



## DESENVOLVIMENTO DE PLATAFORMA PARA ANÁLISE ESTATÍSTICA DE SINAIS ELETROENCEFALOGRÁFICOS

Izabella Nonato Oliveira Lima\*<sup>1</sup>, Camila Davi Ramos<sup>1</sup>, Aurélia A. A. Rodrigues<sup>2</sup> e João-Batista Destro-Filho<sup>1</sup>

<sup>1</sup>FEELT – Universidade Federal de Uberlândia

<sup>2</sup>FAMAT – Universidade Federal de Uberlândia

**Resumo** - A análise estatística de fenômenos biológicos pode auxiliar na identificação de padrões de comportamento de um sinal biomédico, por exemplo. Este trabalho possui o objetivo de desenvolver um software de análise estatística descritiva do sinal eletroencefalográfico por meio de 14 índices que vão desde de média e mediana até valores de curtose e assimetria. O software foi testado com registros de EEG reais de indivíduos normais e pacientes comatosos, levando a resultados significativos em termos de velocidade e praticidade de uso, bem como fornecendo resultados tão precisos quanto outras plataformas de análise estatística já consagradas, que são complexos ao serem usados com dados de EEG.

**Palavras-Chave** – Bioestatística, eletroencefalografia, software.

### DEVELOPMENT OF A SOFTWARE FOR STATISTICAL ANALYSIS OF ELECTROENCEPHALOGRAPHIC (EEG) SIGNS

**Abstract** - Statistical analysis of biological phenomena may help in the identification of patterns of behavior of a biomedical signal, for example. The objective of this work is to develop software for the statistical analysis of the electroencephalographic signal by means of 14 indexes ranging from average and median to values of kurtosis and asymmetry. The software was tested with real EEG records from normal individuals and comatose patients, leading to significant results in terms of speed and practicality of use, as well as providing results as accurate as other established statistical analysis platforms, which are complex when used with EEG data.

**Keywords** - Biostatistics, electroencephalography, software.

#### I. INTRODUÇÃO

A ciência estatística tem como objetivo, observar, coletar, classificar, resumir, organizar, analisar e interpretar um

\*izabella.nonato.lima@outlook.com

conjunto de dados, tendo sua importância no momento de delimitar incertezas advindas dos resultados obtidos do processamento dos dados, podendo ser utilizada, também, na réplica de uma dada pesquisa em mesmas condições, obtendo desde modo, os mesmos resultados em todas as repetições executáveis [1]. A estatística possui diversas ramificações para as mais variadas aplicações, sendo que uma delas é denominada bioestatística, responsável pela análise de dados biológicos, como por exemplo, na medicina a bioestatística auxilia na escolha das situações experimentais e na determinação da quantidade de indivíduos a serem examinados [2].

A bioestatística pode ser ainda, empregada na análise quantitativa de registros de sinal de eletroencefalografia (EEG). Existem uma gama de ritmos de EEG e estes dependem do estado mental do sujeito, como o grau de atenção, despertar e sono. Os ritmos são convencionalmente caracterizados por sua faixa de frequência (atividade neural) e amplitude relativa (interação neural) [3], sendo classificados em cinco bandas de frequências, conhecidas na neurologia como: Delta (<4Hz), Teta (4-7Hz), Alfa (8-13Hz), Beta (14-30Hz) e Gama (>30Hz) [4]. Os ritmos Delta e Teta são constatados no estado de sono profundo e durante a sonolência, respectivamente. Indivíduos normais que estão relaxados, a presença do ritmo Alfa é bastante proeminente, o ritmo Beta normalmente é observado em alguns estágios do sono devido a suas características (baixa amplitude e alta frequência) e o ritmo Gama possui relação a um estado de processamento ativo [5]. Na área clínica o sinal eletroencefalográfico normalmente é registrado utilizando o sistema internacional de posição de eletrodos 10-20 e a aquisição desde sinal é geralmente realizada a uma frequência de no mínimo 200 Hz, para evitar o efeito de aliasing no sinal captado [6].

A análise estatística de sinais EEG é aplicado em diversos estudos que tem como objetivo fornecer informações desde a presença de pequenos artefatos no sinal de EEG captado [7], quanto para avaliação de modo quantitativo da mudança de topografia de campo ao longo do tempo em sinais de EEG multicanal [8] e até mesmo no desenvolvimento de softwares para análise e distinção minuciosa das oscilações que são associadas normalmente com ondas de EEG [9]. Em relação aos softwares de análise estatística, existem aqueles que

fornece diversas ferramentas que complementam a coleta e processamento de dados de EEG, auxiliando no enquadramento e execução de análises estatísticas em dados de múltiplas fontes, como por exemplo, o software *EEGLAB STUDY* [10]. Já o software de eletroencefalografia eletromagnética (EMEGS) é utilizado pré-processamento e visualização de dados eletromagnéticos [11], e o aplicativo *Brainstorm*, tem a finalidade de visualizar e processar dados de MEG, com ênfase em técnicas de estimativa de fonte cortical e integração com dados de ressonância magnética (fMRI) anatômica [12].

Esse trabalho teve como objetivo, levando em consideração a importância da análise do sinal EEG, a construção de um software específico de análise estatística desses sinais, que utilizando três quantificadores de registros de EEG de diferentes níveis de consciência, possibilita por meio da estatística descritiva e testes de hipóteses auxiliar no processamento geral de sinais EEG.

## II. METODOLOGIA

### A. Coleta de Dados

Nesse trabalho foram utilizados dados provenientes de exames de EEG, de população comatosa ou de grupo controle, com três quantificadores previamente calculados [13]. Os quantificadores são baseados em informações diferentes e tratam-se de: A) Porcentagem de Contribuição de Potência (PCP), que leva em consideração os valores de potência relativa dos ritmos cerebrais, a partir de espectros de frequência; B) Frequência Mediana (FM), que mostra a frequência central que se concentra o espectro de potências conforme cada ritmo; e C) Coerência, que calcula o grau de simetria entre os hemisférios cerebrais direito e esquerdo. Cada um desses quantificadores é analisado de forma independente e fornece informações distintas do sinal avaliados, sendo assim, estatisticamente, para cada registro de EEG avaliados são analisadas ao menos três informações distintas.

Para que o processamento estatístico de sinais de EEG possa ocorrer é necessário que estes possuam as seguintes características: apresentar 20 eletrodos, conforme sistema internacional de colocação de eletrodos 10-20, ser segmentado em N trechos (sendo N o número de janelas definidas anteriormente) e ter calculados os quantificadores PCP, FM e Coerência.

### B. Definição de ferramentas

A fim de analisar estatisticamente o sinal EEG, o software desenvolvido permite que o usuário realize o cálculo descritivo. Assim, para a análise de dados quantitativos foram escolhidas 14 ferramentas descritivas que podem caracterizar determinada população. Tais ferramentas foram listadas de I) à XIV). O conjunto de dados avaliados é denominado como Amostra.

I) *Tamanho da Amostra*: É a quantidade de épocas ou trechos que aquele arquivo em análise possui.

II) *Mínimo*: Refere-se ao menor elemento da amostra.

III) *Máximo*: A medida de máximo apresenta o maior valor da amostra.

IV) *Amplitude Total*: É a medida definida como sendo a diferença entre o maior e o menor valor do conjunto de dados.

V) *Mediana*: Valor que separa a metade maior e a metade menor de uma amostra, uma população ou uma distribuição de probabilidade.

VI) *Primeiro Quartil*: É a medida que deixa 25% das observações abaixo desse valor e 75% acima.

VII) *Terceiro Quartil*: É a medida que deixa 75% das observações abaixo desse valor e 25% acima.

VIII) *Média*: Medida que mostra a concentração dos dados de uma distribuição como o ponto de equilíbrio, podendo ser interpretada como um valor significativo de uma lista de números.

IX) *Variância*: Medida de dispersão estatística que indica o “quão longe” em geral os valores encontrados estão do valor esperado.

X) *Desvio Padrão*: Medida de dispersão em torno da média populacional ou da média amostral de uma variável aleatória.

XI) *Desvio Padrão da Mediana*: Medida de dispersão em torno da mediana populacional ou amostral de uma variável aleatória.

XII) *Coefficiente de Variação*: Medida padronizada de distribuição de probabilidade ou de distribuição de frequências, frequentemente expresso em porcentagem (%).

XIII) *Assimetria*: Medida que caracteriza como e quanto a distribuição de dados ou frequências se afasta da condição de assimetria. A faixa de valores para os resultados desse cálculo varia de zero a um.

XIV) *Curtose*: Medida que caracteriza a forma da distribuição da amostra quanto ao seu achatamento, tendo achatamentos como platicúrticas (centro planar), mesocúrticas (distribuição normal) e leptocúrticas (centro pontudo). A faixa obtida para este cálculo varia entre 0 a 3.

### C. Validação dos dados

Para validar o desempenho da plataforma proposta no presente trabalho, foram escolhidos dois softwares estatísticos, já consagrados no uso acadêmico, aqui denominados de A e B para que, os dados analisados na plataforma pudessem também ser analisados nesses dois softwares e posteriormente fosse calculado o erro padrão conforme (1). Nota-se por meio de (1) que quanto menor o resultado dessa, mais parecido são os dados comparados.

$$E = \frac{|Ind_p - Ind_s|}{MAX(Ind_p, Ind_s)} \quad (1)$$

Onde:

$Ind_p$  - Representa o valor do índice retornado por meio do cálculo feito da plataforma desenvolvida no presente trabalho;

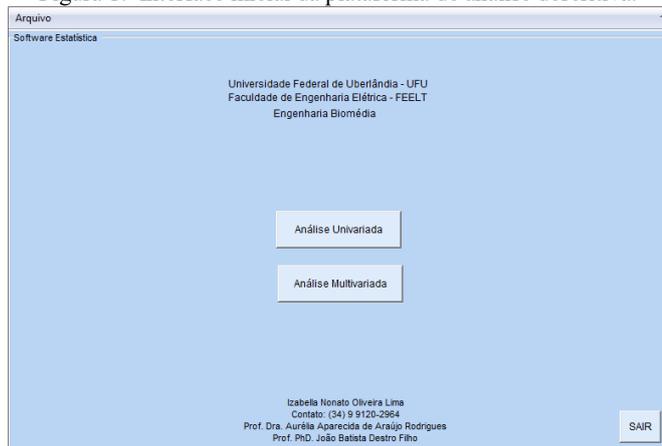
$Ind_s$  - Se refere ao valor do índice retornado pelo software de comparação (podendo ser software A ou software B).

## III. RESULTADOS E DISCUSSÃO

A análise descritiva dos dados proposta nesse trabalho envolveu cálculos de pelo menos 14 índices estatísticos e para

obter tais quantificadores foram elaboradas duas interfaces visando facilitar a utilização pelos usuários. A Figura 1 representa a interface inicial da ferramenta, composta por dois botões indicando Análise Univariada e Análise Multivariada. Para realizar a estatística descritiva basta que o usuário clique no botão “Análise Univariada” para dar sequência ao processamento e assim irá encontrar a interface mostrada na Figura 2. É importante destacar que, ainda na Figura 1, existe o botão “Sair” que pode ser clicado caso o usuário deseje finalizar o procedimento e é composta também por um direcionamento denominado “Arquivo”, localizado no canto superior esquerdo, em que o usuário pode clicar para carregar o manual do usuário.

Figura 1: Interface inicial da plataforma de análise descritiva.



Na Figura 2, interface do cálculo descritivo, existem algumas informações que devem ser preenchidas com bastante atenção para que erros não sejam cometidos ao longo do processamento dos dados. Inicialmente o usuário deverá definir qual grupo deseja fazer o cálculo descritivo: Normal ou Coma. Como descrito na seção 2, o grupo coma se refere aos registros de EEG advindos de pacientes comatosos, registrados na UTI e, portanto, com características especiais do sinal. Já o grupo conhecido como “Normal”, é o conjunto de registros EEG advindo de voluntários saudáveis, sem patologias neurológicas prévias e registrados no setor de neurologia, com condições também específicas. Posterior à escolha do grupo, o usuário deverá informar qual quantificador será avaliado. Existem por enquanto três quantificadores, PCP, FM e Coerência, sendo que cada um apresenta um tipo de informação específica. Em sequência o usuário precisa digitar o nome do arquivo de interesse com o as informações dos registros que serão processados. Esse arquivo deverá conter em forma de vetor coluna os nomes dos dados ou sinais a serem analisados estatisticamente. Ao final, quando clicar no botão “Processar” os cálculos estatísticos serão realizados da seguinte forma:

### I. Caso seja selecionado apenas um sinal

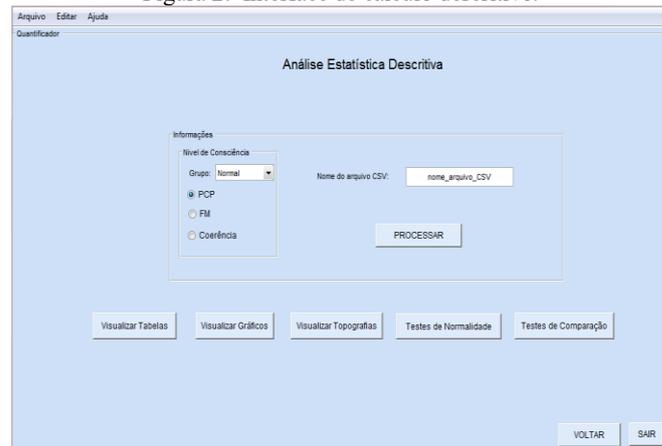
Se o usuário analisar apenas um sinal, a análise descritiva será feita baseada nas informações contidas apenas nesse sinal. Por exemplo, caso esse sinal apresente cinco valores distintos de PCP, os cálculos descritivos serão feitos a partir desses cinco valores. No final desse processamento será criado

um novo arquivo com todas as informações descritivas calculadas.

### II. Caso sejam selecionados dois ou mais sinais

Por outro lado, se o usuário desejar selecionar mais de um sinal para o processamento, os cálculos descritivos serão feitos de maneira individual para cada registro selecionado, conforme disposto no item I. Além disso, também serão realizadas análises descritivas levando em consideração todas as informações dos voluntários. Por exemplo: Se o usuário seleciona três sinais, cada qual com dez informações de FM, a análise descritiva será baseada nos 30 valores de FM.

Figura 2: Interface de cálculo descritivo.



Para facilitar a visualização dos resultados, o usuário tem a possibilidade de verificar em formato de tabelas as informações obtidas clicando no botão “Visualizar Tabelas”. A Figura 3, tal a interface para disponibilização dos dados em tabelas. Nela o usuário deverá informar o nome do arquivo (gerado no processamento da interface mostrada na Figura 2) que se deseja visualizar os dados descritivos. Em seguida informar qual grupo e ainda quantificador a ser avaliado. Ao final, deverá selecionar qual índice deseja visualizar dentre os 14 possíveis (média, mediana, desvio padrão entre outros). Ao clicar em “Processar”, caso o preenchimento dos dados estiver correto, os dados serão mostrados em formato de tabelas.

Figura 3: Interface visualização de dados em forma de tabelas.



É fato que existem na atualidade vários softwares, gratuitos ou pagos, que fazem análise estatística de maneira confiável e eficaz. Contudo, para análise dos dados [13] a utilização de tais softwares mostrou-se ineficiente uma vez que esses sinais

apresentaram número de informações distintas elevado e bastante particularidades. Sendo assim optou-se por produzir essa ferramenta, dedicada inteiramente para análise desses dados. Contudo, a validação dessa plataforma foi necessária e realizada de três maneiras distintas, explanadas nos itens a) à c).

a) Comparação dos resultados obtidos a partir de um sinal EEG para o quantificador PCP, utilizando dois softwares já consagrados na academia, aqui denominados A e B. É importante destacar que o conjunto de amostras utilizado na análise descritiva foi escolhido a partir dos resultados de PCP para um eletrodo e um total de dez valores de PCP. A Tabela 1 contém os resultados obtidos para cada índice descritivo nos três softwares comparados e o valor do erro absoluto.

Tabela 1: Comparação do desempenho da plataforma referente ao teste a).

Índice	SFT	A	B	Erro 1	Erro 2
Tamanho da Amostra	10	10	10	0,0000	0,0000
Mínimo	41,16	41,16	41,16	0,0000	0,0000
Máximo	79,61	79,61	79,61	0,0000	0,0000
Amplitude	38,45	38,45	38,45	0,0000	0,0000
Mediana	67,51	67,51	67,51	0,0000	0,0000
Primeiro Quartil	56,22	54,70	57,47	0,0270	0,0217
Terceiro Quartil	76,77	77,05	76,41	0,0037	0,0046
Média	65,34	65,34	65,34	0,0000	0,0000
Variância	169,37	169,37	169,37	0,0000	0,0000
Desvio Padrão	13,01	13,01	13,01	0,0000	0,0000
Desvio Padrão da Mediana	13,21	13,21	-	0,0000	-
Coefficiente de Variação	19,92	19,92	19,92	0,0000	0,0001
Assimetria	-0,59	-0,69	-0,69	0,1859	0,1858
Curtose	2,14	-0,56	-0,56	1,2637	1,2637

Erro 1: Comparação entre plataforma e software A. Erro 2: Comparação entre plataforma e software B.

b) Comparação dos resultados obtidos a partir de três sinais EEG para o quantificador FM, utilizando os dois softwares A e B mencionados na situação a). É importante destacar que o conjunto de amostras utilizado na análise descritiva foi escolhido a partir dos resultados de FM para um eletrodo e um total de trinta valores de FM. A Tabela 2 contém os resultados obtidos para cada índice descritivo nos três softwares comparados e o valor do erro absoluto.

Tabela 2: Comparação do desempenho da plataforma referente ao teste b).

Índice	SFT	A	B	Erro 1	Erro 2
Tamanho da Amostra	30	30	30	0,0000	0,0000
Mínimo	1,50	1,50	1,50	0,0000	0,0000
Máximo	2,51	2,51	2,51	0,0000	0,0000
Amplitude	1,02	1,02	1,02	0,0000	0,0000
Mediana	1,89	1,89	1,89	0,0000	0,0000

Primeiro Quartil	1,74	1,72	1,75	0,0063	0,0062
Terceiro Quartil	1,99	2,02	1,98	0,0163	0,0017
Média	1,91	1,91	1,91	0,0000	0,0000
Variância	0,06	0,06	0,06	0,0000	0,0003
Desvio Padrão	0,25	0,25	0,25	0,0000	0,0001
Desvio Padrão da Mediana	0,25	0,25	-	0,0077	-
Coefficiente de Variação	12,97	12,97	12,97	0,0000	0,0003
Assimetria	0,67	0,71	0,71	0,0507	0,0508
Curtose	2,95	0,17	0,17	0,9422	0,9422

Erro 1: Comparação entre plataforma e software A. Erro 2: Comparação entre plataforma e software B.

c) Comparação dos resultados obtidos a partir de cinco sinais EEG para o quantificador Coerência, utilizando os dois softwares A e B mencionados na situação a). É importante destacar que o conjunto de amostras utilizado na análise descritiva foi escolhido a partir dos resultados de Coerência para um par de eletrodo em toda faixa de frequência de 1 a 100 Hz. A Tabela 3 contém os resultados obtidos para cada índice descritivo nos três softwares comparados e o valor do erro absoluto.

Tabela 3: Comparação do desempenho da plataforma referente ao teste c).

Índice	SFT	A	B	Erro 1	Erro 2
Tamanho da Amostra	100	100	100	0,0000	0,0000
Mínimo	0,03	0,03	0,03	0,0000	0,0013
Máximo	0,99	0,99	0,99	0,0000	0,0000
Amplitude	0,96	0,96	0,96	0,0000	0,0000
Mediana	0,72	0,72	0,72	0,0000	0,0000
Primeiro Quartil	0,49	0,49	0,49	0,0056	0,0057
Terceiro Quartil	0,85	0,84	0,84	0,0020	0,0020
Média	0,65	0,65	0,65	0,0000	0,0000
Variância	0,07	0,07	0,07	0,0000	0,0007
Desvio Padrão	0,26	0,26	0,26	0,0000	0,0001
Desvio Padrão da Mediana	0,27	0,27	-	0,0000	-
Coefficiente de Variação	39,87	39,87	39,87	0,0000	0,0001
Assimetria	-0,75	-0,76	-0,76	-0,0153	-0,0152
Curtose	2,61	-0,34	-0,34	1,1319	1,1319

Erro 1: Comparação entre plataforma e software A. Erro 2: Comparação entre plataforma e software B.

A observação da Tabela 1, bem como Tabelas 2 e 3, permite concluir que o desempenho da plataforma desenvolvida nesse trabalho é similar aos resultados obtidos em software já consagrados no estudo estatístico de dados científicos uma vez que os valores obtidos de erros são consideravelmente baixos, próximos de zero, com exceção dos índices de assimetria e curtose. Nesses dois últimos casos, deve-se considerar que os softwares A e B utilizam a faixa de limiar de 0 à 3, enquanto que o limiar da plataforma onde o software foi desenvolvido, varia na faixa de 0 à 1, gerando, portanto, a diferença de

valores constatada, contudo a interpretação dos resultados pode ser realizada do mesmo modo para os três softwares.

#### IV. CONCLUSÃO

A análise estatística de dados biomédicos é essencial na geração de conclusões e tomadas de decisão principalmente no que tange à aplicação clínica. Diante disso, a criação de uma plataforma de análise de sinais EEG para cálculos estatísticos se fez necessário uma vez que a utilização de softwares já existentes foi ineficaz e consideravelmente trabalhosa, resultando em elevada perda de tempo. Logo, por meio de interfaces, como se observa nas Figuras 1 – 3, foi possível projetar uma plataforma confiável que utilizasse quantificadores pré-calculados de sinais EEG e por meio de 14 índices estatísticos, resumissem os dados de maneira eficaz e rápida.

Pretende-se como trabalhos futuros, implementar uma plataforma de cálculo de quantificadores e agrupar com essa plataforma, bem como desenvolver uma parte relacionada à análise multivariada, com cálculos de componentes principais e agrupamentos por clusters para tornar essa plataforma inteiramente dedicada ao processamento de sinais EEG especialmente gerados no setor de neurologia do Hospital de Clínicas da Universidade Federal de Uberlândia.

#### AGRADECIMENTOS

Os autores agradecem a equipe de neurologia do HCU-UFU, bem como alunos de iniciação científica e professores que auxiliaram no desenvolvimento e testes da plataforma. Agradecem ainda aos órgãos de fomento pelo auxílio financeiro e estrutural da pesquisa: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ); e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

#### REFERÊNCIAS

- [1] M. J. Fontelles, *BIOESTATÍSTICA Aplicada à Pesquisa Experimental: volume 1*. 2012.
- [2] L. G. Zaros and H. R. De Medeiros, *Bioestatística*. 2. ed. Natal: EDUFRN, 2011.
- [3] S. Sanei and J. A. Chambers, *EEG Signal Processing*. 1<sup>a</sup>. ed. Universidade de Cardiff, Reino Unido: John Wiley & Sons, 2007.
- [4] W. J. Freeman and R. Q. Quiroga, *Imaging brain function with EE: Advanced temporal and spatial analysis of electroencephalographic signals*. 2013.
- [5] D. L. Schomer and F. Lopes da Silva, *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. 2012.
- [6] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. 1<sup>a</sup>. ed. Estados Unidos: Elsevier, 2005.
- [7] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.
- [8] D. Brunet, M. M. Murray, and C. M. Michel, "Spatiotemporal analysis of multichannel EEG: CARTOOL," *Comput. Intell. Neurosci.*, vol. 2011, 2011.
- [9] Z. J. Waldman, S. Shimamoto, I. Song, I. Orosz, A. Bragin, I. Fried, J. Engel, R. Staba, M. R. Sperling, and S. A. Weiss, "A method for the topographical identification and quantification of high frequency oscillations in intracranial electroencephalography recordings," *Clin. Neurophysiol.*, vol. 129, no. 1, pp. 308–318, 2018.
- [10] A. Delorme, T. Mullen, C. Kothe, Z. Akalin Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig, "EEGLAB, SIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG processing," *Comput. Intell. Neurosci.*, vol. 2011, 2011.
- [11] P. Peyk, A. De Cesare, and M. Junghöfer, "Electromagnetic encephalography software: Overview and integration with other EEG/MEG toolboxes," *Comput. Intell. Neurosci.*, vol. 2011, 2011.
- [12] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: A user-friendly application for MEG/EEG analysis," *Comput. Intell. Neurosci.*, vol. 2011, 2011.
- [13] C. D. Ramos, "Caracterização do eletroencefalograma normal em situação de vigília: elaboração da base de dados e análise quantitativa." 2017. 156p. Dissertação (Mestrado). Universidade Federal de Uberlândia, Uberlândia, 2017.