximizar
N	Época de processamento
N_{ρ}^i	Número de épocas de EEG usadas para estimar \mathbf{C}_{ρ}^i
$N_{t,\rho}$	Número total de épocas de EEG para a classe ρ
N_A	Número de autovetores para a classe A
N_B	Número de autovetores para a classe B
N_d	Número de amostras do vetor \mathbf{d}
N_e	Número de eletrodos utilizados
N_h	Quantidade de maiores e menores autovalores associados a \mathbf{Z}
N_{ρ}	Número de épocas pertencentes a classe ρ

o	Ordem desejada para a construção da matriz de covariância
p	Probabilidade de significância
P	Quantidade de amostras por canal
$P(\mathbf{w})$	Função de penalização
$\text{prec}(f)$	Precisão do classificador
q_i	Quartil inferior
q_s	Quartil superior
Q	Número de eletrodos
R	Hiperparâmetro que define quão perto estão dois eletrodos
\mathbb{R}^2	Espaço bidimensional
\mathbb{R}^3	Espaço tridimensional
s_ρ	Parâmetro de escala
S_ρ	Matriz de covariância para a classe ρ
$\text{sens}(f)$	Sensibilidade do classificador
\mathbf{t}	Vetor no espaço de entrada, com coordenadas t_1 e t_2
\mathbf{t}_i	Vetor no espaço de entrada i
T	Espaço de entrada
$\tanh\{.\}$	Tangente hiperbólica do argumento
$\text{tr}\{.\}$	Traço do argumento
\mathbf{v}_i	Vetor com as coordenadas espaciais
$\text{var}\{.\}$	Variância do argumento
\mathbf{w}	Vetor de coeficientes do filtro espacial
\mathbf{w}_h^i	Enésimo filtro espacial h obtido usando CSP para o sujeito i
$\mathbf{w}_G(i)$	Nível de penalização para o canal i
\mathbf{W}	Matriz de filtragem espacial
\mathbf{x}_n	Vetor que contém as amostras do EEG dos Q eletrodos no instante de tempo n

$\tilde{\mathbf{x}}_n$	Vetor com o sinal atrasado em L amostras no instante de tempo n
\mathbf{X}	Época contendo sinais de EEG multicanal
\mathbf{X}^{ρ_n}	Cada uma das N_ρ épocas, associadas à classe ρ , disponíveis para treinamento
$\tilde{\mathbf{X}}_n$	Matriz contendo os sinais originais e os com atraso
$\hat{\mathbf{X}}$	Matriz contendo os sinais originais, com atraso e momentos de ordem superior
y_i	Classes da SVM, -1 e 1
y_l	Variância do vetor para a l -ésima linha de \mathbf{Y}_n
$y_{l,c}$	Elemento da l -ésima linha e c -ésima coluna de \mathbf{Y}_n
\mathbf{y}_n	Sinal filtrado no instante n
\mathbf{Y}_n	Matriz contendo os sinais filtrados por \mathbf{W}
\mathbf{Z}	Autovetores que maximizam a diferença entre as classes
Z_r	Escalar usado no método de CCSP2

SUMÁRIO

<u>1</u>	<u>INTRODUÇÃO.....</u>	<u>23</u>
1.1	JUSTIFICATIVA.....	25
1.2	OBJETIVO.....	26
1.2.1	OBJETIVO GERAL.....	26
1.2.2	OBJETIVOS ESPECÍFICOS	26
1.3	ESTRUTURA DO TRABALHO.....	27
<u>2</u>	<u>FUNDAMENTAÇÃO TEÓRICA</u>	<u>28</u>
2.1	INTERFACE CÉREBRO-MÁQUINA	28
2.2	GERAÇÃO DO SINAL ELÉTRICO NO CÉREBRO	30
2.3	POTENCIAIS ELÉTRICOS CEREBRAIS	33
2.4	ELETOENCEFALOGRAFIA.....	39
2.4.1	RITMOS DA ATIVIDADE CEREBRAL.....	40
2.5	IMAGÉTICA MOTORA E POTENCIAIS ERD/ERS.....	41
2.6	POSICIONAMENTO DOS ELETRODOS PARA AQUISIÇÃO DOS SINAIS.....	42
2.7	CONSIDERAÇÕES FINAIS	43
<u>3</u>	<u>SISTEMA BCI UTILIZANDO IMAGÉTICA MOTORA</u>	<u>45</u>
3.1	PROCESSAMENTO DE SINAIS EM SISTEMAS BCI	45
3.2	ABORDAGEM PARA QUATRO CLASSES	46
3.3	BANCO DE DADOS	47
3.3.1	GERAÇÃO DOS SINAIS DE IMAGÉTICA MOTORA.....	49
3.4	MÉTODO DE CSP.....	50
3.5	CSP COM REGULARIZAÇÃO	52
3.5.1	CSP COMPOSTO	54
3.5.2	CSP REGULARIZADO COM APRENDIZADO GENÉRICO.....	55
3.5.3	CSP REGULARIZADO COM CARREGAMENTO DIAGONAL.....	56
3.5.4	CSP REGULARIZADO ESPACIALMENTE	56

3.5.5	CSP COM REGULARIZAÇÃO DE PESOS DE TIKHONOV.....	57
3.6	EXTRAÇÃO DE CARACTERÍSTICAS	58
3.7	CLASSIFICADORES	59
3.7.1	CLASSIFICADOR FLDA.....	59
3.7.2	CLASSIFICADOR SVM	60
3.7.2.1	SVM linear.....	61
3.7.2.2	SVM não linear	62
3.8	CONSIDERAÇÕES FINAIS	64
4	<u>AVALIAÇÃO DO CLASSIFICADOR</u>	<u>66</u>
4.1	MEDIDAS DE DESEMPENHO	66
4.2	CURVAS ROC	69
4.3	CONSIDERAÇÕES FINAIS	70
5	<u>ANÁLISE ESTATÍSTICA.....</u>	<u>71</u>
5.1	CONCEITOS PRIMÁRIOS DE ESTATÍSTICA	71
5.1.1	HIPÓTESE NULA E ALTERNATIVA.....	71
5.1.2	NÍVEL DE SIGNIFICÂNCIA	72
5.1.3	TESTES UNILATERAIS E BILATERAIS	72
5.2	INFERÊNCIA ESTATÍSTICA USANDO MEDIDAS REPETIDAS	74
5.2.1	INFERÊNCIA ESTATÍSTICA PARA DOIS ESPAÇOS AMOSTRAIS.....	74
5.2.2	INFERÊNCIA ESTATÍSTICA PARA MAIS DE DOIS ESPAÇOS AMOSTRAIS.....	75
5.3	ASSIMETRIA E CURTOSE.....	77
5.4	DIAGRAMA EM CAIXAS.....	78
5.5	CONSIDERAÇÕES FINAIS	79
6	<u>PROPOSTA DE CONTRIBUIÇÃO E PROTOCOLOS PARA AVALIAÇÃO ESTATÍSTICA.....</u>	<u>80</u>
6.1	FUNDAMENTAÇÃO.....	80
6.2	MATRIZ DE INFORMAÇÃO.....	81
6.2.1	MATRIZ DE INFORMAÇÃO EXPANDIDA USANDO MOMENTOS DE SEGUNDA ORDEM	81
6.3	PROTOCOLOS ESTATÍSTICOS	83
6.4	CONSIDERAÇÕES FINAIS	85

7	RESULTADOS.....	86
7.1	DESCRIÇÃO GERAL.....	86
7.2	ANÁLISE DE MOMENTOS DE ORDEM SUPERIOR.....	87
7.3	CLASSIFICADOR CSP-FLDA.....	88
7.4	OTIMIZAÇÃO DE PARÂMETROS INDIVIDUAIS.....	92
7.5	INSERÇÃO DOS ATRASOS EM OUTROS MÉTODOS DE FILTRAGEM ESPACIAL.....	97
7.6	ABORDAGEM USANDO QUATRO CLASSES	101
7.7	CONSIDERAÇÕES FINAIS	104
8	CONCLUSÃO	106
9	PROPOSTAS FUTURAS	107
	REFERÊNCIAS.....	108
	APÊNDICE A – TABELA COMPLETA DE CLASSIFICAÇÃO QUATERNÁRIA DE CLASSES COM ATRASOS E DOIS BANCOS DE DADOS. EM NEGRITO E AZUL AS MAIORES ACURÁCIAS.	115

1 INTRODUÇÃO

A interface homem-computador (HCI, do inglês, *Human-Computer Interface*) é um campo de pesquisa e desenvolvimento em ascensão, de modo que um grande esforço tem sido dedicado para projetar sistemas ergonômicos e de fácil manuseio para os usuários. Isso é feito por meio de interfaces inovadoras para visão, áudio e outros dispositivos de realidade virtual (EBRAHIMI, GARICA e VESIN, 2003). As pesquisas em HCI visam criar novos canais de comunicação entre o homem e o computador para o controle, monitoração e interação dos elementos no seu entorno.

Pessoas com doenças neurológicas, acometidas por elevado grau de paralisia, podem se beneficiar com o uso de tecnologias assistivas para uma melhor comunicação e mobilidade (BELHADJ, BERNOUMASSAT e JRACHAI, 2015). Para alguns casos, o uso da atividade cerebral pode ser a última forma de interação com o meio ambiente (BELHADJ, BERNOUMASSAT e JRACHAI, 2015).

Uma das técnicas para a compensação artificial de deficiências graves de mobilidade é chamada de interface cérebro-máquina (BCI, do inglês, *Brain-Computer Interface*) (EBRAHIMI, GARICA e VESIN, 2003). Uma definição mais abrangente de um sistema de BCI é que ele consiste em um sistema de comunicação no qual mensagens ou comandos cerebrais são enviados para o ambiente externo sem a participação de ações musculares, com a finalidade de acionamento de um determinado sistema (BELHADJ, BERNOUMASSAT e JRACHAI, 2015).

Devido aos incessantes avanços na área de instrumentação biomédica, é possível mensurar e monitorar processos fisiológicos relativos ao interior do corpo humano para os quais não há uma interface natural para esta finalidade como, por exemplo, as ondas cerebrais.

Para a medição das ondas cerebrais, podem ser usados diversos métodos não invasivos, como a eletroencefalografia (EEG), a magneto-eletroencefalografia (MEG), a tomografia por emissão de pósitrons (PET, do inglês, *Positron Emission Tomography*) e a imagem por ressonância magnética funcional (fMRI, do inglês, *Functional Magnetic Resonance Image*). O EEG tem como vantagem a facilidade de realização e o baixo custo do equipamento de aquisição dos sinais em relação aos demais métodos, porém, apresenta como desvantagens a baixa relação sinal ruído e a menor resolução espacial.

Uma importante área de pesquisa em BCI para aplicações biomédicas é o desenvolvimento de sistemas que possam auxiliar pessoas com deficiências graves de

comunicação ou mobilidade a obter uma melhor qualidade de vida. As interfaces cérebro-máquina se tornaram uma área promissora de pesquisa, em diversos campos, como: neurociência, engenharia e processamento de sinais. Uma das razões para isso, é a disponibilidade cada vez maior de recursos computacionais para a implementação em tempo real das técnicas envolvidas (BLANKERTZ, TOMIOKA, *et al.*, 2008). Os sistemas baseados em BCI podem ser usados em uma gama enorme de aplicações, exemplos: ferramenta terapêutica (restauração do controle motor e tratamento de deficiências cognitivas e emocionais) e construção de interfaces de controle (cadeira de rodas motorizada) (WOLPAW e WOLPAW, 2012).

Os sistemas de comunicação BCI permitem aos usuários o envio de comandos ao computador usando somente a sua própria atividade cerebral, que pode ser mensurada por sistemas de aquisição de EEG (LOTTE e GUAN, 2011).

As técnicas de BCI têm diferentes características, apresentando vantagens e desvantagens também distintas (CHOI, RHIU, *et al.*, 2017). De acordo com o modo escolhido para evocar a geração de sinais cerebrais é possível obter maior acurácia na classificação dos dados e facilidade na aplicação do estímulo necessário. Alguns estímulos podem causar desde enjoos até fadiga ocular, cabendo ao projetista escolher o melhor método (CHOI, RHIU, *et al.*, 2017).

Técnicas BCI são geralmente projetadas a partir de uma etapa de extração de características, seguida por um classificador, cuja finalidade é identificar a ação desejada pelo usuário (LOTTE e GUAN, 2011).

A técnica de padrões espaciais comuns (CSP, do inglês, *Common Spatial Patterns*) é um método de extração de características que pode autoprojetar filtros espaciais para maximizar a separação entre duas classes. O CSP é um dos métodos mais populares e eficientes para sistemas de BCI, comumente aplicado em competições de BCI (LOTTE e GUAN, 2011). Há, entretanto, diversas outras técnicas que podem ser utilizadas com a mesma finalidade, como o filtro de CSP regularizado (LOTTE e GUAN, 2011), a STFT (ANNETT, 1996), algoritmos genéticos (YAACOUB, MHANNA, e RIHANA, 2017) e banco de filtros (MAHMOOD, ZAINAB, *et al.*, 2017).

A imagética motora é uma forma particular de sistema BCI, que consiste em relacionar movimentos motores com o pensamento associado a uma determinada imagem (ANNETT, 1996). Neste trabalho pretende-se estudar técnicas de classificação de sinais EEG relacionadas à imagética motora, voltadas para o desenvolvimento de sistemas BCI.

Ao longo desta dissertação, inicialmente é apresentada a abordagem clássica da técnica CSP e, posteriormente, é proposto um método para adição de informação temporal em sua matriz de informações, visando melhorar os índices de acurácia do sistema BCI.

1.1 JUSTIFICATIVA

Lesões medulares podem acarretar limitações como paraplegia e tetraplegia. Em alguns casos, é de suma importância para o incremento da qualidade de vida dessas pessoas a disponibilização de métodos de auxílio para a realização de determinadas ações necessárias ao seu cotidiano (WOLPAW e WOLPAW, 2012). Uma tecnologia de auxílio é a técnica de BCI que, entre outras possibilidades de aplicação, pode ser utilizada no controle de uma cadeira de rodas motorizada (LEEB, FRIEDMAN, *et al.*, 2007).

De acordo com o censo demográfico brasileiro de 2000, apresentado pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2000), a deficiência física pode ser classificada em 4 tipos:

- tetraplegia: a paralisia permanente total de ambos os braços e pernas;
- paraplegia: a paralisia permanente das pernas;
- hemiplegia: a paralisia permanente de um dos lados do corpo;
- falta de um membro ou de parte dele, como perna, braço, mão, pé ou dedo.

Esses 4 tipos de deficiência foram agrupados pelo IBGE, de modo que cerca de 937.463 brasileiros podem ser enquadrados nessa situação (IBGE, 2010). Em 2010, o censo demográfico brasileiro não apresentou essa classificação, apenas uma situação mais ampla, referente a pessoas com dificuldade de locomoção (pequena, média, grande e extrema), sendo que 6,8% (6.352.664) da população feminina total apresenta esse problema contra 4,5% (4.203.969) da masculina. Esses números são referentes aos dados conjuntos dos quatro tipos de classificações apresentadas em 2000, não havendo dados parciais para cada um deles (IBGE, 2010). Apesar da falta de dados mais detalhados sobre o número de pessoas com deficiências profundas de mobilidade, há uma clara necessidade por tecnologias assistivas de mobilidade e comunicação, de forma que o interesse por técnicas BCI tem aumentado (WOLPAW e WOLPAW, 2012).

A área médica não é a única aplicabilidade de sistemas de BCI, eles podem ser usados em aplicações de entretenimento, educação, segurança, desenvolvimento de ambientes inteligentes, entre outros (ABDULKADER, ATIA e MOSTAFA, 2015).

1.2 OBJETIVO

Esta seção apresenta os objetivos gerais e específicos propostos para esta dissertação.

1.2.1 Objetivo geral

Estudar e propor métodos de processamento e classificação de sinais EEG provenientes da técnica de imagética motora para sistemas de BCI com a finalidade de realizar uma contribuição para a matriz de informação utilizada pelo método de CSP e assim aumentar a acurácia de classificação.

1.2.2 Objetivos específicos

Os seguintes objetivos específicos são propostos:

- estudar a influência dos momentos estatísticos de segunda ou maior ordem no problema de classificação dos dados da imagética motora;
- estudar a influência do número de canais de aquisição de sinais cerebrais nos índices de classificação da imagética motora, visando a diminuir o número de eletrodos em sistemas embarcados de BCI;
- comparar o desempenho de diferentes técnicas de filtragem espacial para classificação de tarefas de imagética motora;
- propor e avaliar o desempenho de um sistema BCI baseado em imagética motora para movimentos binários e quaternários.

1.3 ESTRUTURA DO TRABALHO

Esta dissertação é dividida em 9 capítulos que são: introdução, fundamentação teórica, descrição de um sistema de BCI, avaliação do classificador, análise estatística, proposta de contribuição, resultados, conclusão e propostas para continuação do trabalho.

A parte introdutória concentrou-se em apresentar a motivação, justificativa e objetivos da dissertação.

O capítulo 2 apresenta a fundamentação teórica necessária para o entendimento deste trabalho. Inicialmente é apresentado um breve histórico da área de BCI para em seguida apresentar os princípios básicos de geração dos sinais eletroencefalográficos (EEG). A partir da geração do sinal EEG são apresentados os ritmos cerebrais que podem ser relacionados a estados mentais associados à imagética motora. Em sequência, são apresentados os princípios de geração de potenciais relacionados a eventos e o porquê de eles serem usados em sistemas de BCI, além de configurações de eletrodos no escalpo para mensurar esses sinais. Finalizando o capítulo são descritos os bancos de sinais de imagética motora utilizados nesse trabalho.

No capítulo 3 é apresentada a descrição geral das etapas de processamento de um sistema de BCI para classificação de movimentos binários (mão esquerda e direita, nesta dissertação), incluindo as técnicas de CSP, extração de características e classificação. Além disso, é apresentado o método para classificação de movimentos quaternários (mão esquerda, mão direita, pés e língua).

Os capítulos 4 e 5 apresentam os métodos de avaliação de desempenho da técnica proposta. O capítulo 4 trata sobre aprendizado de máquina, por meio de métricas como: sensibilidade, acurácia, precisão, curvas ROC, entre outras. Já o capítulo 5 trata sobre métodos estatísticos para avaliação das métricas do capítulo 4, com a finalidade de determinar a relevância dos resultados.

A contribuição desta dissertação é apresentada no capítulo 6. Ela consiste em um novo método para compor a matriz de informações do método CSP com a inserção de atrasos nos canais EEG. Ainda, são apresentados dois protocolos para avaliação estatística.

O capítulo 7 apresenta os resultados obtidos, avaliados por meio das métricas de aprendizado de máquina e de inferência estatística.

O capítulo 8 apresenta as conclusões finais deste trabalho.

Por fim, o capítulo 9 contém sugestões para o prosseguimento do tema desta dissertação.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os tópicos principais para a compreensão de sistemas de BCI baseados na técnica de imagética motora. Inicialmente é apresentado um breve histórico da interface cérebro máquina para, em seguida, apresentar a forma pela qual são gerados os sinais elétricos captados pela técnica de eletroencefalografia. Por fim, são apresentados métodos para posicionamento dos eletrodos e captação de sinais elétricos cerebrais para utilização em técnicas de BCI que empregam imagética motora, como o sistema 10-20.

2.1 INTERFACE CÉREBRO-MÁQUINA

A área de interfaces cérebro máquina inicia em 1924 quando Hans Berger, professor de psiquiatria da Universidade de Jena na Alemanha, descobre que os sinais elétricos produzidos no cérebro podem ser mensurados no escalpo (WOLPAW e WOLPAW, 2012).

Em 1929, Berger publica o primeiro de seus 14 artigos que estabelecem a eletroencefalografia (EEG) como um campo de pesquisa cerebral e uma ferramenta para auxílio ao diagnóstico (WOLPAW e WOLPAW, 2012).

Em 1968, Kamiya demonstrou que o ser humano poderia controlar as ondas cerebrais do tipo alfa após determinado treinamento. Esse foi o começo do campo de estudo conhecido como neuro-realimentação (*neurofeedback*) que, posteriormente, atraiu o interesse de vários pesquisadores interessados no desenvolvimento de novas tecnologias (LOTTE, NAM e NIJHOLT, 2017).

Em 1969, o pesquisador Fetz e seus colaboradores demonstraram em ratos de laboratório a possibilidade de aprendizado de um único neurônio cortical para o controle de um alimentador de comida (WOLPAW e WOLPAW, 2012).

Em 1970, o termo BCI foi usado pela primeira vez por Jacques Vidal (WOLPAW e WOLPAW, 2012). Vidal descreveu BCI como um sistema baseado em computador capaz de produzir informações detalhadas sobre funções cerebrais (WOLPAW e WOLPAW, 2012). Passados alguns anos, Vidal usou a técnica de potenciais evocados gerados no córtex visual e adquiridos sob o escalpo para determinar o movimento da íris e a direção de deslocamento de um cursor desejado por um voluntário (WOLPAW e WOLPAW, 2012).

A partir de 1971 e até o final de 1988, a área de BCI não obteve avanços significativos, permanecendo estagnada, provavelmente pela dificuldade de implementação dessa nova

tecnologia, que exige uma eletrônica com certo grau de precisão na aquisição e processamento dos sinais (LOTTE, NAM e NIJHOLT, 2017).

A área de BCI muda de patamar quando, no período de 1988 ao começo de 1990, diversos pesquisadores se unem para implementar o primeiro sistema em tempo real de BCI aplicando diversos paradigmas que ainda são usados atualmente (LOTTE, NAM e NIJHOLT, 2017).

Um dos paradigmas mais famosos, o de ortografia baseada em P300 foi desenvolvido em 1988 por Farwell e Donchin. Eles propuseram um sistema de BCI para seleção de sinais ortográficos (letras) em um *grid* apresentado em um monitor. O potencial P300 é uma resposta relacionada a eventos (ERP, do inglês, *Evented Related Potential*), sendo registrada de forma agregada ao EEG (LOTTE, NAM e NIJHOLT, 2017).

Posteriormente, pesquisadores desenvolveram um paradigma, baseado em ritmos sensores-motor (SMR, do inglês, *Sensorimotor Rhythm*), que consiste na análise do ritmo μ_r (7-13 Hz) do EEG, obtido sobre a parte sensório motora do córtex cerebral (LOTTE, NAM e NIJHOLT, 2017). Essa técnica foi utilizada por Jonathan Wolpam para controlar a descida e a subida de uma bola por meio do controle de amplitude da SMR (LOTTE, NAM e NIJHOLT, 2017).

Na mesma época do desenvolvimento da SMR, Gert Pfurscheller e sua equipe desenvolveram outro sistema de BCI baseado também na SMR. Nele, os usuários deveriam imaginar o movimento das mãos esquerda e direita, os quais eram identificados pelo computador usando aprendizado de máquina (LOTTE, NAM e NIJHOLT, 2017). Esse experimento representa o início da área de imagética motora (LOTTE, NAM e NIJHOLT, 2017).

Na década de 90, foram desenvolvidos outros paradigmas de aquisição de sinais como: potenciais corticais lentos (SCP, do inglês, *Slow Cortical Potentials*) e a resposta N400 (LOTTE, NAM e NIJHOLT, 2017).

Com a área bem consolidada, a partir do ano 2000, surgem novos paradigmas de aquisição de sinais, além do desenvolvimento de novas técnicas para processamento. Atualmente, o método mais utilizado e considerado como padrão, é o CSP (LOTTE, NAM e NIJHOLT, 2017). Adicionalmente, formas robustas de classificação usando máquinas de vetor de suporte e redes neurais têm sido costumeiramente empregadas para a classificação dos sinais de EEG em sistemas de BCI (LOTTE, NAM e NIJHOLT, 2017).

A fim de estimular e difundir os projetos em BCI voltados para reabilitação de pessoas, em 2001, um grupo de pesquisadores criou uma competição com a finalidade de desenvolver métodos para predição de movimentos motores, na qual foram utilizados diferentes tipos de potenciais (P300, ERD/EDS, entre outros). Essa competição foi criada pela Universidade de Graz na Áustria e já realizou quatro edições. Há diversos tipos de desafios como: classificação de imagética motora usando duas ou quatro classes, treinamento de classificadores com pequenos bancos de dados, identificação e supressão de artefatos presentes nos sinais de aquisição (movimento dos olhos, por exemplo), entre outros.

Em 2002, a definição de sistemas BCI foi ampliada, sendo descrita então como um sistema que mede a atividade do sistema nervoso central (SNC) e converte-a em informação artificial para substituir, restaurar, aprimorar, suplementar ou melhorar a produção natural do SNC à medida que altera as interações em andamento nos ambientes externo e interno (WOLPAW e WOLPAW, 2012).

Para fomentar a pesquisa e o desenvolvimento na área de BCI, levando a tecnologias que permitam que as pessoas interajam com o mundo por meio de sinais cerebrais, foi criada em 2005 a sociedade internacional de BCI (LOTTE, NAM e NIJHOLT, 2017).

Por meio desse breve histórico consegue-se ter uma ideia do desenvolvimento da área, que começou com Berger medindo ondas cerebrais até o emprego de novos paradigmas de aquisição desses sinais.

Do começo do século 21 até o ano de 2018, outros paradigmas visuais e auditivos para BCIs foram desenvolvidos, além de BCIs híbridas que possibilitam a combinação de dois ou mais paradigmas. Além disso, novas tecnologias de aquisição de sinais foram desenvolvidas como, por exemplo, espectroscopia de infravermelho próxima funcional (fNIRS, do inglês, *functional Near-Infrared Spectroscopy*) e eletrocorticografia.

Finalizando, muitas tecnologias assistivas estão sendo desenvolvidas e exploradas para, por exemplo, reabilitação de pessoas que sofreram acidente vascular cerebral, jogos, decodificadores de movimentos motores, aplicações artísticas, entre outros. (LOTTE, NAM e NIJHOLT, 2017). Conseqüentemente, a área está em constante expansão.

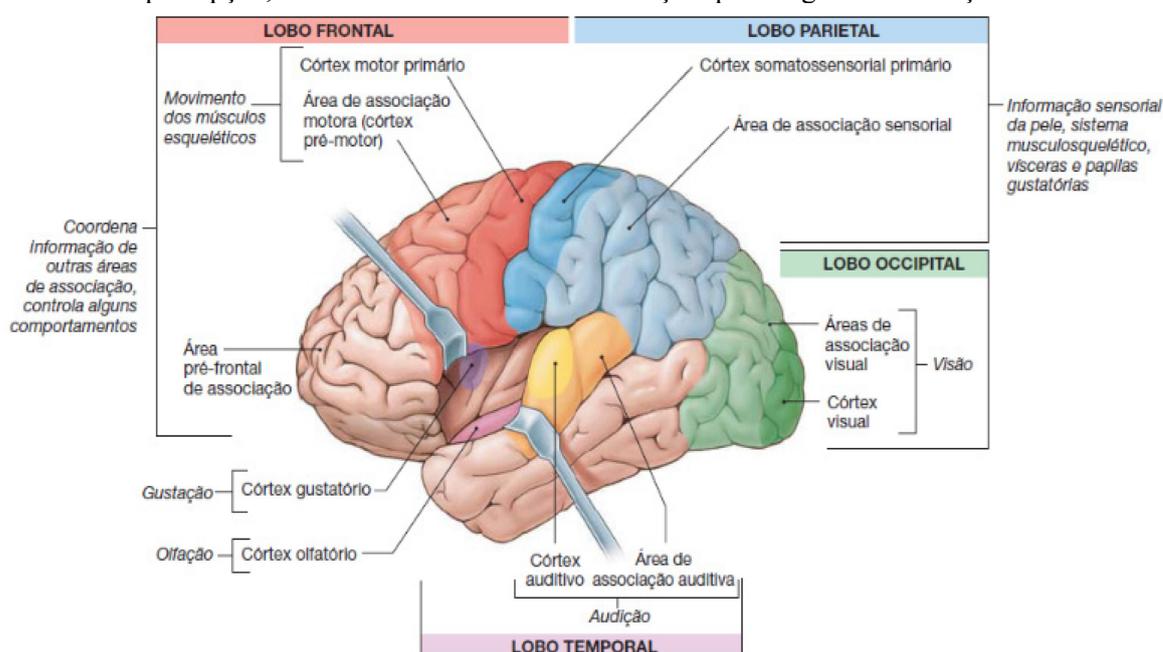
2.2 GERAÇÃO DO SINAL ELÉTRICO NO CÉREBRO

Ao contrário das simplificações populares, o cérebro não pode ser idealizado como um simples computador, no qual um processador central realiza todas as funções. O cérebro é um

conjunto complexo de subsistemas concorrentes, cada um especializado em tarefas particulares (NIJHOLT e TAN, 2010). Estudos dos efeitos de lesões cerebrais e o uso de novas tecnologias de imagens do cérebro possibilitaram que neurocientistas construíssem um mapa topográfico associando diferentes partes do cérebro físico com funções cognitivas distintas (NIJHOLT e TAN, 2010).

De maneira simplificada, o cérebro pode ser dividido em duas partes: o córtex cerebral e regiões subcorticais. As regiões subcorticais controlam as funções básicas como: respiração, frequência cardíaca, temperatura corporal, respostas emocionais e instintivas (tais como medo), reflexos, aprendizado e memória (NIJHOLT e TAN, 2010). O córtex cerebral, por sua vez, inclui a maior parte do processamento sensorial e motor, incluindo funções como: planejamento, processamento da fala e reconhecimento de padrões. A Figura 1 ilustra o córtex cerebral e sua divisão em lobos temporais, de modo que cada região corresponde ao processamento de um tipo de função cognitiva humana.

Figura 1 – Áreas funcionais do córtex cerebral. O córtex cerebral contém áreas sensoriais para percepção, áreas motoras e áreas de associação que integram informações.



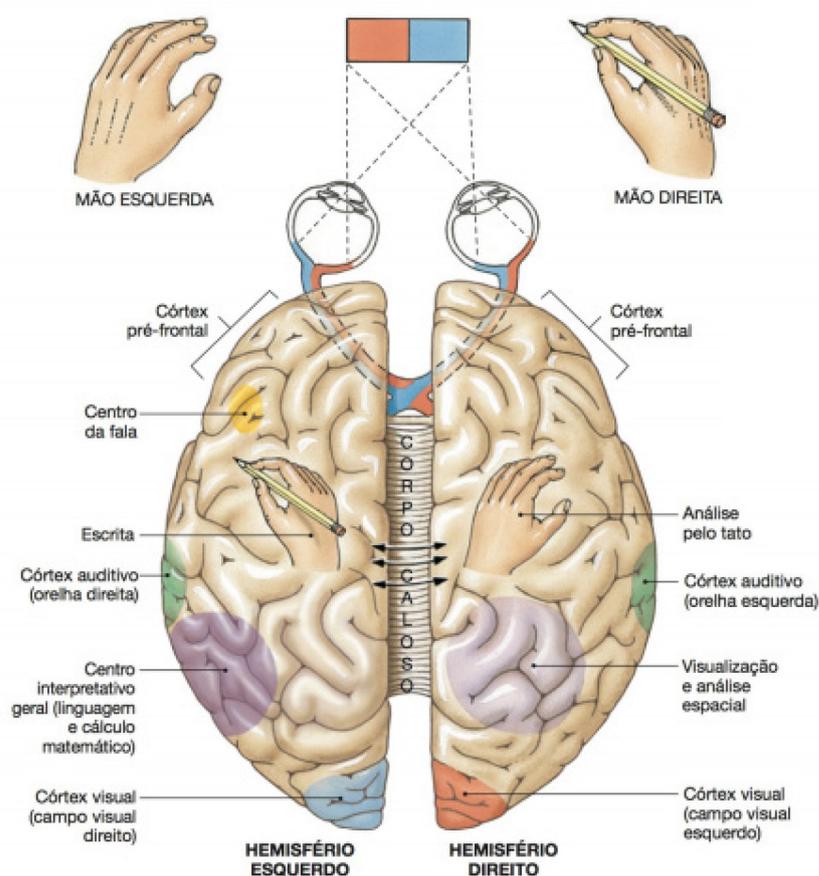
Fonte: (SILVERTHORN, 2010).

Ao examinar o córtex do ponto de vista funcional ele pode ser dividido em três áreas: sensorial, motora e associativa. A área sensorial recebe estímulos sensoriais realizando sua tradução em percepções; a área motora é responsável por comandar os movimentos dos músculos e a área associativa integra informações das anteriores, podendo comandar comportamentos voluntários (SILVERTHORN, 2010). Uma outra classificação para o cérebro

é relacionada à separação dos hemisférios em esquerdo e direito, ao analisá-lo sob uma vista superior (SILVERTHORN, 2010), conforme ilustrado na Figura 2.

Um aspecto importante apontado por Silverthorn (2010) é o fenômeno chamado de lateralização cerebral, que corresponde à dominância de um hemisfério sobre o outro. Essa lateralização implica que certas tarefas cognitivas de um hemisfério são dominantes sobre as do outro. Uma representação dessa dominância é a divisão entre destros e canhotos. Da mesma forma, em caso de ocorrência de um dano cerebral no hemisfério esquerdo, a parte correspondente ao lado direito do corpo também é afetada. Adicionalmente, habilidades associadas a um determinado hemisfério podem ser desenvolvidas pelo outro.

Figura 2 – Hemisférios e lateralização cerebral.



Fonte: (SILVERTHORN, 2010).

Para problemas de classificação binária em imagética motora, os eletrodos devem ser preferencialmente posicionados sobre o córtex motor, o qual está localizado sob o lobo frontal (GE, WANG e YU, 2014). Em pessoas com tetraplegia, o lobo parietal, que usualmente encontrava-se associado à sensibilidade dos dedos passa a se apresentar de forma desativada.

Algumas vezes pode estar ativo, devido a extensões (invasões) dos campos funcionais de regiões adjacentes do córtex (SILVERTHORN, 2010).

2.3 POTENCIAIS ELÉTRICOS CEREBRAIS

Os neurônios são responsáveis pela transmissão e processamento de informação. São células especializadas com terminações que se estendem a partir do corpo celular (SILVERTHORN, 2010). Existem dois tipos de terminações: os dendritos (responsáveis pela recepção dos sinais de outros neurônios) e os axônios (transmissão dos sinais para outros neurônios) (SILVERTHORN, 2010). Essas características permitem que eles se comuniquem entre si e com outras células (SILVERTHORN, 2010). A forma, o comprimento e a quantidade de dendritos e axônios variam de acordo com o neurônio.

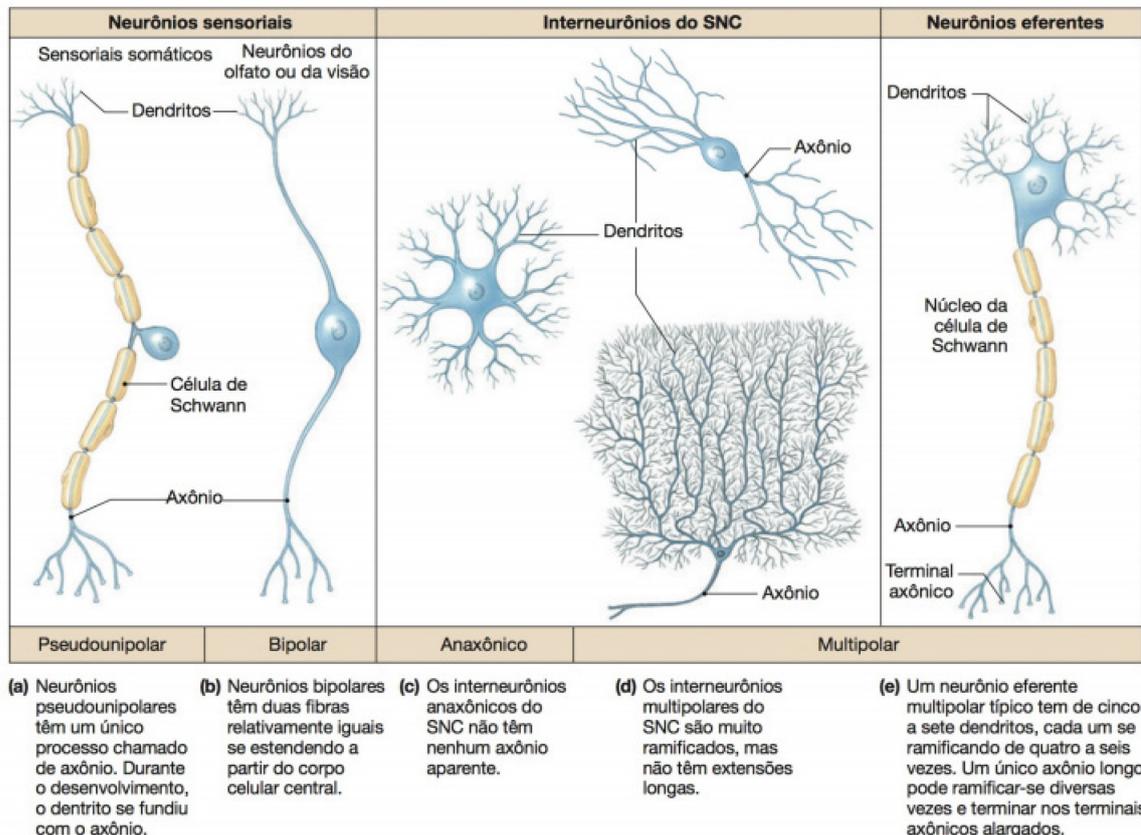
A classificação dos neurônios pode ser feita estruturalmente, de acordo com sua morfologia e número de prolongamentos originados no corpo celular, podendo ser: pseudounipolares (axônios e dendritos se fundem para criar uma única terminação longa), bipolares (um único axônio e outro dendrito), multipolares (vários axônios e dendritos modificados) ou axônicos (sem axônio identificável) (SILVERTHORN, 2010). Funcionalmente, eles são classificados em três tipos: neurônios aferentes (transmitem informações do corpo, como temperatura e pressão, para o sistema nervoso central), interneurônios (neurônios completamente inseridos no sistema nervoso central) e neurônios eferentes ou efetores (transmitem informações do sistema nervoso central para o restante do corpo como, por exemplo, para geração de ações motoras voluntárias) (SILVERTHORN, 2010). A classificação dos neurônios de acordo com sua funcionalidade pode ser visualizada na Figura 3.

A comunicação entre neurônios é feita por meio de sinais elétricos, reações químicas ou eletroquimicamente (SILVERTHORN, 2010).

Os sinais elétricos são gerados pelo movimento de íons. Isso quer dizer que a membrana celular ao aumentar sua permeabilidade ao cátion de sódio (Na^+) faz com que ele acabe entrando na célula, em decorrência do gradiente eletroquímico. Isso acarreta a adição de Na^+ ao líquido intracelular, despolarizando a membrana celular e assim gerando um sinal elétrico (SILVERTHORN, 2010). A permeabilidade iônica é alterada pela abertura e fechamento dos canais iônicos (SILVERTHORN, 2010). Há diferentes tipos de canais, com propriedades distintas entre eles. Para abertura de um canal é necessário atingir determinado

limiar de tensão. Outro aspecto importante é que os canais se abrem em resposta à despolarização e se fecham durante a repolarização da célula (SILVERTHORN, 2010).

Figura 3 – Categorias morfológicas e funcionais dos neurônios.



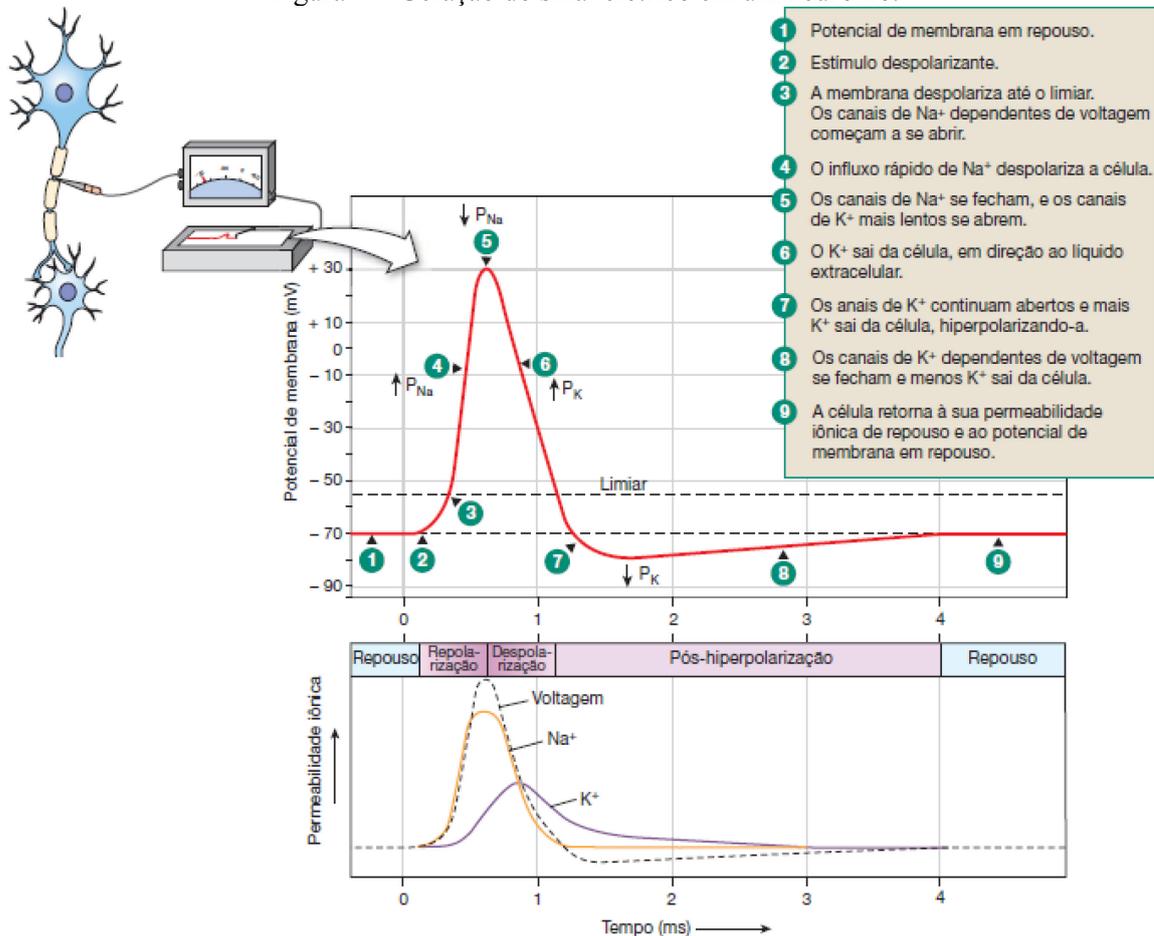
Fonte: (SILVERTHORN, 2010).

Quando os canais iônicos se abrem, há a possibilidade dos íons se moverem para dentro e fora da célula. Esse fluxo de carga gera corrente elétrica. O fluxo de íons dentro da membrana polariza ou hiperpolariza a célula gerando um sinal elétrico, que pode acarretar dois tipos de respostas: o potencial de ação ou o potencial graduado (SILVERTHORN, 2010). A principal diferença entre ambos é que o potencial de ação deve percorrer grandes distâncias por um neurônio sem perder a intensidade, enquanto que o graduado é para pequenas distâncias, perdendo força ao longo do caminho.

Os potenciais de ação ocorrem quando canais iônicos controlados por tensão se abrem, alterando a permeabilidade da membrana para os cátions de sódio (Na^+) e potássio (K^+) (SILVERTHORN, 2010). A Figura 4 ilustra as mudanças de tensão e da permeabilidade iônica que ocorrem em um segmento da membrana durante um potencial de ação, sendo o gráfico dividido em três fases: a ascendente, a descendente e a de pós-hiperpolarização. Um exemplo

apresentado por Silverthorn (2010) para um estímulo de grande distância é a transmissão de um sinal da ponta do dedo da mão até a medula espinhal.

Figura 4 – Geração do sinal elétrico em um neurônio.



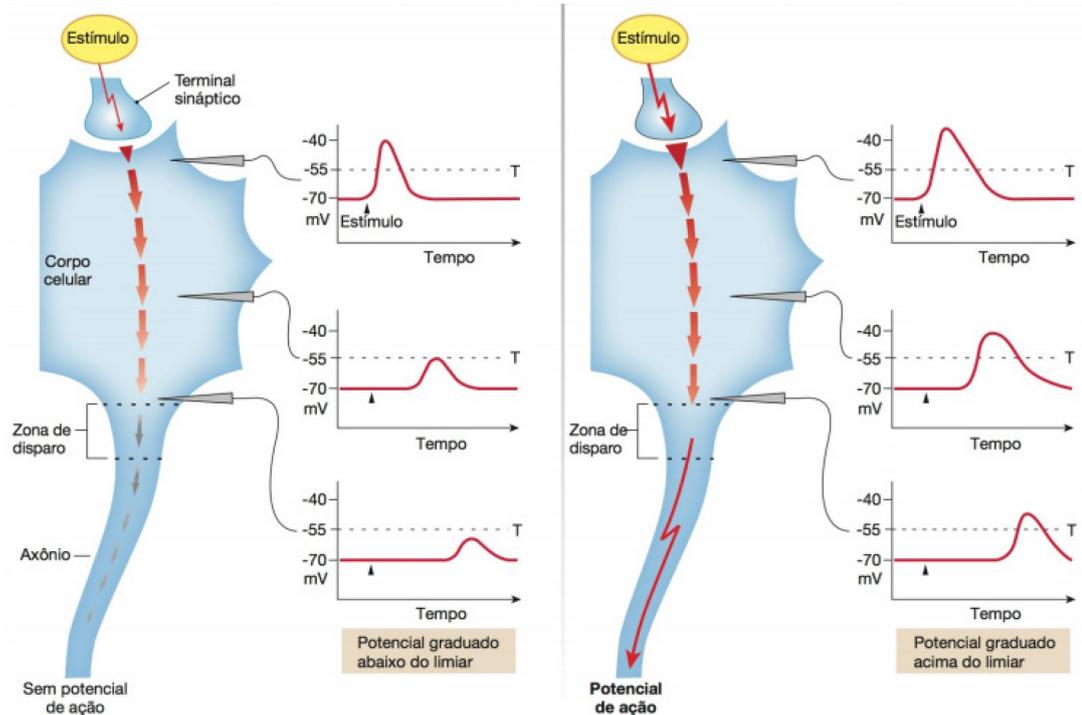
Fonte: (SILVERTHORN, 2010).

Os potenciais de gradação são sinais de força variada que perdem intensidade à medida que percorrem a célula. Se eles forem suficientemente fortes e acima de um limiar de tensão, ao atingir a região integradora do neurônio, inicia-se um potencial de ação, conforme pode ser visto na Figura 5. Esse potencial é utilizado para a realização de comunicações em pequenas distâncias e é conhecido como tudo ou nada, pois depende de um limiar para sua ocorrência.

A visão dada até esse ponto permite a análise de somente um neurônio, mas não permite responder à pergunta: como eles se comunicam entre si e com outras células? O fluxo de informação ocorre na sinapse, que são conexões anatômicas entre os neurônios e seus alvos (SILVERTHORN, 2010). Cada sinapse tem duas partes: o terminal axônico da célula pré-sináptica e a membrana da célula pós-sináptica, conforme apresentado na Figura 6. A

informação se move da célula pré-sináptica para a pós-sináptica (essa não precisa ser necessariamente um neurônio). Por fim, a sinapse pode ser elétrica ou química, dependendo do sinal que passa por cada uma delas.

Figura 5 – Geração de um potencial de ação a partir de um potencial graduado.



Fonte: (SILVERTHORN, 2010).

Segundo Silverthorn: “as sinapses elétricas são responsáveis por transferirem correntes iônicas, diretamente do citoplasma de uma célula para outras por meio de juntas comunicantes (SILVERTHORN, 2010)”.

A principal característica da sinapse elétrica é a velocidade de condução dos sinais entre células (SILVERTHORN, 2010). As sinapses elétricas apresentam as seguintes desvantagens: a falta de capacidade para transformação do sinal excitatório de um neurônio em outro inibitório, como ocorre nas sinapses químicas; não apresentam a propriedade de modulação do sinal e o seu fluxo de íons não pode ser bloqueado (SILVERTHORN, 2010).

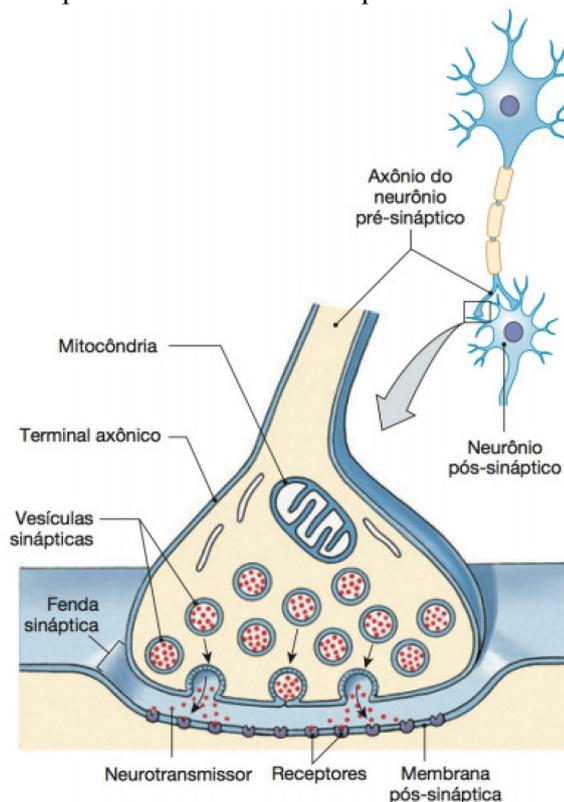
Conforme visto anteriormente, as sinapses elétricas são limitadas. Uma forma do cérebro melhorar a comunicação entre neurônios é usando sinapses químicas.

As sinapses químicas são o meio de comunicação entre neurônios mais utilizado pelo cérebro, isso ocorre devido à sua complexidade e capacidade de proporcionar um controle maior sobre o processamento dos sinais entre neurônios (SILVERTHORN, 2010).

A sinapse química é mais complexa que a elétrica pelo fato dela poder controlar a quantidade de neurotransmissores e de receptores que são liberados, possibilitando assim, o

controle da energia e dos circuitos neuronais (SILVERTHORN, 2010). Outra forma de entender a complexidade de sinapses químicas é por meio da modulação sináptica.

Figura 6 – Os neurotransmissores formam a mensagem em uma sinapse química. O terminal axônico contém mitocôndrias e vesículas sinápticas preenchidas com neurotransmissores. A membrana pós-sináptica possui receptores para o neurotransmissor que se difunde através da fenda sináptica.



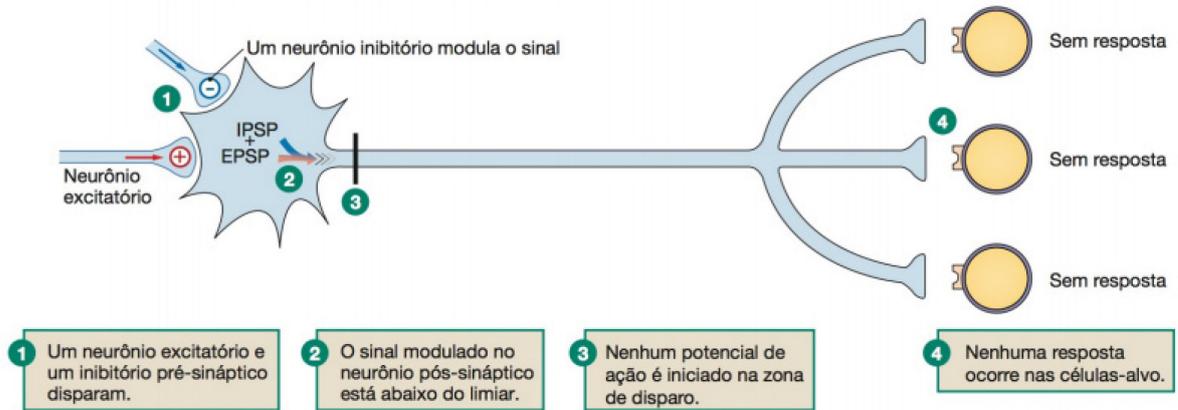
Fonte: (SILVERTHORN, 2010).

A modulação ou plasticidade sináptica é a capacidade de inibir ou potencializar a atividade na sinapse (SILVERTHORN, 2010). O processo de estimular ou inibir é feito pelo neurônio modulador que age tanto nas prés quanto nas pós-sinapses.

A inibição pós-sináptica pode ocorrer quando um neurônio pré-sináptico libera um neurotransmissor inibitório em uma célula pós-sináptica e altera sua resposta, inibindo assim o potencial de ação (SILVERTHORN, 2010). A integração pós-sináptica ocorre quando múltiplos sinais atingem um neurônio, fazendo com que eles sejam unidos e produzam um único sinal com determinada amplitude e frequência. Conforme visto anteriormente, se o sinal integrado estiver acima de um limiar de tensão então é gerado um potencial de ação, caso contrário, o neurônio não dispara. Anatomicamente, na modulação pós-sináptica, quando um neurônio modulatório faz sinapse nos dendritos ou no corpo celular de um neurônio, ocorrem alterações no comportamento de todo o neurônio pós-sináptico, fazendo com que as suas células-alvo sejam afetadas conforme ilustrado na Figura 7.

Figura 7 – Inibição pós-sináptica.

(b) Na inibição pós-sináptica, todos os alvos do neurônio pós-sináptico serão igualmente inibidos.

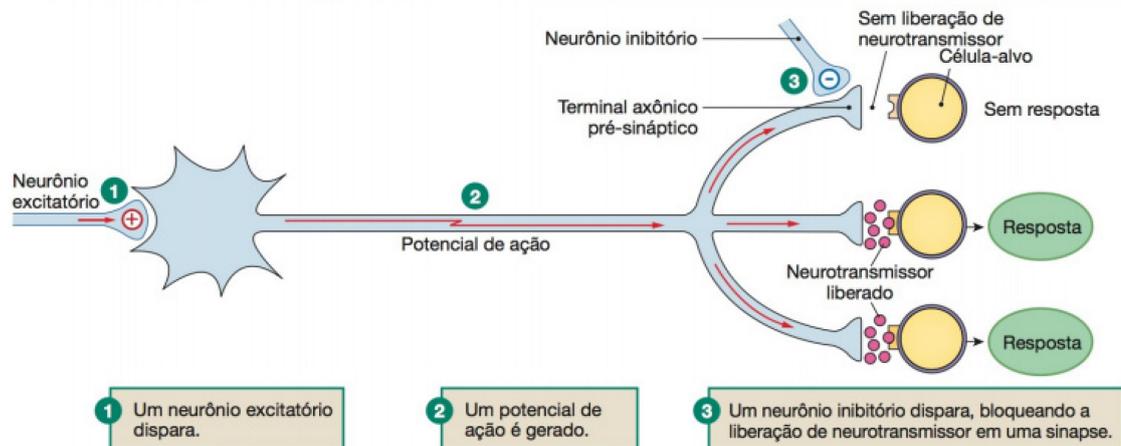


Fonte: (SILVERTHORN, 2010).

A atividade das células pré-sinápticas também pode ser modificada, alterando-se os seus potenciais de ação. Há duas formas de se alterar o potencial de ação e gerar uma modulação pré-sináptica: a forma excitatória e a inibitória (SILVERTHORN, 2010).

Figura 8 – Inibição pré-sináptica.

(a) Na inibição pré-sináptica, um neurônio modulatório faz sinapse em um colateral do neurônio pré-sináptico e seletivamente inibe um alvo.



Fonte: (SILVERTHORN, 2010).

Quando um neurônio modulador (inibitório ou excitatório) se encontra agregado ao terminal axônico de uma célula pré-sináptica ou próximo dela, seus potenciais pós-sinápticos excitatórios (PPSE) e potenciais pós-sinápticos inibitórios (PPSI) podem alterar o potencial de ação que alcança o terminal, gerando assim uma modulação pré-sináptica (SILVERTHORN, 2010).

A modulação por inibição pré-sináptica ocorre quando há diminuição da liberação de neurotransmissores pelo neurônio modulador (SILVERTHORN, 2010). Já a modulação por excitação ocorre quando há aumento na liberação de neurotransmissores devido à entrada

modulatória na célula pré-sináptica, sendo esse processo chamado de facilitação pré-sináptica (SILVERTHORN, 2010).

A inibição pré-sináptica permite a modulação seletiva de seus colaterais (ramificações do axiônio de um neurônio) e de seus alvos (SILVERTHORN, 2010). Um colateral pode ser inibido à medida que os outros não são afetados, possibilitando assim, um controle maior do processo na sinapse pré-sináptica no que na pós-sináptica, conforme ilustrado na Figura 8.

2.4 ELETROENCEFALOGRAFIA

Quando os neurônios são ativados, há a geração de reações químicas dentro do cérebro, fazendo com que surjam correntes sinápticas. Essas correntes geram campos elétricos no couro cabeludo, que podem ser medidos por equipamentos de eletroencefalografia (SILVA, 2017).

O conceito de eletroencefalograma (EEG), é originado da composição de 3 palavras: eletro (referente ao registro de sinais elétricos), encéfalo (referente ao cérebro) e grama (ou grafia que significa escrita ou desenho), de modo que essa técnica surgiu com Richard Caton (1862 – 1926), cientista inglês que fez o primeiro registro desse sinal no escalpo humano em 1875 (SILVA, 2017).

O eletroencefalograma usa eletrodos posicionados diretamente no escalpo do paciente para medir os potenciais elétricos gerados pela atividade cerebral, sendo esses sinais compreendidos entre 5 e 100 μV (NIJHOLT e TAN, 2010). Por causa dos fluidos, ossos e pele que separam os eletrodos da origem dos potenciais elétricos, os sinais adquiridos sofrem espalhamento e a adição de interferências. Embora a medida de EEG tenha boa resolução temporal, com atraso na casa de algumas dezenas de milissegundos, a resolução espacial tende a ser baixa, na faixa de 2 a 3 centímetros de acurácia em sua melhor configuração (NIJHOLT e TAN, 2010).

O EEG permite a medida de variações de potenciais elétricos entre um eletrodo principal (Ch_1) e outro de referência (Ch_2), ou ainda na forma diferencial ($\text{Ch}_1 - \text{Ch}_2$), sendo cada par de eletrodos chamado de derivação (PINTO, 2011).

Existem dois tipos de derivação:

- monopolar: todas as derivações são referidas a um único eletrodo, sendo útil para a visualização de potenciais ao longo do escalpo;
- bipolar: cada derivação tem sua própria referência, sendo útil para a visualização de potenciais em localizações específicas no escalpo.

De acordo com Pinto (2011), para a derivação monopolar os eletrodos principais são posicionados na região onde ocorre a atividade elétrica de interesse, e o de referência onde há pouca atividade elétrica, normalmente longe dos eletrodos principais. Por esse motivo, os potenciais espalhados não atingem o eletrodo de referência, mas podem ser observados nos eletrodos principais (PINTO, 2011). Alguns posicionamentos comuns para o eletrodo de referência são: na posição central do sistema 10-20 (SILVA, 2017); nos mastoídes; na orelha ipsilateral ou contralateral; e na ponta do nariz.

Na derivação bipolar, os dois eletrodos, ficam próximos um do outro, de modo que o potencial espalhado seja eliminado devido à operação de subtração (PINTO, 2011).

Atualmente, o EEG pode ser realizado tanto de forma invasiva como não-invasiva (SILVA, 2017). Por fim, há outros métodos para medição da atividade cerebral como: magnetoencefalografia, eletrocorticografia, imageamento por ressonância magnética funcional e espectroscopia funcional de infravermelho próximo (fNIRS), entre outros.

2.4.1 Ritmos da atividade cerebral

Os sinais de EEG permitem a análise dos ritmos cerebrais, possibilitando a identificação de certos estados mentais, bem como de doenças do sistema nervoso (SILVA, 2017).

Os principais ritmos cerebrais são descritos a seguir:

- **ritmo delta (δ_r):** apresenta faixa de frequência característica de 0,5 a 4 Hertz (Hz) e está associado principalmente ao estado de sono profundo, podendo estar presente quando o indivíduo está acordado. Por se concentrar em frequências relativamente baixas, pode ser confundido com outros sinais de origem não cerebral, como a atividade muscular (SILVA, 2017);
- **ritmo teta (θ_r):** possui frequência na faixa de 4 a 7,5 Hz e aparece na transição do estado de consciência para sonolência (SILVA, 2017);
- **ritmo alfa (α_r):** este ritmo aparece na região occipital do cérebro e compreende a faixa de frequências entre 8 e 13 Hz. Indica estado de relaxamento, sem qualquer atenção ou concentração (SILVA, 2017);

- **ritmo beta (β_r):** compreendido entre 14 e 26 Hz. Está relacionado ao estado de atenção e pensamentos ativos, foco no mundo exterior ou resolução de problemas concretos (SILVA, 2017);
- **ritmo gama (γ_r):** está associado a frequências maiores que 30 Hz. Pode ser indicativo de sincronização associada a um evento, embora seja bastante raro e possua amplitude menor que 2 μ V (SILVA, 2017) e
- **ritmo mu (μ_r):** é um tipo especial de ritmo alfa que é observado na faixa de frequências entre 9 e 11 Hz, estando quase sempre presente quando o sujeito está relaxado. Desaparece quando o sujeito move uma mão ou um dedo no hemisfério oposto, logo essa banda permite a distinção entre mão esquerda e direita (SILVA, 2017).

Vários fatores indicam que os ritmos μ_r e β_r podem fornecer informações confiáveis para sistemas de comunicação baseados em EEG (WOLPAW, BIRBAUMER, *et al.*, 2002). Essa afirmação é suportada pelo fato de que esses ritmos estão associados a áreas corticais conectadas a canais motores de saída do cérebro, podendo ser utilizadas então em aplicações relacionadas a tarefas de imagética motora das mãos.

2.5 IMAGÉTICA MOTORA E POTENCIAIS ERD/ERS

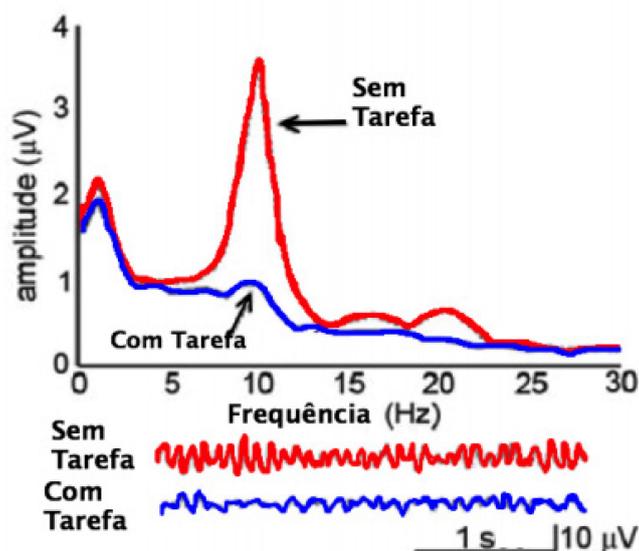
Sinais empregados em sistemas de BCI são oriundos de mudanças na atividade cerebral do indivíduo. Essas mudanças ocorrem devido a respostas a estímulos externos ou ao desenvolvimento de tarefas cognitivas pelo usuário (SILVA, 2017). Basicamente essas mudanças pertencem a dois grupos: potenciais relacionados a eventos (ERP, do inglês, *Event Related Potential*) ou à sincronização e dessincronização associada a eventos (ERD/ERS, do inglês, *Event Related Desynchronization/Synchronization*) (SILVA, 2017).

Segundo Silva (2017), exemplos de ERP são os potenciais evocados visuais em regime permanente (SSVEP, do inglês, *Steady State Visually Evoked Potentials*) e o P300. Esses potenciais são originados no cérebro por meio de estímulos externos, de forma visual, tátil ou audível (SILVA, 2017).

A modificação na amplitude dos ritmos corticais μ_r e β_r durante os movimentos voluntários auto-estimulados apresenta uma área promissora para o desenvolvimento de sistemas inteligentes usando esse sinal eletrofisiológico (YONGWOONG, NAM, *et al.*, 2011). Essas variações nos ritmos corticais são conhecidas como dessincronização/sincronização

relacionada a eventos e podem ser visualizadas na Figura 9. A alteração de potência nos sinais de EEG no domínio do tempo equivale a modificações específicas no domínio da frequência, podendo representar sincronia das populações neuronais (SILVA, 2018).

Figura 9 – A curva em azul representa a redução de energia (ERD) no espectro do EEG em torno de 10 Hz. Em vermelho, o espectro de EEG sem dessincronização.



Fonte: (SILVA, 2017).

A diminuição de amplitude dos ritmos μ_r e β_r , particularmente no hemisfério contralateral ao do movimento é chamada de ERD, sendo ocasionada pelo movimento ou preparação dele (WOLPAW, BIRBAUMER, *et al.*, 2002). Já a ERS ocorre após o movimento ou por meio de relaxamento corporal (WOLPAW, BIRBAUMER, *et al.*, 2002). Tanto a ERD quanto o ERS não exigem movimento, ocorrendo, portanto, em tarefas de imagética motora (movimentos imaginados) (WOLPAW, BIRBAUMER, *et al.*, 2002). Adicionalmente, as ERDs e ERSs indicam o sincronismo ou a falta dele em populações de neurônios, sendo ocasionados por mudanças nas interações locais entre neurônios e interneurônios, alterando o conteúdo em frequência do sinal de EEG (SILVA, 2018).

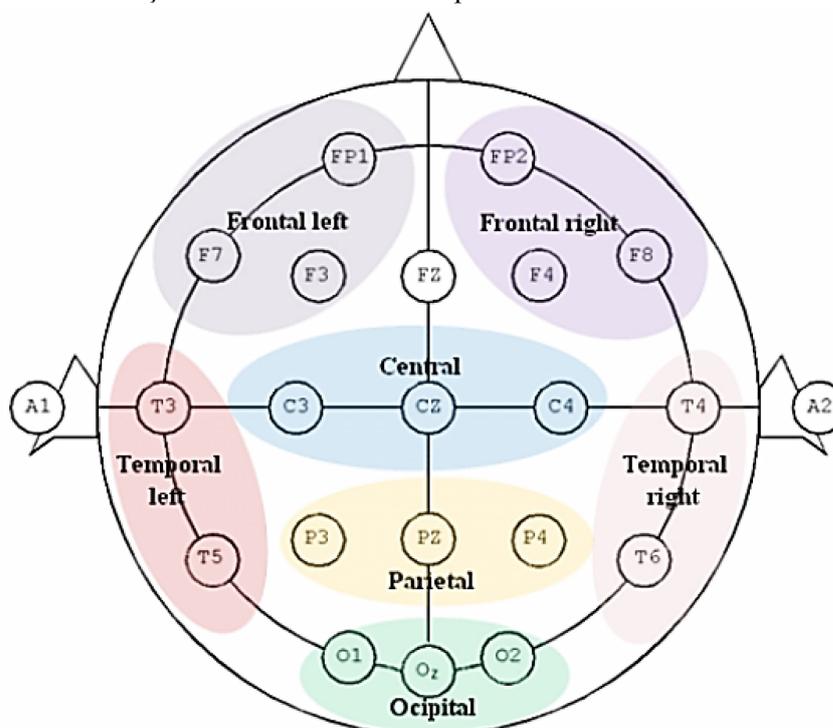
2.6 POSICIONAMENTO DOS ELETRODOS PARA AQUISIÇÃO DOS SINAIS

O sistema 10-20 é uma padronização internacional para o posicionamento de eletrodos no couro cabeludo (SILVA, 2017). Ele recebe esse nome, porque as distâncias reais entre eletrodos adjacentes são referentes a 10% ou 20% das distâncias totais frente-trás e esquerda-direita do crânio.

De acordo com o sistema 10-20, a localização anatômica dos eletrodos é feita usando letras e números, conforme pode ser visualizado na Figura 10, totalizando 21 posições. As letras indicam a região do cérebro onde os eletrodos são posicionados, de acordo com a seguinte nomenclatura: frontal (F), central (C), temporal (T), parietal (P) e occipital (O) (SILVA, 2017). Números ímpares indicam que os eletrodos estão no hemisfério esquerdo do cérebro e pares no hemisfério direito. A exceção são os que estão sob a linha central, denominados então pela letra z. Por exemplo, o eletrodo C₃, está localizado na região central e hemisfério esquerdo do cérebro.

Há outros tipos de configurações, como: 10-10 (74 eletrodos), 10-5 (142 eletrodos), entre outras (WOLPAW e WOLPAW, 2012).

Figura 10 – Posição dos eletrodos no escalpo de acordo com o sistema 10-20.



Fonte: (VALIULIS, 2014).

2.7 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentadas características anatômicas e fisiológicas do cérebro humano que permitem a compreensão do funcionamento do cérebro e da geração de sinais elétricos captados por equipamentos de eletroencefalografia. Adicionalmente, foram descritos os principais ritmos cerebrais gerados que podem ser utilizados em sistemas de BCI para

predição de movimentos motores. Por fim, foi apresentado o sistema internacional 10-20 para posicionamento dos eletrodos no couro cabeludo.

O próximo capítulo introduz uma classe de métodos de processamento para a predição de movimentos motores binários para sistemas BCI, como também os bancos de dados utilizados nesta dissertação.

3 SISTEMA BCI UTILIZANDO IMAGÉTICA MOTORA

Este capítulo descreve uma estrutura de processamento comumente encontrada na literatura científica para a implementação de sistemas BCI baseados em imagética motora. São apresentados os seguintes tópicos: banco de sinais de imagética motora, filtragem espacial, extração de características e classificadores. Além disso, contém protocolos para implementação de sistemas de BCI de imagética motora envolvendo problemas binários e quaternários.

3.1 PROCESSAMENTO DE SINAIS EM SISTEMAS BCI

Diversas estratégias de processamento podem ser utilizadas para a implementação de sistemas BCI baseados em imagética motora, entre elas: CSP (LOTTE e GUAN, 2011), algoritmos genéticos (YAACOUB, MHANNA, e RIHANA, 2017), banco de filtros (MAHMOOD, ZAINAB, *et al.*, 2017), modelos autorregressivos adaptativos (PFURTSCHELLER e NEUPER, 2001) e desigualdades matriciais lineares (SILVA, 2017). Resultados interessantes têm sido obtidos com estruturas de processamento baseadas em uma sequência de três estratégias: filtragem, extração de características e classificação (LOTTE e GUAN, 2011).

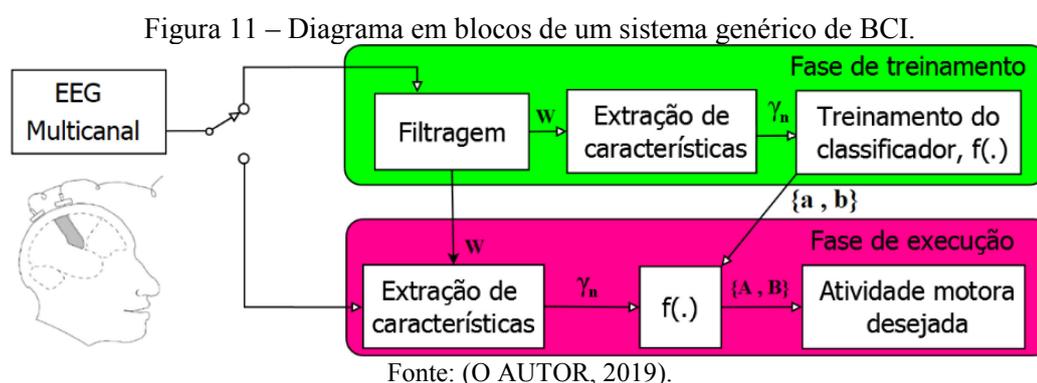
A Figura 11 ilustra o diagrama em blocos de um sistema genérico para BCI baseado em imagética motora para classificação de dois movimentos motores. A utilização do sistema é dividida em duas fases: treinamento e execução. Primeiramente, um conjunto de épocas¹ (\mathbf{X}) de EEG multicanal é utilizada para o treinamento e configuração dos parâmetros do sistema. Em sequência, na fase de execução, os parâmetros obtidos são utilizados para a operação do sistema. Dada a baixa relação sinal-ruído dos sinais de imagética, comumente emprega-se uma etapa de pré-processamento para destacar as informações de interesse (SILVA, 2017).

A fase de treinamento consiste, primeiramente, em calcular uma matriz de filtragem (\mathbf{W}) de forma a maximizar a diferença entre as classes que se deseja diferenciar. Em sequência, é realizada a etapa de extração de características do sinal EEG (γ_n). Por fim, são determinados

¹ Época é o registro dos sinais do EEG multicanal. Matematicamente é representada por uma matriz, na qual cada linha contém as amostras captadas a partir de um determinado eletrodo posicionado no escalpo e cada coluna refere-se ao instante de aquisição do sinal mensurado.

os parâmetros (no caso linear são dois parâmetros: \mathbf{a} é o vetor de pesos e b um escalar de *offset*) do modelo de classificação.

A etapa de execução corresponde à realização da tarefa de classificação. Os sinais de EEG são primeiramente filtrados pelos filtros \mathbf{W} , para então serem extraídas as características de interesse do sinal γ_n . Finalmente, é realizado o procedimento de classificação, que decidirá se o sinal pertence à classe A (movimento da mão esquerda) ou B (movimento da mão direita), no caso de um problema de imagética motora binária.



Finalizando, destaca-se que os sinais EEG multicanal devem ser previamente filtrados para ênfase da banda de frequência de interesse, geralmente 8-30 Hz para o problema de imagética motora de mãos (LOTTE e GUAN, 2011). A escolha da banda de frequência para filtragem dos sinais está relacionada aos ritmos da atividade cerebral (μ_r e β_r). Cada etapa da Figura 11 será discutida mais detalhadamente nas próximas seções.

3.2 ABORDAGEM PARA QUATRO CLASSES

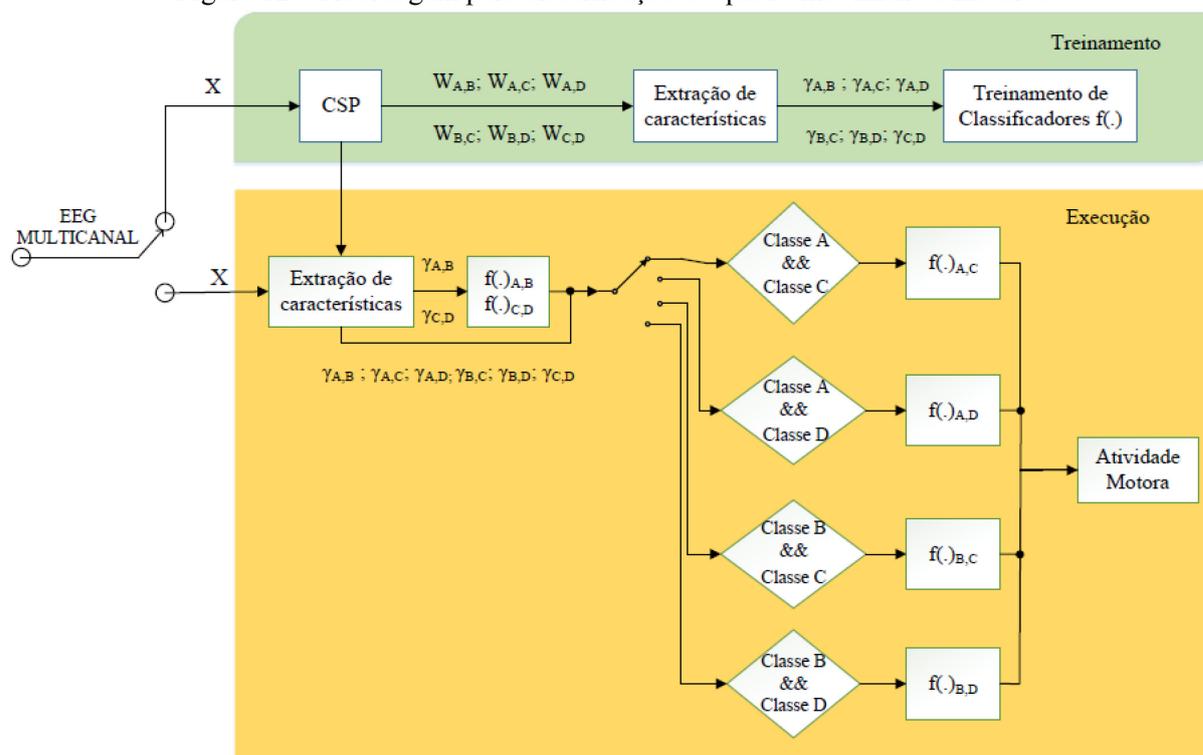
A técnica discutida na seção 3.1 permite somente a classificação de duas tarefas. A fim de ampliar a abrangência do problema de classificação, nesta seção é apresentado um método para realizar a classificação de quatro classes. As mesmas considerações utilizadas anteriormente para os sinais são válidas para esta abordagem.

A ideia de classificação motora em quatro classes é treinar 6 classificadores e determinar a atividade motora por meio de dois confrontos sucessivos entre eles, conforme ilustrado na Figura 12. Por exemplo, a primeira etapa consiste em dois confrontos para uma dada época: classe A \times classe B; e classe C \times classe D. A segunda etapa, consiste em utilizar os resultados dos confrontos anteriores. Por exemplo, se as classes escolhidas foram A e C, então é feito um embate entre elas e determina-se a atividade motora desejada.

O método é dividido em duas fases, treinamento e execução. O método de treinamento consiste em 6 classificadores associados a todas as combinações possíveis entre classes, ou seja, serão gerados os seguintes pares para $f(\cdot)$: AB, AC, AD, BC, BD e CD.

A etapa de execução calcula os vetores de características para todos os classificadores, mas somente alguns são utilizados. Primeiramente a época de EEG é aplicada em dois classificadores, no primeiro é decidido se a época é mais propensa a ser da classe A ou B e o segundo classificador separa entre C ou D. Por fim, de acordo com os dois rótulos obtidos é determinado o classificador adequado. Por exemplo, se na primeira fase for determinado o rótulo A no primeiro classificador e C no segundo, então será usado apenas o classificador AC, determinando-se assim a ação motora da época em processamento.

Figura 12 – Abordagem para classificação de quatro movimentos motores.



Fonte: (O AUTOR, 2019).

3.3 BANCO DE DADOS

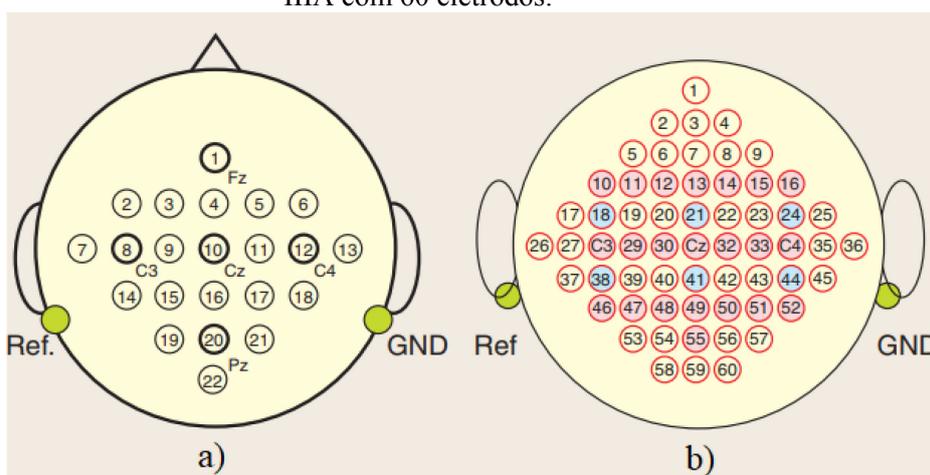
Neste trabalho são utilizados sinais EEG provenientes de dois bancos de dados disponibilizados gratuitamente. Ao longo desse trabalho, esses bancos de dados são denominados como IIA (BRUNNER, LEEB, *et al.*, 2008) e IIIA (GE, WANG e YU, 2014). Ambos contêm tarefas de imagética motora referentes a 4 classes (mão esquerda, mão direita,

ambos os pés e língua) e são amplamente utilizados na literatura científica para comparação entre diferentes técnicas.

O banco de dados IIIA, descrito em Ge, Wang, Yu (2014) é composto por três sujeitos, cujos rótulos adotados para cada um deles são designados por B₁, B₂ e B₃, e 60 eletrodos de EEG. Os sinais foram adquiridos usando um amplificador *neuroscan* e a disposição dos eletrodos no escalpo está ilustrada na Figura 13a. O mastoide² esquerdo é adotado como a referência, enquanto que o direito é utilizado como aterramento. Os sinais de EEG foram amostrados a 250 Hz e processados por um filtro passa-faixa com banda de 1 a 50 Hz. Um filtro *notch* em 50 Hz foi utilizado para supressão do ruído de linha. Esses sinais foram utilizados na competição de 2005 de BCI, sendo disponibilizados pela Universidade de Graz na Áustria (SCHLÖGL, 2005). Nessa competição foi utilizada a mesma quantidade de épocas para treinamento e validação, totalizando 420 (180 para o participante B₁ e 120 para cada um dos participantes, B₂ e B₃) épocas disponíveis.

O segundo banco de sinais, denominado IIA, fez parte da 4ª competição de BCI ocorrida no ano de 2008 (BRUNNER, LEEB, *et al.*, 2008). O banco de dados é composto por 9 participantes (denominados S₁-S₉) e sinais de 22 eletrodos de EEG havendo 2592 épocas para treinamento e a mesma quantidade para validação (288 épocas para cada participante). A disposição dos eletrodos é ilustrada na Figura 13b.

Figura 13 – Disposição de eletrodos: (a) banco de sinais IIA com 22 eletrodos; e (b) banco de sinais IIIA com 60 eletrodos.



Fonte: adaptado de (COYLE, 2009).

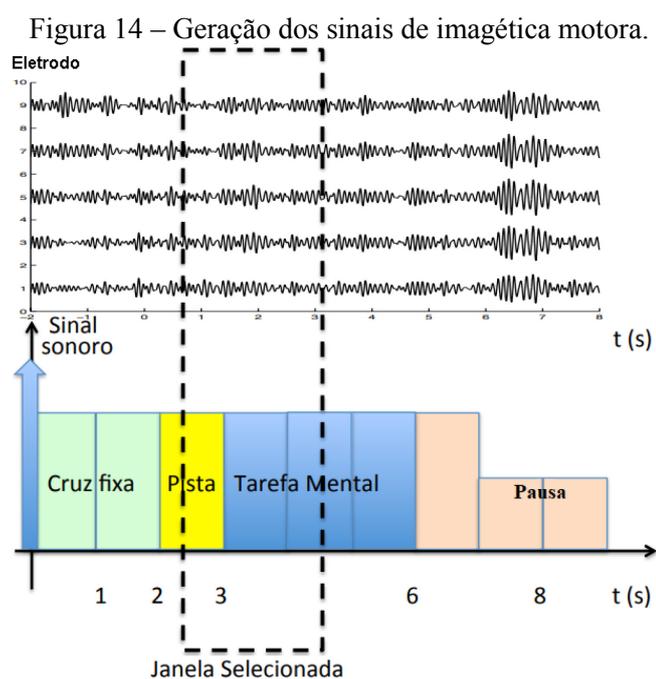
² O mastoide é uma projeção cônica que pode variar de tamanho e forma, localizada na parte posterior do osso temporal.

3.3.1 Geração dos Sinais de Imagética Motora

Os bancos de sinais IIA e IIIA foram gerados seguindo o mesmo paradigma (BRUNNER, LEEB, *et al.*, 2008).

O procedimento consiste no posicionamento do voluntário sentado confortavelmente em uma cadeira em frente a um monitor. No começo do experimento ($t = 0s$), uma cruz de fixação aparece na tela, em conjunto com um aviso sonoro. Em sequência, após dois segundos ($t = 2s$), uma pista na forma de uma flecha apontando para esquerda, direita, cima ou baixo (correspondendo a uma das quatro classes) aparece na tela do monitor e permanece por 1,25 s, voltando então à condição da cruz de fixação por mais 2,75 s. O voluntário deve executar a tarefa de imagética motora desejada (movimento de mão esquerda, mão direita, ambos os pés, ou língua) durante 4 s, ou seja, no período compreendido pelo aparecimento da pista ($t = 2s$) até o instante $t = 6s$. Finalizando, uma pequena pausa é realizada de 6 a 9 s. Ao longo dessa pausa, o monitor é colocado em tela preta novamente. Embora o intervalo de tempo entre 6 e 7 s faça parte da pausa, ele também contém a tarefa mental desejada, pois ainda há resquícios da imagética motora. No intervalo de 7-9 s, assume-se que o usuário está descansando e não há mais resquícios da tarefa motora solicitada. Por fim, o tamanho da janela selecionada pode variar (em cerca de 1 s) dependendo da aplicação (GE, WANG e YU, 2014) (LOTTE e GUAN, 2011).

A Figura 14 apresenta graficamente o paradigma temporal descrito nesta subseção.



Fonte: adaptado de (SILVA, 2017).

3.4 MÉTODO DE CSP

O CSP é um método utilizado para o projeto de filtros espaciais que possui o objetivo de obter uma discriminação ótima entre duas classes de sinais. Isso é obtido pela maximização da variância dos sinais de uma classe em contraposição à minimização da variância da outra (LOTTE e GUAN, 2011). O sinal resultante do processamento pelo filtro espacial CSP é determinado por uma combinação linear dos sinais em um determinado instante de tempo, de forma que:

$$y_n = \mathbf{w}^T \mathbf{x}_n, \quad (3.1)$$

em que $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_Q]^T$ é o vetor de coeficientes do filtro espacial com dimensão $Q \times 1$; Q é o número de sinais; $\mathbf{x}_n = [x_{n,1} \ x_{n,2} \ \dots \ x_{n,Q}]^T$ contém os sinais de interesse no instante de tempo n ; e y_n é o sinal filtrado no instante n . Ao longo deste trabalho será utilizada uma notação na qual letras maiúsculas e minúsculas em negrito referem-se, respectivamente, a matrizes e vetores, enquanto que letras em minúsculas são variáveis escalares e em maiúsculas são constantes.

No caso de sinais de imagética motora, assume-se a existência de Q eletrodos e uma janela de tempo de P amostras de duração em cada canal de aquisição, denominada de época, na qual é esperada a ocorrência da resposta ao processo de imagética motora. Dessa forma, o procedimento de filtragem pode ser representado na forma vetorial, como:

$$\mathbf{y}_n = \mathbf{X}_n^T \mathbf{w}, \quad (3.2)$$

em que $\mathbf{y}_n = [y_{n,1} \ y_{n,2} \ \dots \ y_{n,P}]^T$ é um vetor de dimensão $P \times 1$ contendo os sinais processados para cada instante de tempo da janela de observação (época analisada) e:

$$\mathbf{X}_n = [\mathbf{x}_n \ \mathbf{x}_{n+1} \ \dots \ \mathbf{x}_{n+P-1}], \quad (3.3)$$

é a matriz de sinais EEG da época analisada, com dimensão $Q \times P$.

Os vetores de coeficientes ótimos são obtidos pela maximização/minimização do quociente de Rayleigh generalizado (LOTTE e GUAN, 2011):

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{C}_A \mathbf{w}}{\mathbf{w}^T \mathbf{C}_B \mathbf{w}}, \quad (3.4)$$

em que:

$$\mathbf{C}_\rho = \frac{1}{N_\rho} \sum_{n=1}^{N_\rho} \mathbf{X}_n^\rho \mathbf{X}_n^{\rho T}, \quad (3.5)$$

$\rho \in \{A, B\}$; A e B são as classes associadas ao problema de classificação; \mathbf{C}_ρ é a matriz de informações da classe ρ , convencionalmente constituída pelas potências inter e intra-eletrodos; \mathbf{X}_n^ρ , para $n = 1, 2, \dots, N_\rho$ refere-se a cada uma das N_ρ épocas, associadas à classe ρ , disponíveis para treinamento. Os vetores \mathbf{w} são ótimos no sentido de minimizar a variância para uma classe enquanto maximizam a da outra. (ZHAOJUN, LU e VANDEREYCKEN, 2018).

O problema de minimização/maximização da equação (3.4) pode ser resolvido observando que a função $J(\mathbf{w})$ é invariante ao escalamento de \mathbf{w} , de tal forma que $J(k\mathbf{w}) = J(\mathbf{w})$, sendo k uma constante real. Dessa forma, extremar (3.4) é o mesmo que extremar $\mathbf{w}^T \mathbf{C}_A \mathbf{w}$ sujeito à restrição $\mathbf{w}^T \mathbf{C}_B \mathbf{w} = 1$, de tal forma que:

$$\min/\max_{\mathbf{w}} J(\mathbf{w}) = \mathbf{w}^T \mathbf{C}_A \mathbf{w} \quad (3.6)$$

t. q. $\mathbf{w}^T \mathbf{C}_B \mathbf{w} = 1$

O objetivo da equação (3.6) é obter dois conjuntos de filtros, um que maximiza C_A e minimiza C_B e outro que minimiza C_A e maximiza C_B .

Usando o método do multiplicador de *Lagrange* (NOCEDAL, 2006) na equação (3.6), chega-se em:

$$L(\lambda, \mathbf{w}) = \mathbf{w}^T \mathbf{C}_A \mathbf{w} + \lambda (\mathbf{w}^T \mathbf{C}_B \mathbf{w} - 1). \quad (3.7)$$

Derivando-se a equação (3.7) e igualando-se o resultado a zero, chega-se ao problema clássico de autovalores e autovetores generalizados (LOTTE e GUAN, 2011):

$$\mathbf{C}_B^{-1} \mathbf{C}_A \mathbf{w} = \lambda \mathbf{w}, \quad (3.8)$$

e, portanto, o conjunto de vetores de coeficientes que maximizam a diferença entre as classes é determinado pelos autovetores de

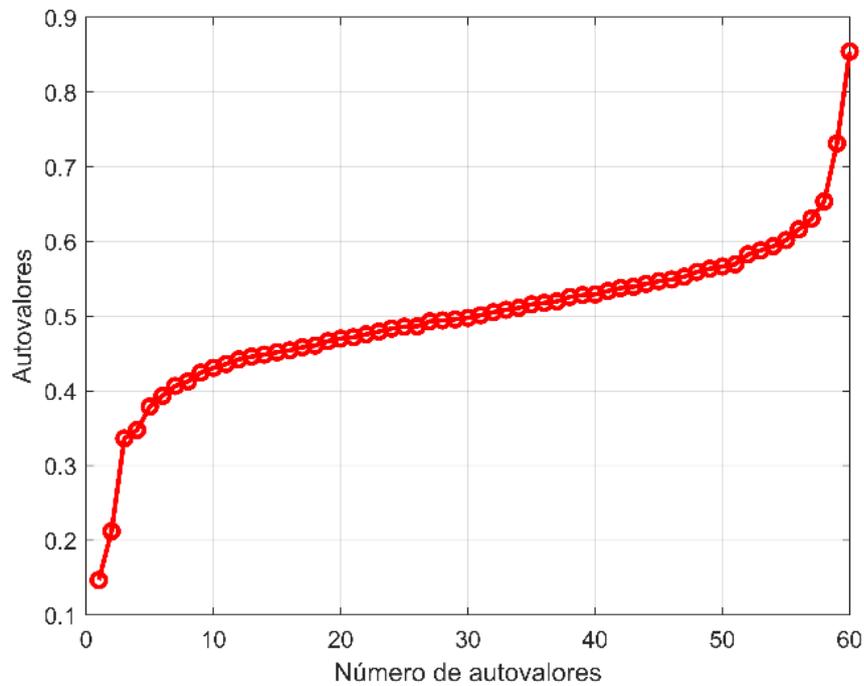
$$\mathbf{Z} = \mathbf{C}_B^{-1} \mathbf{C}_A, \quad (3.9)$$

associados ao maior e menor autovalores (LOTTE e GUAN, 2011).

Conforme discutido no início do capítulo, um dos objetivos da filtragem espacial usando CSP é a diminuição de artefatos presentes nos sinais, para que haja a melhoria de sua relação sinal-ruído (SNR). Além disso, a transformação linear realizada no espaço de eletrodos do EEG faz com que o filtro CSP determine direções em que a variância é máxima para uma

classe e mínima para outra, possibilitando a discriminação entre classes (SILVA, 2017). Uma maneira de se obter essa discriminação entre classes é utilizar os $H/2$ menores e maiores autovalores, sendo H a quantidade de filtros espaciais almejados (Figura 15). Os vetores associados aos autovalores nos dois extremos da Figura 15 produzem componentes com maior e menor variância de acordo com a direção do autovetor (SILVA, 2017). Os autovalores estão limitados ao intervalo de 0 a 1 devido à normalização da matriz de covariância, além disso, ela é semidefinida positiva, fazendo com que \mathbf{W} compartilhe essa propriedade.

Figura 15 – Exemplos de autovalores associados aos autovetores que compõem a matriz \mathbf{W} de filtragem espacial CSP. Os autovalores foram obtidos usando os sinais do sujeito B₁ do banco de dados IIIA.



Fonte: (O AUTOR, 2019).

3.5 CSP COM REGULARIZAÇÃO

Uma maneira de melhorar o desempenho e evitar *overfitting* (excesso de treinamento que reduz o desempenho de classificação final) no método de CSP é por meio de regularizações. Nesta seção, são apresentadas 8 formas de regularizações para o método de CSP (LOTTE e GUAN, 2011). As regularizações são aplicadas nas matrizes de covariância ou na função objetivo do CSP.

A ocorrência de artefatos (e.g. musculares) ou baixa SNR pode acarretar erros na estimação das matrizes de covariância. Uma maior robustez a esses erros pode ser obtida

adicionando-se informação relevante por meio do processo de regularização. Dessa forma, a matriz de covariância pode ser estimada como:

$$\hat{\mathbf{C}}_\rho = (1 - \varphi)\tilde{\mathbf{C}}_\rho + \varphi\mathbf{I}, \quad (3.10)$$

com

$$\tilde{\mathbf{C}}_\rho = (1 - \zeta)s_\rho\mathbf{C}_\rho + \zeta\mathbf{G}_\rho, \quad (3.11)$$

em que \mathbf{C}_ρ é a matriz de covariância inicial para a classe ρ ; $\hat{\mathbf{C}}_\rho$ é matriz de covariância regularizada; \mathbf{I} é matriz identidade; s_ρ é um parâmetro de escala; ζ e φ são parâmetros de regularização (ζ e $\varphi \in [0,1]$); e \mathbf{G}_ρ é uma matriz de covariância genérica. A matriz \mathbf{G}_ρ pode ser associada a aspectos neurofisiológicos ou construída usando sinais de vários sujeitos (LOTTE e GUAN, 2011). O parâmetro φ tem como objetivo diminuir a sensibilidade da estimativa quando se tem um pequeno conjunto de sinais para treinamento (LOTTE e GUAN, 2011). O parâmetro ζ introduz maior robustez aos ruídos aditivos. A partir dessa abordagem, as matrizes regularizadas $\hat{\mathbf{C}}_A$ e $\hat{\mathbf{C}}_B$ substituem as matrizes \mathbf{C}_A e \mathbf{C}_B no método de CSP convencional (LOTTE e GUAN, 2011).

A segunda abordagem de regularização é alcançada alterando-se a função objetivo do método CSP convencional. O procedimento consiste na adição de um termo de regularização para penalizar soluções que não satisfazem determinada condição. Dessa forma, a nova função objetivo a ser minimizada é:

$$J_p(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{C}_A \mathbf{w}}{\mathbf{w}^T \mathbf{C}_B \mathbf{w} + \alpha_c \mathbf{w}^T \mathbf{K} \mathbf{w}} = \frac{\mathbf{w}^T \mathbf{C}_A \mathbf{w}}{\mathbf{w}^T (\mathbf{C}_B + \mathbf{K} \alpha_c) \mathbf{w}}, \quad (3.12)$$

em que α_c é um parâmetro de regularização definido pelo usuário e deve ser maior que zero; e \mathbf{K} é uma matriz de penalização espacial

A equação (3.12) pode ser reescrita como:

$$J_p(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{C}_A \mathbf{w}}{\mathbf{w}^T \mathbf{C}_B \mathbf{w} + \alpha_c P(\mathbf{w})}, \quad (3.13)$$

nesse caso particular, $P(\mathbf{w})$ é uma função quadrática, pois:

$$P(\mathbf{w}) = \|\mathbf{w}\|_{\mathbf{K}}^2 = \mathbf{w}^T \mathbf{K} \mathbf{w}. \quad (3.14)$$

Aplicando o operador Lagrangiano na equação (3.12), obtém-se:

$$L_p(\lambda, \mathbf{w}) = \mathbf{w}^T \mathbf{C}_A \mathbf{w} - \lambda (\mathbf{w}^T (\mathbf{C}_B + \alpha_c \mathbf{K}) \mathbf{w} - 1). \quad (3.15)$$

De modo análogo ao procedimento apresentado para o método CSP convencional, chega-se a um problema clássico de autovalores e autovetores:

$$(\mathbf{C}_B + \alpha_c \mathbf{K})^{-1} \mathbf{C}_A \mathbf{w} = \lambda \mathbf{w}. \quad (3.16)$$

Os filtros \mathbf{w} que maximizam $J_p(\mathbf{w})$ são os autovetores associados aos maiores autovalores de $\mathbf{M}_A = (\mathbf{C}_B + \alpha_c \mathbf{K})^{-1} \mathbf{C}_A$ para a classe A. Para a classe B devem ser usados os maiores autovalores associados aos autovetores de $\mathbf{M}_B = (\mathbf{C}_A + \alpha_c \mathbf{K})^{-1} \mathbf{C}_B$. Várias versões desse método podem ser definidas de acordo com a composição escolhida para \mathbf{K} . Ao contrário do CSP convencional, o método regularizado utiliza somente os maiores autovalores associados aos respectivos autovetores de \mathbf{M}_A e \mathbf{M}_B para compor a matriz de filtragem espacial (LOTTE e GUAN, 2011).

A partir das duas regularizações propostas, são desenvolvidos os oito métodos apresentados a seguir.

3.5.1 CSP composto

O método de CSP composto (CCSP, do inglês, *Composite Common Spatial Patterns*) utiliza a informação dos diversos sujeitos para regularizar a matriz de covariância.

A determinação da matriz de covariância do método de CCSP é realizada por meio da equação (3.11) e utiliza somente o hiperparâmetro ζ , portanto $\alpha_c = \varphi = 0$. A matriz de covariância genérica (\mathbf{G}_ρ) é construída utilizando-se informação das matrizes de covariância de outros sujeitos. Dois métodos, denominados CCSP1 e CCSP2, foram apresentados em (LOTTE e GUAN, 2011) usando essa abordagem para construção de \mathbf{G}_ρ .

No método CCSP1, \mathbf{G}_ρ é construída como uma soma ponderada das matrizes de covariância (correspondente ao mesmo estado mental) de diferentes sujeitos, reduzindo assim erros de estimativa decorrentes da existência de um conjunto pequeno de sinais de treinamento (LOTTE e GUAN, 2011). Dessa forma, \mathbf{G}_ρ é definida como:

$$\mathbf{G}_\rho = \sum_{i \in \Omega} \frac{N_\rho^i}{N_{t,\rho}} \mathbf{C}_\rho^i, \quad (3.17)$$

e s_ρ :

$$s_\rho = \frac{N_\rho}{N_{t,\rho}}, \quad (3.18)$$

em que Ω é um conjunto de sujeitos cujos dados estão disponíveis; \mathbf{C}_ρ^i é a matriz de covariância espacial para a classe ρ e sujeito i ; N_ρ^i é o número de épocas de EEG usadas para estimar \mathbf{C}_ρ^i ; N_ρ é o número de épocas de EEG usadas para estimar \mathbf{C}_ρ (matriz para o sujeito desejado); e $N_{t,\rho}$ é o número total de épocas de EEG para a classe ρ (todos os sujeitos, incluindo o desejado).

O método CCSP2 utiliza uma soma ponderada de matrizes de covariância dos diversos sujeitos disponíveis, sendo os pesos definidos como a divergência de Kullback-Leibler (KL) entre sujeitos (LOTTE e GUAN, 2011). Primeiramente deve-se calcular o escalar Z_r , sendo definido como:

$$Z_r = \sum_{j \in \Omega} \frac{1}{KL(j,t)}, \quad (3.19)$$

Posteriormente, calcula-se a matriz \mathbf{G}_ρ como:

$$\mathbf{G}_\rho = \sum_{i \in \Omega} \frac{1}{Z_r} \frac{1}{KL(i,t)} \mathbf{C}_\rho^i, \quad (3.20)$$

sendo $KL(i,t)$ a divergência de KL entre o sujeito desejado t e os outros participantes denominados por i , sendo definida como:

$$KL(i,t) = \frac{1}{2} \left(\log \left(\frac{\det(\mathbf{C}_\rho)}{\det(\mathbf{C}_\rho^i)} \right) + \text{tr}(\mathbf{C}_\rho^{-1} \mathbf{C}_\rho^i) - N_e \right), \quad (3.21)$$

em que $\text{tr}\{\cdot\}$ é o operador de traço do argumento; $\det\{\cdot\}$ é o determinante do argumento; $\log\{\cdot\}$ é o logaritmo do argumento; e N_e representa o número de eletrodos utilizados.

3.5.2 CSP regularizado com aprendizado genérico

O método CSP regularizado com aprendizado genérico (GLRCSP, do inglês, *Regularized CSP with Generic Learning*) também utiliza a informação de outros sujeitos para estimar a matriz de covariância e assim aplicar a regularização. Esse método utiliza os termos de regularização φ e ζ apresentados nas equações (3.10) e (3.11) para produzir uma estimativa a partir da matriz ponderada de dois estimadores. A matriz \mathbf{G}_ρ é calculada como:

$$\mathbf{G}_\rho = s_G \sum_{i \in \Omega} \mathbf{C}_\rho^i, \quad (3.22)$$

com

$$s_G = s_\rho = \frac{1}{(1-\zeta)M_{C_\rho} + \zeta \sum_{i \in \Omega} M_{C_i}}, \quad (3.23)$$

e M_C é o número de épocas usadas para calcular a matriz de covariância C_ρ .

3.5.3 CSP regularizado com carregamento diagonal

Outra forma de regularização pode ser obtida por meio do método de CSP regularizado com carregamento diagonal (DLCSP, do inglês, *Regularized CSP with Diagonal Loading*). Esse método, como o próprio nome informa, utiliza o método de carregamento na diagonal, o qual consiste no encolhimento (*shrinking*) da matriz de covariância em direção à matriz identidade. Essa proposta usa somente o parâmetro de regularização φ da equação (3.10), sendo que os demais (ζ e α_c) são nulos. O parâmetro φ também pode ser definido por meio do método de Ledoit e Wolf (LOTTE e GUAN, 2011), gerando assim o método de CSP regularizado com carregamento diagonal automático (DLCSPauto, do inglês, *Diagonal Loading Common Spatial Patterns Automatic*). Caso o parâmetro φ seja calculado por meio da validação-cruzada (LOTTE e GUAN, 2011), obtém-se o método de CSP regularizado com carregamento diagonal e validação cruzada (DLCSPcv, em inglês, *Diagonal Loading Common Spatial Patterns com cross-validation*). Finalizando, o método de CSP regularizado com carregamento diagonal e validação cruzada diferente (DLCSPcvd, do inglês, *Diagonal Loading Common Spatial Patterns with cross-validation different*) utiliza dois parâmetros φ distintos, um para a classe A e outro para a B.

3.5.4 CSP regularizado espacialmente

Devido ao algoritmo de CSP ignorar a localização espacial dos eletrodos, foi idealizado um método que leva em conta essa informação, sendo chamado de método de padrões espaciais comuns regularizados espacialmente (SRCSP, do inglês, *Spatially Regularized Common Spatial Patterns*) (LOTTE e GUAN, 2011). Do ponto de vista neurológico, neurônios vizinhos tendem a desempenhar funções semelhantes, isso dá suporte à ideia de que eletrodos próximos devem medir sinais cerebrais semelhantes se estão próximos o bastante um do outro (LOTTE e GUAN, 2011). Para garantir a suavidade espacial dos filtros \mathbf{w} , é usada uma função de penalidade laplaciana $P(\mathbf{w})$, sendo a matriz \mathbf{K} definida como:

$$\mathbf{K} = \mathbf{D}_G - \mathbf{G}, \quad (3.24)$$

em que \mathbf{D}_G é uma matriz diagonal, calculada como:

$$D_G(i, i) = \sum_j G(i, j), \quad (3.25)$$

e a função $G(i, j)$ é dada por:

$$G(i, j) = e^{-\frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{r^2}}, \quad (3.26)$$

em que \mathbf{v}_i é um vetor com as coordenadas espaciais em três dimensões do i -ésimo eletrodo; \mathbf{D}_G é uma matriz diagonal; e r é um hiperparâmetro que define o quão perto estão dois eletrodos. O termo de penalidade $P(\mathbf{w})$ será grande para filtros não suaves, isto é, filtros nos quais os eletrodos têm pesos muito diferentes (LOTTE e GUAN, 2011).

3.5.5 CSP com regularização de pesos de Tikhonov

O método CSP com regularização de pesos de Tikhonov (WTRCSP, do inglês, *Weighted Tikhonov Regularization Common Spatial Patterns*) utiliza o método de Tikhonov para determinar o grau de importância de um canal de EEG na composição da matriz de covariância do método de CSP. Pode apresentar diferentes penalizações para canais distintos. A função de penalização é dada por:

$$P(\mathbf{w}) = \mathbf{w}^T \mathbf{D}_w \mathbf{w}, \quad (3.27)$$

em que \mathbf{D}_w é uma matriz diagonal, tal que $\mathbf{D}_w = \text{diag}(\mathbf{w}_G)$ e $\mathbf{w}_G(i)$ é o nível de penalização para o canal i ; e $\text{diag}\{\cdot\}$ é o operador diagonal da matriz. A matriz \mathbf{D}_w é substituída no lugar da matriz \mathbf{K} em (3.12). Os níveis de penalidade $\mathbf{w}_G(i)$ podem ser definidos de acordo com características anatômicas do cérebro relacionadas à tarefa mental a ser identificada. Os níveis de penalidade também podem ser definidos manualmente, mas é difícil determinar o nível de penalidade para cada canal individualmente (LOTTE e GUAN, 2011). Devido a essa dificuldade, os níveis de penalidade são obtidos usando informações de outros sujeitos, de forma que:

$$\mathbf{w}_G = \left(\frac{1}{2N_h |\Omega|} \sum_{i \in \Omega} \sum_{h=1}^{2N_h} \left\| \frac{\mathbf{w}_h^i}{\|\mathbf{w}_h^i\|} \right\| \right)^{-1}, \quad (3.28)$$

em que \mathbf{w}_h^i é o h -ésimo filtro espacial obtido usando CSP (respectivo aos maiores e menores autovalores associados aos autovetores de \mathbf{Z} na equação (3.9)) para o i -ésimo sujeito adicional disponível; e N_h é a quantidade de maiores e menores autovalores associados a \mathbf{Z} .

3.6 EXTRAÇÃO DE CARACTERÍSTICAS

Uma vez determinados os vetores que maximizam a discriminação das classes desejadas (utilizando os sinais \mathbf{X}^{ρ_n}) é realizada a filtragem das épocas de acordo com a equação (3.29), logo:

$$\mathbf{Y}_n = \mathbf{W}^T \mathbf{X}_n, \quad (3.29)$$

em que $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_H]$ é a matriz de vetores de coeficientes, de dimensão $Q \times H$; H é o número de vetores (características extraídas); e $\mathbf{Y}_n = [\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_H]^T$ é a matriz de sinais processados. Os vetores de coeficientes que compõem a matriz \mathbf{W} são determinados pela escolha dos autovetores de \mathbf{Z} associados aos $H/2$ maiores e menores autovalores. Se o número H de vetores for muito pequeno, a discriminação entre as classes poderá não ser efetiva. Por outro lado, valores elevados de H podem levar à condição de *overfitting* (BLANKERTZ, TOMIOKA, *et al.*, 2008). Finalmente, é realizada uma transformação não-linear na matriz \mathbf{Y}_n de forma a aproximar a função densidade de probabilidade das amostras de cada linha a uma distribuição gaussiana (RAMOSER, PFURTSCHELLER e MULLER-GERKING, 2000):

$$\gamma_n = \log(\text{var}(\mathbf{Y}_n)), \quad (3.30)$$

em que $\text{var}(\mathbf{Y}_n)$ resulta em um vetor coluna em que cada linha (l) é dada por $y_l = (1/P) \sum_{c=1}^P (y_{l,c} - m_l)^2$; $m_l = (1/P) \sum_{c=1}^P (y_{l,c})$ é a média da l -linha; e $y_{l,c}$ é o elemento da l -ésima linha e c -ésima coluna da matriz \mathbf{Y}_n .

De acordo com o diagrama em blocos apresentado na Figura 11, essa etapa de processamento está presente nas duas fases, treinamento e execução. Finalizando, segundo Lotte e Guan (2011), a configuração com 3 pares de filtros espaciais é a que produz melhores taxas de classificação usando o classificador FLDA.

3.7 CLASSIFICADORES

Classificadores têm como objetivo, realizar a separação dos sinais em grupos pré-definidos. Podem ser baseados em modelos lineares ou não lineares e utilizam uma fase prévia de treinamento para determinação de seus parâmetros.

Os classificadores mais difundidos na literatura são: os lineares, os de redes neurais, os Bayesianos não lineares e os que utilizam a técnica dos k vizinhos mais próximos (KNN, do inglês, *K Nearest Neighbor*) e combinações entre eles (LOTTE, CONGEDO, *et al.*, 2007).

Nesta seção são apresentados somente dois deles, um linear (FLDA) e outro que pode ser linear ou não linear (SVM). Foram escolhidos em decorrência da grande frequência com que são utilizados em classificação de imagética motora binária (LOTTE e GUAN, 2011) e quaternária (GE, WANG e YU, 2014).

3.7.1 Classificador FLDA

O classificador de Análise de Discriminantes Lineares de Fisher (FLDA, do inglês, *Fisher Linear Discriminant Analysis*) também denominado simplesmente por LDA é um classificador linear de baixo custo computacional que apresenta bom desempenho quando os dados apresentam distribuição normal (LOTTE, CONGEDO, *et al.*, 2007). Tem sido utilizado com sucesso em diversas aplicações de BCI que usam P300 e imagética motora (LOTTE, CONGEDO, *et al.*, 2007).

A classificação binária é realizada em um hiperplano definido por um vetor de pesos $\mathbf{a} \in \mathbb{R}^N$ e um *offset* $b \in \mathbb{R}$ determinados empiricamente a partir de um conjunto de treinamento de $N_A + N_B$ vetores de características associados a ambas as classes. O critério de otimização utilizado é o critério de Fisher (BELHADJ, BERNOUMASSAT e JRACHAI, 2015).

O hiperplano é definido como:

$$f(\boldsymbol{\gamma}_n) = \begin{cases} \mathbf{a}^T \boldsymbol{\gamma}_n - b \geq 0 & \boldsymbol{\gamma}_n \in \text{classe } A \\ \mathbf{a}^T \boldsymbol{\gamma}_n - b < 0 & \boldsymbol{\gamma}_n \in \text{classe } B \end{cases}, \quad (3.31)$$

em que $\boldsymbol{\gamma}_n$ é o vetor de características de entrada para a época n . O vetor de pesos é dado por:

$$\mathbf{a} = \left(\frac{\mathbf{S}_A + \mathbf{S}_B}{2} \right)^{-1} (\mathbf{m}_A - \mathbf{m}_B), \quad (3.32)$$

em que \mathbf{m}_A e \mathbf{m}_B denotam as médias para as classes A e B, respectivamente, calculadas como:

$$\mathbf{m}_\rho = \frac{1}{N_\rho} \sum_{n=1}^{N_\rho} \boldsymbol{\gamma}_{\rho,n}, \quad (3.33)$$

em que $\rho \in \{A, B\}$, $\boldsymbol{\gamma}_{\rho,n}$ refere-se ao vetor de característica da época n relativa a classe ρ e N_ρ é a quantidade de épocas pertencentes a cada classe. As matrizes de covariância são definidas como:

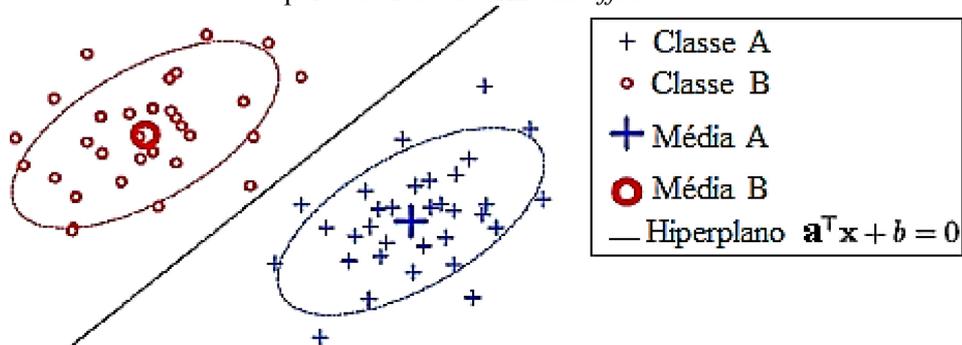
$$\mathbf{S}_\rho = \frac{1}{N_\rho} \sum_{n=1}^{N_\rho} (\boldsymbol{\gamma}_{\rho,n} - \mathbf{m}_{\rho,n})(\boldsymbol{\gamma}_{\rho,n} - \mathbf{m}_{\rho,n})^T. \quad (3.34)$$

O parâmetro de *offset* é calculado como:

$$b = \frac{1}{2} \mathbf{a}^T (\mathbf{m}_A + \mathbf{m}_B). \quad (3.35)$$

A Figura 16 ilustra o método de classificação linear usando FLDA. Nota-se pela imagem a equação da reta calculada, nesse caso bidimensional, separando os dados entre as classes A e B. Dados à esquerda da reta são classificados como pertencentes à classe A, enquanto que à direita são da classe B.

Figura 16 – FLDA, hiperplano ótimo (linha sólida) que separa duas classes. O hiperplano é descrito pelo vetor \mathbf{a} e o termo de *offset* b .



Fonte: adaptado de (BELHADJ, BERNOUMASSAT e JRACHAI, 2015).

3.7.2 Classificador SVM

As máquinas de vetores de suporte (SVM, do inglês, *Support Vector Machines*) são embasadas na teoria de aprendizado estatístico (FACELI, LORENA, *et al.*, 2011). Essa teoria estabelece alguns princípios que devem ser seguidos para que se obtenham classificadores com adequada capacidade de generalização (FACELI, LORENA, *et al.*, 2011).

O SVM é uma técnica para classificação binária que tem como objetivo a determinação de um hiperplano de separação para duas classes discretas. Isso é alcançado em uma etapa denominada de treinamento, por meio de um processo iterativo de ajuste utilizando exemplos (PEIXOTO, ARAKI e CENTENO, 2016). Há dois tipos de SVMs: lineares (margens rígidas ou suaves) e não lineares.

Finalizando, escolher entre o FLDA e SVM dependerá do tamanho do espaço de características. O hiperplano calculado pelo FLDA somente será ótimo quando as matrizes de covariância para as classes forem idênticas, o qual é violado na prática (GOKCEN e PENG, 2002). Além disso, computacionalmente, com o aumento do tamanho do espaço de características, o FLDA tende a produzir matrizes de covariância mal condicionadas (GOKCEN e PENG, 2002). Em contrapartida, o hiperplano ótimo para o classificador SVM é calculado por meio da otimização da separação de margens das classes, logo, não assume igualdade das matrizes de covariância e, portanto, não sofrerá do problema de mau condicionamento. Ademais, a capacidade de generalização do SVM aumenta com o incremento do tamanho do espaço de características (GOKCEN e PENG, 2002).

Resumidamente, para pequenos espaços de características, usa-se o classificador FLDA, caso contrário, SVM.

3.7.2.1 SVM linear

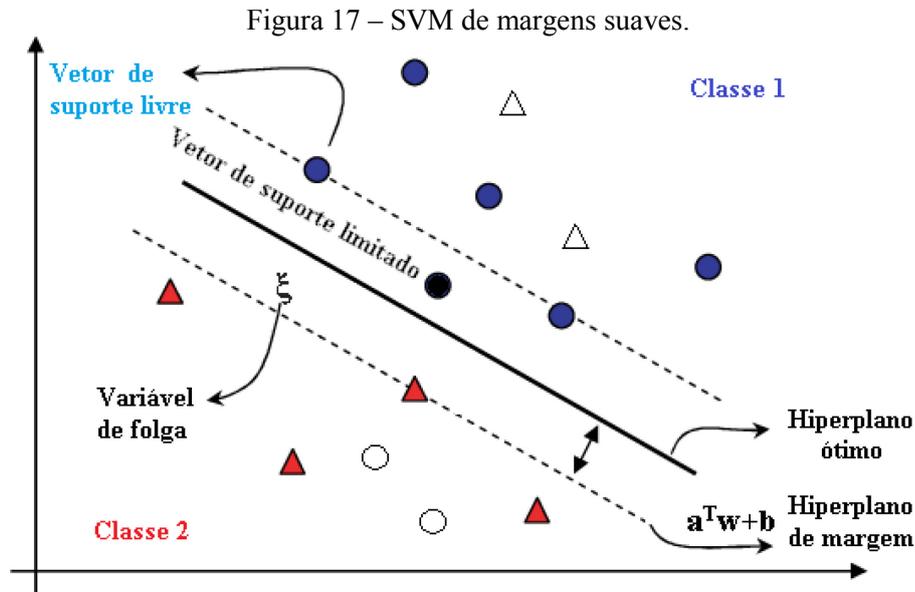
As SVMs lineares são eficazes na classificação de conjuntos de dados linearmente separáveis ou que possuam uma distribuição aproximadamente linear, de acordo com o tipo de margem escolhida (FACELI, LORENA, *et al.*, 2011). As SVMs lineares de margens suaves apresentam maior robustez em relação às de margens rígidas, pois permitem a solução de problemas mais gerais devido à introdução de variáveis de folga (ξ).

A SVM linear tem como objetivo minimizar a seguinte função custo:

$$\begin{aligned} \min_{\mathbf{a}, b, \xi} \quad & \frac{1}{2} \|\mathbf{a}\|^2 + C_s \left(\sum_{i=1}^n \xi_i \right) \\ \text{s.a} \quad & y_i (\mathbf{a} \cdot \mathbf{x}_i) + b \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, n \end{aligned} \quad (3.36)$$

em que \mathbf{a} e b são os parâmetros utilizados para construção do hiperplano; C_s é um termo de regularização que impõe um peso à minimização dos erros no conjunto de treinamento em relação à minimização da complexidade do modelo; e y_i são as possíveis classes $\{-1, +1\}$.

As variáveis de folga impostas na restrição permitem que haja um relaxamento nas restrições, fazendo com que alguns objetos possam pertencer à região delimitada pelos hiperplanos de margem (veja a Figura 17) e também que ocorram alguns erros de classificação (FACELI, LORENA, *et al.*, 2011).



Fonte: adaptado de (BONIFÁCIO, 2017).

Na Figura 17 os objetos que fazem parte e foram usados para a construção do hiperplano de margem são chamados de vetores de suporte livres, ou seja, estão sob as linhas pontilhadas. O objeto em preto nessa imagem representa um vetor de suporte limitado. Objetos em azul e vermelho são os objetos que foram corretamente classificados e estão fora das margens. Objetos em branco são vetores de suporte limitados que foram incorretamente classificados. Por fim, os objetos que estão dentro da margem foram corretamente classificados.

3.7.2.2 SVM não linear

Faceli et al (2011) afirmam que em situações reais é difícil encontrar aplicações em que os dados sejam linearmente separáveis, devido a fatores como: ruído, artefatos e a natureza não linear do problema.

As SVMs não lineares mapeiam o conjunto de dados de treinamento originais para um novo espaço de maior dimensão, chamado de espaço de características. Seja $\phi: T \rightarrow F$ o mapeamento em que T é o espaço de entrada e F o de características, de modo que a escolha adequada de ϕ resulta um conjunto de treinamento em F que pode ser separado por uma SVM linear (FACELI, LORENA, *et al.*, 2011). Para que a separação seja adequada, devem-se

obedecer duas regras: a transformação deve ser não linear e a dimensão do espaço de características deve ser suficientemente alta (FACELI, LORENA, *et al.*, 2011). Para ilustrar esses conceitos é considerado o exemplo da . Transformando os objetos do \mathbb{R}^2 para o \mathbb{R}^3 por meio de

$$\phi(\mathbf{t}) = \phi(t_1, t_2) = (t_1^2, 2\sqrt{t_1}t_2, t_2^2), \quad (3.37)$$

faz-se com que o conjunto de dados não linear em \mathbb{R}^2 torne-se linearmente separável em \mathbb{R}^3 .

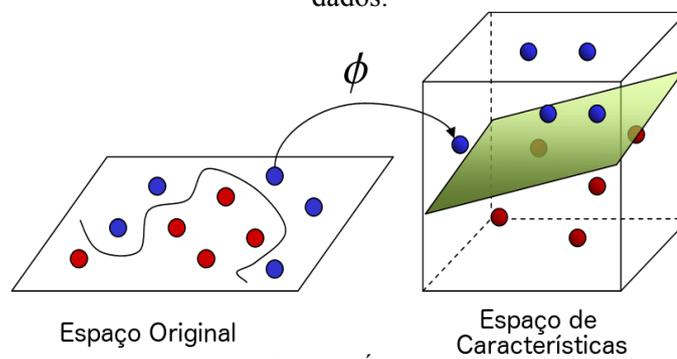
Então, é possível encontrar um hiperplano em \mathbb{R}^3 , dado por

$$h(\mathbf{t}) = \mathbf{a}\phi(\mathbf{t}) + b = a_1t_1^2 + a_2\sqrt{t_1}t_2 + a_3t_2^2 + b = 0, \quad (3.38)$$

em que \mathbf{t} é um vetor em \mathbb{R}^2 com coordenadas t_1 e t_2 . Embora a função encontrada em \mathbb{R}^3 seja

linear nesse espaço, em \mathbb{R}^2 ela é não linear.

Figura 18 – SVM não linear. A primeira imagem não é linearmente separável por um hiperplano. A segunda imagem representa o espaço aumentado por meio de um Kernel, no qual é possível separar os dados.



Fonte: (BONIFÁCIO, 2017).

O mapeamento dos objetos em um espaço de maior dimensão, por meio de $\phi(\mathbf{t})$ e a utilização de um hiperplano para separação linear (SVM) pode resultar em uma classificação ótima, ou seja, apresentando maior margem de separação (FACELI, LORENA, *et al.*, 2011). Nessas situações, o SVM utilizado é do tipo de margens suaves, devido à sua maior capacidade de generalização em função de robustez a ruídos e artefatos (FACELI, LORENA, *et al.*, 2011).

O espaço \mathcal{F} pode ter dimensão muito elevada ou até mesmo infinita, de modo que o cálculo de $\phi(\mathbf{t})$ pode ser extremamente custoso ou mesmo inviável para uma dada aplicação. Porém, o processo de separação requer somente a definição do produto escalar entre os objetos no espaço de características e o vetor de coeficientes (FACELI, LORENA, *et al.*, 2011). O

cálculo é feito usando funções chamadas de Kernels (FACELI, LORENA, *et al.*, 2011). O Kernel deve ser reprodutivo e capaz de gerar um espaço de características de Hilbert.

“O produto interno no espaço de Hilbert que conduz à construção do SVM implica que para qualquer função Kernel satisfazendo a condição de Mercer existe um espaço de características onde as funções geram um produto interno (ABREU, 2016)”.

A definição de Kernel consiste em uma função que recebe dois pontos t_i e t_j no espaço de entrada e calcula o produto escalar desses objetos no espaço de características (FACELI, LORENA, *et al.*, 2011):

$$K(t_i, t_j) = \langle \phi(t_i) \phi(t_j) \rangle. \quad (3.39)$$

Não é necessário o conhecimento do mapeamento $\phi(t)$ para o emprego do Kernel, sendo ele gerado implicitamente (FACELI, LORENA, *et al.*, 2011). Sua utilidade está em sua simplicidade e capacidade de gerar espaços abstratos (FACELI, LORENA, *et al.*, 2011).

Um Kernel deve obedecer às condições de Mercer (FACELI, LORENA, *et al.*, 2011). Alguns dos Kernels mais utilizados na prática são os polinomiais, os de função base radial (RBF, do inglês, *radial basis function*) e os sigmodais, todos eles necessitam a definição de determinados parâmetros. A Tabela 1 ilustra esses Kernels e quais parâmetros devem ser definidos.

Tabela 1 – Funções Kernel mais comuns.

Tipo de Kernel	Função $K(t_i, t_j)$	Parâmetros escolhidos
Polinomial	$\delta_s ((t_i \cdot t_j) + \kappa_s)^d$	δ_s, κ_s e d
RBF	$e^{(-\sigma \ t_i - t_j \ ^2)}$	σ
Sigmoidal	$\tanh(\delta_s (t_i \cdot t_j) + \kappa_s)$	δ_s e κ_s

Fonte: (FACELI, LORENA, *et al.*, 2011).

Para implementação da SVM linear é necessária então a definição do Kernel e de seus parâmetros, além de uma constante de regularização C_s .

3.8 CONSIDERAÇÕES FINAIS

Este capítulo apresentou de forma detalhada a estrutura de processamento de um sistema de BCI para classificação de um problema binário e quaternário de imagética motora.

O próximo capítulo apresenta métodos quantitativos comumente utilizados em aprendizado de máquina, para que seja feita a avaliação dos resultados obtidos pela classificação das épocas usando o método CSP-FLDA.

4 AVALIAÇÃO DO CLASSIFICADOR

Este capítulo aborda diferentes formas para análise dos resultados obtidos pelo classificador de imagética motora. São apresentadas medidas de desempenho comumente usadas para avaliação de métodos de aprendizado de máquina, como: acurácia, sensibilidade, especificidade, medida F e curva ROC.

A avaliação de um método de aprendizado de máquina supervisionado é normalmente realizada por meio da análise do desempenho do classificador gerado por ele na rotulação de novos objetos não apresentados previamente em seu treinamento. Existem diversas formas para isso, como análise por matriz de confusão e curvas ROC (do inglês, *Receiving Operating Characteristics*) (FACELI, LORENA, *et al.*, 2011).

4.1 MEDIDAS DE DESEMPENHO

As medidas de desempenho são métodos para avaliação do classificador, algumas delas são: acurácia, precisão, sensibilidade, especificidade, precisão e taxa de erro (FACELI, LORENA, *et al.*, 2011).

Primeiramente, começa-se definindo a matriz de confusão para um problema de classificação binária, em que uma classe foi denominada positiva (+) correspondendo à classe A (mão esquerda) e a outra negativa (-), referente à classe B (mão direita):

Tabela 2 – Matriz de confusão para um problema envolvendo duas classes.

		Classe predita	
		+	-
Classe verdadeira	+	VP	FN
	-	FP	VN

Fonte: (O AUTOR, 2019).

em que *VN* refere-se ao percentual de verdadeiros negativos, *VP* aos verdadeiros positivos, *FP* aos falsos positivos e *FN* aos falsos negativos. Os resultados *VP* e *VN* podem ser vistos como os dados que foram corretamente classificados. Um falso negativo indica que o vetor de características, ao invés de ser classificado como pertencente à classe A, foi classificado como oriundo da B. No caso de um falso positivo, o vetor de características foi classificado como pertencente à classe A ao invés da B.

A partir da matriz de confusão, ilustrada na Tabela 2, medidas de desempenho são derivadas como (FACELI, LORENA, *et al.*, 2011):

- Taxa de erro na classe positiva: conhecida como taxa de falsos negativos (TFN). Avalia a proporção de dados incorretamente classificados na classe positiva. É calculada como:

$$\text{err}_+(f) = \frac{FN}{VP + FN}; \quad (4.1)$$

- Taxa de erro na classe negativa: conhecida como taxa de falsos positivos (TFP). Avalia a proporção de dados incorretamente classificados na classe negativa. É calculada como:

$$\text{err}_-(f) = \frac{FP}{FP + VN}; \quad (4.2)$$

- Taxa de erro total: avalia somente os dados que foram incorretamente classificados. É complementar à acurácia, sendo calculada como:

$$\text{err}(f) = \frac{FP + FN}{FP + FN + VP + VN}; \quad (4.3)$$

- Taxa de acerto ou acurácia total: avalia somente os dados que foram corretamente classificados. É calculada como:

$$\text{ac}(f) = \frac{VP + VN}{FP + FN + VP + VN}; \quad (4.4)$$

- Precisão: proporção de positivos classificados corretamente entre todos aqueles classificados como positivos, sendo calculada como:

$$\text{prec}(f) = \frac{VP}{VP + FP}; \quad (4.5)$$

- Sensibilidade ou revocação: corresponde à taxa de acerto na classe positiva, ou seja, a classificação correta dos dados na classe positiva. Conhecida também como taxa de verdadeiros positivos (TVP).

$$\text{sens}(f) = \frac{VP}{VP + FN}; \quad (4.6)$$

- Especificidade: corresponde à taxa de acerto na classe negativa, ou seja, a classificação correta dos dados na classe negativa.

$$\text{esp}(f) = \frac{VN}{VN + FP} = 1 - \text{TFP}(f); \quad (4.7)$$

- Medida F: é uma combinação das medidas de sensibilidade e precisão, de modo a corrigir o problema apresentado ao avaliá-las isoladamente. O problema de avaliar somente a precisão pode ser entendido através do seguinte exemplo: considerar que a precisão máxima é de 1 implica que cada item rotulado pertencente a uma determinada classe, realmente pertence a ela. Entretanto, não há informação sobre os dados que foram incorretamente classificados. De forma análoga, esse raciocínio é estendido para a sensibilidade. Devido a serem medidas complementares, a precisão e a exatidão podem ser combinadas em uma única medida, chamada de medida F, que corresponde à média harmônica ponderada entre elas. Assim, temos que:

$$F_M(f) = \frac{(k + 1)\text{sens}(f)\text{prec}(f)}{\text{sens}(f) + k\text{prec}(f)}; \quad (4.8)$$

em que k é uma constante que possibilita aumentar o peso para a sensibilidade ou para a precisão. Fazer a constante k assumir o valor unitário equivale a atribuir o mesmo peso para a sensibilidade e para a precisão. Nesse caso tem-se a medida F_1 . As medidas F_M e F_1 variam de 0 a 1. Quando elas atingem 1, significa que a sensibilidade e a precisão estão perfeitas. A medida F_1 é calculada como:

$$F_1(f) = \frac{2\text{sens}(f)\text{prec}(f)}{\text{sens}(f) + \text{prec}(f)}; \quad (4.9)$$

- Coefficiente de correlação de Matthews: usado para medir a qualidade do classificador binário, retornando um valor de -1 a $+1$. O valor de $+1$ representa uma classificação perfeita, enquanto que o valor 0 indica que o classificador em questão não é melhor que um classificador aleatório. O valor -1 indica um modelo totalmente desconexo entre predição e observação. Pode ser calculado como (BOUGHORBEL, JARRAY e EL-ANBARI, 2017):

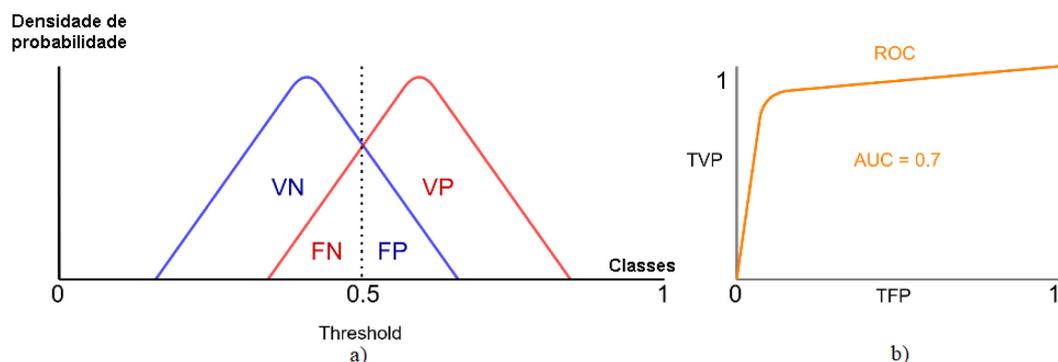
$$\text{MCC}(f) = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (4.10)$$

4.2 CURVAS ROC

Uma forma alternativa de avaliar classificadores em problemas binários é através de curvas ROC (FACELI, LORENA, et al., 2011). Geometricamente, a curva ROC é construída por um conjunto de pares de coordenadas (TVP e TFP) num plano bidimensional variando-se o limiar de decisão (em inglês, *threshold*) (FACELI, LORENA, et al., 2011). A linha vertical pontilhada na Figura 19a ilustra um exemplo de *threshold* para a separação de classes. O melhor ponto da curva ROC encontra-se no canto superior esquerdo. Ele indica que há uma sensibilidade e especificidade de 100%, levando à correta classificação de todos os sinais envolvidos.

Segundo Vaz (2009), o desempenho do classificador é medido pela área abaixo da curva (AUC, em inglês, Area Under Curve) ROC, sendo o ideal com valor unitário de 1. Áreas menores que 0,5 não são adequadas, pois caracterizam o desempenho similar ao de um classificador aleatório. Uma das alternativas para o cálculo da AUC, quando não se tem o modelo teórico da ROC, é usar a regra do trapézio. Caso contrário, o resultado é obtido por meio de uma integral definida (VAZ, 2009). Considerando que esteja-se comparando duas ou mais curvas ROC, uma outra forma de avaliação consiste em analisar qual delas mais se aproxima do ponto (0,1) representado na Figura 19b, sendo esse o melhor desempenho (FACELI, LORENA, et al., 2011).

Figura 19 – Curva ROC.



Fonte: (PACHECO, 2019)

Para ilustrar os conceitos apresentados, será analisada a Figura 19. A curva ROC ilustrada na Figura 19a apresenta medidas de probabilidade, variando de 0 a 1, em conjunto com as suas respectivas densidades para os verdadeiros negativos (VN) e os verdadeiros positivos (VP). Variando-se o limiar de classificação, será gerada a respectiva curva ROC em relação aos dados e com ela é calculada a sua AUC. A AUC de 0,7 representada na Figura 19b

indica que o classificador terá 70% de chance em distinguir corretamente a que classe pertencem as épocas. Uma AUC de 0,5 ilustra um classificador que não consegue distinguir a separação entre classes, isso é equivalente a classificar as épocas usando um modelo aleatório, sendo esse problema ocasionado pela sobreposição das densidades de probabilidade na Figura 19a.

4.3 CONSIDERAÇÕES FINAIS

Este capítulo apresentou formas de avaliação dos resultados obtidos pela etapa de classificação usando os dados de sua matriz de confusão e curva ROC.

O próximo capítulo apresenta formas de análise estatística para os dados, visando demonstrar se os resultados obtidos são relevantes ou não quando comparados com outros métodos.

5 ANÁLISE ESTATÍSTICA

Este capítulo apresenta métodos de inferência estatística e as premissas para a aplicação de cada uma das técnicas. Além disso, aborda conceitos teóricos comumente utilizados pelos métodos estatísticos, como: nível de significância, hipóteses e tipos de testes de hipóteses.

A seção de inferência abordará métodos para detecção de diferença estatística entre dois ou mais espaços amostrais. Adicionalmente, é apresentada a estimação de momentos centrais. Por fim, o gráfico de caixas que é comumente utilizado em análises estatísticas é apresentado.

5.1 CONCEITOS PRIMÁRIOS DE ESTATÍSTICA

Para o entendimento das técnicas de inferência estatística é necessário o entendimento de alguns conceitos básicos, que são: como construir a hipótese para o problema, o tipo de teste de hipótese que deve ser empregado e que nível de significância deve ser utilizado.

5.1.1 HIPÓTESE NULA E ALTERNATIVA

A aplicação de um teste estatístico (ou teste de significância) permite verificar se os dados fornecem evidência suficiente para que se possa aceitar como verdadeira a hipótese alternativa (H_1). Isso é realizado de acordo com um determinado nível de segurança para comprovar que as diferenças observadas nos dados não são meramente casuais (BARBETTA, REIS e BORNIA, 2010).

Em geral, a hipótese alternativa (H_1) é formulada em termos de desigualdade (\neq , $<$ ou $>$) e comumente corresponde ao que se quer provar (BARBETTA, REIS e BORNIA, 2010). A hipótese nula, geralmente é escrita em forma de igualdade, sendo a hipótese aceita como verdadeira até que a prova estatística seja realizada e demonstre o contrário (BARBETTA, REIS e BORNIA, 2010).

De acordo com Barbetta et al (2010), para a aplicação de um teste estatístico, deve-se saber quatro coisas:

- como formular H_0 e H_1 em termos de parâmetros populacionais;

- como obter a estatística do teste (exemplos: número de caras de uma moeda ou acurácia de classificação no caso da imagética motora);
- qual é a distribuição de referência adequada para calcular o valor da probabilidade de significância;
- quais as suposições básicas para o uso do teste escolhido (número amostras ou se o teste foi realizado sob a mesma condição).

5.1.2 NÍVEL DE SIGNIFICÂNCIA

O nível de significância (α) “indica o grau de raridade exigido de um resultado observado no sentido de rejeitar a hipótese nula (H_0)”. (WITTE e WITTE, 2005).

O valor de α deve ser comparado com a probabilidade de significância (p) para aceitar ou rejeitar a hipótese levantada. O valor p é definido como:

“O valor p (ou probabilidade de significância) é definido como a probabilidade da estatística do teste acusar um resultado tão ou mais distante do esperado, como o resultado ocorrido na particular amostra observada, supondo H_0 como a hipótese verdadeira (BARBETTA, REIS e BORNIA, 2010)”.

Barbetta et al (2010) apontam que após definido o nível de significância, a seguinte regra geral de decisão para um teste estatístico pode ser aplicada:

$$\text{Regra} = \begin{cases} p > \alpha & \rightarrow \text{aceita } H_0 \\ p \leq \alpha & \rightarrow \text{rejeita } H_0 \end{cases}, \quad (5.1)$$

sendo comum adotar níveis de significância de 5% e 1%, de acordo com o nível de segurança em que se quer afirmar H_1 . O valor de $\alpha = 0,01$ permite maior segurança do que $\alpha = 0,05$.

5.1.3 TESTES UNILATERAIS E BILATERAIS

Os testes de hipóteses podem ser bilaterais e unilaterais. A Figura 20 ilustra os três tipos de testes, sendo a área destacada em azul chamada de região crítica, ou seja, o conjunto de todos os valores de teste que fazem com que se rejeite a hipótese nula (TRIOLA, 2014).

O teste bilateral indica que a região crítica está nas duas regiões extremas (caudas) sob a curva representada na Figura 20c, e as hipóteses são:

$$\text{Hipóteses} = \begin{cases} \mu_r = \mu_p \rightarrow \text{aceita } H_0 \\ \mu_r \neq \mu_p \rightarrow \text{rejeita } H_0 \end{cases} \quad (5.2)$$

sendo μ_r a média do espaço amostral produzida pelo método de referência e μ_p a média do espaço amostral produzida pelo método proposto.

O teste unilateral à esquerda indica que a região crítica está na região extrema (cauda) esquerda sob a curva, conforme representado na Figura 20b e as hipóteses são:

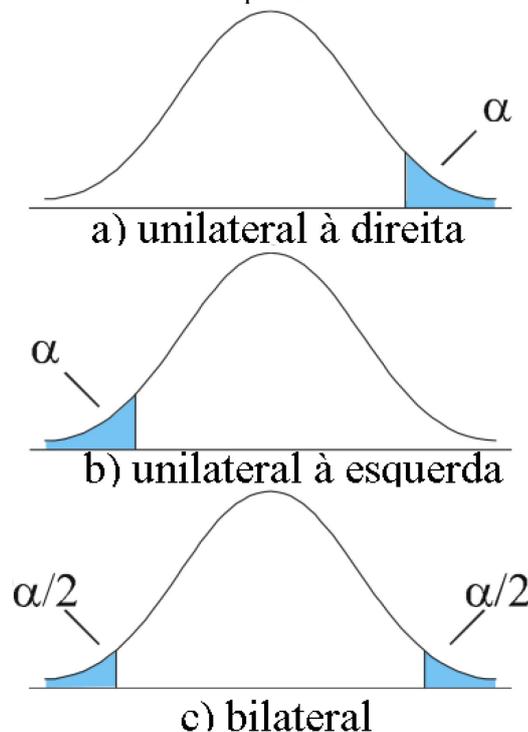
$$\text{Hipóteses} = \begin{cases} \mu_r = \mu_p \rightarrow \text{aceita } H_0 \\ \mu_r < \mu_p \rightarrow \text{rejeita } H_0 \end{cases} \quad (5.3)$$

O teste unilateral à direita indica que a região crítica está na região extrema (cauda) direita sob a curva, conforme representado na Figura 20a e as hipóteses são:

$$\text{Hipóteses} = \begin{cases} \mu_r = \mu_p \rightarrow \text{aceita } H_0 \\ \mu_r > \mu_p \rightarrow \text{rejeita } H_0 \end{cases} \quad (5.4)$$

O valor p é calculado como a integral da área destacada em azul na Figura 20 de acordo com o teste de hipótese escolhido. A aceitação ou rejeição da hipótese nula é feita comparando-se o nível de significância com a probabilidade de significância, de acordo com a equação (5.1).

Figura 20 – Testes de hipótese unilateral e bilateral.



Fonte: (O AUTOR, 2019).

5.2 INFERÊNCIA ESTATÍSTICA USANDO MEDIDAS REPETIDAS

É necessária a aplicação do método de inferência adequado para a correta detecção da diferença estatística. As medidas repetidas ocorrem sempre que o mesmo sujeito é mensurado mais de uma vez (WITTE e WITTE, 2005). Esta seção aborda somente os métodos de inferência que usam medidas repetidas, ou seja, as amostras são dependentes do sujeito.

5.2.1 Inferência estatística para dois espaços amostrais

Há duas abordagens para detecção de diferença estatística entre dois espaços amostrais: o paramétrico (teste T pareado) e o não paramétrico (teste de Wilcoxon de postos sinalizados). A escolha de qual método utilizar depende das premissas de ambos.

O teste T de amostras dependentes (teste T pareado) é utilizado para detectar se dois espaços amostrais apresentam ou não a mesma média. Segundo (WITTE e WITTE, 2005), ele tem como premissas:

- a distribuição das diferenças entre os espaços amostrais é gaussiana e
- existem 10 amostras (WITTE e WITTE, 2005).

As hipóteses para o teste T pareado são:

$$\text{Hipóteses} = \begin{cases} \mu_r = \mu_p, & \text{aceita } H_0 \\ \mu_r \neq \mu_p, & \text{rejeita } H_0 \end{cases} \quad (5.5)$$

Caso a premissa de gaussianidade seja violada, o teste a ser utilizado é o de Wilcoxon de postos sinalizados, (WITTE e WITTE, 2005). Isso é corroborado pela seguinte citação: “o teste de postos sinalizados de Wilcoxon nunca será muito pior do que o teste T e em muitos casos em que a população é não normal, ele pode ser superior (HINES, MONTGOMEY, *et al.*, 2013)”.

A eficiência do teste de postos sinalizados de Wilcoxon é de aproximadamente 85% para populações com distribuições não normais. Outro aspecto importante, é que o teste de postos sinalizados de Wilcoxon utiliza a informação de magnitude e sinais dos desvios entre os espaços amostrais, produzindo assim, uma eficiência maior do que o teste de sinais de Wilcoxon (HINES, MONTGOMEY, *et al.*, 2013). A única premissa para o teste de Wilcoxon de postos

sinalizados é de que as amostras sejam dependentes, além disso, ele não depende do formato e simetria das distribuições das amostras (HINES, MONTGOMEY, *et al.*, 2013). A falta de simetria de cada espaço amostral garante que a distribuição das diferenças seja contínua e simétrica (HINES, MONTGOMEY, *et al.*, 2013).

As hipóteses para o teste de Wilcoxon de postos sinalizados são:

$$\text{Hipóteses} = \begin{cases} \tilde{u}_r = \tilde{u}_p, & \text{aceita } H_0 \\ \tilde{u}_r \neq \tilde{u}_p, & \text{rejeita } H_0 \end{cases} \quad (5.6)$$

em que $\tilde{u}_r\{\cdot\}$ é a mediana do argumento, caracterizado pelo espaço amostral do método de referência; e $\tilde{u}_p\{\cdot\}$ é a mediana do argumento, caracterizado pelo espaço amostral do método proposto. Caso a distribuição seja simétrica, as hipóteses apresentadas na equação (5.6) tornam-se as mesmas da equação (5.5).

5.2.2 Inferência estatística para mais de dois espaços amostrais

Para a detecção de diferença estatística em mais de dois espaços amostrais, é necessária a utilização de duas técnicas. A primeira, consiste em identificar se há ou não diferença entre os grupos e a segunda, chamada de post-hoc, consiste em identificar em quais pares encontram-se as diferenças apontadas. Novamente, há duas abordagens, a paramétrica (ANOVA de medidas repetidas e post-hoc de Bonferroni) e a não paramétrica (Friedman e post-hoc de Dunn-Bonferroni).

O método paramétrico para a constatação da diferença estatística entre os espaços amostrais é a ANOVA com medições repetidas. Ela é definida como: “a ANOVA com medições repetidas testa as diferenças entre médias aritméticas de populações quando as medições para todas as populações são baseadas nos mesmos sujeitos (WITTE e WITTE, 2005)”.

As hipóteses para o teste ANOVA são:

$$\text{Hipóteses}(G) = \begin{cases} \mu_i = \mu_j, & \text{aceita } H_0 \text{ para } \forall_{i,j} \\ \mu_i \neq \mu_j, & \text{rejeita } H_0 \text{ para } \forall_{i,j} \end{cases} \quad (5.7)$$

em que G é o número total de grupos; e i e j são os índices para os grupos. A equação (5.7) estabelece que a hipótese nula é aceita quando todas as médias são iguais e H_1 é aceita quando há pelo menos um par de médias diferentes. Entretanto, a ANOVA não é capaz de detectar qual

é esse par, somente indica que ele existe. Para a identificação das diferenças entre os pares de médias, devem-se utilizar métodos de comparações múltiplas.

As premissas utilizadas para ANOVA, segundo (WITTE e WITTE, 2005) são:

- amostras dependentes;
- gaussianidade dos resíduos e
- esfericidade.

Se os resíduos, considerando todos os espaços amostrais, apresentarem gaussianidade, então cada espaço amostral terá distribuição normal também (KOZAK e PIEPHO, 2018).

A esfericidade indica a igualdade entre todos os possíveis coeficientes de correlação das populações (WITTE e WITTE, 2005). A esfericidade pode ser checada por meio do teste de Mauchly (ELLIOTT e WOODWARD, 2007). Caso a esfericidade não seja atendida, é necessário aplicar o método ANOVA com correções de Greenhouse-Geisser ou de Huynh-Feldt. A escolha de qual correção utilizar obedece ao seguinte critério:

$$\text{Critério} = \begin{cases} \varepsilon_g \leq 0,75 & \rightarrow \text{Greenhouse-Geisser} \\ \varepsilon_g > 0,75 & \rightarrow \text{Huynh-Feldt} \end{cases}, \quad (5.8)$$

em que ε_g é a estimativa de esfericidade de Greenhouse-Geisser (ELLIOTT e WOODWARD, 2007).

Como a ANOVA detecta somente se há ou não diferença entre os grupos, é necessário usar um teste posterior a ela (post-hoc) para a identificação de quais espaços amostrais são diferentes uns dos outros.

O teste post-hoc de comparações múltiplas de Bonferroni, assume como premissa, que os resíduos apresentam distribuição normal, já que é baseado no teste T. O objetivo do teste de Bonferroni é detectar quais pares de médias dentro do grupo diferem entre si. Há outros testes paramétricos de comparações múltiplas, como Tukey e Duncan.

Ao não cumprir as premissas do teste ANOVA de medidas repetidas, o método utilizado passa a ser o de Friedman (ELLIOTT e WOODWARD, 2007), sendo ele livre das propriedades da distribuição dos resíduos. Para a detecção da diferença de medianas é usado o teste post-hoc de Dunn-Bonferroni (ELLIOTT e WOODWARD, 2007).

5.3 ASSIMETRIA E CURTOSE

Além das medidas de tendência central (média, mediana e moda) e de dispersão (desvio padrão), há mais dois parâmetros importantes que caracterizam uma distribuição, assimetria e curtose.

A assimetria (*skewness*) reflete o grau de simetria em torno da média da distribuição de uma variável aleatória (HINES, MONTGOMEY, *et al.*, 2013). Para distribuições assimétricas de dados, uma das caudas será dominante sobre a outra. Uma assimetria à esquerda indica que um conjunto maior de dados está presente nessa direção. De forma análoga, observação semelhante pode ser feita para assimetria à direita. Caso a distribuição seja simétrica, a metade esquerda do histograma é uma imagem espelhada de sua metade direita (TRIOLA, 2014).

Triola (2014) diz que para uma distribuição assimétrica à esquerda, a média e a mediana da distribuição estão à esquerda da moda, porém a ordem em termos de valores numéricos não é sempre previsível. Analogamente, a assimetria à direita faz com que a mediana e a média estejam à direita da moda. Para uma distribuição simétrica (assimetria zero), a média, a mediana e a moda são iguais.

A assimetria é calculada como:

$$\beta_3 = \frac{\mu_3}{\sigma^3}, \quad (5.9)$$

em que β_3 é a assimetria; σ o desvio padrão; e μ_3 é o momento central de terceira ordem. O k -ésimo momento central pode ser calculado como:

$$\mu_k = \frac{\sum_{i=1}^{N_d} (d_i - \mu_d)^k}{N_d}, \quad (5.10)$$

em que d_i é o i -ésimo elemento do vetor de entrada de amostras \mathbf{d} ; μ_d é a média de \mathbf{d} ; N_d é o número de amostras de \mathbf{d} ; e k indica o k -ésimo momento desejado.

De acordo com o valor de β_3 é possível fazer as seguintes conclusões sobre a assimetria:

$$\begin{aligned} \beta_3 < 0 &\rightarrow \text{assimetria à esquerda} \\ \beta_3 = 0 &\rightarrow \text{assimetria nula} \\ \beta_3 > 0 &\rightarrow \text{assimetria à direita} \end{aligned}, \quad (5.11)$$

O momento de quarta ordem, chamado de curtose, é muito utilizado para a medição do estreitamento ou alargamento de uma distribuição comparativamente à distribuição

gaussiana (HINES, MONTGOMEY, *et al.*, 2013). Além disso, indica a propensão ou ocorrência de artefatos (*outliers*) na distribuição (WESTFALL, 2014). Ela pode ser estimada como:

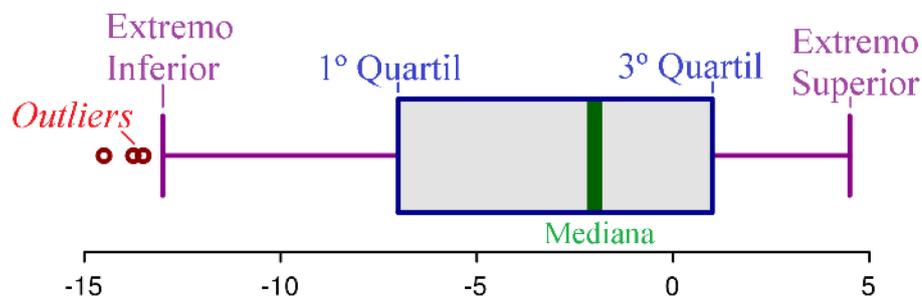
$$\beta_4 = \frac{\mu_4}{\sigma^4} - 3, \quad (5.12)$$

em que β_4 é a curtose normalizada; σ é o desvio padrão; e μ_4 é o momento central de quarta ordem. A normalização implica que β_4 seja nulo para distribuições gaussianas.

5.4 DIAGRAMA EM CAIXAS

O diagrama em caixa (*boxplot*) é uma forma de apresentar graficamente propriedades estatísticas como: dispersão, assimetria da distribuição de frequências e artefatos (BARBETTA, REIS e BORNIA, 2010). Um exemplo genérico de *boxplot* está representado na Figura 21, ele é construído usando informações da mediana, quartis e extremos da função densidade de probabilidade.

Figura 21 – Diagrama em caixa (*boxplot*).



Fonte: (O AUTOR, 2019).

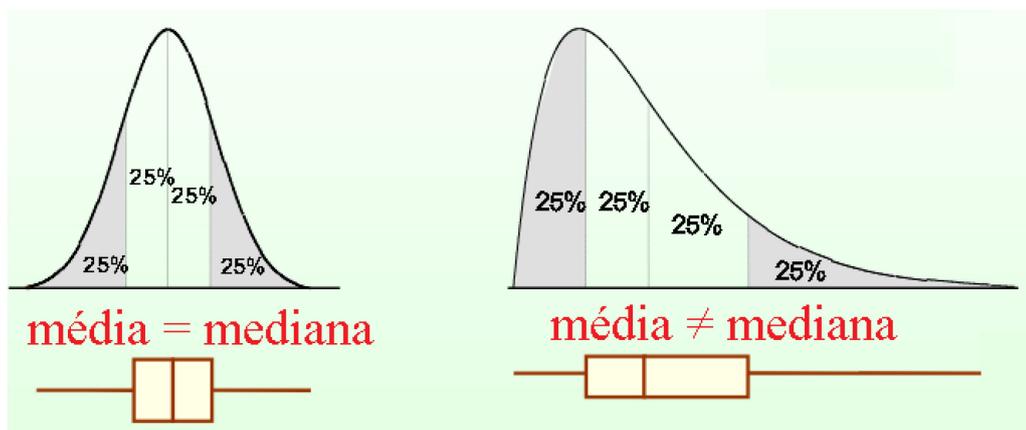
O *boxplot* é um retângulo que representa 50% dos valores mais típicos da distribuição, referentes ao primeiro e terceiro quartis. Os quartis são divididos pela mediana, que representa o segundo quartil. Valores que ultrapassem os extremos (*whiskers*) representados na Figura 21 são classificados como artefatos (*outliers*). Valores que ultrapassem em mais de três vezes o desvio interquartilício são representados de maneira diferente aos artefatos no gráfico contendo o *boxplot* (BARBETTA, REIS e BORNIA, 2010). Os extremos são calculados como:

$$\text{extremos} = \begin{cases} \text{ext}_{\text{inferior}} = q_1 - 1,5d_q \\ \text{ext}_{\text{superior}} = q_3 + 1,5d_q \end{cases}, \quad (5.13)$$

em que o quartil inferior (q_i) é o valor que separa os 25% menores valores da distribuição representada na Figura 22; o quartil superior (q_s) corresponde ao valor que separa os 25% maiores valores; e d_q é o desvio interquartílico.

Para uma melhor compreensão do *boxplot*, sua informação é comparada com curvas de distribuição representadas na Figura 22. Analisando a Figura 22, nota-se que em distribuições simétricas a média e a mediana são iguais, enquanto que em assimétricas a média tende a deslocar-se para o lado da cauda mais longa.

Figura 22 – Diagrama em caixa e sua distribuição de probabilidade equivalente.



Fonte: (BARBETTA, REIS e BORNIA, 2010).

5.5 CONSIDERAÇÕES FINAIS

Neste capítulo, entre outros assuntos, foram apresentados alguns métodos de inferência estatística. A partir deles é possível verificar se os dados obtidos apresentam ou não significância estatística.

O próximo capítulo apresentará a proposta de inovação desta dissertação.

6 PROPOSTA DE CONTRIBUIÇÃO E PROTOCOLOS PARA AVALIAÇÃO ESTATÍSTICA

Este capítulo apresenta as contribuições desta dissertação. Inicialmente é apresentada uma nova proposta para a construção da matriz de informação do método CSP, caracterizada pela inclusão de novas correlações inter e intra canais de EEG. Além disso, são propostas duas metodologias para avaliação estatística dos resultados.

6.1 FUNDAMENTAÇÃO

Os sistemas BCI têm como princípio básico a classificação de eventos de sincronização e dessincronização associados à imaginação de diferentes tipos de movimentos (BERNARDI, PIMENTA e MORENO, 2019) (ALOMARI, SAMAHA e ALKAMHA, 2013). O processo conhecido como imagética motora aciona circuitos neurais em áreas associadas ao córtex motor, de forma muito similar à realizada pela execução de movimentos reais (SELFSLAGH, SHOKUR, *et al.*) (WILLIAMS, 2019).

Os parâmetros usualmente utilizados para a classificação da imagética motora são as potências dos sinais EEG nos diversos eletrodos de aquisição e as correlações entre os diversos canais em um mesmo instante de tempo (LOTTE e GUAN, 2011) (GE, WANG e YU, 2014) (THIYAM e RAJKUMAR, 2016).

Apesar do reconhecido sucesso dos sistemas convencionais de imagética motora, que atingem elevadas taxas de acurácia na classificação dos eventos (ANG, ZHENG, *et al.*, 2008), trabalhos científicos referentes ao estudo de atividades cognitivas têm reportado alterações estatisticamente significantes nas correlações entre os sinais EEG captados em regiões relativamente distantes do escalpo (tanto ipsilateral quanto contralateral) (KLEEM, LI e HERNANDEZ, 2000). Esses achados podem ser decorrentes do espalhamento elétrico ocasionado pela atividade elétrica das pós-sinapses no córtex cerebral, como também de vias neurais ainda desconhecidas. Em ambas as situações, entende-se que atrasos de propagação estejam envolvidos.

Adicionalmente, novos estudos têm apresentado indícios de comportamento dinâmico não-linear do cérebro e, portanto, informações relevantes relativas aos processos sensório-motores podem estar sendo negligenciadas (RODRIGUEZ-BERMUDEZ e GARCIA-

LAENCINA, 2015). A partir dessa constatação, novas ferramentas e técnicas de processamento têm sido propostas.

A partir desses avanços nas observações e no entendimento do funcionamento do cérebro, existem indícios de que informações relevantes ao processo de classificação de imagética motora possam ter sido ignoradas até o momento e possam vir a ser incorporadas aos métodos convencionais, como é o caso do classificador CSP-FLDA, resultando em aumento de eficiência. A inclusão dessas informações poderia elevar as taxas de classificações corretas, trazendo maior confiabilidade aos sistemas implementados.

Dessa forma, neste trabalho, propõe-se a investigação do impacto da inclusão de novas informações, referentes às correlações inter e intra-eletrodos na matriz de informação do classificador CSP-FLDA, e suas variações, em sistemas BCI baseados em imagética motora.

6.2 MATRIZ DE INFORMAÇÃO

Nesta seção é apresentada uma proposta para expansão da matriz de informações do classificador CSP-FLDA. Ela refere-se ao uso de momentos de segunda ordem adicionais, para atraso não nulo, inter e intra canais de EEG.

6.2.1 Matriz de informação expandida usando momentos de segunda ordem

O método CSP-FLDA, amplamente utilizado na literatura em aplicações de imagética motora (WOLPAW e WOLPAW, 2012) faz uso de uma matriz de informações espaciais associadas à distribuição dos sinais de EEG ao longo do escalpo.

Particularmente, a imagética motora das mãos está associada aos ritmos μ_r e β_r (BELHADJ, BERNOUMASSAT e JRACHAI, 2015). Esses ritmos possuem características de amplitude e frequência relativamente bem definidas, o que possibilita a sua identificação por sistemas especialistas.

A informação comumente utilizada em classificadores CSP-FLDA é composta por estimativas de um conjunto de momentos de segunda ordem inter-eletrodos dentro de uma determinada época. Essas correlações podem ser visualizadas pela substituição de (3.3) em (3.5) e tomando-se seu valor esperado, de forma que:

$$E\{\mathbf{C}_\rho\} = P \times \begin{bmatrix} E\{x_{\rho,n,1}^2\} & E\{x_{\rho,n,1}x_{\rho,n,2}\} & \cdots & E\{x_{\rho,n,1}x_{\rho,n,Q}\} \\ E\{x_{\rho,n,2}x_{\rho,n,1}\} & E\{x_{\rho,n,2}^2\} & \cdots & E\{x_{\rho,n,2}x_{\rho,n,Q}\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{x_{\rho,n,Q}x_{\rho,n,1}\} & E\{x_{\rho,n,Q}x_{\rho,n,2}\} & \cdots & E\{x_{\rho,n,Q}^2\} \end{bmatrix}, \quad (6.1)$$

ou ainda:

$$E\{\mathbf{C}_\rho\}_{l,c} = P \times E\{x_{\rho,n,l}x_{\rho,n,c}\}, \quad (6.2)$$

em que l e c correspondem à l -ésima linha e c -ésima coluna da matriz \mathbf{C}_ρ ; e $x_{\rho,n,l}$ é o sinal do l -ésimo canal de EEG ao longo do tempo discreto n ; e $\rho \in \{A, B\}$ é a classe a que pertence.

Dessa forma, verifica-se que as informações espaciais estão associadas às potências dos sinais de cada canal e às suas correlações cruzadas sincronizadas em um determinado instante de tempo.

Nesta subseção, propõe-se expandir o conjunto de informações contidas na matriz de informação original do método CSP, por meio da inclusão de momentos de segunda ordem contendo correlações para (L) atrasos de tempo não nulos entre os diferentes canais, para uma mesma época. Essas correlações adicionais são obtidas diretamente das matrizes de épocas \mathbf{X}_n^ρ , originalmente utilizadas para a obtenção de \mathbf{C}_ρ , ou seja, sem acréscimo de novos sinais. Com esse procedimento deseja-se não apenas utilizar informações referentes à sincronização e dessincronização (potência) de eventos, como também informação temporal relevante, associada às estatísticas dos ritmos envolvidos em um mesmo eletrodo, assim como às correlações entre eletrodos no escalpo do paciente para diferentes atrasos (*lags*).

Dessa forma, propõe-se uma nova matriz de épocas, definida como:

$$\tilde{\mathbf{X}}_n = [\tilde{x}_{n,1,0} \ \tilde{x}_{n,1,1} \ \cdots \ \tilde{x}_{n,1,L} \ \tilde{x}_{n,2,0} \ \tilde{x}_{n,2,1} \ \cdots \ \tilde{x}_{n,2,L} \ \cdots \ \tilde{x}_{n,Q,0} \ \cdots \ \tilde{x}_{n,Q,L}]^T, \quad (6.3)$$

em que $\tilde{\mathbf{X}}_n$ possui dimensão $Q(L+1) \times (P-L)$; e $\tilde{x}_{n,q,l} = [x_{n,q(1+l)} \ x_{n,q(2+l)} \ \cdots \ x_{n,q(P-L+l)}]^T$.

Como resultado, a nova matriz de informações é descrita por:

$$\tilde{\mathbf{C}}_\rho = \sum_{n=1}^{N_\rho} \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T. \quad (6.4)$$

Tomando-se o valor esperado de (6.4), a matriz de correlação do sinal de entrada é dada por:

$$E\{\tilde{\mathbf{C}}_\rho\} = P \times \begin{bmatrix} E\{\tilde{x}_{\rho,n,0}^2\} & E\{\tilde{x}_{\rho,n,0}\tilde{x}_{\rho,n,1}\} & \cdots & E\{\tilde{x}_{\rho,n,0}\tilde{x}_{\rho,n,Q,L}\} \\ E\{\tilde{x}_{\rho,n,1}\tilde{x}_{\rho,n,0}\} & E\{\tilde{x}_{\rho,n,1}^2\} & \cdots & E\{\tilde{x}_{\rho,n,1}\tilde{x}_{\rho,n,Q,L}\} \\ E\{\tilde{x}_{\rho,n,1,2}\tilde{x}_{\rho,n,0}\} & E\{\tilde{x}_{\rho,n,1,2}\tilde{x}_{\rho,n,1}\} & \cdots & E\{\tilde{x}_{\rho,n,1,2}\tilde{x}_{\rho,n,Q,L}\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{\tilde{x}_{\rho,n,Q,L}\tilde{x}_{\rho,n,0}\} & E\{\tilde{x}_{\rho,n,Q,L}\tilde{x}_{\rho,n,1}\} & \cdots & E\{\tilde{x}_{\rho,n,Q,L}^2\} \end{bmatrix}, \quad (6.5)$$

ou ainda

$$E\{\tilde{\mathbf{C}}_\rho\}_{l,c} = P \times E\{\tilde{x}_{\rho,n,a_1,b_1}\tilde{x}_{\rho,n,a_2,b_2}\}, \quad (6.6)$$

em que

$$\begin{aligned} a_1 &= \left\lfloor \frac{(l-1)}{(L+1)} \right\rfloor + 1 \\ b_1 &= \begin{cases} L, & u_1 = 0 \\ u_1 - 1, & u_1 > 0 \end{cases} \text{ com } u_1 = \text{mod}(l, (L+1)) \\ a_2 &= \left\lfloor \frac{(c-1)}{(L+1)} \right\rfloor + 1 \\ b_2 &= \begin{cases} L, & u_2 = 0 \\ u_2 - 1, & u_2 > 0 \end{cases} \text{ com } u_2 = \text{mod}(c, (L+1)) \end{aligned}, \quad (6.7)$$

em que a_1 e a_2 correspondem aos eletrodos; b_1 e b_2 correspondem aos atrasos nas posições dadas respectivamente pelo par (l, c) na matriz $\tilde{\mathbf{C}}_\rho$; $\text{mod}\{\cdot\}$ é o resto da divisão do argumento; e $\lfloor \cdot \rfloor$ é a parte inteira do argumento, ou seja, o seu arredondamento para baixo.

A equação (6.6) apresenta uma nova forma de construir a matriz de informação.

6.3 PROTOCOLOS ESTATÍSTICOS

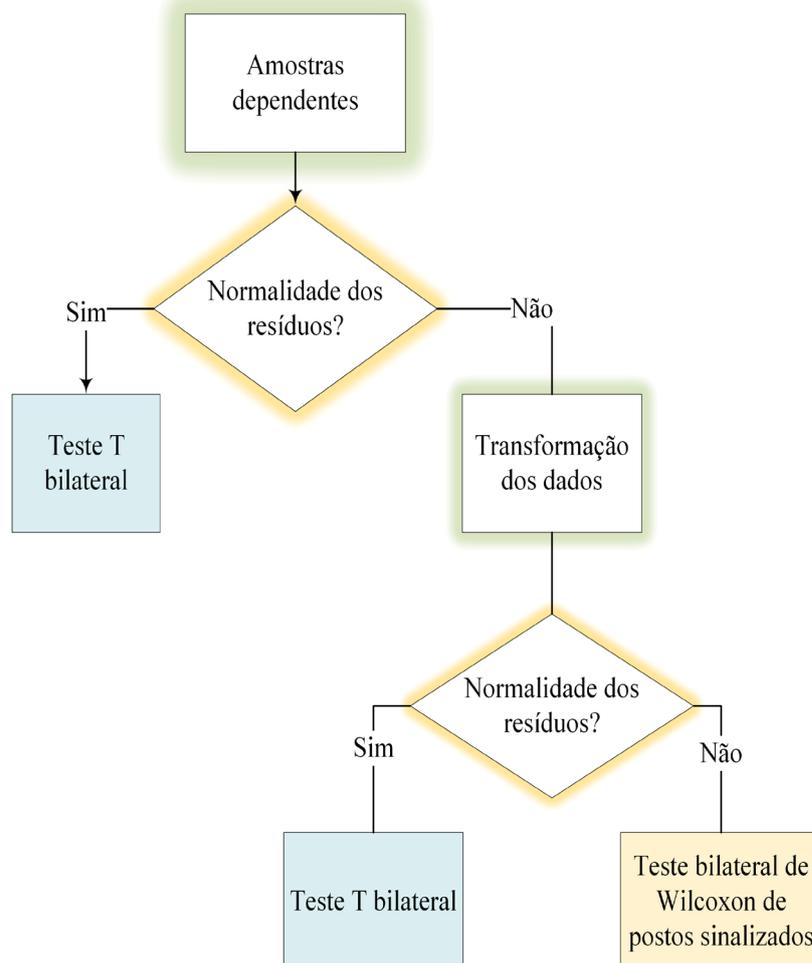
A partir dos conceitos estatísticos e métodos de inferência apresentados no capítulo 5, utilizaram-se dois procedimentos propostos para a identificação da significância estatística entre espaços amostrais, com a finalidade de verificar o aumento de desempenho dos classificadores avaliados em relação às novas matrizes de informações.

Considerando-se somente dois espaços amostrais, o procedimento de avaliação estatística com duas distribuições amostrais dependentes é ilustrado na Figura 23. Nota-se a dependência em relação à característica de normalidade da distribuição amostral para a escolha dos métodos.

Analisando o procedimento apresentado na Figura 23, verifica-se que caso a distribuição seja normal, então é aplicado o teste T pareado bilateral. Caso seja violada a premissa de gaussianidade, então utilizam-se transformações do tipo *boxcox*, logarítmica ou raiz quadrada (TRIOLA, 2014). A transformação é realizada, por exemplo, substituindo-se cada valor de amostra de um determinado espaço amostral x_i por $(x_i)^{1/2}$. Além da normalidade, a transformação amostral pode satisfazer a exigência de mesma variância entre espaços amostrais distintos (TRIOLA, 2014). Caso não seja obtida normalidade com o uso das transformações então é aplicado o teste bilateral não paramétrico de Wilcoxon de postos sinalizados (WITTE e WITTE, 2005).

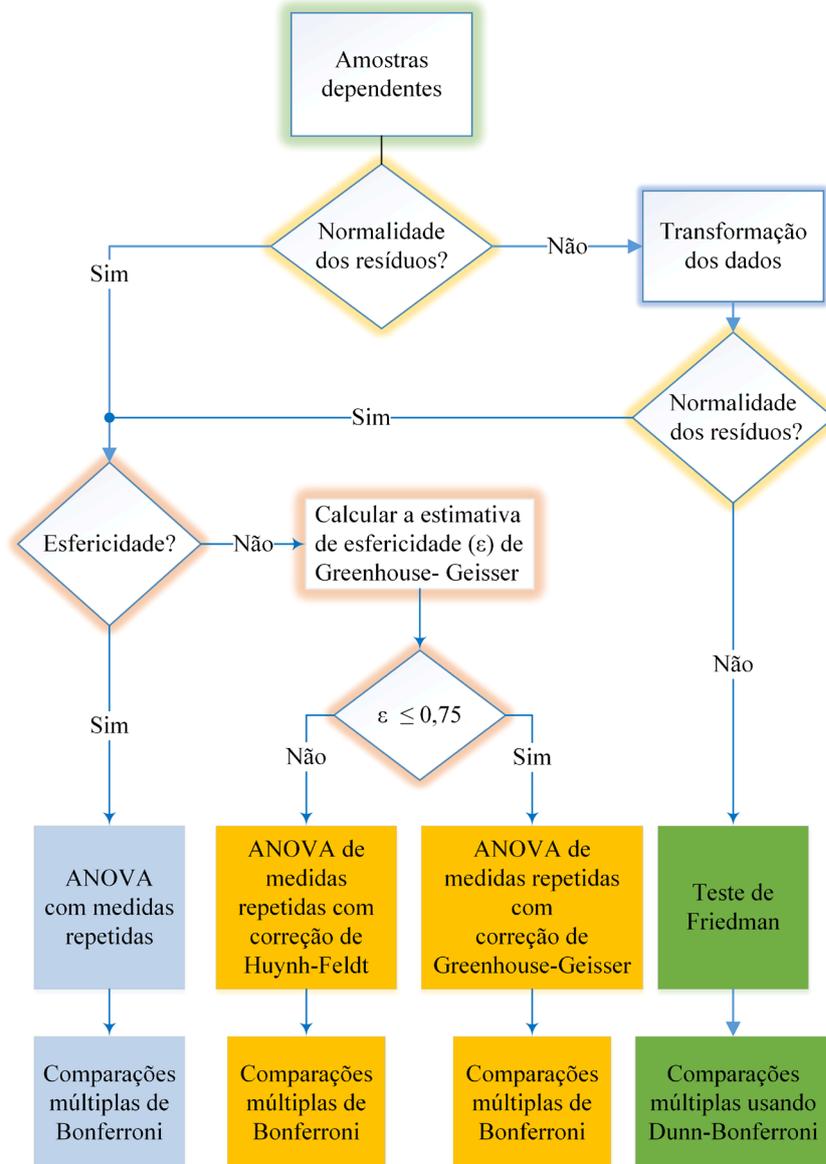
A Figura 24 apresenta o procedimento utilizado para a análise de três ou mais espaços amostrais. Para fazer a detecção de diferenças entre os pares possíveis dentro do grupo, pode-se usar a abordagem paramétrica (teste de Bonferroni) ou a não paramétrica (teste de Dunn-Bonferroni) (ZWICK, NEUHOFF, *et al.*, 1982).

Figura 23 – Teste pareado para dois espaços amostrais.



Fonte: (O AUTOR, 2019).

Figura 24 - Teste para mais de 2 espaços amostrais.



Fonte: (O AUTOR, 2019).

6.4 CONSIDERAÇÕES FINAIS

Este capítulo apresentou uma proposta para a expansão da matriz de informações do método de CSP. Além disso, traz o desenvolvimento de dois protocolos estatísticos para a identificação da significância estatística para dois ou mais espaços amostrais. O próximo capítulo apresentará os resultados obtidos com a nova matriz de informações em problemas de imagética motora.

7 RESULTADOS

Este capítulo apresenta simulações estatísticas e os resultados obtidos para avaliação do impacto da nova proposta de matriz de informação em problemas de imagética motora baseadas no classificador CSP-FLDA e suas variações. Os resultados são obtidos e depois analisados usando métricas objetivas de desempenho, inferência estatística e ferramentas gráficas.

O objetivo é realizar uma avaliação dos resultados obtidos com o método proposto em comparação com o uso da matriz de informação convencional amplamente utilizada na literatura.

O método clássico (ou convencional) consiste no classificador com matriz de informação convencional ($L = 0$), enquanto que o método proposto consiste no mesmo classificador, porém com uso da matriz de informação apresentada na equação (6.5) ($L > 0$) ou nas equações (6.9) e (6.11).

7.1 DESCRIÇÃO GERAL

Sinais de imagética motora provenientes dos bancos de dados IIA (BRUNNER, LEEB, *et al.*, 2008) e IIIA (GE, WANG e YU, 2014) da Universidade de Graz na Áustria foram utilizadas para avaliação dos desempenhos (acurácias) dos métodos em questão.

Cada acurácia foi calculada a partir de $N = 100$ realizações do processo de classificação. Em cada uma das realizações foi realizado o processo de separação aleatória dos bancos de dados IIA e IIIA, em conjuntos de treinamento e validação. Cada um desses conjuntos foi composto por 50% das épocas de cada banco, sendo: (a) 1296 épocas do banco IIA e 210 épocas do banco IIIA; para o caso de classificação binária; e (b) 2592 épocas do banco IIA e 420 épocas do banco IIIA para classificação quaternária. Garantiu-se que cada conjunto aleatório de treinamento possuísse uma diversidade de épocas de, pelo menos, 20% em relação aos demais. O valor $N > 30$ foi escolhido para que fosse possível assumir uma distribuição gaussiana e permitir o uso de testes estatísticos mais robustos nas comparações entre classificadores e voluntários (HINES, MONTGOMEY, *et al.*, 2013).

A janela de tempo selecionada (representada na Figura 14) corresponde a um intervalo de tempo de 2 segundos, totalizando 500 amostras.

Comparações envolvendo dois espaços amostrais seguem o procedimento apresentado na Figura 23. Para três ou mais espaços amostrais utiliza-se o procedimento da Figura 24.

O teste de aderência para gaussianidade é o de Shapiro-Wilk, pois apresenta maior poder estatístico considerando todos os tipos de distribuições e quantidades de amostras (RAZALI e WAK, 2011).

Ao longo dos experimentos, os sujeitos do banco de dados IIA são denominados como S₁, S₂, S₃, S₄, S₅, S₆, S₇, S₈ e S₉, enquanto que os do IIIA, são nomeados como B₁, B₂ e B₃. O classificador utilizado será o FLDA em todas as simulações realizadas. O termo “atraso” será utilizado para representar a diferença de tempo (*lag*) entre os sinais e o termo “realização” para indicar um processo de classificação, que resulta no cálculo de uma acurácia.

7.2 ANÁLISE DE MOMENTOS DE ORDEM SUPERIOR

Nesta seção é avaliada a relevância dos momentos de ordem superior na matriz de informações.

O procedimento consiste na avaliação da assimetria e curtose das distribuições amostrais dos sinais EEG presentes nos bancos de dados IIA e IIIA.

A Tabela 3 apresenta valores de assimetria e curtose referentes aos canais C₃, C_z e C₄ que se localizam sobre o córtex motor de imagética motora das mãos. Esses canais foram escolhidos devido a estarem sob o córtex motor que é a área de interesse para classificação e para ilustrarem as características estatísticas dos sinais EEG no escalpo. Os dados da Tabela 3 foram calculados conforme equações (5.9) e (5.12).

Tabela 3 – Assimetria e curtose média para os canais C₃, C_z e C₄.

Canal	Assimetria	Curtose
C ₃	0,06	-0,18
C _z	-0,11	0,06
C ₄	-0,07	-0,13

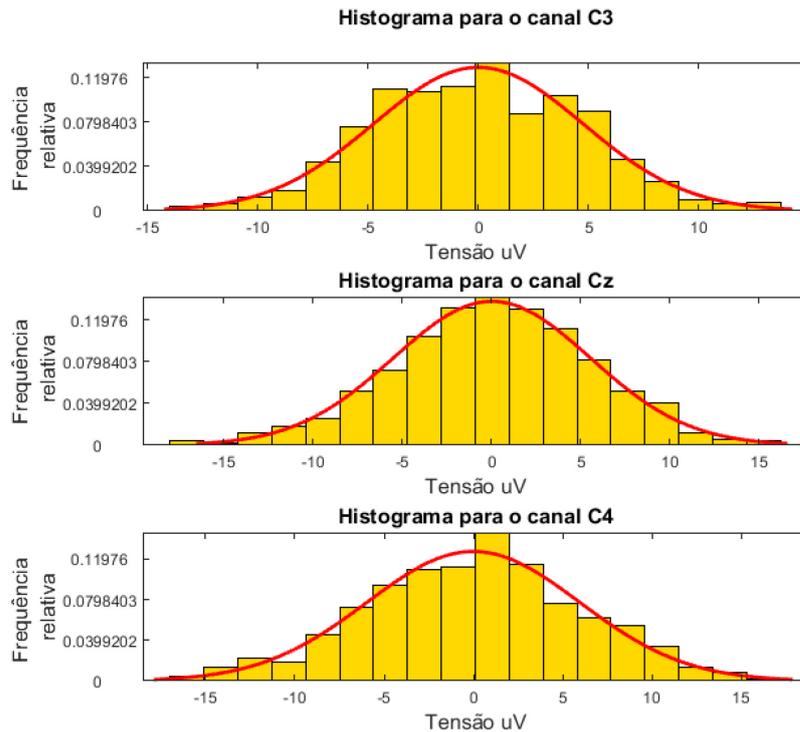
Fonte: (O AUTOR, 2019).

Analisando-se os resultados da Tabela 3, verifica-se que a assimetria e a curtose normalizada são muito pequenas, indicando que as distribuições são aproximadamente simétricas e sem artefatos (*outliers*), conforme ilustrado nos histogramas da Figura 25.

Como conclusão, verificou-se de forma preliminar que aparentemente, em função da baixa relação sinal ruído, não há informação significativa referente aos momentos de ordem

superior do EEG, o que concorda com a consideração de Gaussianidade para sinais EEG utilizada comumente na literatura (BLANCO, 1995) (GONEN e TCHESLAVSKI, 2012). Dessa forma, ao longo do restante desse trabalho não serão utilizados momentos de ordem superior na matriz de informações.

Figura 25 – Histograma para diferentes canais de eletrodos em conjunto com a distribuição normal sobreposta.



Fonte: (O AUTOR, 2019).

7.3 CLASSIFICADOR CSP-FLDA

Nesta seção é realizada uma comparação de desempenho entre os resultados obtidos pelo método CSP-FLDA utilizando a matriz de informações clássica e a proposta na equação (6.5).

O parâmetro de *offset* utilizado para os bancos de dados IIA é de 0,5s, conforme sugerido por (BRUNNER, LEEB, *et al.*, 2008). Para o banco de dados IIIA foi utilizado um *offset* de 3,5 segundos, conforme (SCHLÖGL, 2005). Em todas as simulações foram utilizados 3 pares de filtros para extração de características, conforme sugerido no artigo de Lotte e Guan (2011).

Foram calculadas as acurácias médias e desvios padrões para $L = \{ 0, 1, 2, 3, 4, 5 \}$ e suas significâncias estatísticas foram avaliadas de acordo com o procedimento descrito na Figura 24, utilizando 100 realizações.

Na Tabela 4 o método CSP-FLDA com matriz de informação convencional é identificado por $L = 0$ e o CSP-FLDA proposto pelas demais colunas $\{L > 0\}$.

Os resultados do teste de *Saphiro-Wilk* com $\alpha = 1\%$ demonstram que as acurácias obtidas para os sujeitos S_1, S_6, S_7, B_1 e B_3 não apresentam distribuição normal. Foi então aplicada a transformação de *boxcox*, resultando em amostras com distribuição Gaussiana. Em sequência, com a premissa de gaussianidade atendida, foi aplicado o teste ANOVA de medidas repetidas com a correção de *Greenhouse-Geisser* para os participantes: S_1, S_2, S_4, S_7, S_9 e B_1 , e de *Huynh-Feldt* para: S_3, S_5, S_6, S_8, B_2 e B_3 . A utilização de um procedimento de correção decorre do fato de que a premissa de esfericidade requerida pela ANOVA de medidas repetidas não foi atendida. A correção de esfericidade de *Huynh-Feldt* foi utilizada nos casos em que a ANOVA com a correção de esfericidade de *Greenhouse-Geisser* resultou em valor menor ou igual a 0,75.

Tabela 4 – Acurácia do classificador CSP-FLDA para imagética motora de mão direita e esquerda. A primeira coluna apresenta o método clássico ($L = 0$) e as demais o método proposto para $L = \{1, 2, \dots, 5\}$. A última linha apresenta a média de cada coluna em conjunto com o desvio padrão. A coluna $L = 0$ é a de referência para comparações múltiplas. Em negrito e azul são destacadas as distribuições de maior média amostral.

	$L = 0$	$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$
S_1	88,7±2,4	93,6±2,0	93,1±1,8	93,8±1,8	93,2±1,9	92,8±2,2
S_2	58,1±4,3	58,9±4,2	57,2±4,5	57,7±5,1	57,2±4,9	56,4±4,6
S_3	97,1±1,1	97,1±1,1	97,6±1,0	97,6±1,0	97,8±0,9	97,6±1,1
S_4	68,6±4,0	69,0±3,7	68,4±4,5	66,5±4,2	66,6±4,7	65,5±4,8
S_5	58,7±4,0	66,8±3,6	83,6±2,7	89,5±2,2	90,2±2,2	90,4±2,4
S_6	67,1±3,1	71,1±3,5	70,0±4,7	68,6±4,9	66,6±4,5	65,3±4,8
S_7	80,1±3,0	83,5±3,5	83,6±4,3	80,8±4,5	81,9±4,9	80,9±5,0
S_8	96,9±1,3	97,0±1,5	96,4±1,5	96,6±1,4	96,8±1,3	97,2±1,2
S_9	93,8±1,4	93,1±1,6	93,3±1,6	93,3±1,6	93,1±1,5	93,0±1,6
B_1	95,2±2,6	98,5±1,3	98,8±1,2	98,9±1,0	98,5±1,4	97,8±2,2
B_2	58,8±6,5	62,2±6,2	61,2±5,9	57,7±6,9	59,1±7,1	93,2±4,3
B_3	95,8±2,7	96,7±2,8	95,3±4,0	95,1±3,4	93,2±4,3	93,9±3,7
$\mu \pm \sigma$	79,9±12	82,3±15	83,2±15	83,0±16	82,9±16	82,4±16

Fonte: (O AUTOR, 2019).

Os resultados obtidos indicaram a existência de pelo menos um par de médias diferente das demais, sendo a probabilidade de erro menor que 5% ($p < 0,05$). Dessa forma, aplicou-se o

teste de comparações múltiplas de Bonferroni para a identificação dos pares que apresentam diferença estatística, sendo os resultados obtidos apresentados na Tabela 4. A coluna contendo as acurácias médias para $L = 0$ foi utilizada como base para as comparações do método post-hoc (comparações múltiplas). Para cada linha, a coluna com a maior acurácia média é destacada em negrito e azul. Em caso de semelhança estatística, mais de uma coluna é destacada.

Analisando os dados da Tabela 4 e comparando-se os resultados globais para todos os voluntários (última linha da Tabela 4), verifica-se que o método proposto para $L = 2$ apresenta um aumento médio de 3,3% na acurácia de classificação, em relação ao método convencional.

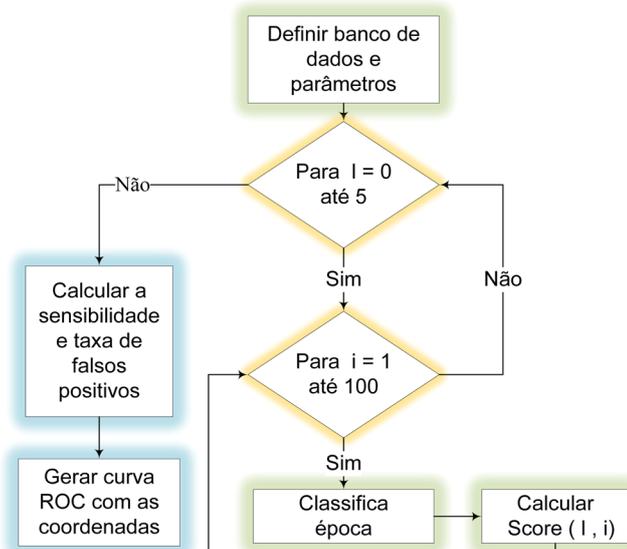
Para o caso $L = 2$ o método proposto apresenta desempenho superior ou pelo menos idêntico para todos os sujeitos dos dois bancos de dados, exceto para S_8 . Além disso, com um nível de confiança de 95% pode-se afirmar que as 12 maiores acurácias são obtidas usando o método proposto.

Para $L = 1$ há melhora de acurácia em todos os participantes em relação ao método clássico. Em 7 participantes há melhora de pelo menos 3% na acurácia média usando o método proposto.

Destaca-se nos resultados da Tabela 4, o participante S_5 , que aumentou sua acurácia média de 58,7% para 90,4%.

Uma forma alternativa para se comparar o desempenho propiciado pelo método proposto é a construção de curvas ROC.

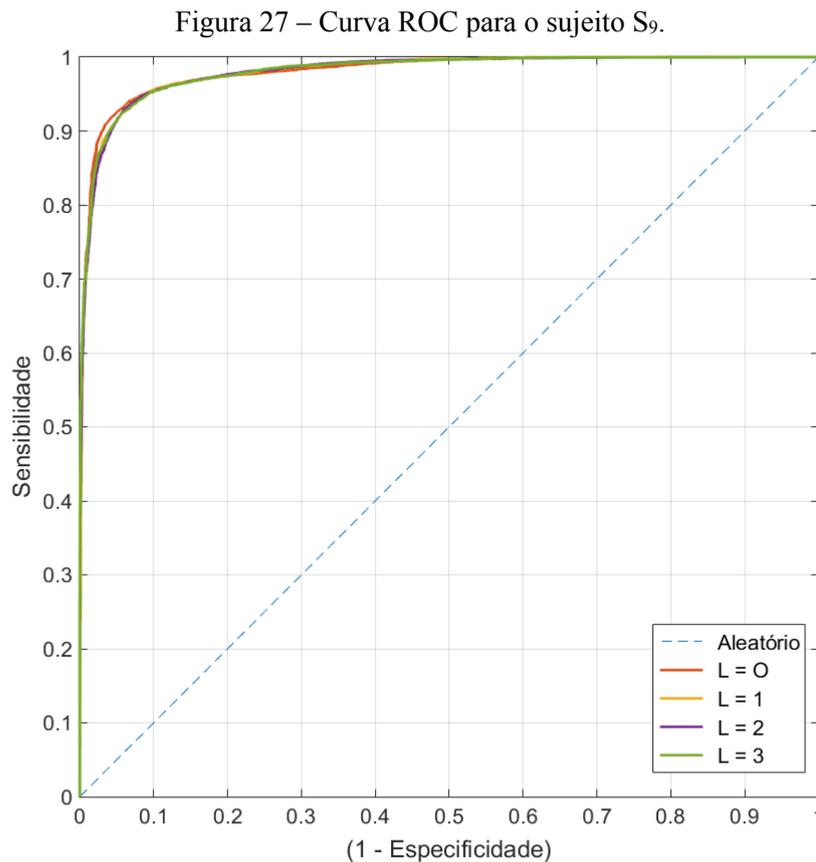
Figura 26 – Protocolo para geração da curva ROC. A variável l representa o atraso adicionado para a construção da matriz de informação. A variável i representa o número da realização. $\text{Score}(l, i)$ é o cálculo da pontuação da época em análise.

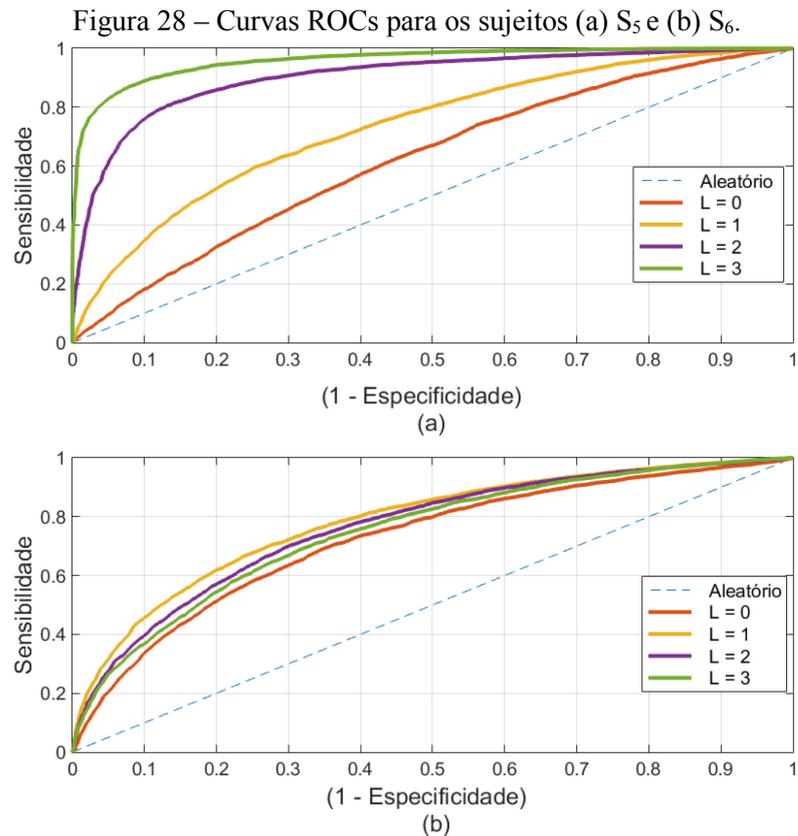


Fonte: (O AUTOR, 2019).

O procedimento de obtenção da curva ROC é apresentado na Figura 26. Uma vez selecionado o banco de dados e o conjunto de parâmetros de simulação, para cada um dos valores de atraso (L), são geradas $N = 100$ realizações, permitindo gerar os pares de coordenadas (especificidade, sensibilidade) para construção da curva ROC. Ao final dos laços de repetições, será gerada uma variável contendo as pontuações (em inglês, scores). Os *scores* representam a probabilidade à posteriori da época pertencer a determinada classe.

A Figura 27 apresenta a curva ROC para o participante S_9 , variando-se o atraso (L) de 0 a 3. O intuito da Figura 27 é demonstrar que as acurácias médias para o participante S_9 , produzem curvas ROC quase idênticas. Analisando a Figura 27, percebe-se que as curvas quase se sobrepõem, indicando a semelhança encontrada pela etapa de inferência estatística, além disso, o parâmetro AUC (integral da curva ROC) é de aproximadamente 0,98 para todos os métodos, isso indica um classificador quase ótimo, visto que o máximo é 1.





Fonte: (O AUTOR, 2019).

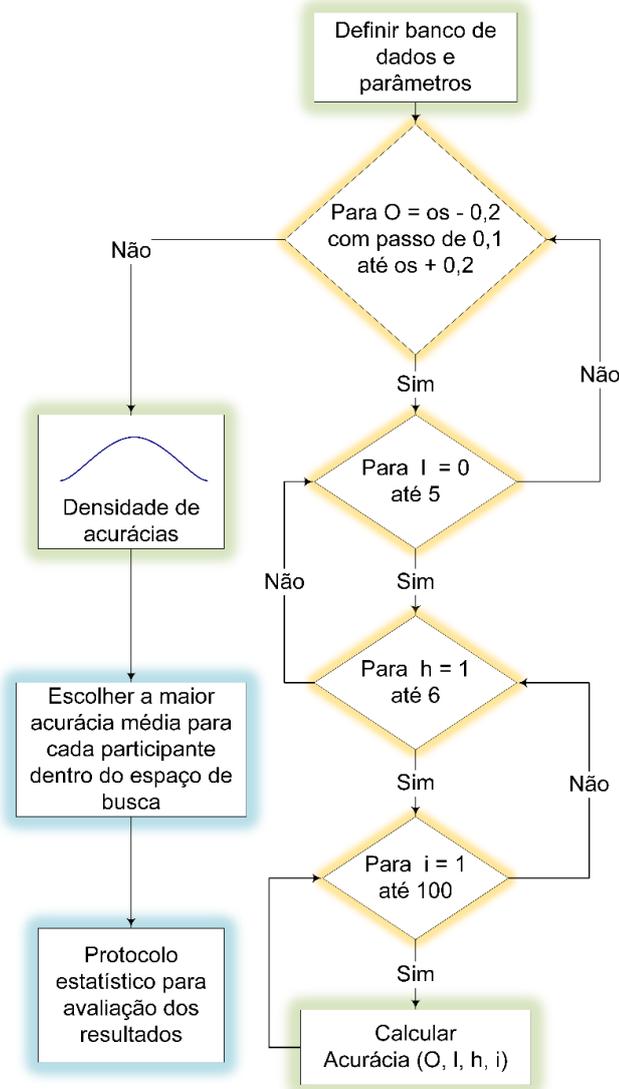
A Figura 28 ilustra a curva ROC para os participantes S_5 e S_6 . A AUC para $L = 0$ é de 0,62, enquanto para $L = 3$ esse valor sobe para 0,96. Para o participante S_6 , em $L = 0$ a AUC é de 0,72 e para $L = 1$ esse valor sobe para 0,78.

Conclui-se, pelos resultados, que a adição de versões atrasadas dos sinais para a composição da nova matriz de informação (método proposto) no CSP aumenta o desempenho do processo de classificação.

7.4 OTIMIZAÇÃO DE PARÂMETROS INDIVIDUAIS

A análise comparativa entre o método clássico e o método proposto, apresentada na seção anterior, foi baseada na implementação do classificador CSP-FLDA utilizando parâmetros pré-definidos indicados na literatura. Nesta seção é realizada a análise comparativa de desempenho entre ambos os métodos utilizando conjuntos de parâmetros otimizados para cada sujeito. O protocolo utilizado para geração dos resultados desta subseção está ilustrado na Figura 29.

Figura 29 - Protocolo para otimização de parâmetros e geração das maiores acurácias médias. O parâmetro os depende do banco de dados, sendo definido como 3,5s para o IIIA e 0,5s para o IIA. A variável l representa o atraso adicionado para a construção da matriz de informação. A variável h representa a quantidade de pares de filtros. A variável i representa o número de realizações da etapa de classificação. Acurácia (O, l, h, i) é o cálculo da acurácia de classificação das épocas de execução disponíveis de acordo com os parâmetros internos da função.



Fonte: (O AUTOR, 2019).

Foram utilizados todos os canais disponíveis. Os parâmetros de projeto: *offset*, atraso e número de pares de filtros, são otimizados dentro de um espaço de busca, de forma a obter a máxima acurácia média de classificação para cada sujeito de cada banco de dados. A Figura 29 apresenta o procedimento realizado, no qual o primeiro laço de repetição representa a escolha do *offset*, que depende do banco de dados escolhido: $O = \{3,3; 3,4; 3,5; 3,6; 3,7\}$ (em segundos) para o banco de dados IIIA e $O = \{0,3; 0,4; 0,5; 0,6; 0,7\}$ (em segundos) para o banco de dados IIA. O segundo laço de repetição refere-se ao atraso $l = \{0, 1, 2, 3, 4, 5\}$; e o terceiro, refere-se à quantidade de pares de filtros $h = \{1, 2, 3, 4, 5, 6\}$.

Ao final do processo é calculada a acurácia média para as 100 realizações de cada um dos 180 conjuntos de parâmetros, para cada sujeito.

O procedimento apresentado na Figura 23 é utilizado para definir se existe diferença estatística entre os resultados.

A Tabela 5 apresenta os resultados de simulação. De acordo com as premissas do método paramétrico de amostras repetidas, foi avaliada a gaussianidade dos resíduos dos pares com $\alpha = 1\%$. Os testes indicaram que os pares contendo os sujeitos S₉, B₁ e B₃ não apresentam gaussianidade, sendo que os restantes são normalmente distribuídos. A transformação de *boxcox* fez com B₁ e B₃ se tornassem gaussianos. Dessa forma, para S₉ é utilizado o teste de Wilcoxon de postos sinalizados e para os demais sujeitos é utilizado o teste T pareado, ambos com $\alpha = 5\%$. As maiores acurácias, definidas por meio da probabilidade de significância (p), são destacadas em negrito e azul. Quando estatisticamente semelhantes ambos os métodos são destacados.

Tabela 5 – Acurácias médias para cada sujeito de acordo com os seus parâmetros ótimos dentro do espaço de busca. A última linha representa a média e a variância global. As duas últimas colunas apresentam os parâmetros ótimos encontrados dentro do espaço de busca. Em negrito e azul, as maiores acurácias. Na tabela, O representa o *offset*, L o atraso e W é referente à quantidade de pares de filtros espaciais selecionados.

	CSP	CSP	p	Parâmetros ótimos	
	proposto	clássico		Proposto	/ Clássico
S ₁	94,1±1,7	88,7±2,4	1×10^{-39}	$O = 3,4; L = 1; W = 3$	$O = 3,5; W = 3$
S ₂	61,3±4,0	60,0±4,6	2×10^{-3}	$O = 3,3; L = 1; W = 6$	$O = 3,3; W = 6$
S ₃	97,8±0,9	97,2±1,1	7×10^{-9}	$O = 3,5; L = 4; W = 3$	$O = 3,5; W = 2$
S ₄	71,6±4,2	71,3±4,0	2×10^{-1}	$O = 3,7; L = 1; W = 2$	$O = 3,7; W = 5$
S ₅	93,0±1,8	61,8±4,3	4×10^{-85}	$O = 3,3; L = 4; W = 2$	$O = 3,3; W = 6$
S ₆	72,7±3,4	68,0±3,1	1×10^{-22}	$O = 3,4; L = 1; W = 3$	$O = 3,3; W = 3$
S ₇	85,1±3,4	81,4±4,0	6×10^{-18}	$O = 3,3; L = 5; W = 6$	$O = 3,3; W = 2$
S ₈	97,8±1,3	97,2±1,4	2×10^{-8}	$O = 3,5; L = 1; W = 1$	$O = 3,5; W = 1$
S ₉	94,0±1,4	93,9±1,5	2×10^{-1}	$O = 3,5; L = 1; W = 1$	$O = 3,5; W = 1$
B ₁	99,2±0,8	96,1±2,2	7×10^{-29}	$O = 0,4; L = 3; W = 3$	$O = 0,6; W = 2$
B ₂	67,4±6,1	61,6±4,7	6×10^{-13}	$O = 0,6; L = 1; W = 1$	$O = 0,6; W = 2$
B ₃	97,8±2,0	96,5±2,5	5×10^{-6}	$O = 0,4; L = 2; W = 5$	$O = 0,5; W = 2$
$\mu \pm \sigma$	86,0±14	81,1±15	3×10^{-99}	-	-

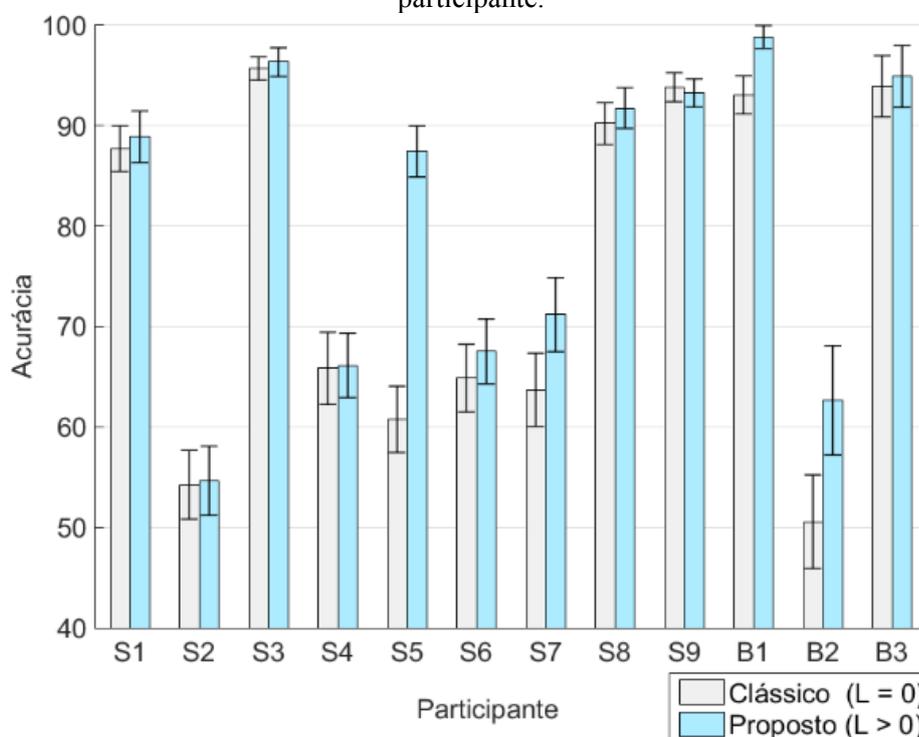
Fonte: (O AUTOR, 2019).

Analisando os resultados da Tabela 5, nota-se que os sujeitos S₄ e S₉, estatisticamente, apresentam o mesmo desempenho para ambos os métodos, pois p é maior que o α estipulado. Para os demais participantes, o método proposto é superior em todos os casos, pois p é menor

que α . Ao comparar o resultado médio global, que se refere à última linha da Tabela 5, conclui-se que há um aumento de acurácia média de 5% do método proposto em relação ao método convencional. Este resultado é significativo, de acordo com o valor de p .

A Figura 30 complementa a apresentação dos resultados da Tabela 5. Nela, são apresentadas as acurácias médias e respectivos desvios padrões para cada sujeito da Tabela 5. Analisando a imagem, nota-se claramente o aumento da acurácia média.

Figura 30 – Comparação de acurácias médias entre o CSP proposto e o clássico, usando todos os eletrodos disponíveis. Em conjunto com a acurácia está o respectivo desvio padrão para cada participante.



Fonte: (O AUTOR, 2019).

Em Ge, Wang e Yu (2014), os autores apontam que o método CSP-FLDA é capaz de produzir bons resultados usando poucos canais de aquisição (no caso específico, de 1 a 3 eletrodos). Partindo dessa ideia, realizou-se o mesmo procedimento anterior de otimização de parâmetros, utilizando-se apenas os 7 canais posicionados sobre o córtex motor (banco de dados IIIA { 7, C₃, 9, C_z, 11, C₄, 13} e banco de dados IIA {C₃, 29, 30, C_z, 32, 33, C₄}) de acordo com a Figura 13. A Tabela 6 apresenta as acurácias obtidas.

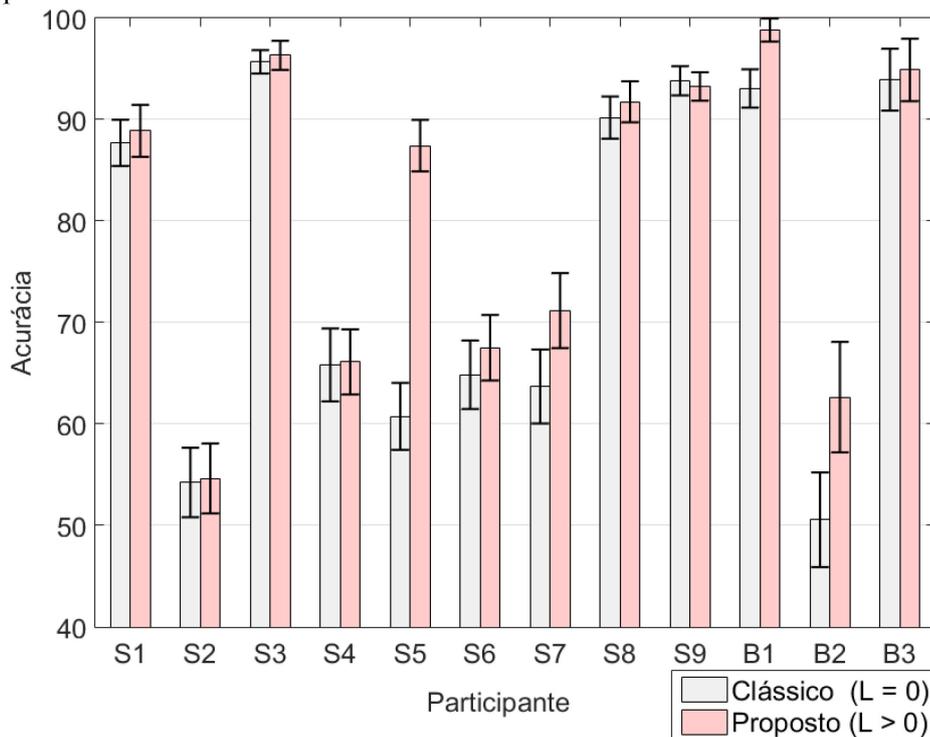
O requisito de gaussianidade foi respeitado para todos os sujeitos, após a aplicação da transformação de *boxcox*. Dessa forma, foi aplicado o teste T bilateral com $\alpha = 5\%$. Para o teste do par de acurácias médias globais (última linha da Tabela 6) foi usado o método de Wilcoxon de postos sinalizados.

Tabela 6 – Acurácias médias para cada sujeito de acordo com os seus parâmetros ótimos dentro do espaço de busca utilizando apenas 7 canais. A última linha representa a média e o desvio padrão para todos os sujeitos. As duas últimas colunas, representam os parâmetros ótimos obtidos dentro do espaço de busca. Em negrito e azul, as maiores acurácias. Na tabela, *O* representa o *offset*, *L* o atraso e *W* é referente à quantidade de pares de filtros espaciais selecionados.

	CSP	CSP	<i>p</i>	Parâmetros ótimos	
	proposto	clássico		Proposto	Clássico
S1	88,9±2,6	87,7±2,3	2×10^{-7}	<i>O</i> = 3,4; <i>L</i> = 1; <i>W</i> = 3	<i>O</i> = 3,4; <i>W</i> = 2
S2	54,6±3,4	54,2±3,4	2×10^{-1}	<i>O</i> = 3,3; <i>L</i> = 1; <i>W</i> = 4	<i>O</i> = 3,4; <i>W</i> = 5
S3	96,3±1,4	95,7±1,1	4×10^{-6}	<i>O</i> = 3,5; <i>L</i> = 5; <i>W</i> = 3	<i>O</i> = 3,6; <i>W</i> = 6
S4	66,1±3,2	65,8±3,6	2×10^{-1}	<i>O</i> = 3,4; <i>L</i> = 1; <i>W</i> = 5	<i>O</i> = 3,3; <i>W</i> = 3
S5	87,4±2,5	60,8±3,3	3×10^{-84}	<i>O</i> = 3,3; <i>L</i> = 5; <i>W</i> = 2	<i>O</i> = 3,3; <i>W</i> = 6
S6	67,5±3,2	64,9±3,4	2×10^{-11}	<i>O</i> = 3,4; <i>L</i> = 2; <i>W</i> = 3	<i>O</i> = 3,3; <i>W</i> = 1
S7	71,2±3,7	63,7±3,6	1×10^{-36}	<i>O</i> = 3,3; <i>L</i> = 1; <i>W</i> = 4	<i>O</i> = 3,3; <i>W</i> = 1
S8	91,8±2,0	90,2±2,1	3×10^{-10}	<i>O</i> = 3,6; <i>L</i> = 1; <i>W</i> = 3	<i>O</i> = 3,5; <i>W</i> = 1
S9	93,3±1,4	93,8±1,4	1×10^{-1}	<i>O</i> = 3,5; <i>L</i> = 1; <i>W</i> = 1	<i>O</i> = 3,5; <i>W</i> = 1
B1	98,8±1,1	93,1±1,9	5×10^{-46}	<i>O</i> = 0,4; <i>L</i> = 4; <i>W</i> = 3	<i>O</i> = 0,5; <i>W</i> = 1
B2	62,6±5,4	50,6±4,7	7×10^{-30}	<i>O</i> = 0,3; <i>L</i> = 1; <i>W</i> = 1	<i>O</i> = 0,7; <i>W</i> = 6
B3	94,9±3,1	93,9±3,0	2×10^{-2}	<i>O</i> = 0,3; <i>L</i> = 5; <i>W</i> = 3	<i>O</i> = 0,4; <i>W</i> = 2
$\mu \pm \sigma$	81,1±15,1	76,2±17,1	1×10^{-92}	-	-

Fonte: (O AUTOR, 2019).

Figura 31 – Comparação de acurácias médias entre o CSP proposto e o clássico, usando 7 eletrodos posicionados sobre o córtex motor. Em conjunto com a acurácia, está o respectivo desvio padrão para cada participante.



Fonte: (O AUTOR, 2019).

Por meio dos testes estatísticos, afirma-se com 95% de confiança que o método proposto é superior para todos os sujeitos, exceto para S_4 e S_9 , pois $p < \alpha$. Em S_9 , o método clássico apresenta desempenho superior, corroborado por $p < \alpha$. Para S_4 , estatisticamente, ambos os métodos apresentam o mesmo desempenho, pois $p > \alpha$. Para ilustrar os resultados da Tabela 6, apresenta-se os dados na forma pareada na Figura 31.

7.5 INSERÇÃO DOS ATRASOS EM OUTROS MÉTODOS DE FILTRAGEM ESPACIAL

Nessa seção é analisada a influência da nova proposta de matriz de informação, dada pela equação (6.5), no seguinte conjunto de métodos (A_i para $i = 1, 2, \dots, 7$) CSP:

- A_1 : CSP;
- A_2 : GLRCSP ($\alpha_c = [10^{-10} \ 10^{-9} \ \dots \ 10^{-1}]^T$; $\zeta = [0 \ 0.01 \ 0.02 \ \dots \ 0.09]^T$; $\varphi = 0$);
- A_3 : CCSP1 ($\alpha_c = 0$; $\zeta = [0 \ 0.01 \ 0.02 \ \dots \ 0.09]^T$; $\varphi = 0$);
- A_4 : DLCSPauto ($\alpha_c = 0$; $\zeta = 0$; $\varphi = [0 \ 0.01 \ 0.02 \ \dots \ 0.09]^T$);
- A_5 : DLCSPcv ($\alpha_c = 0$; $\zeta = 0$; $\varphi = [0 \ 0.01 \ 0.02 \ \dots \ 0.09]^T$);
- A_6 : DLCSPcvd ($\alpha_c = 0$; $\zeta = 0$; $\varphi = [0 \ 0.01 \ 0.02 \ \dots \ 0.09]^T$);
- A_7 : WTRCSP ($\alpha_c = [10^{-10} \ 10^{-9} \ \dots \ 10^{-1}]^T$; $\zeta = 0$; $\varphi = 0$).

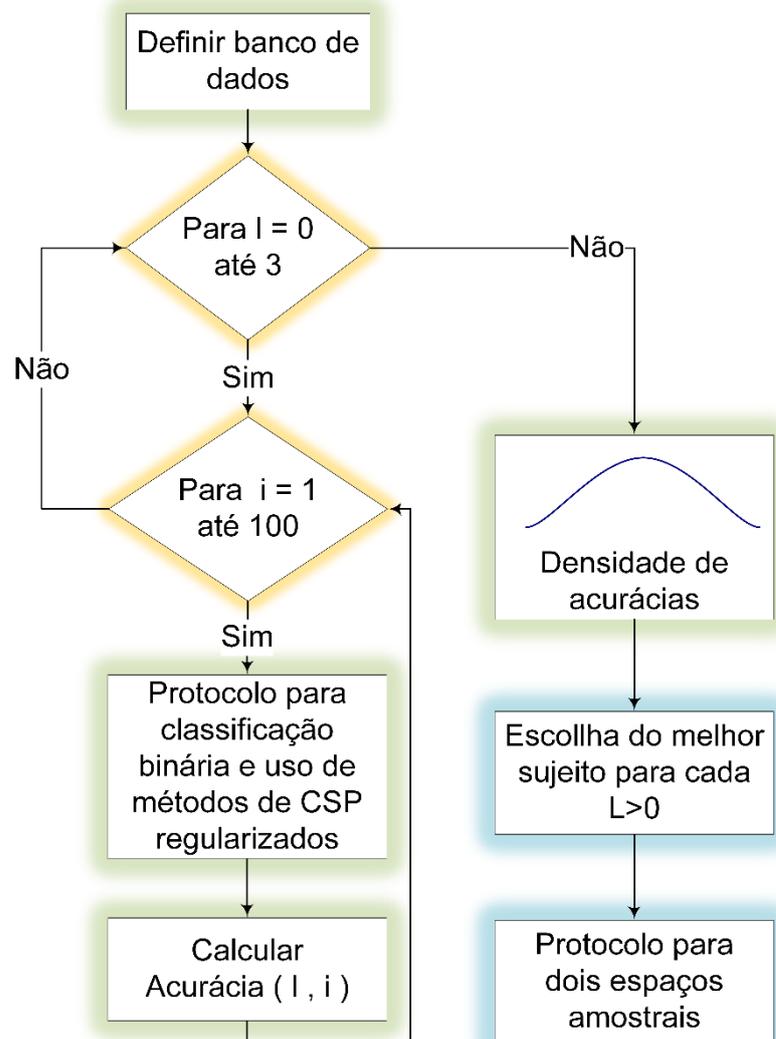
Os parâmetros foram configurados para a geração de hiperespaços ótimos para cada um dos métodos, usando os mesmos valores apresentados em (LOTTE e GUAN, 2011), adicionalmente, insere-se os atrasos em suas matrizes de informação. Além disso, foram utilizados 3 pares de filtros e *offset* de 3,5s para o banco de dados IIIA e de 0,5s para o banco de dados IIA.

A Figura 32 apresenta o procedimento utilizado para a análise da influência dos atrasos nos métodos regularizados de CSP. O procedimento utilizado é semelhante ao da seção 7.3, na qual é variado o número de atrasos. Para $L = 0$ obtém-se o método original e para $L = \{1, 2, 3\}$ obtém-se o proposto.

Os métodos CSP regularizados submetidos à nova matriz de informação são denominados de A_{1L} , A_{2L} , A_{3L} , A_{4L} , A_{5L} , A_{6L} e A_{7L} . Foi utilizado o atraso que resulta no melhor desempenho dentro de um espaço de busca dado por $L = \{1, 2, 3\}$, ou seja, cada sujeito foi processado utilizando um valor específico de L .

A Tabela 7 apresenta os resultados referentes ao banco de dados IIIA, no qual foram utilizados os seguintes parâmetros de simulação: 3 pares de de filtros espaciais e *offset* de 3,5s (LOTTE e GUAN, 2011).

Figura 32 – Protocolo para avaliação de métodos regularizados. A variável l representa o atraso adicionado para a construção da matriz de informação. A variável i representa o número de realizações da etapa de classificação. Acurácia (l, i) é o cálculo da acurácia de classificação das épocas de execução disponíveis de acordo com os parâmetros internos da função.



Fonte: (O AUTOR, 2019).

Foi empregado o teste T nas comparações entre os métodos sem e com atraso. Por exemplo, foi confrontada a distribuição para o sujeito B_1 usando a técnica A_1 contra A_1L e assim por diante, com nível de significância de 5%. Para a última coluna da Tabela 7, foi empregado o teste de Wilcoxon de postos sinalizados e $\alpha = 5\%$. Para o participante B_1 , os métodos A_2L , A_3L , A_4L e A_5L tiveram suas acurácias transformadas pela transformação de *boxcox*; para B_3 isso foi feito para os métodos A_4L , A_5L , A_6L e A_7L . Para o restante dos métodos

e sujeitos, não houve nenhuma transformação, pois, o requisito de gaussianidade dos resíduos foi atendido. Na Tabela 7 foram destacadas em negrito e azul as maiores acurácias médias ($p < \alpha$). No caso de mais de uma acurácia ter sido destacada para um mesmo sujeito e método compreende-se que são iguais estatisticamente ($p > \alpha$).

Pelas médias globais da Tabela 7, encontradas na penúltima coluna, conclui-se que houve aumento de desempenho entre 2% e 4,6%, sendo esses dados obtidos a partir da menor (média global de A₃L subtraído da média global de A₃) e da maior diferença (média global de A₂L subtraído da média global de A₂) de acurácias médias globais. Individualmente, o método proposto apresenta desempenho igual ou melhor em todas as comparações em relação ao método clássico.

A Figura 33 apresenta uma visão geral dos resultados apresentados na Tabela 7. Nota-se um acentuado ganho de desempenho para o participante B₁ com a nova matriz de informação para todos os métodos. Em menor escala, isso ocorre também para B₂ e B₃.

Tabela 7 – Comparação entre os métodos com e sem atrasos para os sujeitos do banco de dados IIIA. Em azul e negrito os melhores resultados. A penúltima coluna, representa a média e o desvio padrão, considerando todos os participantes da respectiva linha. A última coluna indica o atraso utilizado na matriz de informação respectiva a cada sujeito: $L = \{B_1, B_2, B_3\}$.

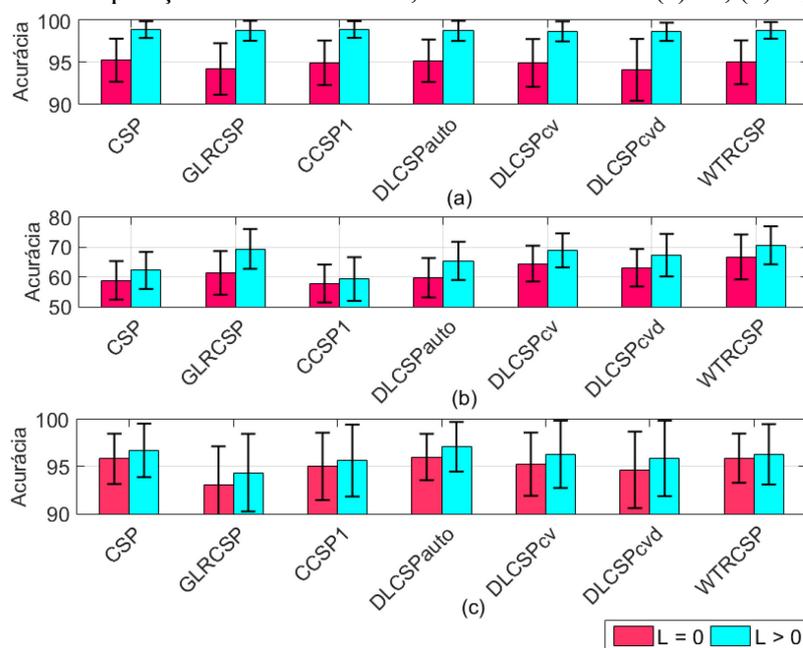
	B ₁	B ₂	B ₃	$\mu \pm \sigma$	L ótimo
A ₁	95,2±2,6	58,8±6,5	95,8±2,7	83,3±17,8	$L = 0$
A ₁ L	98,9±1,0	62,2±6,2	96,7±2,8	85,9±17,3	$L = \{3, 1, 1\}$
A ₂	94,1±3,1	61,4±7,3	93,1±4,1	82,9±16,1	$L = 0$
A ₂ L	98,7±1,2	69,4±6,7	94,4±4,1	87,5±13,7	$L = \{2, 3, 1\}$
A ₃	94,9±2,7	57,8±6,4	95,0±3,6	82,6±18,1	$L = 0$
A ₃ L	98,9±1,0	59,3±7,3	95,6±3,8	84,6±18,6	$L = \{3, 1, 1\}$
A ₄	95,1±2,5	59,8±6,6	96,0±2,5	83,6±17,5	$L = 0$
A ₄ L	98,7±1,2	65,4±6,4	97,1±2,6	87,1±15,9	$L = \{2, 2, 1\}$
A ₅	94,9±2,9	64,5±6,0	95,3±3,4	84,9±15,1	$L = 0$
A ₅ L	98,7±1,2	69,0±6,0	96,3±3,6	88,0±14,1	$L = \{3, 2, 1\}$
A ₆	94,0±3,7	63,1±6,3	94,7±4,1	83,9±15,5	$L = 0$
A ₆ L	98,6±1,1	67,3±7,1	95,9±4,0	87,3±14,9	$L = \{3, 1, 1\}$
A ₇	94,9±2,6	66,7±7,5	95,9±2,6	85,9±14,4	$L = 0$
A ₇ L	98,8±1,0	70,7±6,4	96,3±3,2	88,6±13,4	$L = \{2, 1, 1\}$

Fonte: (O AUTOR, 2019).

De forma análoga às simulações anteriores, os resultados para o banco de dados IIA foram obtidos para os seguintes parâmetros de simulação: 3 pares de filtros espaciais e *offset*

de 0,5s (LOTTE e GUAN, 2011). A mesma nomenclatura (A_i , para $i=1, 2, \dots, 7$) para referência aos métodos regularizados foi aplicada na Tabela 8 que apresenta os resultados obtidos.

Figura 33 – Comparação entre os métodos, com e sem atrasos. (a) B_1 , (b) B_2 e (c) B_3 .



Fonte: (O AUTOR, 2019).

Tabela 8 – Comparação entre os métodos com e sem atraso para estimação da matriz de informação. Banco de dados IIA. Em negrito e azul são destacados os melhores resultados. A última coluna, apresenta a média e o desvio padrão do conjunto de sujeitos.

	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	$\mu \pm \sigma$
A ₁	88,7±2,4	58,1±4,3	97,1±1,1	68,7±4,0	58,7±4,0	67,1±3,1	80,1±3,0	96,9±1,3	93,8±1,4	78,8±15,3
A _{1L}	93,8±1,8	58,9±4,2	97,6±1,0	69,0±3,7	89,5±2,2	71,1±3,5	83,6±4,3	97,0±1,5	93,3±1,6	83,8±13,6
A ₂	88,1±2,8	57,4±4,2	97,0±1,1	70,0±3,3	57,8±4,0	68,3±3,4	78,9±4,2	96,6±1,4	93,4±1,6	78,6±15,4
A _{2L}	94,0±1,7	60,0±3,5	97,1±1,5	70,7±3,3	89,0±2,5	70,6±4,0	83,1±3,1	96,7±1,4	93,4±1,6	83,9±13,1
A ₃	88,2±2,7	57,0±4,0	96,9±1,2	70,0±3,4	57,1±4,7	68,5±3,6	79,1±4,0	96,7±1,4	93,6±1,5	78,6±15,5
A _{3L}	94,0±1,7	57,7±3,7	97,5±1,0	69,6±3,9	89,1±2,6	71,9±3,4	83,3±2,9	96,7±1,5	93,5±1,4	83,7±13,6
A ₄	88,9±2,4	58,1±4,4	97,1±1,1	68,7±3,9	58,7±4,1	67,0±3,2	80,0±3,0	96,9±1,2	93,9±1,4	78,8±15,4
A _{4L}	93,5±2,4	59,5±3,7	97,6±1,0	69,7±3,8	80,8±3,0	71,3±3,6	84,2±3,5	97,0±1,4	93,7±1,4	83,1±13,3
A ₅	88,2±2,9	58,2±4,0	96,9±1,2	68,1±3,5	57,3±4,2	67,4±3,7	79,9±3,2	96,7±1,3	93,7±1,5	78,5±15,6
A _{5L}	94,1±1,6	59,5±3,7	97,6±1,1	69,4±3,8	89,5±2,3	70,4±3,9	83,6±3,6	96,5±1,4	93,3±1,7	83,8±13,5
A ₆	87,9±3,0	56,6±4,6	96,5±1,6	67,1±4,1	57,0±4,6	66,7±4,4	79,2±4,1	96,3±1,7	93,5±1,6	77,9±15,9
A _{6L}	94,0±1,8	58,6±4,0	97,1±1,3	69,3±4,1	89,5±2,3	70,4±3,9	83,6±3,6	96,5±1,4	93,3±1,7	83,3±13,8
A ₇	89,1±2,6	59,4±4,3	96,9±1,3	68,4±3,5	58,0±3,9	67,7±3,9	79,8±3,6	96,8±1,4	93,9±1,6	78,9±15,3
A _{7L}	93,5±2,0	60,1±3,6	97,6±0,9	69,9±3,7	89,0±2,4	71,2±3,6	84,2±4,1	96,9±1,4	93,5±1,7	84,0±13,2

Fonte: (O AUTOR, 2019).

Os testes estatísticos foram realizados da mesma forma anterior. Para exemplificar, no método CSP original (A_1L e A_1) foi aplicada a transformação de boxcox em S_3 e S_9 , tornando sua distribuição gaussiana. Para o sujeito S_8 a transformação não surtiu efeito e será usada a sua distribuição original. Lembrando que a mesma transformação é aplicada nos dois métodos aplicados ao mesmo sujeito, pois conforme a subseção 5.2.1, o teste T pareado é calculado a partir do resíduo produzido pela diferença das duas distribuições. Para os participantes S_1 , S_7 e S_9 foi aplicado o teste T pareado, enquanto que para S_8 e para a média global (última coluna da Tabela 8) foi aplicado o teste de Wilcoxon, ambos com $\alpha = 5\%$.

Na última coluna da Tabela 8, a menor variação da média global ocorre para o método GLRCSP ($83,9\% - 78,6\% = 4,3\%$), enquanto que a maior variação ocorre no DLCSPcvd ($83,3\% - 77,9\% = 5,4\%$). Na Tabela 8, desconsiderando a sua última coluna, há 63 comparações, das quais, 60 apresentam desempenho superior ($p < \alpha$) ou igual ao método clássico ($p > \alpha$) usando a nova matriz de informação proposta.

Tabela 9 – Atrasos ótimos para os métodos propostos da Tabela 8: $L = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9\}$.

Método proposto	L ótimo
A_1L	$L = \{3, 1, 2, 1, 3, 1, 2, 1, 2\}$
A_2L	$L = \{3, 2, 2, 1, 3, 1, 1, 1, 3\}$
A_3L	$L = \{3, 3, 2, 2, 3, 1, 2, 1, 1\}$
A_4L	$L = \{1, 2, 3, 2, 3, 1, 1, 1, 1\}$
A_5L	$L = \{3, 2, 2, 2, 3, 1, 1, 1, 1\}$
A_6L	$L = \{3, 3, 2, 2, 3, 1, 2, 1, 1\}$
A_7L	$L = \{1, 2, 3, 1, 3, 2, 2, 1, 1\}$

Fonte: (O AUTOR, 2019).

A Tabela 9 complementa Tabela 8, apresentando os atrasos ótimos utilizados para cada participante e método.

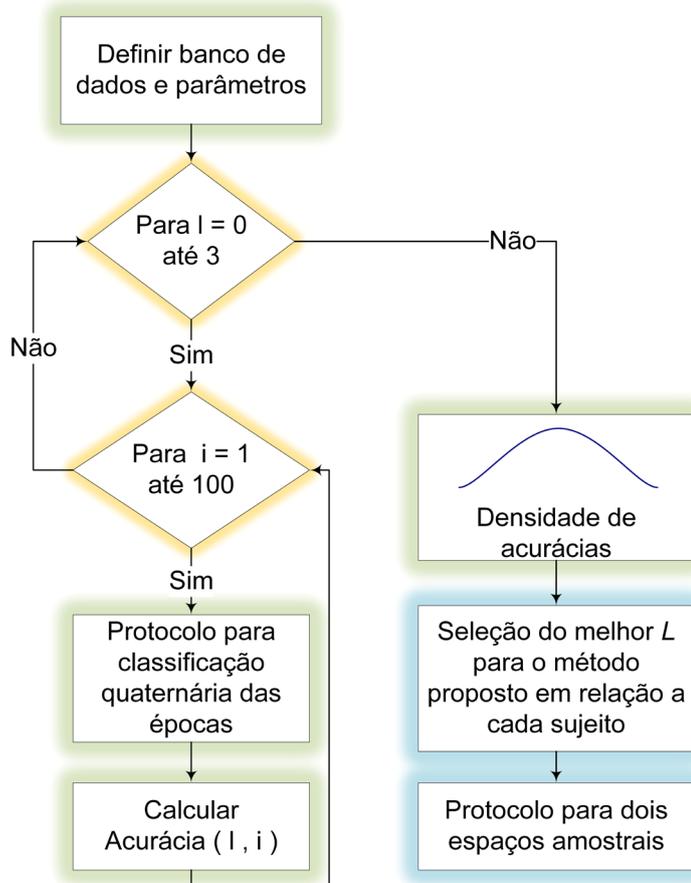
7.6 ABORDAGEM USANDO QUATRO CLASSES

Diferentemente das seções anteriores que abordaram a classificação de imagética motora binária, a presente seção expande o processo de classificação para problemas quaternários. Os quatro movimentos em questão são: mão esquerda, mão direita, pés e língua. A probabilidade de classificação aleatória de 4 movimentos é de 25%, se equiprováveis.

O procedimento para avaliação dos resultados é ilustrado na Figura 34. Nela, os parâmetros l e i representam, respectivamente, o atraso e a realização. Ao final do laço de

repetição calcula-se a acurácia média e seu desvio padrão para cada L . A distribuição de maior acurácia média entre todos os atrasos é comparada com o método CSP-FLDA convencional por meio do protocolo para dois espaços amostrais dado pela Figura 23.

Figura 34 – Protocolo para avaliação de problemas quaternários. A variável l representa o atraso adicionado para a construção da matriz de informação. A variável i representa o número de realizações da etapa de classificação. Acurácia (l, i) é o cálculo da acurácia de classificação das épocas de execução disponíveis de acordo com os parâmetros internos da função.



Fonte: (O AUTOR, 2019).

Esse procedimento tem como objetivo verificar se a inserção de atrasos para a geração da nova matriz de informação para o método CSP-FLDA proporciona vantagem estatística na classificação de problemas quaternários.

Seguiu-se o mesmo protocolo geral usado para problemas binários, no qual foram realizadas 100 iterações com pelo menos 20% de diferença entre cada conjunto de treinamento. Os mesmos parâmetros foram aplicados para o método clássico e o proposto (*offset* = 3,5s para o banco de dados IIIA e 0,5s para o IIA; 3 pares de filtros espaciais). A Tabela 10 apresenta os resultados obtidos. O método proposto apresentou os melhores resultados para $L = 1$ em todos

os casos, tendo sido utilizado um espaço de busca de $L = \{1, 2, 3\}$, enquanto que o método clássico utiliza $L = 0$.

Foram selecionados apenas os sujeitos, de ambos os bancos de dados, que atingiram pelo menos 70% de acurácia, pelo método clássico, no problema de classificação binária (veja Tabela 4), sendo eles: S₁, S₃, S₇, S₈, S₉, B₁ e B₃. Uma tabela contendo todos os participantes e resultados é apresentada no Apêndice A.

Primeiramente foi verificada a normalidade dos resíduos, sendo que somente S₁ não é inicialmente considerado gaussiano. Aplicando a transformação de *boxcox* nas amostras de S₁, resultou em distribuição normal para os resíduos. Como todos os sujeitos apresentam distribuição normal para os resíduos, usou-se o teste T de amostras dependentes com $\alpha = 5\%$, conforme indicado pela Figura 23. Na última linha da Tabela 10, referente à média global, utiliza-se o teste de Wilcoxon com $\alpha = 5\%$.

Tabela 10 – Experimento de classificação quaternária para os métodos CSP clássico ($L = 0$) e CSP proposto ($L = 1$). Em negrito e azul são apresentadas as maiores acurácias. Em caso de duplicidade, significa que não há diferença estatística entre os métodos.

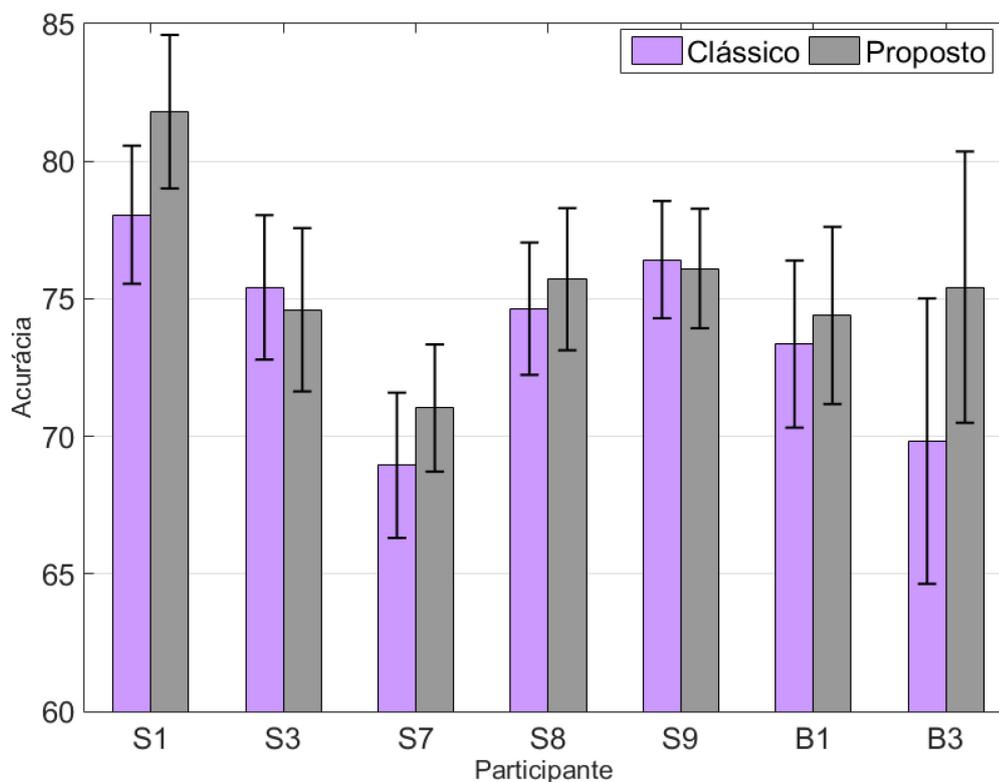
	CSP clássico	CSP proposto	p	L ótimo
S ₁	78,1±2,5	81,8±2,8	$7,4 \cdot 10^{-23}$	1
S ₃	75,4±2,6	74,6±3,0	$6,4 \cdot 10^{-4}$	1
S ₇	68,9±2,6	71,0±2,3	$8,3 \cdot 10^{-15}$	1
S ₈	74,6±2,4	75,7±2,6	$1,0 \cdot 10^{-4}$	1
S ₉	76,4±2,1	76,1±2,6	$7,0 \cdot 10^{-2}$	1
B ₁	73,4±3,0	74,4±3,2	$1,5 \cdot 10^{-5}$	1
B ₃	69,8±5,2	75,4±4,9	$2,4 \cdot 10^{-16}$	1
$\mu \pm \sigma$	73,8±4,4	75,6±4,2	$1,6 \cdot 10^{-30}$	-

Fonte: (O AUTOR, 2019).

Analisando os dados da Tabela 10, pode-se afirmar com um nível de significância de 95% que 5 dos 7 sujeitos apresentam desempenho superior usando o método proposto ($p < \alpha$) e que todos eles são passíveis de implementação prática, pois a acurácia média é igual ou maior que 70% (WOLPAW, BIRBAUMER, *et al.*, 2002). Para o sujeito S₉, ambos os métodos apresentam desempenho similar ($p > \alpha$) e para S₃ o método clássico é melhor, pois apresenta diferença estatística ($p < \alpha$).

A Figura 35 apresenta os resultados da Tabela 10 em forma gráfica. Através dela, verifica-se claramente que o método proposto apresenta 5 dos 7 melhores resultados.

Figura 35 – Comparação do método proposto com o clássico para a classificação de 4 movimentos motores.



Fonte: (O AUTOR, 2019).

7.7 CONSIDERAÇÕES FINAIS

Os resultados obtidos da classificação binária de imagética motora, considerando os dois bancos de dados e parâmetros indicados na literatura, apresenta uma melhora de acurácia média global de 3,3%. Ao realizar a otimização de parâmetros, esse valor aumenta para 4,9%. Com a restrição do número de canais para 7, a acurácia média global do método proposto é 4,9% maior do que a clássica. Comparando as duas abordagens, independentemente do número de canais utilizado, o desempenho global do método proposto é 4,9% maior que o do método clássico. De forma absoluta, aproximadamente a mesma acurácia média de classificação é obtida para os casos de 7 canais (método proposto) e todos os canais (método clássico).

Em relação aos métodos regularizados, verifica-se que, globalmente, para o banco de dados IIA houve um aumento de 2% a 4,6% e para o IIIA de 4,3% a 5,4%, de acordo com o método utilizado.

Para a classificação quaternária, a acurácia média global apresentou um aumento de 1,8% usando o método proposto, considerando os dois bancos de dados disponíveis.

Os resultados obtidos neste capítulo comprovam o aumento de desempenho propiciado pela abordagem proposta para estimação da matriz de informação do CSP, tanto para duas quanto para quatro classes, considerando a acurácia média global. Ao realizar a comparação entre métodos para um mesmo sujeito, os resultados demonstraram que a grande maioria das comparações apresentou melhora estatística significativa usando o método proposto e em somente alguns poucos casos houve piora ou desempenho similar.

Técnicas de inferência estatística foram utilizadas para suportar a conclusão de melhoria, deterioração ou manutenção de desempenho entre os métodos analisados. Adicionalmente, através da AUC da curva ROC também pode-se verificar que o método proposto apresenta desempenho superior ou similar ao clássico, dependendo do sujeito em questão.

Uma limitação da técnica proposta, deve-se ao fato de que ao aumentar o número de atrasos (L) para compor a matriz de informação, isso pode acarretar num mau condicionamento dela, ou seja, fazer com que ela deixe de ser semidefinida positiva e conseqüentemente não ser mais uma matriz de covariância. O valor de L máximo para que isso ocorra, depende do número de amostras e da quantidade de canais utilizados.

Finalizando, os próximos capítulos apresentam as conclusões finais da dissertação e sugestões para trabalhos futuros.

8 CONCLUSÃO

Ao longo dessa dissertação foram apresentados os conceitos essenciais para a compreensão do funcionamento e implementação de um sistema de BCI baseado no método CSP-FLDA. Além disso, foi realizado um breve levantamento histórico dos principais acontecimentos da área, para em seguida apresentar a definição de BCI e outros conceitos como: ritmos cerebrais, potenciais ERD e ERS, e a localização dos eletrodos no escalpo. Em seguida, foi apresentado um sistema de BCI para classificação da imagética motora binária, elencando e explicando cada uma de suas partes constituintes.

Foi apresentada uma proposta de contribuição à área de BCI aplicada à imagética motora. Essa nova proposta consiste na utilização de momentos de segunda ordem inter e intra-eletrodos adicionais, obtidos por meio da inclusão de atrasos entre os sinais disponíveis. A possibilidade de inserção de momentos de terceira e quarta ordem foi descartada, pois não foi encontrada informação relevante, provavelmente em função da relação sinal-ruído. O método proposto foi analisado de forma gráfica, por inferência estatística e por métricas de aprendizado de máquina. Essas ferramentas, indicam aumento de desempenho nas taxas de classificação quando utilizado o método proposto, tanto em comparações com o CSP clássico quanto com suas versões regularizadas. Um aspecto interessante em relação ao método proposto é que é possível superar o desempenho de classificadores convencionais mesmo com um menor número de canais. De forma secundária, foi proposta uma heurística para classificação quaternária de movimentos motores.

A partir do exposto, conclui-se que o desempenho de classificadores de imagética motora pode ser aumentado utilizando-se informações adicionais que até a data presente (no conhecimento do autor) não haviam sido utilizadas na literatura.

9 PROPOSTAS FUTURAS

Como perspectiva futura para a continuação desse trabalho pode-se elencar: a determinação de métodos para a escolha do atraso ótimo nas matrizes de covariância; e a análise de impacto de desempenho da técnica proposta usando outros classificadores.

Analisando duas bases de dados da área (IEEE e SCOPUS³) até o presente ano, 2019, selecionaram-se alguns temas que podem ser utilizados para o prosseguimento do assunto proposto.

A primeira proposta, consiste em fazer a expansão do método de CSP usando um tensor de entrada ao invés de uma matriz. Assim podem ser incluídos, concomitantemente, diversos atrasos ($L = 0$, $L = 1$, $L = 2$, etc), utilizando uma técnica similar à encontrada em (CHAUDHARY, TARAN, *et al.*, 2019).

A segunda proposta consiste em empregar um classificador baseado em probabilidade tridimensional para verificar se a taxa de classificação é aumentada (ASHLEY, ARVANEH e MIHAYLOVA, 2019).

A última proposta consiste numa nova abordagem para a normalização da matriz de covariância do método CSP (OLIAS, MARTIN-CLEMENTE, *et al.*, 2019).

Essas três propostas, são algumas possibilidades de continuação desse trabalho na área de BCI que se apresenta em grande expansão.

³ Palavras utilizadas na busca: BCI, EEG, MI.

REFERÊNCIAS

ABDULKADER, S. N.; ATIA, A.; MOSTAFA, M. Brain computer interfacing: Applications and challenges. **Egyptian Informatics Journal**, v. 16, n. 2, p. 213-230, 2015.

ABREU, A. L. E. D. **Bootstrap e modelos de Support Vector Machine-SVM**. UFPR. Curitiba, p. 117. 2016.

ALOMARI, M. H.; SAMAHA, A.; ALKAMHA, K. Automated Classification of L/R Hand Movement. **International Journal of Advanced Computer Science and Applications**, v. 4, n. 6, 2013.

ANG, K. K. et al. Filter bank common spatial pattern (FBCSP) in brain-computer interface. **IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)**. [S.l.]: [s.n.]. 2008. p. 2390-2397.

ANNETT, J. On knowing how to do things: a theory of motor imagery. **Cognitive Brain Research**, v. 3, n. 2, p. 65-69, 1996.

ASHLEY, A. L.; ARVANEH, ; MIHAYLOVA, L. S. A Novel Three Dimensional Probability-based Classifier for Improving Motor Imagery-based BCI. **IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.]: [s.n.]. 2019. p. 1150-1153.

BARBETTA, A.; REIS, M. M.; BORNIA, C. A. **Estatística para cursos de engenharia e informática**. 3^a. ed. São Paulo: Atlas, 2010.

BELHADJ, S. A.; BERNOUMASSAT, N.; JRACHAI, M. D. CSP features extraction and FLDA classification of EEG based motor imagery for Brain Computer interaction. **Electrictrical Engineering(ICEE)**, 2015 4 th International Conference on. IEEE, p. 1-6, 2015.

BERNARDI, G. B.; PIMENTA, T. C.; MORENO, R. L. A Simplified Tool for Testing of Feature Selection and Classification Algorithms in Motor Imagery of Right and Left Hands of

EEG Signals. **IEEE 10th Latin American Symposium on Circuits & Systems (LASCAS)**. [S.l.]: IEEE. 2019. p. 197-200.

BLANCO, S. Stationarity of the EEG series. **Engineering in medicine and biology Magazine**, v. 14, n. 4, p. 395-399, 1995.

BLANKERTZ, B. et al. Optimizing spatial filters for robust EEG single-trial analysis. **IEEE Signal processing magazine**, v. 25, p. 41-56, 2008.

BONIFÁCIO, A. L. **Estratégia computacional para avaliação de propriedades mecânicas de concreto de agregado leve**. [S.l.]. 2017.

BOUGHORBEL , S.; JARRAY, F.; EL-ANBARI, M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. **PloS one**, v. 12, n. 6, 2017.

BRUNNER, C. et al. **BCI Competition 2008 – Graz data set A**, 2008. Disponível em: <http://www.bbc.de/competition/iv/desc_2a.pdf>. Acesso em: 25 set. 2018.

CHAUDHARY, S. et al. Convolutional Neural Network Based Approach Towards Motor Imagery Tasks EEG Signals Classification. [S.l.]: **IEEE Sensors Journal**, 2019.

CHOI, I. et al. A systematic review of hybrid brain-computer interfaces: Taxonomy and usability perspectives. **PLoS ONE**, v. 12, n. 4, 2017.

COYLE, D. Neural network based auto association and time-series prediction for biosignal processing in brain-computer interfaces. **IEEE Computational Intelligence Magazine**, v. 4, n. 4, 2009.

EBRAHIMI, T.; GARICA, G.; VESIN, J. M. Brain computer interface in multimedia communication. **IEEE signal processing magazine**, p. 14-24, 2003.

ELLIOTT, A. C.; WOODWARD, W. A. **Statistical Analysis Quick Reference Guidebook: With SPSS examples**. [S.l.]: Sage, 2007.

FACELI, K. et al. **Inteligência Artificial: uma abordagem por aprendizado de máquina**. Rio de Janeiro: LTC, 2011.

GE, S.; WANG, YU,. Classification of four-class motor imagery employing single-channel electroencephalography. **PloS ONE**, v. 9, n. 6, 2014.

GOKCEN, I.; PENG, J. Comparing linear discriminant analysis and support vector machines. **International Conference on Advances in Information Systems**, Berlin, p. 104-113, 2002.

GONEN, F. F.; TCHESLAVSKI, G. V. Techniques to assess stationarity and gaussianity of EEG: An overview. **International Journal of Bioautomation**, v. 16, n. 2, p. 135-142, 2012.

HINES, W. W. et al. **Probabilidade e estatística na Engenharia**. Rio de Janeiro. 2013.

IBGE. **Censo demográfico brasileiro 2000**, 2000. Disponível em:
<https://ww2.ibge.gov.br/home/estatistica/populacao/censo2000/populacao/censo2000_p>.
Acesso em: 07 julho 2018.

IBGE. **Censo demográfico brasileiro 2010**, 2010. Disponível em:
<https://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.p>.
Acesso em: 10 julho 2018.

KLEEM, W. R.; LI, T. H.; HERNANDEZ, J. H. Coherent EEG indicators of cognitive binding during ambiguous figure tasks. **Consciousness and Cognition**, v. 9, n. 1, p. 66-85, 2000.

KOZAK, M.; PIEPHO, H. P. What's normal anyway? Residual plots are more telling than significance tests when checking ANOVA assumptions. **Journal of Agronomy and Crop Science**, p. 86-98, 2018.

LEEB, R. et al. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. **Computational Intelligence and Neuroscience**, v. 2007, 2007.

LOTTE, F. et al. A review of classification algorithms for EEG-based brain–computer interfaces. **Journal of neural engineering**, v. 4, 2007.

LOTTE, F.; GUAN, C. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. **IEEE Transactions on Biomedical Engineering**, v. 58, n. 2, p. 355-362, 2011.

LOTTE, F.; GUAN, C. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. **IEEE Transactions on biomedical Engineering**, v. 58, n. 2, p. 355-362, 2011.

LOTTE, F.; NAM, C. S.; NIJHOLT, A. **Brain–Computer Interfaces Handbook**. [S.l.]: CRC Press, 2017.

MAHMOOD, A. et al. Classification of multi-class motor imagery EEG using four band common spatial pattern. **Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE**, p. 1034-1037, 2017.

NIJHOLT, A.; TAN, D. S. **Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction**. [S.l.]: Springer, 2010.

NOCEDAL, J., WRIGHT, S. J.; **Numerical Optimization**, Springer Series in Operations Research, n. 2, 2006.

OLIAS, J. et al. EEG Signal Processing in MI-BCI Applications with Improved Covariance Matrix Estimators. [S.l.]: **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, 2019.

PACHECO, A. Medida de desempenho de classificadores - Parte 2. **Computação inteligente**, 03 junho 2019. Disponível em: <<http://computacaointeligente.com.br/artigos/medidas-classificadores-2/>>. Acesso em: 03 junho 2019.

PEIXOTO, B. D. A.; ARAKI, H.; CENTENO, J. **Utilização do algoritmo SVM (SUPPORT VECTOR MACHINE) e árvores de decisão para classificação de imagens de alta resolução do sensor ADS-40**. UFPR. Curitiba. 2016.

PFURTSCHELLER, G.; NEUPER, C. Motor imagery and direct brain-computer communication. **Proceedings of the IEEE**, v. 89, n. 7, p. 1123-1134, 2001.

PINTO, M. A. S. **Estudo do potencial evocado visual em regime permanente baseado em LED para interface cérebro máquina**. UFMG. Belo Horizonte, p. 140. 2011.

RAMOSER, H.; PFURTSCHELLER, G.; MULLER-GERKING, J. Optimal spatial filtering of single trial EEG during imagined hand movement. **IEEE transactions on rehabilitation engineering**, v. 8, n. 4, p. 441-446, 2000.

RAZALI, N. M.; WAK, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. **Journal of Statistical Modeling and Analytics**, v. II, p. 21-31, 2011.

RODRIGUEZ-BERMUDEZ, G.; GARCIA-LAENCINA, P. H. Analysis of EEG signals using nonlinear dynamics and chaos: a review. **Applied mathematics & information sciences**, v. 9, n. 5, p. 2309, 2015.

SCHLÖGL, A. **Dataset IIIA: 4-class EEG data**., 2005. Disponível em: <<http://www.bbc.de/competition/iii/>>. Acesso em: 25 set. 2018.

SELFSLAGH, A. et al. Non-invasive, Brain-controlled Functional electrical stimulation for Locomotion Rehabilitation in Individuals with paraplegia. **Scientific reports**, v. 9, n. 1, p. 6782.

SILVA, C. D. **Processamento de Sinais de EEG para Classificação de Tarefas Motoras de Interface Cérebro-Máquina**. UFSC. Florianópolis, p. 191. 2017.

SILVA, L. C. P. D. **Análise de sinais de EEG após reabilitação com imagética motora**. UFRJ. [S.l.]. 2018.

SILVERTHORN, D. U. **Fisiologia humana: uma abordagem integrada**. [S.l.]: Artmed, 2010.

THIYAM, D.; RAJKUMAR, E. Common Spatial Pattern Algorithm Based Signal Processing Techniques for Classification of Motor Imagery Movements: A Mini Review. **IJCTA**, v. 9, n. 36, p. 53-65, 2016.

TRIOLA, M. F. **Introdução à estatística: atualização de tecnologia**. Rio de Janeiro. 2014.

VALIULIS, V. **The effect of transcranial magnetic stimulation on brain bioelectrical activity**. Vilniaus universitetas. [S.l.]. 2014.

WESTFALL, P. H. Kurtosis as peakedness. **The American Statistician**, v. 68, n. 3, p. 191-195, 2014.

WILLIAMS, S. E. Comparing movement imagery and action observation as techniques to increase imagery ability. **Psychology of Sport and Exercise**, v. 44, p. 99-106, 2019.

WITTE, J. S.; WITTE, R. S. **Estatística**. 7^a. ed. Rio de Janeiro: LTC, 2005.

WOLPAW, J. R. et al. Brain-computer interfaces for communication and control. **Clinical Neurophysiology**, 2002.

WOLPAW, J.; WOLPAW, E. W. **Brain Computer Interfaces: principles and practice**. OUP USA, 2012.

YAACOUB, C.; MHANNA, G.; RIHANA, S. A Genetic-Based Feature Selection Approach in the Identification of Left/Right Hand Motor Imagery for a Brain-Computer Interface. **Brain sciences**, v. 7, n. 1, p. 12, 2017.

YONGWOONG, J. et al. Event-related (De)synchronization (ERD/ERS) during motor imagery tasks. **International Journal of Industrial Ergonomics**, 2011.

ZHAOJUN, B.; LU, D.; VANDEREYCKEN, B. Robust Rayleigh quotient minimization and nonlinear eigenvalue problems. **SIAM Journal on Scientific Computing**, 2018.

ZWICK, R. et al. Statistical tests for correlated proportions: some extensions. *Psychological Bulletin*, v. 92, n. 1, p. 258, 1982.

APÊNDICE A – TABELA COMPLETA DE CLASSIFICAÇÃO QUATERNÁRIA DE CLASSES COM ATRASOS E DOIS BANCOS DE DADOS. EM NEGRITO E AZUL AS MAIORES ACURÁCIAS.

	$L = 0$	$L = 1$	$L = 2$	$L = 3$
S ₁	78,1±2,5	81,8±2,8	81,2±2,8	81,3±2,7
S ₂	42,4±2,3	43,1±2,6	43,2±2,8	43,8±2,6
S ₃	75,4±2,6	74,6±3,0	74,3±2,9	74,1±2,9
S ₄	49,0±3,0	55,5±2,8	55,3±3,2	53,8±3,0
S ₅	37,3±2,5	39,8±2,9	46,5±3,2	50,0±3,2
S ₆	41,9±2,9	43,5±3,0	43,0±3,1	43,4±3,3
S ₇	68,9±2,6	71,2±2,3	71,0±2,3	70,4±2,5
S ₈	74,6±2,4	75,7±2,6	74,5±2,9	72,1±2,9
S ₉	76,4±2,1	76,1±2,2	75,3±2,2	74,7±2,2
B ₁	73,4±3,0	74,4±3,2	73,3±3,3	73,1±3,2
B ₂	52,9±3,6	56,8±4,3	59,4±4,7	57,1±4,2
B ₃	69,8±5,2	75,4±4,9	74,9±4,3	74,6±4,4
$\mu \pm \sigma$	61,7±15,3	64,0±15,0	64,4±13,8	64,0±13,3