

# Relatório Seminário - Análise de Imaginação Motora em EEG por LDA

Matheus Berbel Barusso

**Abstract**—Ondas cerebrais são sinais elétricos que podem ser capturados e estudados. Este trabalho busca explicar a análise de um Canal de EEG do dataset BCI IV 2a realizada por meio do pré-processamento dos dados e então a aplicação de Aprendizado de Máquina.

**Index Terms**—EEG, LDA, Aprendizado de Máquina

## I. INTRODUÇÃO

A análise de sinais eletroencefalográficos (EEG) é uma das principais ferramentas na análise e investigação das maneiras de funcionamento do cérebro, porém, por ser um sinal de amplitude muito baixa e altamente suscetível a ruídos processos robustos de pré-processamento são necessários para a extração real das suas características.

Com essas informações em mão podemos então aplicar técnicas de Aprendizado de Máquina para a classificação e estudo desses sinais, para isso, a Análise Discriminante Linear será empregada devido a sua eficiência computacional e capacidade de discriminar classes mesmo em situações de grande dimensionalidade. O pré-processamento e aplicação da LDA será realizado no software *Matlab*.

## II. FUNDAMENTAÇÃO TEÓRICA

Os sinais elétricos de um EEG são obtidos por meio de eletrodos acoplados em diferentes posições da cabeça humana, descritas pelo Sistemas Internacional 10 20, visto na Figura 1.

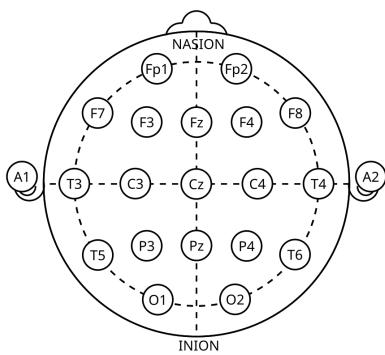


Fig. 1. Sistema Internacional 10 20 de distribuição de Eletrodos

Com o sinal capturado e extraído digitalmente por meio de Interfaces Cérebro-Computador (BCI) o sinal bruto é obtido, pronto para passar por diferentes processos de filtragem e normalização para ser utilizado em um estudo estatístico.

Matheus Berbel Barusso, graduando em Engenharia de Computação da Universidade Tecnológica Federal do Paraná - UTFPR - Rua Cristo Rei 19, Vila Becker, Toledo-PR, CEP 85902-490 Brasil e-mail: (barusso@utfpr.edu.br).

## III. EXPERIMENTO

Sinais de EEG possuem dimensionalidades variáveis de acordo com a quantidade de indivíduos, quantidade de classes de ação capturadas, quantidade de canais capturados entre outros. Para o presente trabalho será utilizado um dataset que contém:

- 9 Indivíduos.
- 22 Canais de EEG capturados (exceto 3 canais de EOG).
- 72 *trials* dos 9 sujeitos.
- Frequência de Amostragem de 250Hz, com *trials* de 4 segundos de duração.
- Classes de ação de números:
  - 1) Movimento Imaginário da Mão Esquerda.
  - 2) Movimento Imaginário da Mão Direita.
  - 3) Movimento Imaginário de Ambos Pés.
  - 4) Movimento Imaginário da Língua.

Os canais utilizados nesse dataset são os seguintes: Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, POz.

Dessa maneira é obtido a seguinte Matriz Multi-canais:

$$\mathbf{C} = 9 \times 22 \times 1000 \times 288$$

Dessa forma, a análise real dos dados se inicia e é dividida nas seguintes etapas:

### A. Visualização

A análise se inicia então com a visualização do sinal bruto de um sujeito, época e um canal individual escolhido arbitrariamente para checar informações como todos os canais se comportam em função do tempo, Amplitudes Médias e um mapa da Densidade Espectral de Potência para cada um dos canais do dataset.

### B. Pré-processamento

Após isso é dado início ao Pré-Processamento dessas informações, esta etapa é dividida em 9 subetapas:

1) *Filtragem*: Os sinais de um EEG são altamente suscetíveis a ruídos externos como ruídos não fisiológicos, como o proveniente da rede elétrica que o equipamento de captura está ligado (50/60 Hz) ou também ruídos fisiológicos provenientes de sinais elétricos de ações corporais, principalmente faciais. Esses ruídos geram artefatos no sinal obtido que precisam ser removidos para uma análise coerente. Para isso, foram aplicados três filtros em conjunto:

- Filtro FIR Passa-banda 0,5 - 30 Hz de ordem 100.
- Filtro Notch 60 Hz.
- Filtro Passa-alta maior que 0,5 Hz.

2) *Remoção de Tendências:* Com o passar das instâncias de tempo o sinal possui uma tendência de lentamente ser influenciado por uma inclinação, a série segue linhas inclinadas ou uma curva. Para essa tendência então ser removida é utilizada a função `detrend()`, que remove a tendência, nesse caso, linear presente no sinal.

3) *Remoção da Linha Base:* Para termos um período de referência que será utilizado nos processos de ICA (Análise de componentes Independentes) é necessário remover o deslocamento DC (*offset*) e centralizar o sinal do EEG em torno de zero, para dessa forma corrigir variações lentas e não neurais que podem estar presentes na amplitude. Isso é feito por meio da subtração da média ou mediana do período pré-evento ou ensaio global, conforme observado na Equação 1.

$$x_{corr}[n] = x[n] - \frac{1}{K} \sum_{n=1}^K x[n] \quad (1)$$

4) *Reamostragem:* Essa seção do pré-processamento tem como objetivo a redução do tamanho dos dados, e consequentemente, o uso de memória para um processamento de maior velocidade. Serve também para garantir uma compatibilidade entre outros datasets ou até mesmo para métodos de análise que necessitam de frequências específicas. Isso pode ser obtido por diferentes métodos numéricos, o escolhido é por meio de Interpolação (*upsample*), demonstrada na Equação 2.

$$x_{upsampled}[n] = \begin{cases} x[n/p], & n = kp, k \in \mathbb{Z} \\ 0, & \text{caso contrário} \end{cases} \quad (2)$$

Após a reamostragem é obtido um total de 512 amostras.

5) *Detecção de Artefatos:* Nos sinais de EEG capturados ainda existem sinais não neurais que contribuem com a sua contaminação como os fisiológicos já citados, entre eles temos por exemplo, movimentos oculares (EOG), movimentos musculares (EMG), movimentação da cabeça e/ou dos eletrodos durante a captura do sinal. Com a remoção desses artefatos a qualidade do sinal aumenta e a chance de reconhecimento de falsos padrões diminui, além de melhorar a performance de análises espectrais, conectividade e classificação. A Figura 2 apresenta a diferença entre os sinais registrados e em contra partida, a participação de cada tipo de sinal na composição do que é capturado pelos eletrodos.

Dentre os métodos possíveis para a detecção e remoção desses artefatos o escolhido foi a Análise de Componentes Independentes (ICA), que separa os sinais independentes, identificando e removendo os artefatos encontrados. Para isso é calculado o *Threshold simples (Peak-to-Peak)*, que mede a amplitude pico-a-pico do sinal e decide se algo é artefato com base no visto na Equação

$$\text{Liminar em } \mu\text{V} \begin{cases} < 80\mu\text{V} & \text{Normal} \\ 80 - 120\mu\text{V} & \text{Artefato Leve} \\ > 120\mu\text{V} & \text{Artefato Grave} \end{cases} \quad (3)$$

Utilizando ICA para a remoção dos artefatos temos que:

$$\mathbf{X}_{\text{limpo}} = \mathbf{A}\mathbf{S}_{\text{limpo}}$$

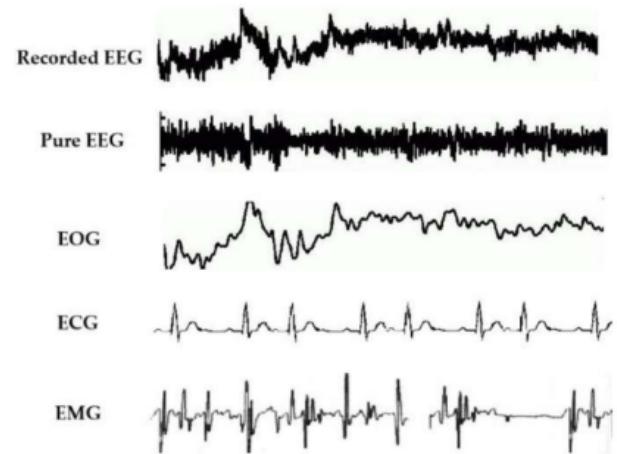


Fig. 2. Sinal registrado de EEG decomposto.

Para encontrar o  $\mathbf{A}$  é utilizado o algoritmo *whitening*, que consiste em três etapas:

- 1) Centralização (*zero mean*): Remove o valor médio do sinal por meio de:

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbb{E}[\mathbf{X}] \quad (4)$$

Deixando os dados ao redor de 0, facilitando os cálculos das matrizes de covariância.

- 2) *Whitening* ou Esferização: Faz com que  $\mathbf{X}$  tenha suas dimensões descorrelacionadas e com a mesma variância:

$$\mathbf{V} = \mathbf{D}^{-1/2} \mathbf{E}^\top \quad (5)$$

Com  $\mathbf{E}$  contendo os autovetores da matriz de covariância de  $\mathbf{X}$  e  $\mathbf{D}$  com os autovalores.

- 3) ICA por Decorrelação: Busca uma rotação que maximiza a independência estatística entre os componentes por meio de:

$$\mathbf{S} = \mathbf{R}\mathbf{Z} \quad e \quad \mathbf{W} = \mathbf{R}\mathbf{V} \quad (6)$$

Onde  $\mathbf{Z}$  é o sinal após o *whitening*,  $\mathbf{R}$  é uma rotação encontrada pelo ICA e  $\mathbf{W}$  é a matriz de separação final. Com isso temos que:

$$\mathbf{W} = \mathbf{A}^{-1}$$

E portanto a reconstrução do EEG limpo é:

$$\mathbf{X}_{\text{limpo}} = \mathbf{W}^{-1} \mathbf{S}_{\text{limpo}} \quad (7)$$

Dessa maneira, os artefatos de outros sinais que compõem o que foi registrado inicialmente no EEG foram retirados, deixando o sinal com uma maior qualidade.

6) *Normalização:* O processo de normalização transforma os dados para uma escala comum, para que dessa maneira todos os canais tenham uma magnitude comparável, evitar que canais de alta amplitude dominem a classificação ou a mineração, reduzir variações entre canais e sujeitos e também acelerar a convergência em algoritmos de aprendizado de máquina. Para isso o método escolhido é o *Z-Score*, medindo

quantos desvio-padrão um valor  $x$  está acima ou abaixo da média por meio da fórmula:

$$x' = \frac{x - \mu}{\sigma} \quad (8)$$

Onde  $x$  é o valor original,  $\mu$  é a média dos dados analisados e  $\sigma$  é o desvio-padrão dos dados.

7) *Re-referenciação*: É o processo de redefinir o ponto de zero elétrico em todos os canais, subtraindo de cada canal uma referência comum com o intuito de reduzir artefatos globais, interferência de linha e variações lentas que são comuns a todos os canais, como visto na Equação 9.

$$x_i^{\text{ref}}(t) = x_i(t) - R(t) \quad (9)$$

Onde  $R(t)$  é a referência em comum que buscamos. Ele pode ser encontrado por diversos métodos, a Média Comum (CAR), foi o escolhido, e apresentado a seguir.:

$$\frac{1}{n_c} \sum_{j=1}^{n_c} x_j(t)$$

8) *Segmentação em Épocas*: Consiste na divisão dos sinais contínuos em segmentos temporais de tamanho menor para uma análise que será específica. Com isso é obtido um sinal mais estável para análise e que facilita a remoção de artefatos pois rejeita apenas as épocas corrompidas. Além disso isso é melhor para utilização em aprendizado de máquina pois aumenta o número de amostras e facilita o processo de mineração. O número total de épocas é observado na Equação 10.

$$E = \left\lfloor \frac{T - L}{L_{\text{step}}} \right\rfloor + 1 \quad (10)$$

Com  $T$  representando a duração total do sinal,  $L$  o tamanho da janela e  $L_{\text{step}}$  o deslocamento entre janelas.

### C. Mineração de Dados

Após o pré-processamento ser finalizado a etapa de Mineração é iniciada. Essa etapa tem como objetivo transformar os sinais de EEG em valores representativos que podem então, ser utilizados em análise e classificação. A mineração será realizada para três medidas, todas no domínio do tempo:

1) *Estatísticas*: As medidas estatísticas utilizadas são a Assimetria, que analisa se os dados "tendem" mais para um dos lados em relação a um referencial, e a Curtose, que analisa o quanto a distribuição é achatada ou pontiaguda em relação a distribuição normal. Ambas equações utilizadas são apresentadas nas Equações 11 e 12, respectivamente.

$$v = \frac{m_3(x)}{\sigma_x^3} \quad (11)$$

$$c = \frac{m_4(x)}{\sigma_x^4} \quad (12)$$

2) *Parâmetros de Hjorth*: Esses parâmetros são medidas estatísticas no domínio do tempo que descrevem o comportamento dinâmico de um sinal, especialmente em sinais como os de EEG, EMG, entre outros. São utilizados devido a sua facilidade de serem calculados e possuem informações sobre Atividade, que é a variância do sinal, medindo a oscilação global do sinal, mobilidade, que analisa o quão rápido o sinal muda, indicando a frequência média do sinal e a complexidade do movimento, que analisa o quão parecido o sinal é de uma senoidal pura. As equações dos respectivos parâmetros são apresentadas nas Equações 13, 14 e 15.

$$HA(x) = \sigma_x^2 \quad (13)$$

$$HM(x) = \sqrt{\frac{HA(x')}{HA(x)}} = \frac{\sigma_{x'}}{\sigma_x} \quad (14)$$

$$HC(x) = \frac{M(x')}{M(x)} = \frac{\sigma_x \sigma_{x''}}{\sigma_x^2} \quad (15)$$

3) *Energia e Entropia*: São medidas baseadas principalmente na energia, isso é, medem o quanto de intensidade existe em um sinal. Medir a energia de um sinal é útil para por exemplo, detectar mudanças bruscas e classificar padrões de vibração em sinais de EEG. As medidas utilizadas são a Energia Média, que mede o nível de energia do sinal independentemente se essa potência é positiva ou negativa, Entropia do logaritmo de energia (ELA), que aplica logaritmo no sinal da energia ponto a ponto dando mais peso a pequenas variações de sinal e a Entropia de Shannon, que é a medida clássica de entropia baseada em probabilidades usada em diversas áreas, medindo o grau de incerteza, aleatoriedade ou o quão desorganizado um sinal é. Essas medidas são obtidas respectivamente pelas Equações 16, 17 e 18.

$$e_1 = \frac{1}{T} \sum x^2 \quad (16)$$

$$e_2 = \sum \log(x^2) \quad (17)$$

$$e_3 = - \sum p_i \log_2(p_i) \quad (18)$$

### D. Conectividade Funcional

Após a mineração dos dados do sinal de EEG chega o momento de quantificar o grau de sincronização do sinal em diferentes canais por meio da mineração da matriz de conectividade cerebral, isso é, se duas áreas cerebrais mantêm uma relação de fase constante e caso essa relação seja confirmada isso indica que uma coordenação entre essas áreas. Diversos métodos podem ser utilizados, ambos no domínio do tempo e no domínio da frequência para essa quantificação, dentro os existentes o escolhido foi *Phase Locking Value* (PLV). Esse método foi aplicado para construir uma matriz de adjacência que representa a conectividade funcional entre diferentes eletrodos (canais). A equação para o uso de PLV é:

$$PLV_{xy}(\omega) = |E\{e^{j(\phi_x(\omega) - \phi_y(\omega))}\}| \quad (19)$$

Onde  $\phi_x(\omega)$  e  $\phi_y(\omega)$  são as fases de diferentes eletrodos na mesma frequência  $\omega$ ,  $e^{j(\phi_x - \phi_y)}$  é a representação complexa da diferença de fase e  $E\{\cdot\}$  é a média (esperança) com o passar do tempo, ou dos trials. Para essa etapa diversos métodos já explorados e outros são utilizados para que sejam encontradas medidas estatísticas e topológicas, que são métricas da teoria dos grafos aplicadas no estudo do cérebro o considerando como uma rede cerebral com o intuito de encontrar os graus de relações nos nós (eletrodos) desse grafo, que serão utilizadas na próxima etapa.

#### E. Seleção de Características

Consiste em diferentes subetapas que consistem principalmente na junção e manipulação de todos os dados previamente encontrados:

1) *Montagem de Matriz de Medidas*: Como o nome sugere, consiste na união de todas as medidas previamente calculadas em uma matriz que irá então ser normalizada com as medidas concatenadas, novamente, o método utilizado é o Z-Score, apresentado previamente na Equação 8.

2) *Redução de dimensionalidade*: Com todos os processos realizados anteriormente são montados dados com muitas dimensões, porém muitas dessas dimensões usualmente são correlacionadas entre si ou redundantes. Dessa maneira o *Principal Component Analyses* (PCA) reduz o número de *features* mas mantém maior parte da variância dos dados, transformando features correlacionadas em componentes principais ortogonais. Realizar esse processo evita que classificadores possuam um desempenho devido a *overfitting*, isso é, quando um modelo de aprendizado de máquina se ajusta excessivamente aos dados utilizado no seu treinamento, se acostumando com ruídos e detalhes que farão com que esse modelo tenha um desempenho ruim ao encontrar novos dados. Nesse contexto o utilizado para a redução da dimensionalidade é:

$$\tilde{X} = X - \mathbf{1}_n \tilde{X} \quad (20)$$

Onde  $X$  é a matriz original dos dados,  $\tilde{X}$  é o vetor de médias das *features*,  $\mathbf{1}_n$  é um vetor coluna de 1's de tamanho, e quando multiplicados replicam a média para cada linha. O objetivo desta equação é centralizar os dados entorno de zero, pois encontrar variância precisa que os dados estejam centralizados em torno de uma referência.

$$C = \frac{1}{n-1} \tilde{X}^T \tilde{X} \quad (21)$$

Que forma a matriz de covariância ( $d \times d$ ), que presenta como as variáveis irão variar juntas. O PCA busca componentes principais em que a variância é máxima, e por isso a matriz de covariância é necessária.

$$Cv_i = \lambda_i v_i \quad (22)$$

Em que  $v_i$  representa os autovetores da matriz de covariância e  $\lambda_i$  são os autovalores associados, quanto maior os autovalores, mais importante é o componente principal.

$$Z = \tilde{X} V_k \quad (23)$$

Onde  $V_k$  contém os  $k$  autovetores principais, isso é, associado aos maiores autovalores encontrados e  $Z$  é a nova matriz de dados reduzida.

Com isso, é plotado dois gráficos, apresentados nas Figuras 3 e 4, que representam quantas componentes principais realmente importam para representar os dados, mantendo a variância desses dados

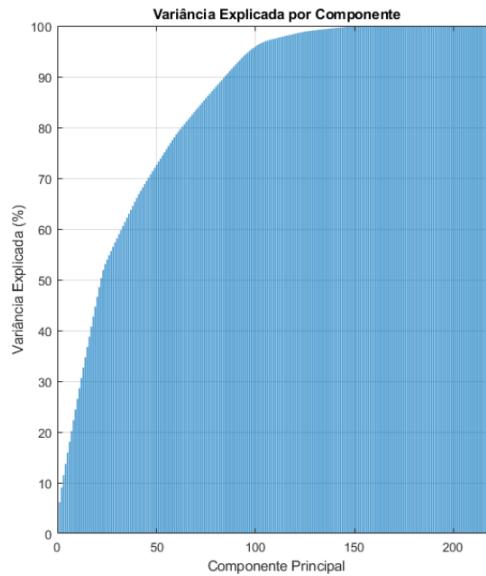


Fig. 3. Variância explicada por componente.

No gráfico é possível observar que as primeiras componentes explicam a maior parte da variância e conforme você avança, cada nova componente adiciona menos informação nova até chegar no limite, a região achata da curva, que indica que os novos componentes não adicionam mais informações significativas.

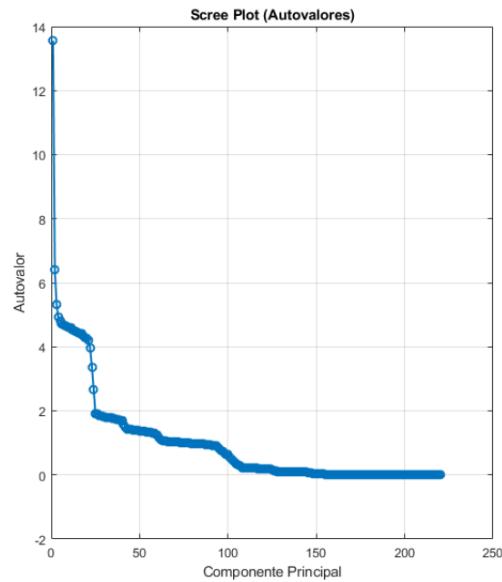


Fig. 4. Autovalores associados a cada componente principal.

Este gráfico apresenta os autovalores associados as componentes principais dos dados, onde autovalor alto indica que o componente tem muita variância e vice-versa. O que é encontrado no gráfico é que os primeiros componentes tem autovalor muito alto em relação aos demais, e conforme mais componentes são adicionados a sua variância vão diminuindo até chegar em 1, onde possuem menos variância que uma variável original, dessa forma, corroborando com o obtido no gráfico da Figura 3.

#### F. Classificação

A etapa de classificação se inicia com a montagem de um vetor de rótulos, associando sujeitos, *trial*, época e rótulo (associado ao movimento realizado). Após isso as instâncias são divididas em dois grupos, uma para treino e uma para teste, que será validada por meio de Validação Cruzada, que pode ser realizada de diferentes maneiras, neste trabalho o método de validação foi o *K-Fold*. Esse método consiste na divisão dos dados em *k folds* e em cada iteração 1 fold é utilizado como teste e os outros *k-1 folds* são usados como treino, esse processo se repete *k* vezes, alterando qual *fold* é o conjunto de teste. No final é calculado a média da performance do modelo.

Porém, o *K-Fold* é apenas a maneira que define como é o treinamento e teste do modelo de aprendizado de máquina, enquanto o *Linear Discriminant Analysis* (LDA) é a técnica utilizada dentro de cada etapa do treino para reduzir a dimensionalidade de forma **supervisionada**. No software *Matlab* isso é implementado utilizando a função `fitcdiscr(X, Y)`. O funcionamento analítico é dado da seguinte forma: Em cada *fold* se roda o LDA só com os dados de treino do *fold* da vez e após isso é aplicada a transformação LDA no *fold* de teste. O funcionamento do LDA dentro do *K-Fold* é o seguinte, considerando  $k = A$  para fins de explicação e simplificação:

- 1) 1 *fold* é escolhido para teste e os outros  $A - 1$  serão treino.
- 2) Utiliza as médias e dispersões das classes, caso utilize o *fold* reservado para teste acontecerá *data leakage* (vazamento de dados). Nesse processo o LDA calcula médias por classe (Equação 24), por exemplo, classe  $c$ : e média total (Equação 25):

$$\mu_c = \frac{1}{N_c} \sum_{x \in c} x \quad (24)$$

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (25)$$

É calculado pelo LDA também, a Matriz de Dispersão Intra-Classe  $S_W$  (Equação 26), que representa o quanto os dados estão espalhados dentro de cada classe, ou seja, a variabilidade que não ajuda na separação.

$$S_W = \sum_{c=1}^C \sum_{x \in c} (x - \mu_c)(x - \mu_c)^T \quad (26)$$

Com a Matriz de Dispersão Intra-Classe calculada é calculado então a Matriz de Dispersão Entre Classes  $S_B$

(Equação 27), que irá representar o quanto as médias das classes estão distantes entre si.

$$S_B = \sum_{c=1}^C (\mu_c - \mu)(\mu_c - \mu)^T \quad (27)$$

Então é iniciado a Busca das direções ótimas, tentando encontrar uma ou mais direções no espaço dos dados que separe as classes de melhor maneira. As direções encontradas são vetores, chamados autovetores que maximizam a razão

$$\frac{\text{separação entre classes}}{\text{dispersão dentro das classes}}$$

tentando se afastar ao máximo das médias das classes e compactar o máximo de dados dentro de cada classe, no contexto a razão a ser maximizada é:

$$\frac{|W^T S_B W|}{|W^T S_W W|}$$

Levando a resolver o autoproblema que possui  $\lambda$  representando os valores associados à capacidade discriminativa e  $v$  que são os vetores discriminantes:

$$S_W^{-1} S_B v = \lambda v$$

Com isso se inicia a Projeção dos Dados para as novas direções das classes, com  $W$  representando a matriz das direções da LDA e  $Z$  os dados transformados, por meio do apresentado na Equação 28:

$$Z = X_{treino} W \quad (28)$$

Dessa maneira LDA não vai mais trabalhar nas features originais, mas sim em  $Z$  que é onde já se sabe que as classes estão mais separáveis. Após isso ocorre a projeção do *fold* de teste, reaplicando a mesma lógica, mas sem recalcular  $W$  no teste, pois ele deve vir apenas do treinamento para evitar *data leakage*.

Se inicia então a etapa de Classificação, onde o LDA utiliza o modelo Bayesiano Gaussiano, apresentado na Figura ??, com mesma covariância para todas as classes, onde  $\sum$  representa a covariância compartilhada estimada nos dados projetados anteriormente e  $P(c)$  é a frequência da classe no treino.

$$\text{classe}(z) = \text{argmax}[z^T \sum_{c=1}^C \mu_c - \frac{1}{2} \mu_c^T \sum_{c=1}^C \mu_c + \ln P(c)] \quad (29)$$

1) *Cálculo das médias*: Com o treinamento concluível utilizando a validação cruzada são calculadas Médias de Acurácia (Equação 30) que indica o desempenho global do modelo ao longo de todos os *folds*, Média de Precisão (Equação 31) que indica quantas das previsões positivas estavam corretas, Média de Sensibilidade (Equação 32) que fornece quantos positivos reais o modelo detecta e Média de Especificidade (Equação 33) que o quanto o modelo erra ao confundir negativos como positivos.

$$\overline{\text{Acc}} = \frac{1}{K} \sum_{k=1}^K \frac{TP + TN}{TP + TN + FP + FN} \quad (30)$$

$$\overline{Prec} = \frac{1}{K} \sum_{k=1}^K \frac{TP}{TP + FP} \quad (31)$$

$$\overline{Sens} = \frac{1}{K} \sum_{k=1}^K \frac{TP}{TP + FN} \quad (32)$$

$$\overline{Esp} = \frac{1}{K} \sum_{k=1}^K \frac{TN}{TN + FP} \quad (33)$$

Onde:

- **TP** significa *True Positive*, isso é, Verdadeiro Positivo.
- **TN** significa *True Negative*, isso é, Verdadeiro Negativo.
- **FP** significa *False Positive*, isso é, Falso Positivo.
- **FN** significa *False Negative*, isso é, Falso Negativo.

#### G. Classificação

A etapa de Classificação é a apresentação e avaliação do desempenho do modelo, apresentando a acurácia calculada, apresentando a Matriz de Confusão que indica as porcentagens relacionadas aos acertos e erros, conforme apresentado logo acima, permitindo uma análise visual de como o modelo performou com os dados apresentados.

#### IV. RESULTADOS E DISCUSSÃO

Os códigos utilizados para a realização do experimento estão disponíveis em <https://github.com/MatheusBarusso/sistemas-inteligentes> juntamente com suas instruções de execução local para replicação. Ao utilizarmos o script, que possui uma pipeline explicitada na seção acima, obtemos os seguintes resultados observáveis:

1) *Visualização*: Para propósito de análise de visualização foram escolhidas arbitrariamente as variáveis com os seguintes valores: Sujeito = 1, Epoca = 1, Canal = 1. Com isso foram obtidas um total de 5 Figuras plotadas. Iniciando pela Figura 5, que apresenta as variações de voltagem em  $\mu V$  com o passar do tempo para o canal Fz do sinal bruto do EEG.

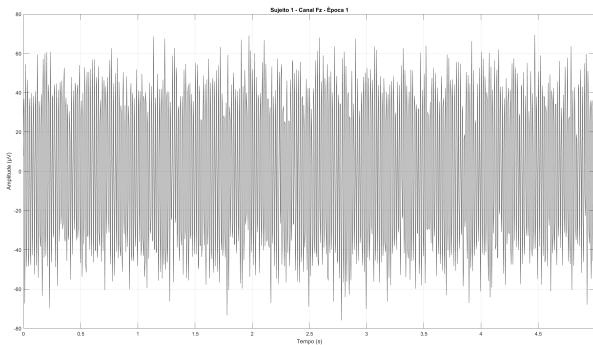


Fig. 5. Amplitude do sinal bruto de EEG para o canal Fz.

Após isso é apresentada a Figura 6, com a amplitude de todos os canais ao passar do tempo. É notável a aparição de

artefatos (que foram processados posteriormente) nos canais FCz e C1.

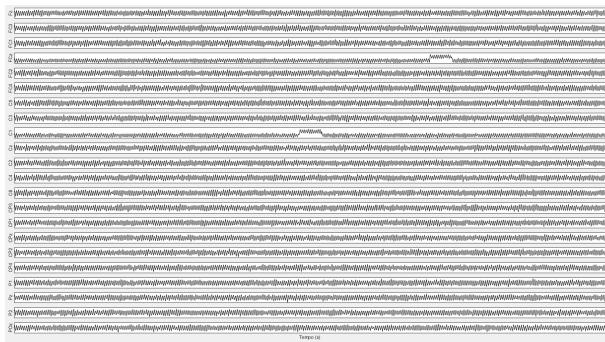


Fig. 6. Amplitude do sinal bruto de EEG para todos canais.

É apresentada também a comparação das características do sinal de EEG das 4 classes de ação em diferentes *Trials*, conforme apresentado na 7.

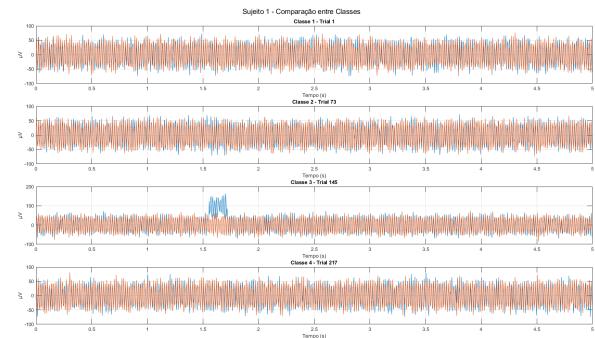


Fig. 7. Diferentes Classes em Diferentes *Trials*.

Então é plotado o mapa topológico, apresentando uma visão que simula o posicionamento dos canais (eletrodos) do EEG do Sujeito 1 para o *Trial 1*, com uma mapa de calor que mostra os canais com maior atividade na ocasião (Figura 8).

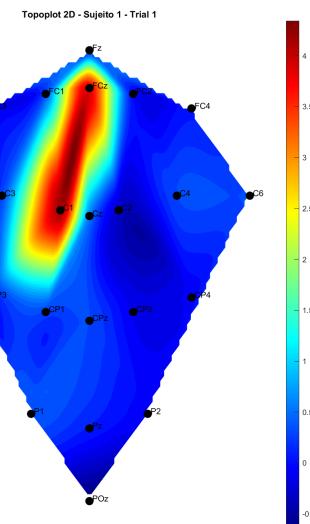


Fig. 8. Topologia de Canais Cerebrais do Sujeito 1 na *Trial 1*.

Foram plotados e apresentados o gráfico com a Densidade Espectral de Potência para todos os canais do dataset para o Sujeito e *Trial* escolhidos, que é apresentado na Figura 9.

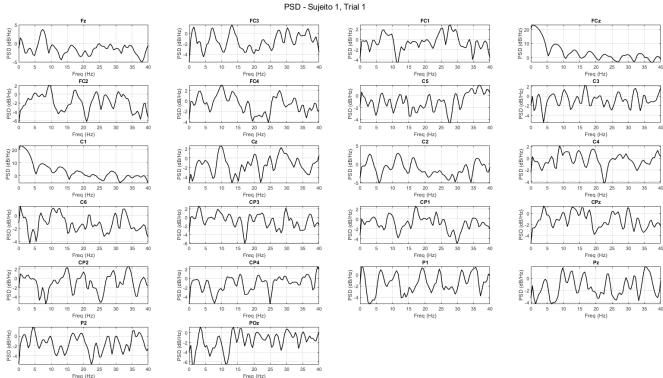


Fig. 9. Densidade Espectral de Potência do Sujeito 1 na *Trial* 1 para todos canais.

2) *Redução de Dimensionalidade*: A próxima etapa que apresenta (plota) resultados gráficos é a seção de Redução de Dimensionalidade, que utiliza PCA para diminuir as dimensões das matrizes geradas. As Figuras geradas já foram apresentadas na Figuras 3 e 4.

3) *Classificação e Avaliação de Desempenho*: Essas são as duas últimas etapas do experimento, que apresentam os resultados do modelo treinado. Como apenas uma medida de resultado pode criar uma falsa impressão de um resultado desejável (com boa confiança) foram utilizados por meio de diferentes métricas entre eles estão a montagem de uma Matriz de Confusão (Figura 10), apresentação de precisão global e local (por classe), *Recall* e *F1-Score* por classes. Além disso é apresentado também o *Matthews Correlation Coefficient* (MCC), que diferentemente da Precisão Global que pode se enganar caso uma classe seja muito mais presente que as outras, considera os quatro quadrantes da matriz de confusão para uma análise mais detalhada.

#### 1) Classe de Movimento Imaginário da Mão Esquerda.

- Precisão: 63,41%
- *Recall*: 70,71%
- *F1-Score*: 66,86%

#### 2) Classe de Movimento Imaginário da Mão Direita.

- Precisão: 48,63%
- *Recall*: 44,86%
- *F1-Score*: 46,67%

#### 3) Classe de Movimento Imaginário de Ambos Pés.

- Precisão: 45,26%
- *Recall*: 58,73%
- *F1-Score*: 51,12%

#### 4) Classe de Movimento Imaginário da Língua.

- Precisão: 82,62%
- *Recall*: 54,92%
- *F1-Score*: 65,98%

- Precisão Global: 57,30%
- MCC Global: 0,435

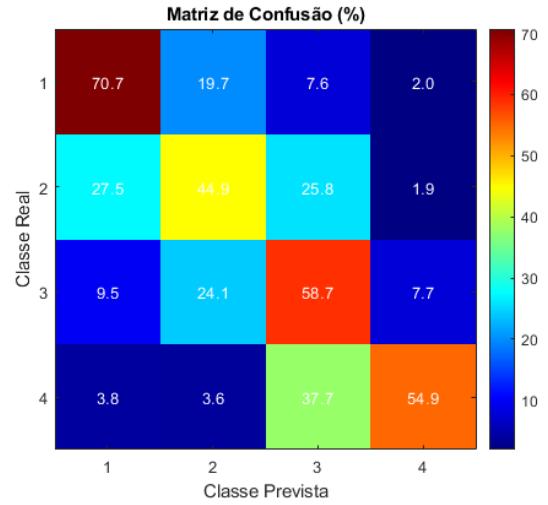


Fig. 10. Matriz de Confusão para o modelo criado.

## V. CONCLUSÕES

O presente seminário demonstrou a implementação do fluxo completo do processamento de sinais de Eletroencefalograma, contendo desde os processos de pré-processamento até a classificação supervisionada utilizando Análise Discriminante Linear, um método de aprendizado de máquina supervisionado.

A análise dos resultados revela que o modelo atingiu uma precisão global de 57,30% e um MCC de 0,435, mostrando uma capacidade de aprendizado e correlação moderada, considerando a probabilidade de acerto para quatro classes em um chute ser 25%. No entanto, o modelo teve um desempenho que variou significativamente entre as classes analisadas.

A classe de Movimento Imaginário de Mão Esquerda (Classe 1) apresentou o comportamento mais equilibrado e robusto, em contrapartida, a Classe 2 (Mão Direita) provou ser um ponto crítico do modelo, com baixa capacidade de detecção e alta confusão com a mão oposta. A classe de Movimento Imaginário da Língua apresentou alta confiança mas ao mesmo tempo uma baixa sensibilidade, obtendo 82,62% de precisão com poucos falsos positivos mas falhou na detecção de quase metade das intenções reais do usuário, conforme visto no seu *Recall* de 54,92%, confundindo essa classe frequentemente com a Classe de Movimento Imaginário de Ambos Pés (Classe 3), apontando uma possível sobreposição topológica ou espectral nos canais centrais que o PCA e o LDA não conseguiram separar totalmente.

Conclui-se que a abordagem utilizando LDA e PCA oferece uma solução computacionalmente eficaz e com resultado moderado. Porém trata-se de um experimento inicial, para resultados aplicáveis as métricas devem ser elevadas até um patamar estabelecido para fornecer uma melhor precisão.