Iniciação Científica. [Relatório Parcial, 03/21 - 10/21]

Departamento de Estatística (MAE) - IME/USP

Profª Florencia Graciela Leonardi

Classificação de sinais de EEG com modelos de regressão funcional

Rodrigo Marcel Araujo Oliveira (Nº USP 9299208)

✉ rodrigo.marcel.oliveira@usp.br

✉ rodmarcel92@gmail.com

04 de Outubro de 2021

Sumário

Página

Introdução

A eletroencefalografia é um dos melhores métodos para avaliar a atividade elétrica cortical, por ser um método barato, não invasivo e confiável [6].O sinal do EEG pode ser resultado da atividade espontânea do cérebro ou pode estar relacionado com eventos cerebrais sensoriais, motores e cognitivos [6]. No entanto, os problemas de classificação dos sinais do EEG [2] têm sido um grande desafio para comunidade científica, mas com os avanços nas tecnologias de Inteligência Artificial [20][15] e técnicas de Machine Learning [17][12] e Deep Learning [16] muitos estudos vem sendo feitos para entender o comportamento dessas estruturas.[1]

Este projeto de pesquisa tem como objetivo estudar técnincas de processamento de sinais [19], tais como as transformadas de Fourier e transformadas de Wavelet [10], para decomposição do sinal do EEG, e avaliar o desempenho de modelos de regressão fucional [3][4][5] para predição de novos dados [18].

Conforme o desenvolvimento do trabalho, os códigos serão comentados e podem ser acessados no repositório do `GitHub` dedicado a esse projeto:

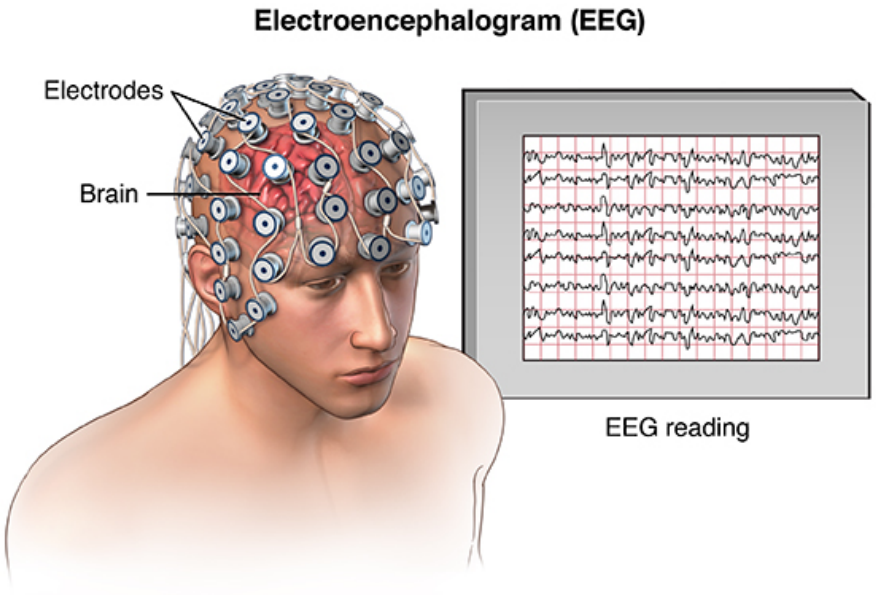
`https://github.com/roaraujo/Neuroscience-Statistical-Learning`

Eletroencefalograma (EEG)

A eletroencefalografia consiste no registo e avaliação dos potenciais elétricos gerados pelo cérebro. O eletroencefalograma (EEG) é o exame que registra a atividade elétrica cerebral, é uma técnica muito importante para avaliação neurofisiológica de pacientes com distúrbios do sono [9], morte cerebral, tumores, infecções cerebrais, epilepsia [7], predisposição genética ao alcoolismo, etc. O sinal do EEG pode ser resultado da atividade espontânea do cérebro ou pode estar relacionado com eventos cerebrais sensoriais, motores e cognitivos [6]. Este registo tem formas muito complexas, que variam em função da localização dos elétrodos, do número de interconexões que têm os neurónios e pelo facto de o encéfalo não ter uma estrutura uniforme. [6]

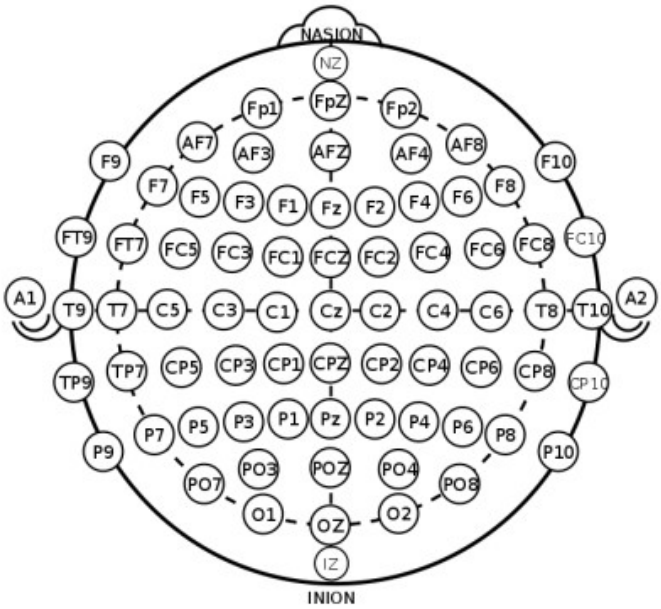
Geralmente o equipamento de EEG é da forma de um boné com vários eletrodos que são projetados para se fixar na superfície da cabeça, conforme ilustrado na Figura 1.

Figura 1: EEG



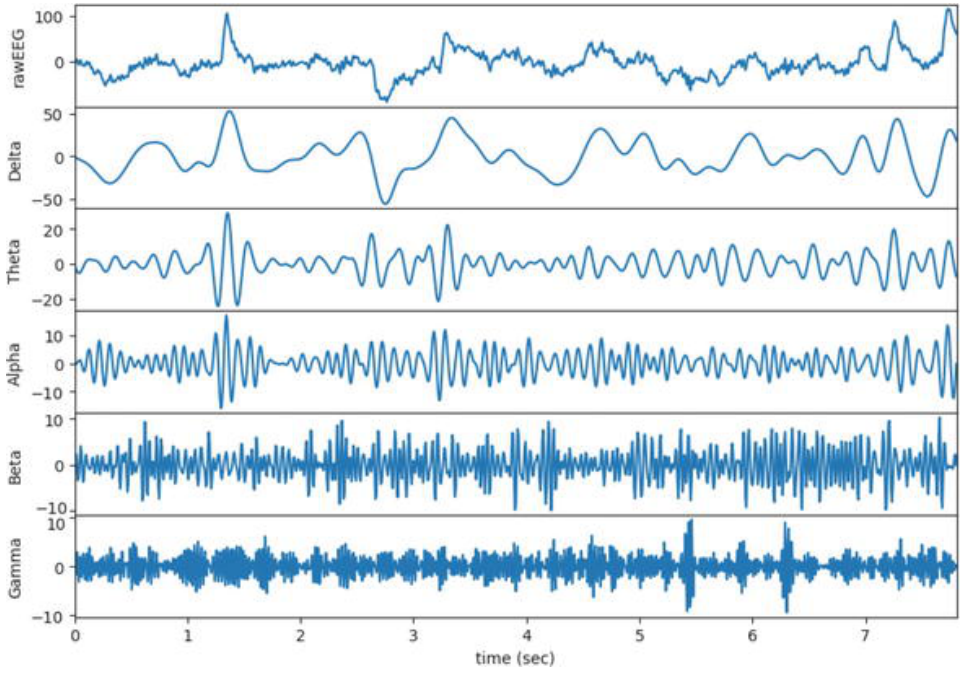
O número de eletrodos usados em gravações típicas variam com o tipo de experimento e o orçamento que está sendo feito. Contudo, independentemente do número de eletrodos usados, sua colocação sobre a cabeça geralmente segue um protocolo padrão internacional denominado Sistema 10–20, ilustrado na Figura 2 [21]:

Figura 2: Protocolo padrão internacional denominado Sistema 10–20



Normalmente, os sinais de EEG são categorizados em cinco categorias principais de ondas cerebrais de acordo com sua frequência, ou seja, o Delta (δ, <4 Hz), Theta (θ, 4-8 Hz), Alpha (α, 8-12 Hz), Beta (β, 12–35 Hz) e ondas Gama (γ,> 35 Hz), conforme mostrado na Figura 3.

Figura 3: Ondas cerebrais de acordo com sua frequência



As intensidades dessas ondas variam dependendo de quais atividades uma pessoa está realizando, e pode fornecer informações importantes sobre a saúde e o estado de espírito de uma pessoa. A Beta está associada à atenção redobrada, concentração, melhor acuidade visual e coordenação. A Teta ocorre predominantemente durante o sono leve de adultos. A Delta está associadas ao sono profundo. E por último a Gamma, está associada à percepção e à consciência.[21]

Os sinais de EEG são frequentemente contaminados por vários artefatos, os tipos mais comuns são de movimentos: musculares; oculares e cardíacos. Esses fatores são causados pelo movimento físico do corpo da pessoa, movimento produzem um pico repentino de alto valor em todos os canais de registro de EEG. Os movimentos musculares, como ranger de dentes, produz vários picos de alta frequência na gravação de EEG, inclusive fatores cardíacos, que são causados pelas atividades elétricas do coração, além dos oculares que são ondas oscilantes lentas que aparecem no lobo frontal, causadas pelos movimentos dos olhos ou olhos fechados.

Em virtude desses fatores, suas magnitudes podem corromper os registros EEG e consequentemente nos levar a interpretações errôneas dos resultados de análises.

Análise espectral

A análise espectral desempenha um papel importante na extração de informações do sinal, constitui uma forma alternativa de identificar, descrever e analisar sinais, permite a identificação de fontes de interferência e proporciona uma forma rápida e eficiente de identificar as componentes de um sinal.

Transformada de Fourier

Uma função periodica $S(t)$ que satisfaça as condições de Dirichlet pode ser expressa como uma série de Fourier, com termos seno e coseno harmonicamente relacionados [1].

$$S(t) = a_0 + \sum_{n\in\mathbb{N}} a_n \cos n\omega t + b_n \sin n\omega t$$

com a_0, a_n e b_n corresponde aos coeficientes de Fourier, e T sendo o período, e a frequência angular expressa por $\omega = \frac{2\pi}{T}$, dessa forma temos:

$$a_0 = \frac{1}{T}\int_0^T S(t)dt$$

$$a_n = \frac{2}{T}\int_0^T S(t)\cos n\omega t dt$$

$$b_n = \frac{2}{T}\int_0^T S(t)\sin n\omega t dt$$

As condições de Dirichlet constitui-se em $S(t)$ ser contínua, monotónica e integrável, e portanto:

$$\int_0^T |S(t)|dt < \infty$$

Usando a identidade de Euler, isto é, $e^{ix} = cos(x) + isin(x)$, podemos definir a Transformada de Fourier de uma função $f(t)$ por:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{i\omega t}dt.$$

em que $i = \sqrt{-1}$ e ω é a frequência angular dara por $\omega = 2\pi f$, onde f é a frequência.

O algoritmo da Transformada Rápida de Fourier (FFT), é uma maneira de obtermos a função $F(\omega)$ de uma série temporal. Os valores das abcissas são valores de frequência e os valores das ordenadas são números complexos de modo que:

$$F(\omega) = A(\omega) + iB(\omega) = R(\omega)e^{i\phi(\omega)}$$

em que $R = \sqrt{A^2 + B^2}$ e a fase de $F(\omega)$ é $\phi = \tan^{-1}(B/A)$.

Para fazer o processo inverso de modo a trazer de volta essa nova série temporal para o domínio do tempo, obtendo uma nova série $f'(t)$, podemos utilizar a Transformada Inversa de Fourier:

$$f'(t) = \int_{-\infty}^{\infty} F(\omega)e^{i\omega t}d\omega.$$

E portanto, temos:

$$f'(t) = \int_{-\infty}^{\infty} R(\omega)e^{i\phi(\omega)}e^{i\omega t}d\omega.$$

Transformada de Wavelet

Uma Wavelet é uma oscilação semelhante a uma onda localizada no tempo, cuja a idéia fundamental é analisar em função da escala, isto é, funções de base curta em altas freqûencias, e longa em baixas frequências. As propriedades propiciam a decomposição de outras funções. Isso permite que descontinuidades no sinal possam ser isaladas e analisadas por funções de base curtas, e para funções de base longas é possível obter uma análise de frequência mais detalhada.

As funções de Wavelets separam dados em diversas componentes de frequência, e extraem cada componentes de acordo com sua escala. As bases da análise de Fourier são ondas senoidais, e portanto o sinal é analisado como um todo, são suaves e previsíveis, isso se torna um problema quando nos deparamos com sinais que contêm descontinuidades e veriações bruscas. As Walevets decompõe o sinal em versões escalonadas e deslocadas de sua Wavelet original, elas tendem a serem irregulares e assimétricas.

As propriedades das funções de Wavelets permitem deslocamento no tempo, dilatação e compressão para variação da largura da janela, deslocamento no espectro, ou seja, o filtro que passa na banda é ajustável. A função de Wavelet no tempo pode ser escrita da seguinte forma:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s})$$

A expressão $\psi(\frac{t-\tau}{s})$ é denominada como Wavelet mãe, constituída por versões deslocadas do sinal através do seu atraso ou avanço em realação ao ponto inicial e dilatadas ou comprimidas. O parâmetro s da função de Wavelet corresponde a largura da janela e a faixa de frequências capturadas, isso possibilita a dilatação ou compressão da escala de tempo, além disso, o fator $\sqrt{s}$ corresponde a normalização. Já o parâmetro τ é um deslocamento da função no tempo.

Considerando o domínio da frequência a função geral é escrita como:

$$\Psi_{s,\tau}(\omega) = \sqrt{s}e^{-i\tau\omega}\Psi(s\omega)$$

As funções devem satisfazer algumas condições de admissibilidade, isto é, devem ter médias iguais a zero para não adicionar tendência a transformação, além disso temos:

$$C_\psi = \int \frac{\mid \Psi(\omega) \mid}{\mid \omega \mid} d\omega < \infty$$

$$\int \psi(t) dt = 0$$

A transformada de Wavelet Contínua (CWT, *ContinuosWaveletTransform*) é definida por:

$$W_{s,\tau} = \frac{1}{\sqrt{s}} \int x(t) \psi^*(\frac{t-\tau}{s}) dt = \int x(t) \psi^*_{s,\tau}(t) dt$$

onde $W_{s,\tau}$ é a transformada de Wavelet da função $x(t)$ no tempo t, considerando uma eacala s, e ψ^* denota o conjugado complexo.

O sinal original pode ser recuperado usando a transformação inversa, escrita como:

$$x(t) = \frac{1}{C_\psi} \int \int W_{s,\tau} \psi_{s,\tau}(t) \frac{ds d\tau}{s^2}$$

Para sinais discretos a CWT pode ser escrita como:

$$W_s(\tau) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \Psi^*_s(k) e^{i\omega_0 k\tau}, \ com \ \tau = 0, 1, ..., N$$

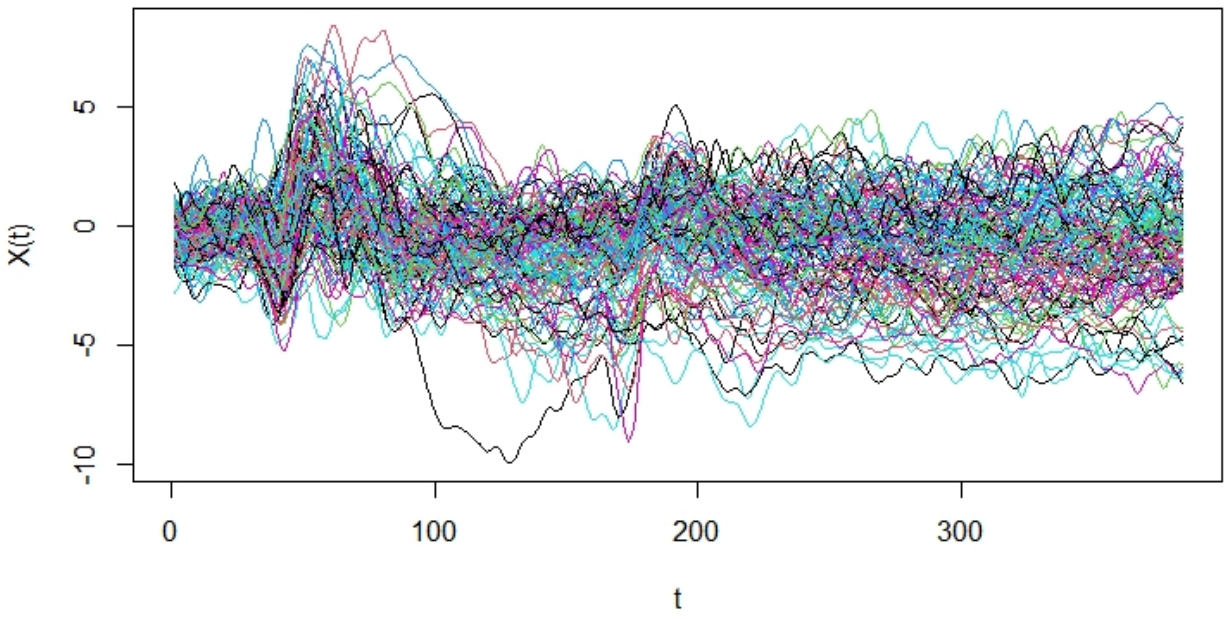
Onde $X(k) = F\{x(n)\}$ e $\psi_s(k) = F\{\psi_s(n)\}$ e $\omega_0 = \frac{2\pi}{N}$

Análise de dados funcionais

Dados Funcionais

A análise de dados funcionais é um ramo da estatística preocupado com a análise de dados na forma de funções. Com advento de novas tecnologias propiciou a produção de dados cada vez mais complexos, isso desencadeou o desenvolvimento de novas áreas da estatística, entre elas temos a análise de dados funcionais (FDA). Os dados observados consistem em uma amostra de funções de uma população em que cada função amostrada é de uma grade discreta. A grade pode ser esparsa ou finita, regular ou irregular, etc. Normalmente, essas funções são definidas no domínio euclidiano unidimensional, contudo há funções definidas em domínios de dimensões superiores, tais como dados de imagem ou dados de temporais, além de funções em outros domínios não euclidianos.[22] A seguir temos um exemplo de sinais de EEG ao longo do tempo.

Figura 4: Sinais de EEG



O surgimento deste campo na estatística foi motivado por dados longitudinais, curvas de séries temporais. A ideia principal que caracteriza o FDA é pensar que para cada função tem-se um único objeto estruturado, em vez de uma coleção de pontos de dados, isso permite a construção de modelos que podem lidar simultaneamente com estrutura dentro das funções e entre as funções.[22]

Assim como qualquer método estatístico, o FDA combina informações através de observações de alguma forma para fazer inferência sobre as populações. Para isso, considera-se cada função como a unidade de amostra, e a replicação envolve como nossos modelos combinam informações entre funções, assim podemos fazer inferências sobre as populações das quais foram extraídas. A maioria dos métodos do FDA, a regularização envolve suavização, isso implica uma suposição de que as similariedades dentro das funções está estritamente relacionada com à distância entre as observações. A regularização pode ser aplicado na probabilidade dos dados funcionais para capturar a correlação dentro da função, a ideia principal pode-se obter estimadores mais eficientes e interpretáveis, melhores previsões e, até mesmo cálculos que são potencialmente mais rápido e estável.[23]

Funções de Base

As funções de bases são os blocos de construção do FDA e determinam o mecanismo pelo qual a regularização é feita. Para cada função de base define-se uma combinação linear entre os locais dentro da função,isso induz uma correlação entre as regiões funcionais e seu uso permite uma quantidade de números finitos de coeficientes da base para produzir estimativas e inferências em uma função de dimensão infinita espaço. Algumas das funções de base mais comumente usadas no FDA são splines, série de Fourier, etc. Cada uma é adequada para funções com certas características.[23]

As bases de B-splines são adequadas para modelar funções suaves, elas funcionam melhor quando o dimensão da grade de observação não é muito alta. A série de Fourier transforma as funções no domínio da frequência e funcionam bem para funções com características periódicas estacionárias, essa série possui algoritmos rápidos para calcular coeficientes de base, é adequada para funções amostradas em grades regulares de alta dimensão.[23]

Se assumirmos que nossos dados funcionais Y_t são observados através do modelo:

$$Y_{ti} = X_{ti} + \epsilon_{ti}$$

onde os resíduos ϵ_t são independentes de X_t, podemos obter de volta o sinal original X_t usando uma suavização linear,

$$\hat{x} = \sum_{i=1}^{n} s_{ij} y_i$$

em que s_{ij} é o peso que o ponto t_j atribui ao ponto t_i e $y_i = Y_{t_i}$.

Uma curva pode ser representada por uma base quando os dados pertencem ao espaço $L2$. A base é um conjunto de funções conhecidas $\{\phi_k\}_{k \in N}$ em que a função pode ser aproximada por uma combinação linear de um número suficientemente grande de k_n dessas funções, o valor da função X_t é recuperado usando uma expansão de base truncada fixa em termos dos k_n elementos da base. [23]

$$X(t) = \sum_{k \in \mathbb{N}} c_k \phi_k(t) \approx \sum_{k=1}^{k_n} c_k \phi_k(t) = \mathbf{c}^T \mathbf{\Phi}(\mathbf{t})$$

A matriz de suavização é dada por: $\mathbf{S} = \mathbf{\Phi}(\mathbf{\Phi^T\Phi})^{-1}\mathbf{\Phi^T}$, com graus de liberdade do ajuste $df = tr(S_v) = k_n$. [23]

Podemos ainda assumir um modelo de regressão funcional dado por:

$$Y = \langle X, \beta \rangle + \epsilon = \frac{1}{\sqrt{T}} \int_T X(t)B(t)dt + \epsilon$$

onde $\langle X, \beta \rangle$ denota o produto interno, já ϵ são erros aleatórios com média zero e variância finita. O modelo linear funcional pode ser estimado pela expressão:[23]

$$\hat{Y} = \langle X, \hat{\beta} \rangle = \mathbf{C^T}\psi(\mathbf{t})\phi^{\mathbf{T}}(\mathbf{t})\hat{\mathbf{b}} = \tilde{\mathbf{X}}\hat{\mathbf{b}}$$

com $\tilde{\mathbf{X}}(\mathbf{t}) = \mathbf{C}^T\psi(\mathbf{t})\phi^T(\mathbf{t})$ e $\hat{\mathbf{b}} = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-\mathbf{1}}\tilde{\mathbf{X}}^T Y$ e portanto:

$$\hat{Y} = \tilde{\mathbf{X}}\hat{\mathbf{b}} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-\mathbf{1}}\tilde{\mathbf{X}}^T Y = \mathbf{H}\,Y$$

Com $\mathbf{H}$ sendo a matriz com graus de liberdade $df = tr(\mathbf{H})$.[23]

- **Base de Fourier**

  A base mais apropriada para funções periódicas definidas em um intervalo T é a Base de Fourier onde os $\phi_k's$ tem a seguinte forma:

$$\phi_0(t) = \frac{1}{\sqrt{\mid T \mid}}$$
$$\phi_{2r-1}(t) = \frac{\sin r\omega t}{\sqrt{\frac{|T|}{2}}}$$
$$\phi_{2r}(t) = \frac{\cos r\omega t}{\sqrt{\frac{|T|}{2}}}$$

  com $r = 1, ..., \frac{\mathbf{K-1}}{2}$ em que é o $\mathbf{K}$ número de bases da função, e a fequência ω determina o período e a duração do intervalo $\mid T \mid = \frac{2\pi}{\omega}$

- **Base de Spline**

  As bases de B-splines são usados como funções de base para ajustar curvas de suavização a grandes conjuntos de dados. Esse método consiste em, dividir o eixo das abscissas em alguns intervalos, para cadas intervalo os pontos finais são chamados de pontos de interrupção e com isso esses pontos são convertidos em nós, o que impõem várias condições de suavização e continuidade para cada interface. Assim, as bases de B-spline de ordem k são definidas por:

$$B_{i,1}(x) = \begin{cases} 1, \ t_i \leq x < t_{i+1} \\ 0, \ c.c \end{cases}$$

$$B_{i,k}(x) = \frac{(x - t_i)}{t_{i+k-1} - t_i} B_{i,k-1}(x) + \frac{(t_{i+k} - x)}{t_{i+k} - t_{i+1}} B_{i+1,k-1}(x)$$

  com $i = 0, ..., n-1$, em que t é um vetor de nó não decrescente, ou seja, $t = \{t_0, t_1, ..., t_{n+k-1}\}$. Comumente B-splines cúbicos são usados com k = 4.

Regressão Funcional

Um modelo de regressão é considerado "funcional"quando uma variável preditora é funcional. O Modelo Linear Funcional (FLM) [24] foi introduzido como:

$$Y_i = B_0 + \int X_i(t)B(t)dt + \epsilon_i$$

em que Y_i, $i = 1, ..., N$ é uma resposta contínua, $X_i(t)$ é um preditor funcional, $B(t)$ coeficiente funcional, com B_0 sendo o intercepto e $\epsilon_i \sim N(0, \sigma^2)$ erros residuais. Já para respostas não gaussianas foi proposto o Modelo Linear Funcional Generalizado (GFLM), introduzido pela primeira em 1999 [25], em que a distribuição pertence a familia exponencial:

$$g\{Y_i\} = B_0 + \int X_i(t)B(t)dt$$

para alguma função de ligação $g(\cdot)$, com $X_i(t) = \sum_{k=1}^{K_X} X_{ik}^* \phi_k(t)$, $B(t) = \sum_{k=1}^{K_B} B_k^* \psi_k(t)$.

A regularização pode ser feita no espaço de base por meio de truncamento, penalidades ou esparsidade. Para regressão de preditor funcional, esta regularização pode ser aplicado aos coeficientes funcionais $B(t)$ e também para os preditores $X_i(t)$. A regularização de $B(t)$ reduz a colinearidade no ajuste de regressão, aumenta a interpretabilidade de estimativas de coeficiente, aumenta potencialmente a estimativa e a eficiência de predição. Já para uso de funções de base e regularização em $X_i(t)$ contribui para reduzir o erro de medição em preditores $X_i(t)$, que para respostas gaussianas aumenta a eficiência de estimativa e para modelos não lineares envolvendo resultados não gaussianos reduz viés induzido por erro de medição. [22] [26]

- **Regressão Logística**

  Na regressão logística, a probabilidade p_i da ocorrência de um evento $Y_i = 1$ ao invés do evento $Y_i = 0$ condicional a um vetor de covariáveis $X_i(t)$ é expressa como:

$$p_i = \mathbf{P}[\mathbf{Y} = 1 \mid \mathbf{X_i}(t) : t \in T] = \frac{exp\{\alpha + \int_T \mathbf{X_i}(\mathbf{t})\mathbf{B}(\mathbf{t})dt\}}{1 + exp\{\alpha + \int_T \mathbf{X_i}(\mathbf{t})\mathbf{B}(\mathbf{t})dt\}}$$

  com $i = 1, ..., n$

Critérios de Desempenho dos Modelos

Para avaliar o desempenho dos modelos ajustados e posteriormente eleger os melhores, vamos nos basear principalmente em métricas derivadas da matriz de confusão, a seguir temos uma breve explicação do objetivo dessas métricas e suas respectivas formulas de cálculo.

Métricas Baseadas na Matriz de Confusão

A matriz de confusão consiste em uma matriz (2x2) onde as respostas reais e preditas são alocadas de forma que possamos avaliar o desempenho de um modelo de classificação, nesta matriz as linhas correspondem a classe real do indivíduo e as colunas a classe predita. O nome matriz de confusão deriva do fato de tornar fácil verificar se o sistema (predito x real) esta se confundindo, a seguir temos uma ilustração de uma matriz de confusão genérica.

Tabela 1: Matriz de Confusão

		Valor Predito	
		Positivo	**Negativo**
Valor Real	**Positivo**	n_{11}	n_{12}
	Negativo	n_{21}	n_{22}

a partir dessa matriz podemos definir algumas métricas utilizadas na avaliação de modelos de classificação.

- **Sensibilidade (SEN):** Calcula a proporção de observações positivas classificadas corretamente pelo modelo, quanto maior a sensibilidade de um modelo melhor será a capacidade de ele classificar as observações positivas corretamente.

$$SEN = \frac{n_{11}}{n_{11} + n_{12}}$$

- **Especificidade (ESP):** Semelhante a sensibilidade essa medida verifica a proporção de negativos classificados corretamente pelo modelo, em geral um modelo bom tem alta sensibilidade e alta especificidade.

$$ESP = \frac{n_{22}}{n_{21} + n_{22}}$$

- **Acurácia (ACU):** Essa medida mede a proporção de predições corretas do modelo, quanto maior seu valor maior a taxa de acerto do modelo [**?**].

$$ACU = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

- **Acurácia Balanceada (ACUB):** Essa medida é semelhante a acurácia mas leva em conta o desbalanceamento dos dados.

$$ACUB = \frac{SEN + ESP}{2}$$

- **Prevalência (PRE):** É a proporção de individuos positivos, serve para avaliarmos o desbalanceamento dos dados.

$$PRE = \frac{n_{11} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

- **Valor Predito Positivo (VPP):** É a probabilidade que individuos classificados como positivo sejam de fato positivo.

$$VPP = \frac{n_{11}}{n_{11} + n_{12}}$$

- **Valor Predito Negativo (VPN):** É a probabilidade que individuos classificados como negativo sejam de fato negativo.

$$VPN = \frac{n_{22}}{n_{21} + n_{22}}$$

- **Taxa de detecção (TD):** É a proporção de individuos positivos detectados pelo modelo dentre todos os individuos.

$$TD = \frac{n_{11}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

- **prevalência da detecção (PD):** É a proporção de individuos preditos positivos dentre todos os individuos preditos.

$$PD = \frac{n_{11} + n_{12}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

- **Estatistica Kappa (Kappa):** Esta estatística verifica a proporção da concordância entre as categorias preditas e as categorias verdadeiras.

$$Kappa = \frac{\theta_o - \theta_e}{1 - \theta_e}$$

onde, $\theta_o = n_{11} + n_{11}$ é a concordância total observada e θ_e é a concordância total esperada, obtida pela formula a seguir:

$$\theta_e = \sum_{i=1}^{2} \theta_i,\ \theta_i = \frac{1}{(n_{11} + n_{12} + n_{21} + n_{22})^2} \sum_{j=1}^{2} n_{ji} \sum_{j=1}^{2} n_{ij}$$

Outras Métricas

Além das medidas derivadas da matriz de confusão, existem também medidas que avaliam a qualidade do ajuste como AIC, BIC e log-verossimilhança, essas medidas foram utilizadas em alguns modelos para regulariza-los, a seguir falamos de forma sucinta sobre essas métricas.

- **AIC, BIC, Log-verossimilhança:** Essas métricas são utilizadas em modelos de classificação, auxiliando na escolha das variáveis relevantes para o modelo e na estimativa dos parâmetros.
  - **log-verossimilhança:** O log da verossimilhança mede o quão bem o modelo se ajusta aos dados, como essa medida varia entre $[-\infty, 0]$ quanto mais próximo de 0 melhor o modelo se ajusta.
  - **AIC, BIC:** Essas duas medidas seguem o mesmo princípio do log da verossimilhança, no entanto aplicam penalidades de acordo com o número de parâmetros. A formula de cálculo para essas métricas podem aparecer de diversas formas dependendo das suposições adotadas na modelagem, a seguir apresentamos a forma de calculo mais genérica adota por [27]

$$AIC = -2l + 2p$$

$$BIC = -2l + log(n)p + C$$

    onde, l é o logaritmo da verossimilhança, p é o número de parâmetros, n é o tamanho amostral e C é uma constante que depende somente da amostra.

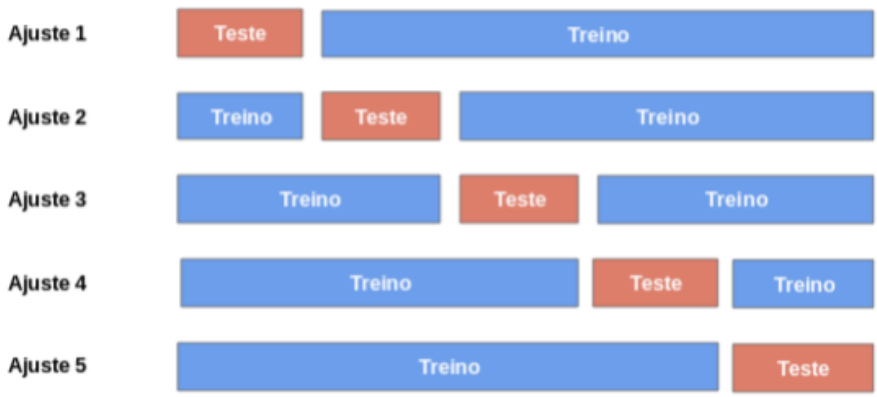
Validação Cruzada

A situação ideal para avaliação da capacidade preditiva de um modelo ocorre quando existem dois conjuntos de dados, o primeiro, chamado conjunto de treinamento, que será usado para ajustar o modelo e o segundo, chamado conjunto de teste, onde serão realizadas as predições do modelo e posterior cálculo das medidas de qualidade das predições. No entanto, esse cenário quase nunca ocorre, o que faz com que modelos sejam ajustados e testados no mesmo conjunto de dados, tornando as medidas de avaliação mais otimistas do que deveriam ser.

Devido a isso, se fazem necessárias as técnicas de validação cruzada, que dividem o conjunto de dados em subconjuntos a serem usados como dados de treinamento e de teste, para os quais o mesmo modelo é treinado e testado gerando estimativas melhores para o desempenho do modelo e tornando as medidas da capacidade do modelo mais condizentes com seu verdadeiro desempenho. A seguir apresentamos algumas dessas técnicas de validação cruzada, para um leitor que tenha interesse em se aprofundar no tema recomendamos [28], que ao final apresenta uma recomendação de como escolher a melhor técnica de validação cruzada

- ***Hold-out*:** Essa técnica consiste simplesmente na divisão do conjunto de dados de forma aleatória em dois grupos, um de treinamento e outro de teste. É comum que o grupo de treinamento seja maior que o de teste, geralmente o tamanho do grupo de treinamento é 2/3 do banco de dados.

- ***K-fold*:** Nesta técnica o conjunto de dados é dividido em k partes, em cada ajuste uma dessas partes é considerada como conjunto de testes e as demais como conjunto de treinamento. A cada ajuste são coletadas as estimativas dos parâmetros e as medidas de desempenho, ao final do processo são calculadas suas médias que serão as estimativas e medidas finais. A seguir temos uma imagem ilustrativa do processo retirada de [18].



- ***Leave-One-Out Cross-Validation* (LOOCV):** Essa técnica é um caso extremo do *k-fold* quando k é igual ao tamanho do banco de dado, ou seja, todo banco de dados será usado tanto para treinar quanto para testar o modelo. Uma desvantagem dessa técnica é seu alto custo computacional para conjuntos grandes de dados.

Bootstrap

Essa técnica é utilizada nos modelos de árvores e tem por objetivo gerar diversos conjuntos para treinamento, através de uma amostragem com reposição do banco de dados, esses conjuntos tornam mais precisas as estimativas dos parâmetros do modelo e facilitam a obtenção dos erros das estimativas sem a necessidade de uma teoria mais complicada. A seguir apresentamos o algoritmo para o cálculo do erro padrão de um estimador $\hat{\theta} = t(x)$ descrito por [18].

1. Selecione B amostras independentes $x_1^*, \ldots, x_B^*$ cada uma consistindo de n (tamanho do conjunto de dados) valores do conjunto de dados. Tome $B \approx 200$
2. Para cada amostra *bootstrap* x_b^* calcule a réplica *bootstrap* do estimador

$$\widehat{\theta}^*(b) = t(x_b^*),\ b = 1, \ldots, B$$

3. o erro padrão de $\hat{\theta}$ é estimado pelo desvio padrão das B réplicas.

$$\widehat{e.p.}_B = \left[\frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^*(b) - \hat{\theta}^*(.))^2 \right]^{\frac{1}{2}}$$

com

$$\widehat{\theta}^*(.) = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*(b)$$

Metodologia e Resultados

Metodologia

O nosso estudo está focado em avaliar quantitativamente os métodos de Regressão Funcional para classificação de EEG, e para isso foram aplicadas técnicas de processamento de sinais para decomposição do sinal do EEG, tais como as transformadas de Fourier e as transformadas de Wavelet. Para avaliação do nosso estudo, estudamos 4 métodos diferentes, que serão discutidos adiante. Esses 4 métodos foram avaliados em 2 conjuntos de dados de EEG, relacionados a problemas de reconhecimento de emoções e detecção de movimentos motores. A seguir apresentamos as metodologias utilizadas no estudo:

1. Transformada de Wavelet + Função de Base de Fourier + Regressão Logística Funcional
2. Transformada de Wavelet + Função de Base de Spline + Regressão Logística Funcional
3. Transformada de Fourier + Função de Base de Fourier + Regressão Logística Funcional
4. Transformada de Fourier + Função de Base de Spline + Regressão Logística Funcional

Os conjuntos de dados utilizados no estudo são:

- **Emotion**

  Para analisar a relação funcional entre eletroencefalografia e eletromiografia facial (EMG), foram registrados simultaneamente sinais de EEG e EMG de 24 participantes enquanto eles jogavam uma tarefa de jogo computadorizada. O subconjunto fornecido contém observações agregadas de 23 participantes. As curvas foram calculadas em média sobre cada sujeito e cada uma das 8 configurações de estudo, resultando em 23 vezes 8 curvas. Durante as rodadas de jogos de azar, três condições binárias de jogo foram variadas, resultando em um total de 8 estudos diferentes, as definições são: a condutividade da meta correspondente ao resultado monetário (ganho ou perda) no final de cada rodada do jogo; a configuração de potência, que determinava se o jogador era capaz ou não de alterar o resultado a seu favor (alto ou baixo, respectivamente); e a configuração de controle, que foi manipulada para mudar o sentimento

subjetivo do participante sobre sua capacidade de lidar com o resultado do jogo. Para nosso estudo o objetivo será explicar os potenciais no sinal EEG para condutividade da meta correspondente ao resultado monetário e a configuração de potência do jogador[1]. A base contém uma dimensão de (184, 384) para as covariáveis e (1, 184) para as variáveis categóricas. Para modelagem, separamos a base em treino (70%) e teste (30%), desse modo a base de treino contém uma dimensão de (128, 384) para covariáveis e (1, 128) para variável resposta, já para base de teste temos respectivamente, (56, 384) e (1, 56).

Tabela 2: Descrição das Variáveis

Variável	**Descrição**
gameoutcome	Variável categórica correspondente ao resultado monetário (ganho ou perda).
power	Variável categórica correspondente a configuração de potência (alto ou baixo).
control	Variável categórica correspondente a configuração de controle (alto ou baixo).
subject	identificação dos sujeitos.
EEG	Média dos sinais de EEG para cada sujeito
EMG	Média dos sinais de EMG para cada sujeito.
s	Pontos de tempo para a covariável funcional.
t	Pontos de tempo para a resposta funcional.

[1]https://rdrr.io/github/fdboost/FDboost/man/emotion.html

- **SelfRegulationSCP1**

  O experimento deste conjunto de dados consiste em avaliar se o sujeito está aumentando ou diminuindo sua lentidão cortical potencial, isto é, se o sujeito moveu o cursor para cima ou para baixo. Para isso, um sujeito saudável foi solicitado a mover o cursor para cima e para baixo em um computador de tela, enquanto seus potenciais corticais foram medidos e durante a gravação o sujeito recebeu uma resposta visual de seus potenciais corticais lentos (Cz-Mastoides). A positividade cortical leva a um movimento descendente do cursor na tela, já a negatividade cortical leva a um aumento movimento do cursor. Há um total de 561 tentativas, cada uma com duração de 6 segundos. Assim, durante cada tentativa, a tarefa era visualmente apresentado por uma meta destacada na parte superior ou inferior da tela para indicar negatividade ou positividade do segundo 0,5 até o final do ensaio. A resposta visual foi apresentado a partir de 2 segundos a 5,5 segundos, porém apenas o intervalo de 3,5 segundos de cada tentativa é fornecido. As gravações foram feitas com 6 canais de EEG em 256 Hz, o que resultou em 896 amostras por canal para cada tentativa. [2]. A base de treino contém uma dimensão de (268, 896) para covariáveis e (1, 268) para variável resposta, já para base de teste temos respectivamente, (293, 896) e (1, 293).

Tabela 3: Descrição das Variáveis

Variável	**Descrição**
Potenciais corticais	Variável categorica correspondente aos rótulos de classe (negatividade ou positividade)
Canal 1	Variável A1-Cz (sistema 10/20, A1 = mastóide esquerda)
Canal 2	A2-Cz
Canal 3	2 cm frontal de C3
Canal 4	2 cm parietal de C3
Canal 5	2 cm frontal de C4
Canal 6	2 cm parietal de C4

[2]http://www.timeseriesclassification.com/description.php?Dataset=SelfRegulationSCP1

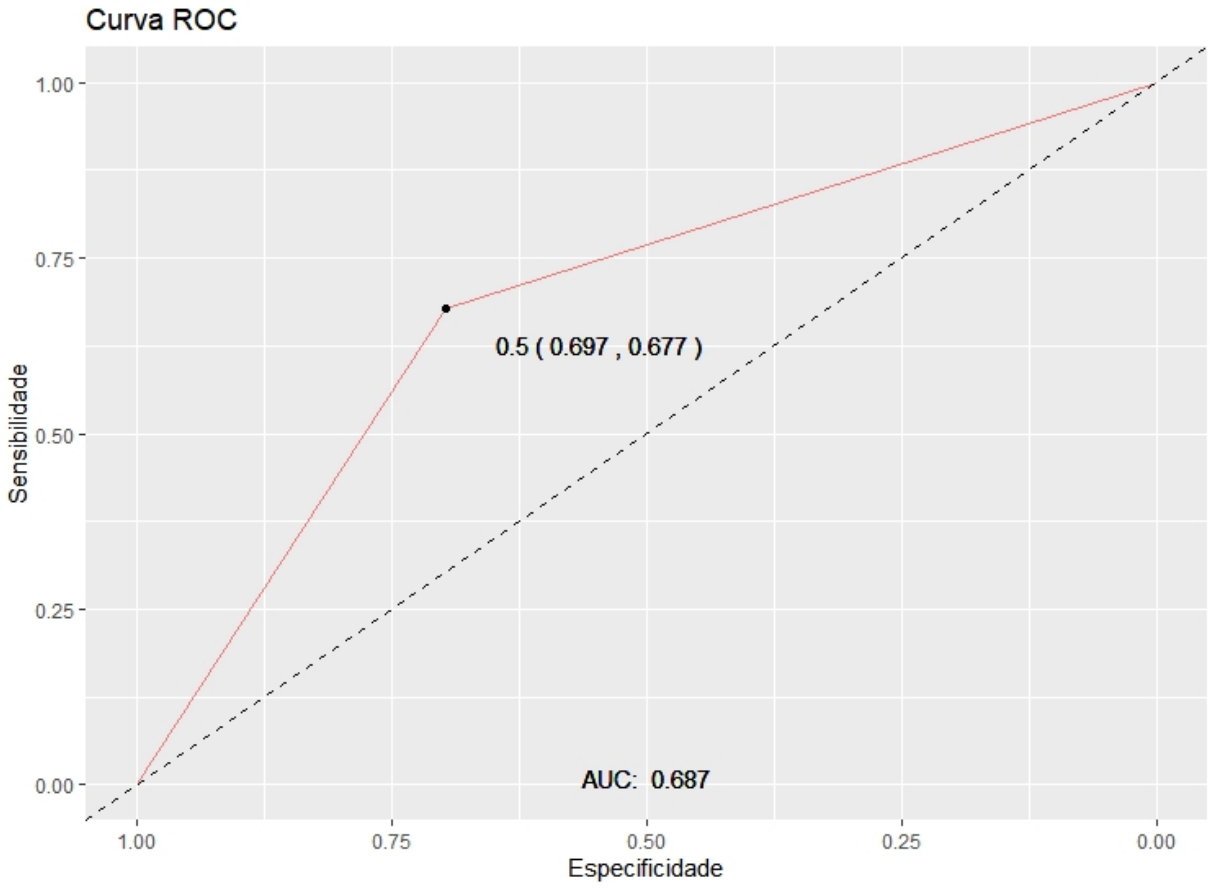
Resultados

- **Emotion - gameoutcome**

  A seguir temos o resultado dos modelo considerando a condutividade da meta correspondente ao resultado monetário (ganho ou perda) como variável resposta. Para o conjunto de treinamento utilizando a transformada de Wavelet com a função de Base de Fourier (composta por 8 componentes) e posteriormente a Regressão Logística Funcional, obteve-se a matriz de confusão da tabela seguinte.

	Referência	
Predito	**No**	**Yes**
No	46	20
Yes	20	42

O desempenho na curva ROC é mostrado no gráfico seguinte. E o AUC foi de 0.687.



Finalmente o desempenho geral do modelo sobre o conjunto de validação é apresentado na tabela seguinte.

Desempenho do Modelo			
Accuracy	0.5536	Sensitivity	0.6154
95% CI	(0.4147, 0.6866)	Specificity	0.5000
No Information Rate	0.5357	Pos Pred Value	0.5161
P-Value [Acc >NIR]	0.4480	Neg Pred Value	0.6000
		Prevalence	0.4643
Kappa	0.1139	Detection Rate	0.2857
		Detection Prevalence	0.5536
Mcnemar's Test P-Value	0.4237	Balanced Accuracy	0.5577

Referências

[1] Guimarães, L. S. A. Desvendando o Cérebro: um Estudo Quantitativo de Técnicas para Classificação de EEG.. **Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo**, São Paulo, 2020

[2] Gannaz, Irène. Classification of EEG recordings in auditory brain activity via a logistic functional linear regression model. **International Workshop on Functional and Operatorial Statistics**, Université de Lyon France, 2014, *hal-00830313v2*

[3] Yihong Zhao, R. Todd Ogden, Philip T. Reiss. Wavelet-based LASSO in functional linear regression. **J Comput Graph Stat. Author manuscript; available in PMC 2013 Jun 19.**, Published in final edited form as: J Comput Graph Stat. 2012 Jul 1; 21(3): 600–617. Published online 2012 Aug 16. doi: 10.1080/10618600.2012.679241, *PMCID: PMC3685865*

[4] Wang, Jane-Ling; Chiou, Jeng-m. Functional Data Analysis. **Department of Statistics, University of California, Davis, California 956116; Institute of Statitical Science, Academia Sinica, Taipei 11529, Taiwan, Republic of China**, *Annual Reviews Further*

[5] Greven, Sonja; Scheipl, Fabian. A general framework for functional regression modelling. **Department of Statistics, Ludwig-Maximilians-Universitat Munchen, Germany**, *Statistical Modelling 2017; 17(1-2): 1-35*

[6] Rodrigues, L. D. S. Diagnóstico da Doença de Alzheimer em Intervalos de Curta Duração utilizando o EEG.. **Dissertação (Mestrado em Tecnologia Biomédica) - Escola Superior de Tecnologia e Gestão Instituto Politécnico de Bragança**, Portugal, 2012

[7] Sturzbecher, MJ. Métodos clássicos e alternativas para a análise de dados de fMRI e EEG-fMRI simultâneo em indivíduos assintomáticos, pacientes com epilepsia e com estenose carotídea. **Tese (Doutorado em Física Aplicada a Medicina e Biologia) - Faculdade de Filosofia, Ciência e Letras de Ribeirão Preto, Universidade de São Paulo**, Ribeirão Preto, 2011

[8] Zhang Xiang Zhang, Lina Yao, Xianzhi Wang, Jessica Monagthan, David Mcalpine e Yu Zang. A survey on deep learning based brain computer interface: Recent advances and new frontiers. **airXiv prepint arXiv:1905.04149**, 2019

[9] Silveira, T. L. T. Classificação de estágios de sono atráves da aplicação de transformada WAVELET discreta sobre um único canal de eletroencefalograma. **Tese (Mestre em Ciência da Computação) - Universidade Federal de Santa Maria, Rio Grande do Sul**, Santa Maria, 2016

[10] Silveira, T; Kozakevicius, A; Rodrigues, R. C. Detecção de sonolência através de aproximação não-linear associada às transformadas wavelet e de Fourier. **(Grupo de Microeletrônica) - Universidade Federal de Santa Maria, Rio Grande do Sul**, Santa Maria, 2014

[11] ARLOT, Sylvain; CELISSE, Alain. *A Survey of Cross-Validation Procedures for Model Selection.* **Statistics Surveys**. Vol. 4, 2010, pp. 40–79.

[12] GARETH James, Daniela Witten, Trevor Hastie, Robert Tibshirani *An Introduction to Statistical Learning: with Applications in R.* **Springer, New York.** 2014.

[13] Snodgrass e Vanderwart(1980) *Joan G Snodgrass e Mary Vanderwart. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity.* **Journal of experimental psychology: Human learning and memory, 6(2):174.** 2014.

[14] **https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/29472-impulsionado-pelas-mulheres-consumo-de-alcool-cresce-entre-brasileiros-em-2019** *Acesso: 09/03/2021*

[15] HASTIE, Trevor , Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* **Springer** , 2 ed. New York, 2009.

[16] JAKE VanderPlas *Python Data Science Handbook: Essential Tools for Working with Data.* 1ª Edição, eBook Kindle, 2017.

[17] JOHNSON, Richard Arnold, Dean W. Wichern. *Applied Multivariate Statistical Analysis.* **Pearson Prentice Hall.** 6 ed, 2007.

[18] SINGER, Julio da Motta; MORETTIN, Pedro Alberto. Introdução à Ciência de Dados: *Fundamentos e Aplicações*. Versão parcial e preliminar. **Instituto de Matemática e Estatística (IME)**, São Paulo, Maio de 2020.

[19] Audun Eltvik. *Deep Learning for the Classification of EEG Time-Frequency Representations* **Norwegian University of Science and Technology, Department of Engineering Cybernetics**, Noruega, 2018.

[20] TIBISHIRANI, R., James, G., Witten D., Hastie, T. *An Introduction to Statistical Learning: with Applications in R.* **Springer**, 8 ed. New York, 2013.

[21] Saeid, S., Chambers, J. A. *EEG SIGNAL PROCESSING.* **Candif University, UK**, 2007.

[22] Morris, S. J. *Functional Regression.* **The University of Texas, MD Anderson Cancer Center**, 2014.

[23] Bande, M. F.; Fuente, M. O. *Statistical Computing in Functional Data Analysis: The R Package fda.usc* **Journal of Statistical Software**, 2012.

[24] Hastie T.; Mallows C. *Discussion of: A statistical view of some chemometrics regression tools. Techno., 35(2):140-143.*, 1993.

[25] Marx B. D. ; Eilers P. H. C. *Generalized linear regression on sampled signals and curves: A P-spline approach. Techno., 41(1):1-13.*, 1999.

[26] Carroll R. J.; Ruppert D.; Stefanski L. A. *Measurement Error in Nonlinear.* **Models: A Modern Perspective, New York: Springer-Verlag.**, 1995.

[27] AKAIKE, Hirotugo. *A Bayesian Analysis of The Minimum AIC Procedure.* **Annals of the Institute of Statistical Mathematics**. Vol.30, 1978, pp. 9–14.

[28] ARLOT, Sylvain; CELISSE, Alain. *A Survey of Cross-Validation Procedures for Model Selection.* **Statistics Surveys**. Vol. 4, 2010, pp. 40–79.