



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DA COMPUTAÇÃO
CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

ANTÔNIO MARCOS DA SILVA OLIVEIRA

EXTRAÇÃO DE CARACTERÍSTICAS DE SINAIS DE
ELETROENCEFALOGRAMA USANDO MATRIZ DE COVARIÂNCIA PARA
IDENTIFICAÇÃO DE CRISES EPILÉPTICAS

FORTALEZA

2021

ANTÔNIO MARCOS DA SILVA OLIVEIRA

EXTRAÇÃO DE CARACTERÍSTICAS DE SINAIS DE ELETROENCEFALOGRAMA
USANDO MATRIZ DE COVARIÂNCIA PARA IDENTIFICAÇÃO DE CRISES
EPILEPTICAS

Monografia apresentada ao Curso de Engenharia de Computação do Departamento de Engenharia da Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Guilherme Barreto.

FORTALEZA

2021

Página reservada para ficha catalográfica.

Utilize a ferramenta *online* [Catalog!](http://www.fichacatalografica.ufc.br/) para elaborar a ficha catalográfica de seu trabalho acadêmico, gerando-a em arquivo PDF, disponível para download e/ou impressão.

(<http://www.fichacatalografica.ufc.br/>)

ANTÔNIO MARCOS DA SILVA OLIVEIRA

EXTRAÇÃO DE CARACTERÍSTICAS DE SINAIS DE ELETROENCEFALOGRAMA
USANDO MATRIZ DE COVARIÂNCIA PARA IDENTIFICAÇÃO DE CRISES
EPILÉPTICAS

Monografia apresentada ao Curso de Engenharia de Computação do Departamento de Engenharia da Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Guilherme Barreto.

Aprovada em: ____/____/____.

BANCA EXAMINADORA

Prof. Dr. Guilherme de Alencar Barreto (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Samir Câmara Magalhães
Universidade de Fortaleza (Unifor)

Prof. Dr. José Maria Pires de Menezes Júnior
Universidade Federal do Piauí (UFPI)

A Deus.

Aos meus pais, Antônio Fraga e Maria Paula.

AGRADECIMENTOS

À Universidade Federal do Ceará, por toda a estrutura disponibilizada.

Ao Prof. Dr. Guilherme Barreto, por sua notória dedicação na formação de conhecimento e pela excelente orientação.

Aos professores participantes da banca examinadora Dr. Samir Câmara Magalhães e Dr. José Maria Pires de Menezes Júnior pelo tempo, pelas valiosas colaborações e sugestões.

Aos professores de Graduação, pela valiosa contribuição de cada um na formação de conhecimento.

Aos colegas da turma de graduação, pelas reflexões, incentivos, críticas e sugestões recebidas, principalmente meu colega e amigo Moisés Caetano.

Por fim, agradeço à minha família e amigos, com quem dividi, de várias formas, as dificuldades durante o período da graduação.

“O cientista não é o homem que fornece as verdadeiras respostas; é quem faz as verdadeiras perguntas.”

Claude Lévi-Strauss

RESUMO

A epilepsia é uma condição física anômala que ocorre no cérebro e afeta o sistema nervoso central provocando o envio de impulsos elétricos em excesso que podem causar eventos chamados de crises epiléticas. Esta doença pode ser diagnosticada através de exames de eletroencefalograma (EEG de escalpo) onde eletrodos são colocados no couro cabeludo para capturar e gravar os sinais elétricos do cérebro. Utilizou-se de um conjunto de dados de EEG de uso público para treinar e testar um classificador linear de mínimos quadrados para classificar épocas de dois segundos de duração em convulsivas (ictais) ou não convulsivas (não-ictais). Devido ao alto desbalanceamento das classes, realizou-se a seleção de amostras ictais/não-ictais na proporção 1:1 ou 1:3. Para este fim, sendo as amostras ictais localizadas em posições fixas, foram levantadas sete hipóteses sobre a seleção de amostras não-ictais a partir da posição das amostras ictais. A matriz de covariância foi aplicada diretamente aos dados para a formação dos vetores de características. Após experimentos realizados com seis pacientes, retirados do conjunto de dados público mencionado, concluiu-se que o classificador apresentou melhoras na classificação quando as amostras não-ictais são obtidas de uma localização distante das amostras ictais (hipóteses quatro a sete).

Palavras-chave: Epilepsia. Eletroencefalograma. Extração de características. Classificador. Matriz de covariância. Vetores de características.

ABSTRACT

Epilepsy is an anomalous physical condition that occurs in the brain and affects the central nervous system causing excess electrical impulses that can cause events called epileptic seizures. This disease can be diagnosed through electroencephalogram (scalp EEG) exams where electrodes are placed on the scalp to capture and record electrical signals from the brain. A public use EEG database was used to train and test the least squares linear classifier to classify periods of two seconds in duration as convulsive (ictal) or non-convulsive (non-ictal). Due to the high imbalance of the classes, the selection of ictal / non-ictal samples in the proportion 1: 1 or 1: 3 was carried out. For this purpose, since the ictal samples were located in fixed positions, seven hypotheses were raised about the selection of non-ictal samples from the position of the ictal samples. The covariance matrix was applied directly to the data for the formation of the characteristic vectors. After experiments carried out with six patients, taken from the mentioned public data set, it was concluded that the classifier showed improvements in the classification when the non-ictal samples are obtained from a distant location from the ictal samples (hypotheses four to seven).

Keywords: Epilepsy. Electroencephalogram. Extraction of characteristics. Classifier. Covariance matrix. Feature vectors.

LISTA DE FIGURAS

Figura 1 – Exemplo de EEG.....	22
Figura 2 – Disposição física dos eletrodos segundo o sistema internacional 10-20	23
Figura 3 – Etapas do processamento do EEG de um paciente	25
Figura 4 – Modelagem do experimento	33
Figura 5 – Matrizes X e Y após Extração dos vetores de características	37
Figura 6 – Hipóteses mostrando a proporção ictal/não-ictal	40

LISTA DE GRÁFICOS

Gráfico 1 – Resultados do paciente chb01	47
Gráfico 2 – Resultados do paciente chb03	50
Gráfico 3 – Resultados do paciente chb05	53
Gráfico 4 – Resultados do paciente chb07	55
Gráfico 5 – Resultados do paciente chb14	56
Gráfico 6 – Resultados do paciente chb18	57

LISTA DE TABELAS

Tabela 1 – Proporção de amostras não-ictal/ictal das hipóteses	36
Tabela 2 – Pacientes selecionados para teste das hipóteses	44
Tabela 3 – Arquivos do paciente chb01 contendo crises	45
Tabela 4 – Resultados para as hipóteses 1 e 2 do paciente chb01	45
Tabela 5 – Resultados para as hipóteses 3 e 4 do paciente chb01	46
Tabela 6 – Resultados para as hipóteses 5 e 6 do paciente chb01	46
Tabela 7 – Resultados para a hipótese 7 do paciente chb01	47
Tabela 8 – Resumo dos resultados do paciente chb01	48
Tabela 9 – Arquivos do paciente chb03 contendo crises	48
Tabela 10 – Resultados para as hipóteses 1 e 2 do paciente chb03	49
Tabela 11 – Resultados para as hipóteses 3 e 4 do paciente chb03	49
Tabela 12 – Resultados para as hipóteses 5 e 6 do paciente chb03	49
Tabela 13 – Resultados para a hipótese 7 do paciente chb03	50
Tabela 14 – Resumo dos resultados do paciente chb03	51
Tabela 15 – Arquivos do paciente chb05 contendo crises	51
Tabela 16 – Resultados para as hipóteses 1 e 2 do paciente ch05	52
Tabela 17 – Resultados para as hipóteses 3 e 4 do paciente ch05	52
Tabela 18 – Resultados para as hipóteses 5 e 6 do paciente ch05	52
Tabela 19 – Resultados para a hipótese 7 do paciente ch05	53
Tabela 20 – Resumo dos resultados do paciente chb05	53
Tabela 21 – Arquivos do paciente chb07 contendo crises	54
Tabela 22 – Resumo dos resultados do paciente chb07	55
Tabela 23 – Arquivos do paciente chb14 contendo crises	56
Tabela 24 – Resumo dos resultados do paciente chb14	56

Tabela 25 – Arquivos do paciente chb18 contendo crises	57
Tabela 26 – Resumo dos resultados do paciente chb18	58

LISTA DE ABREVIATURAS E SIGLAS

CHB	Children's Hospital Boston
EDF	Formato Europeu de Dados (European Data Format)
EEG	Eletroencefalograma
MIT	Massachusetts Institute of Technology
OLS	Ordinary Least Squares

LISTA DE SÍMBOLOS

λ	Autovalor
ρ	Coeficiente de correlação
σ	Covariância
\hat{R}_x	Matriz de correlação
Σ	Matriz de covariância

SUMÁRIO

1 INTRODUÇÃO.....	17
1.1 Objetivo geral	17
1.2 Objetivos específicos.....	18
1.3 Organização do restante do trabalho	18
2 EPILEPSIA E ELETROENCEFALOGRAMA.....	20
2.1 Definições básicas sobre epilepsia e EEG.....	20
2.2 Fundamentos do EEG de Escalpo	21
2.3 Resumo do capítulo	24
3 EXTRAÇÃO DE CARACTERÍSTICAS.....	25
3.1 Extração de características.....	25
3.2 Matriz de covariância como extrator de características.....	26
3.2.1 Definição de covariância	26
3.2.2 Definição de matriz de covariância	26
3.2.3 Estimação da matriz de covariância	28
3.2.3.1 <i>Estimação da matriz de covariância no Matlab/Octave</i>	29
3.3 Montagem dos vetores de características	30
3.4 Resumo do capítulo	32
4 MATERIAIS E MÉTODOS	33
4.1 Conjunto de dados	33
4.1.1 Subconjunto de dados.....	34
4.2 Leitura dos arquivos	35
4.3 Balanceamento das classes	35
4.4 Construção dos Vetores de Características	36
4.5 Hipóteses desenvolvidas para seleção de amostras de treinamento	39
4.5.1 Hipótese 1	40
4.5.2 Hipótese 2.....	40
4.5.3 Hipótese 3	41
4.5.4 Hipótese 4.....	41
4.5.5 Hipótese 5	41
4.5.6 Hipótese 6.....	41
4.5.7 Hipótese 7.....	42
4.6 Classificador para avaliação das hipóteses.....	42
4.7 Resumo do capítulo	43
5 RESULTADOS.....	44
5.1 Paciente chb01	44

5.2 Paciente chb03	48
5.3 Paciente chb05	51
5.4 Pacientes chb07, chb14 e chb18	54
5.4.1 Paciente chb07	54
5.4.2 Paciente chb14	55
5.4.3 Paciente chb18	57
5.7 Resumo do capítulo	58
6 CONCLUSÕES E TRABALHOS FUTUROS	59
6.1 Conclusões.....	59
6.2 Trabalhos Futuros	59
REFERÊNCIAS	61
APÊNDICE A – RESULTADOS POR PACIENTE	62

1 INTRODUÇÃO

A epilepsia é uma condição física que ocorre no cérebro e afeta o sistema nervoso do indivíduo provocando o envio, pelas células nervosas do cérebro, de impulsos elétricos em excesso que causam eventos chamados de convulsões. Estas crises epiléticas podem ser focais (apenas um segmento do cérebro é afetado) ou generalizadas (todo o cérebro é afetado) e podem durar desde alguns segundos a alguns minutos (MARCUSE, FIELDS and YOO, 2016).

Esta patologia pode ser diagnosticada através de exames de Eletroencefalograma (EEG de escalpo) onde eletrodos são colocados no couro cabeludo para capturar e gravar os sinais elétricos do cérebro. No entanto, a análise dos sinais de EEG por um profissional treinado consome muito tempo e pode, eventualmente, induzir a erros de diagnóstico. Por isso, a utilização de técnicas semiautomatizadas com o intuito de auxiliar esses profissionais é bastante interessante.

Uma dessas técnicas utiliza algoritmos de aprendizado supervisionado onde, a partir de um conjunto de dados já analisados e classificados por um especialista, o algoritmo aprende em uma etapa de treinamento com esses dados e torna-se capaz de realizar a previsão de novos dados a ele submetidos posteriormente em uma etapa de teste.

No entanto, devido à complexidade da atividade cerebral, os sinais de EEG são ruidosos, não lineares e não estacionários (SHOEB, 2010). Tais características tornam desafiadora a sua classificação. Por serem sinais não lineares, é esperado que a extração de características seja realizada através de métodos não lineares. Neste trabalho, porém, utilizamos um método de extração de características linear por partes, onde selecionamos um pequeno intervalo fixo de amostras (épocas de 2 segundos ou 512 amostras) e aplicamos um método conhecido como matriz de covariância a cada época.

Uma convulsão ou crise epilética (período ictal) contida em um exame de EEG convencional com período de uma hora apresenta duração muito curta se comparada ao intervalo não convulsivo (período não ictal), mostrando o enorme desbalanceamento entre as classes. Tal desbalanceamento pode levar a um baixo desempenho nos classificadores. Para contornar esse problema, extraímos períodos não-ictais proporcionais aos períodos ictais.

1.1 Objetivo geral

Neste trabalho analisaremos se a qualidade das amostras selecionadas para a

construção do conjunto de treinamento, e consequentemente dos classificadores, pode ser influenciada pela localização dos períodos não ictais com relação aos períodos ictais. Para isso, foram extraídas amostras não-ictais localizadas em várias posições em relação às amostras ictais. Cada uma das diferentes formas de extração das amostras não-ictais será uma hipótese, que será testada com relação a qualidade de suas amostras. As diferentes hipóteses a serem testadas nesse trabalho serão explicadas posteriormente, no Capítulo 5.

1.2 Objetivos específicos

Além do objetivo geral listado anteriormente, para que seja possível testar cada uma das hipóteses, faz-se necessário implementar diversas tarefas, listadas a seguir:

- Proposição e avaliação de métodos de construção do conjunto de treinamento por meio de diferentes estratégias de seleção de amostras positivas (ictais) e negativas (não ictais) baseada na localização e na proporção das amostras de uma classe em relação à outra;
- Avaliação da Matriz de Covariância como método de extração de características para formação dos vetores de características que serão utilizados pelos classificadores;
- Implementação do classificador linear de mínimos quadrados e aplicação do mesmo nas fases de treino e teste, para que se possa tirar conclusões acerca das diversas hipóteses;
- Avaliar a metodologia proposta em um banco de dados de EEG de amplo uso na literatura especializada.

1.3 Organização do restante do trabalho

Este trabalho está organizado da seguinte forma: O Capítulo 2 tratará sobre a epilepsia, suas principais consequências e um dos principais meios utilizados para diagnosticá-la, o eletroencefalograma de escalpo (EEG).

No Capítulo 3 será apresentada a Matriz de Covariância, um recurso matemático que será utilizado para realizar a extração dos vetores de características.

No Capítulo 4 será detalhado o conjunto de dados de EEG a ser utilizado neste trabalho e a metodologia utilizada para realizar a leitura dos dados de EEG dos pacientes, o

balanceamento das classes, a montagem dos vetores de características e por fim, como o classificador linear de mínimos quadrados será utilizado para treinamento e teste a partir dos dados extraídos.

No Capítulo 5 serão apresentados os resultados obtidos após a aplicação da metodologia nos arquivos de seis pacientes distintos. Por fim, no Capítulo 6 serão discutidos os resultados e sugestões de continuidade para este trabalho.

2 EPILEPSIA E ELETROENCEFALOGRAMA

Neste capítulo abordará a epilepsia, uma doença antiga que atinge milhões de pessoas ao redor do mundo e um dos principais exames utilizados para diagnosticá-la, o EEG de escalpo.

2.1 Definições básicas sobre epilepsia e EEG

De acordo com a Organização Mundial da Saúde (*World Health Organization - WHO*), a epilepsia afeta mais de 50 milhões de pessoas em todo o mundo, sendo que cerca de 80% dessas pessoas vivem em países com baixa e média renda. O risco de morte prematura de uma pessoa com epilepsia é até três vezes maior do que o de uma pessoa normal.

A epilepsia é uma alteração ou desordem temporária e reversível do funcionamento do cérebro que pode causar crises que duram segundos ou minutos. Uma crise pode ser definida como sendo a ocorrência temporária de atividades excessivas ou anormais no cérebro conforme afirmam Marcuse, Fields and Yoo (2016, p. 121). Essas crises podem acarretar vários problemas ao indivíduo tais como desorientação, perda de consciência, lapso de memória, espasmos musculares.

As pessoas com epilepsia muitas vezes sofrem ainda com discriminação por parte daqueles que desconhecem a doença. Afinal, mesmo sendo uma doença não contagiosa, os efeitos causados durante uma crise podem assustar algumas pessoas. Por outro lado, desde que diagnosticada a tempo e com tratamento adequado, cerca de 70% das pessoas com epilepsia conseguem controlar a doença.

Nem sempre uma crise representa epilepsia. Na verdade, cerca de 10% das pessoas sofrerão pelo menos uma crise ao longo da vida. A hipoglicemia, o aumento súbito da temperatura corporal, o excesso de drogas e álcool ou mesmo alguns medicamentos em pessoas sensíveis também podem causar convulsão.

Diante dos riscos causados pela epilepsia, torna-se muito importante o seu diagnóstico precoce. Dessa forma, o indivíduo poderá receber o tratamento adequado, ter suas crises controladas e levar uma vida praticamente normal, com pouca ou nenhuma limitação.

O diagnóstico da epilepsia é clínico. Exames complementares são utilizados para confirmação do diagnóstico e pesquisa das causas. Os seguintes passos podem ser usados para a investigação e diagnóstico da epilepsia:

1. Histórico pessoal e familiar;
2. Eletroencefalograma (EEG);
3. Videoeletroencefalograma (vídeo-EEG);
4. Tomografia computadorizada;
5. Ressonância magnética.

A seguir falaremos sobre o eletroencefalograma, um dos exames complementares utilizados no diagnóstico da epilepsia.

2.2 Fundamentos do EEG de Escalpo

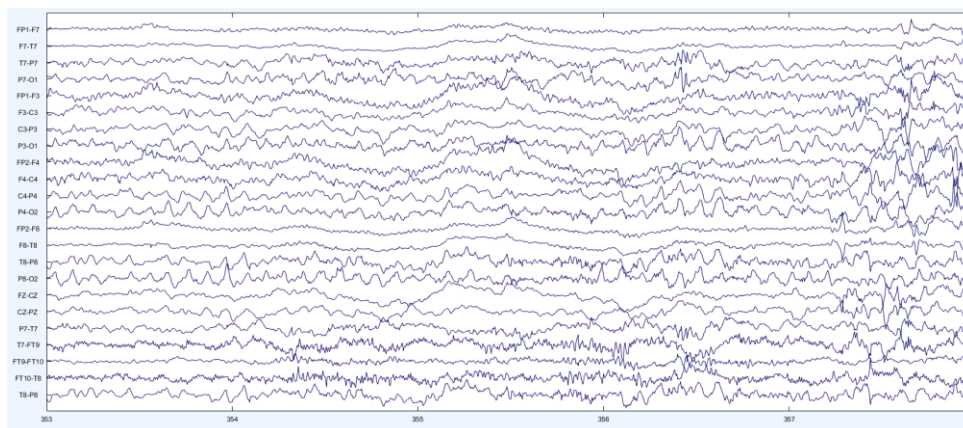
Hans Berger, um professor de psiquiatria da Universidade de Jena, Alemanha gravou em julho de 1924 o primeiro EEG humano de um garoto de 17 anos. Ele também registrou 73 EEG de escalpo em seu filho, Klaus que precisou estar com o cabelo bem curto para facilitar no processo de gravação. Por esta razão, Berger preferia realizar o EEG em pacientes carecas (MISRA and KALITA, 2018).

Em suas pesquisas Berger descobriu dois tipos de ondas cerebrais as quais denominou alfa e beta e publicou suas descobertas em 1929. Mais tarde suas descobertas foram confirmadas por outros pesquisadores que utilizaram equipamentos com amplificadores de alto ganho e alta frequência, muito mais sensíveis que os utilizados por Berger.

O EEG de escalpo, é um exame de custo relativamente baixo e não invasivo, capaz de registrar a atividade elétrica do córtex cerebral através da montagem simétrica de eletrodos no couro cabeludo do paciente (MARCUSE, FIELDS and YOO, 2016). Esses eletrodos capturam as atividades elétricas geradas no cérebro e permitem assim o registro gráfico dessas atividades. A Figura 1 mostra o EEG de um paciente plotado no aplicativo EEGLAB¹.

¹ <https://sccn.ucsd.edu/eeglab>

Figura 1 – Exemplo de EEG.



Fonte: O autor.

Como essa atividade é medida em microvolts (μV), ela deve ser amplificada por um fator de 1.000.000 para que possa ser exibida em uma tela de computador. No entanto, ocorre que potenciais extra cerebrais são também amplificados. Esses potenciais indesejados, chamados de artefatos, devem ser identificados e corrigidos, pois podem mascarar o EEG e levar a interpretações erradas.

Grande parte desses artefatos podem ser identificados, evitados ou corrigidos durante a realização do EEG pelo técnico responsável. Através de monitoramento constante do paciente e do ambiente, esse profissional consegue evitar muitos cenários causadores de artefatos.

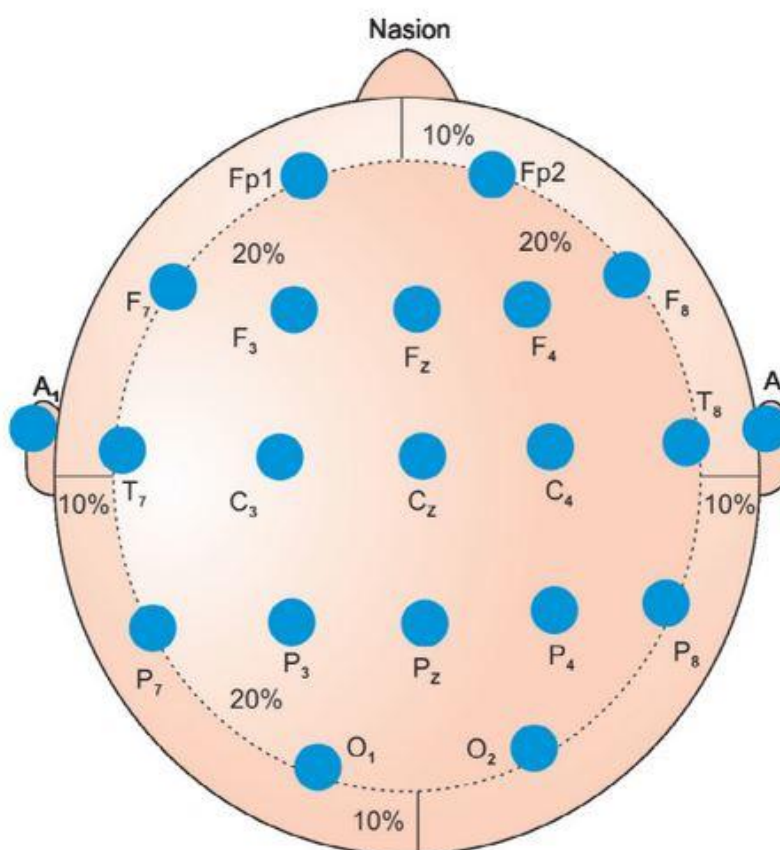
A seguir, listamos alguns artefatos que podem ser encontrados em um EEG convencional:

- **Artefatos de eletrodos:** gerados por mau contato ou mesmo soltura de algum dos eletrodos;
- **Artefatos de tremor:** gerados por um paciente com Parkinson, por exemplo;
- **Artefatos de rede:** gerados pela frequência fundamental da rede elétrica;
- **Artefatos de arrepio:** sala muito fria, por exemplo;
- **Artefatos de ranger de dentes:** geralmente ocorrem no estado de sono;
- **Artefatos musculares:** gerados por movimentos, involuntários ou não;
- **Artefatos de movimentos oculares:** gerados por vários tipos de movimentos dos olhos;
- **Artefatos de movimentos da língua:** gerados por movimentos da língua.

A montagem dos eletrodos no couro cabeludo do paciente segue uma padronização internacional garantindo uniformização no processo de obtenção de registros e que exames realizados em determinado laboratório possam ser corretamente interpretados em outro.

O padrão utilizado durante a gravação do conjunto de dados a ser usado neste trabalho foi o sistema internacional 10-20 (MISRA; KALITA, 2018). Esse padrão é assim chamado porque todos os eletrodos são colocados a distâncias compreendidas entre 10% e 20% das referências utilizadas no crânio. Posteriormente, no Capítulo 4, serão apresentados mais detalhes sobre o conjunto de dados. A Figura 2 mostra a disposição dos eletrodos utilizando esse padrão de montagem.

Figura 2 - Disposição física dos eletrodos segundo o sistema internacional 10-20.



Fonte: (MISRA; KALITA, 2018).

A posição de cada eletrodo no escalpo é designada por uma letra e um número. Os números pares correspondem ao hemisfério direito e os números ímpares correspondem ao hemisfério esquerdo.

O cérebro humano pode produzir diversos ritmos cerebrais que podem ser captados por um exame de EEG e representados por suas faixas de frequência (SANEI; CHAMBERS, 2007). Entre elas podemos citar:

- **Alfa:** atinge frequências de 8 a 13 Hz em EEG de adulto normal. É melhor observada sobre as regiões occipitais;
- **Beta:** definida na faixa de frequência de 13-30 Hz. Aumenta em amplitude durante a sonolência;
- **Teta:** Abrange a faixa de frequência de 4-8 Hz.
- **Delta:** corresponde à faixa de frequências menores que 4 Hz. Não está presente em adultos durante o estado de vigília.

Um registro de EEG conterà todas essas frequências, no entanto, terá maior ou menor predominância de um tipo já que os ritmos cerebrais variam de indivíduo para indivíduo, dependendo da idade, comportamentos como nível de atenção, sono ou vigília ou mesmo patologias associadas.

No caso da epilepsia, apesar da grande importância do EEG, o diagnóstico normalmente se baseia em fundamentos clínicos. De qualquer forma, o EEG fornece suporte ao diagnóstico, apontando evidências como, por exemplo, identificando se o tipo de epilepsia é focal ou generalizado. A correta interpretação do EEG é uma tarefa desafiadora. Mesmo um profissional experiente pode ser induzido a erros, seja pela complexidade do EEG, pela natureza repetitiva da atividade ou pelos artefatos incluídos no exame durante sua gravação.

2.3 Resumo do capítulo

Neste capítulo falou-se sobre a epilepsia e a importância do EEG de escalpo como exame auxiliar e bastante utilizado em seu diagnóstico. Foi apresentada uma breve explicação sobre com ênfase nos desafios que o EEG representa desde sua coleta, eliminação de artefatos e análise por um profissional.

No próximo capítulo será apresentada a matriz de covariância como método a ser utilizado para extração de características do conjunto de dados a ser utilizado.

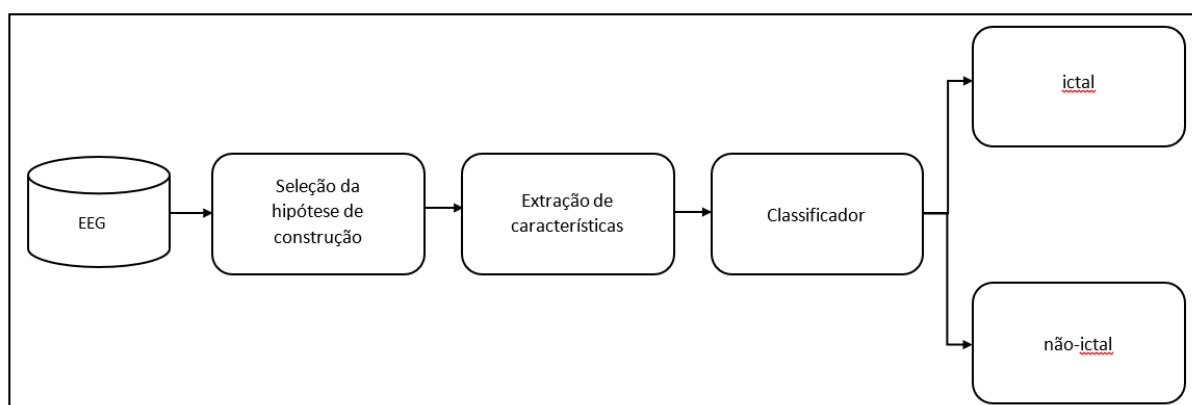
3 EXTRAÇÃO DE CARACTERÍSTICAS

Um sistema de auxílio a diagnóstico baseado em aprendizado de máquina depende diretamente da qualidade dos dados de entrada, tais como números de casos clínicos, presença de ruídos e da técnica de extração de atributos utilizados.

Neste capítulo apresentaremos as ferramentas que serão utilizadas para testar as hipóteses de construção do conjunto de treinamento. Explicaremos o uso da matriz de covariância como método a ser utilizado para realizar a extração de características do conjunto de dados de EEG. Falaremos sobre o classificador linear de mínimos quadrados, que será utilizado para classificar as épocas em crises (ictais) e não crises (não-ictais).

A Figura 3 resume as etapas a serem implementadas para cada hipótese. A seguir explicaremos a etapa de extração de características.

Figura 3 - Etapas de processamento do EEG de um paciente.



Fonte: O autor.

3.1 Extração de características

Esta etapa consiste na utilização de técnicas de transformação do sinal original do EEG em uma representação matemática que permita a identificação de um comportamento. No caso deste trabalho, realiza a transformação dos dados (amostras) do EEG em vetores de características que serão utilizados pelo classificador linear de mínimos quadrados para identificar se uma época representa ou não uma crise epilética.

Um dos efeitos imediatos dessa transformação sobre os dados originais é a redução da dimensionalidade do espaço de dados. Ou seja, um segmento do sinal de EEG com 512

amostras de cada um dos 23 canais será transformado primeiramente em uma matriz de covariância de dimensão 23x23 e, posteriormente, em um vetor de dimensão 276. Os detalhes desta transformação serão apresentados na próxima seção. Embora ocorra a redução de dimensionalidade, ela ocorre de tal forma que não haja perda significativa de informação nos dados originais.

A matriz de covariância, um método de análise estatística em que se aplica diretamente a matriz de covariância às amostras de EEG, consegue essa redução de dimensionalidade. A seguir falaremos sobre a matriz de covariância.

3.2 Matriz de covariância como extrator de características

Nesta seção abordaremos o uso da matriz de covariância na extração do vetor de características de um sinal, em nosso caso, do EEG. Ela representa a dispersão do sinal em torno do valor médio, ou seja, representa a energia do sinal.

3.2.1 Definição de covariância

A covariância é uma medida estatística que quantifica a variação conjunta de duas variáveis aleatórias. Assim, dadas N observações de duas variáveis aleatórias \mathbf{x} e \mathbf{y} , a covariância entre essas duas variáveis é definida como:

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}}), \quad (1)$$

onde $\bar{\mathbf{x}}$ e $\bar{\mathbf{y}}$ são as médias de \mathbf{x} e \mathbf{y} respectivamente.

3.2.2 Definição de matriz de covariância

Costuma-se dispor as covariâncias entre todas as variáveis, tomadas duas a duas, $\sigma_{i,j}$, $i, j = 1, \dots, p$, em uma matriz de covariância, $\Sigma = [\sigma_{ij}]_{p \times p}$:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}. \quad (2)$$

Os elementos da diagonal principal desta matriz são as variâncias dos p atributos envolvidos no problema. Fora da diagonal principal tem-se as covariâncias das variáveis X_i e X_j , para $i \neq j$.

A matriz de covariância também pode ser escrita como

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1p}\sigma_1\sigma_p \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2p}\sigma_2\sigma_p \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1}\sigma_p\sigma_1 & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}, \quad (3)$$

se lembrarmos que o coeficiente de correlação entre X_i e X_j é dado por

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (4)$$

de onde tiramos que a covariância entre X_i e X_j pode ser expressa como

$$\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j. \quad (5)$$

Algumas propriedades da matriz de covariância são listadas a seguir:

1. A matriz $\mathbf{\Sigma}$ é *simétrica*, pois $\sigma_{ij} = \sigma_{ji}$:

$$\mathbf{\Sigma} = \mathbf{\Sigma}^T$$

2. A matriz $\mathbf{\Sigma}$ é *definida positiva*:

$$\mathbf{x}^T \mathbf{\Sigma} \mathbf{x} > 0,$$

para qualquer vetor \mathbf{x} real e não nulo.

3. Os autovalores λ_i de Σ são sempre positivos.
4. O determinante de Σ é sempre positivo.

Considerando p igual a 2, o vetor aleatório e o vetor de médias passam a ser representados como:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ e } \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}.$$

Assim, a matriz de covariância Σ reduz-se a uma matriz quadrada de ordem 2:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}_{2 \times 2}. \quad (6)$$

3.2.3 Estimação da matriz de covariância

A seguir serão descritas algumas técnicas de estimação da matriz de covariância (G. A. Barreto (2020). Estimação da matriz de covariância. Notas de Aulas. Disciplina de Introdução ao Reconhecimento de Padrões. Curso de Engenharia de Computação da UFC.).

Seja um conjunto de N observações de uma variável aleatória $\mathbf{x} \in \mathbb{R}^p$:

$$\{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)\}. \quad (7)$$

A matriz de covariância pode ser estimada diretamente através da seguinte expressão:

$$\Sigma = \frac{1}{N} \sum_{n=1}^N [\mathbf{x}(n) - \mathbf{m}][\mathbf{x}(n) - \mathbf{m}]^T. \quad (8)$$

Já o vetor médio do conjunto de observações é estimado como

$$\mathbf{m} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n). \quad (9)$$

Pode-se também usar a seguinte expressão alternativa para estimar a matriz de covariância:

$$\mathbf{\Sigma}_x = \hat{\mathbf{R}}_x - \mathbf{m}\mathbf{m}^T. \quad (10)$$

Com a matriz de correção $\hat{\mathbf{R}}_x$ sendo estimada como

$$\hat{\mathbf{R}}_x = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n)\mathbf{x}^T(n). \quad (11)$$

Pode-se derivar a Equação 10 a partir da Equação 8:

$$\begin{aligned} \mathbf{\Sigma} &= \frac{1}{N} \sum_{n=1}^N [\mathbf{x}(n) - \mathbf{m}][\mathbf{x}(n) - \mathbf{m}]^T \\ &= \frac{1}{N} \sum_{n=1}^N [\mathbf{x}(n)\mathbf{x}^T(n) - \mathbf{x}(n)\mathbf{m}^T - \mathbf{m}\mathbf{x}^T(n) - \mathbf{m}\mathbf{m}^T] \\ &= \frac{1}{N} \left[\left(\sum_{n=1}^N \mathbf{x}(n)\mathbf{x}^T(n) \right) - \left(\sum_{n=1}^N \mathbf{x}(n) \right) \mathbf{m}^T - \mathbf{m} \left(\sum_{n=1}^N \mathbf{x}^T(n) \right) - \left(\sum_{n=1}^N \mathbf{m}\mathbf{m}^T \right) \right] \\ &= \frac{1}{N} \left[\left(\sum_{n=1}^N \mathbf{x}(n)\mathbf{x}^T(n) \right) - (N\mathbf{m})\mathbf{m}^T - \mathbf{m}(N\mathbf{m}^T) - N\mathbf{m}\mathbf{m}^T \right] \\ &= \frac{1}{N} \left[\left(\sum_{n=1}^N \mathbf{x}(n)\mathbf{x}^T(n) \right) - N\mathbf{m}\mathbf{m}^T \right] \\ &= \frac{1}{N} \left[\left(\sum_{n=1}^N \mathbf{x}(n)\mathbf{x}^T(n) \right) - \mathbf{m}\mathbf{m}^T \right] \\ &= \hat{\mathbf{R}}_x - \mathbf{m}\mathbf{m}^T. \end{aligned}$$

3.2.3.1 Estimação da matriz de covariância no Matlab/Octave

Os comandos a seguir ilustram como determinar o vetor de médias e a matriz de covariância para um conjunto de N observações de um vetor aleatório gaussiano de dimensão 2, utilizando comandos do *Matlab/Octave*:

```

>> p=2; % dim. vetor aleatório
>> N=5000; % total de observações
>> X=randn(N, p); % gera vetores aleatórios
>> M=mean(X)
      M = 0.0097 -0.0027
>> Cx=cov(X)
      Cx = 1.0224 -0.0124
           -0.0124 0.9887

```

3.3 Montagem dos vetores de características

Uma grande vantagem da matriz de covariância é que a mesma pode ser aplicada a todas as variáveis p (canais do EEG) simultaneamente. Neste trabalho, ela será aplicada diretamente a um segmento de dados de dois segundos de duração que chamaremos de época.

Conforme será explicado mais a frente, os exames de EEG do conjunto de dados que usaremos apresentam em sua maioria 23 canais e duração de uma hora (3600 segundos). Como a taxa de amostragem usada no processo de gravação do EEG que será utilizado foi de 256 Hz, tem-se o registro de 256 amostras por segundo. Portanto, para cada arquivo de EEG tem-se uma matriz de dimensão $23 \times (256 \times 3600)$, ou seja, 23×921600 amostras.

Ao aplicarmos a matriz de covariância a uma época de dois segundos, tem-se como resultado, a seguinte matriz de dimensão 23×23 ($p \times p$), ou seja, 529 elementos:

$$\Sigma = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,p} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p,1} & \sigma_{p,2} & \dots & \sigma_{p,p} \end{bmatrix}_{23 \times 23} \quad (12)$$

Ocorre que sendo essa matriz simétrica, os seus elementos se repetem em relação a diagonal principal ($\sigma_{1,2} = \sigma_{2,1}, \sigma_{1,3} = \sigma_{3,1}, \dots$). Assim, para a construção do vetor de características, seleciona-se apenas os elementos pertencentes à parte triangular superior da matriz de covariância conforme destacado na Equação 12. Juntamente com a seleção dos elementos é aplicado um operador de vetorização $\text{vect}(\Sigma)$ que transformará cada matriz de alta dimensão ($23 \times 23 = 529$ elementos) em um vetor de dimensão reduzido (276 elementos), conforme mostrado a seguir:

$$vect(\Sigma) = \begin{bmatrix} \sigma_{1,1} \\ \sigma_{2,1} \\ \sigma_{p,1} \\ \vdots \\ \sigma_{1,p} \\ \sigma_{2,p} \\ \sigma_{p,p} \end{bmatrix}_{529 \times 1} . \quad (13)$$

Durante o processo de vetorização será aplicado um coeficiente de ajuste ($\sqrt{2}$) aos elementos não pertencentes a diagonal principal com o objetivo de conservar a igualdade entre as normas de $\|\Sigma\|_F = vect\|\Sigma\|_2$ em que $\|\cdot\|_F$ denota a norma de Frobenius (BARACHANT, 2013). Depois de aplicar o operador de vetorização e eliminar os elementos repetidos devido a propriedade de simetria da matriz Σ , o vetor de características será dado por:

$$vect(\Sigma) = \begin{bmatrix} \sigma_{1,1} \\ \sqrt{2}\sigma_{1,2} \\ \sigma_{2,2} \\ \vdots \\ \sqrt{2}\sigma_{1,p} \\ \sqrt{2}\sigma_{2,p} \\ \sigma_{p,p} \end{bmatrix}_{276 \times 1} \quad (14)$$

O número de elementos Ne do vetor coluna formado na expressão acima, após o processo de vetorização pode ser obtido aplicando a expressão abaixo, onde p corresponde ao número de canais:

$$Ne = p(p + 1)/2. \quad (15)$$

Para o exemplo fornecido o número de elementos Ne será dado por $23(23 + 1)/2 = 276$ elementos.

Conforme será explicado no próximo capítulo, após aplicar o processo mencionado a todas as épocas de determinado arquivo, a matriz \mathbf{X} que representa os vetores de características terá a seguinte forma:

$$\mathbf{X} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,1} & \sigma_{1,1} \\ \sigma_{1,2} & \sigma_{1,2} & \sigma_{1,2} \\ \sigma_{2,2} & \sigma_{2,2} & \sigma_{2,2} \\ \vdots & \vdots & \vdots \\ \sigma_{1,p} & \sigma_{1,p} & \sigma_{1,p} \\ \sigma_{2,p} & \sigma_{2,p} & \sigma_{2,p} \\ \sigma_{p,p} & \sigma_{p,p} & \sigma_{p,p} \end{bmatrix}_{276 \times E}, \quad (16)$$

onde E representa o número de épocas que foram vetorizadas.

3.4 Resumo do capítulo

Neste capítulo foi apresentado o método de extração de características em que a abordagem é a vetorização da matriz de covariância. Este método realiza a redução de dimensionalidade da matriz de covariância $p \times p$ original, em que p representa o número de canais da época de 2 segundos que está sendo trabalhada.

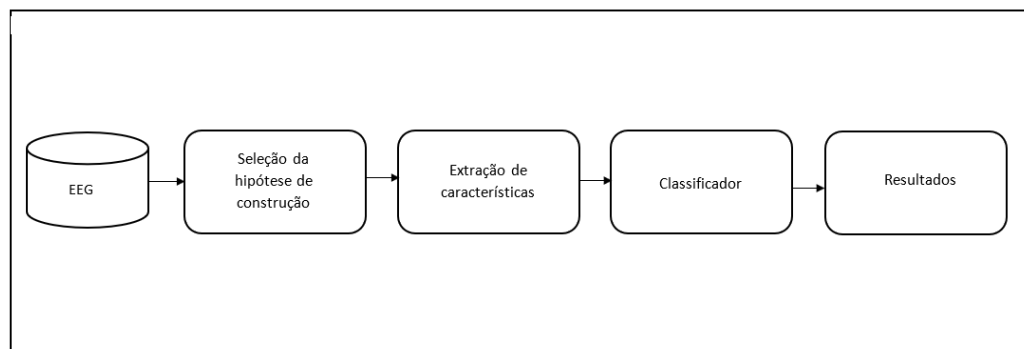
No próximo capítulo será apresentada a metodologia a ser utilizada neste trabalho, o conjunto de dados e cada uma das sete hipóteses levantadas neste trabalho.

4 MATERIAIS E MÉTODOS

Neste capítulo serão descritos todos os procedimentos a serem realizados com os dados a fim de prepará-los para treino e teste do classificador OLS. Serão detalhados os passos para seleção dos intervalos de amostras de forma a balancear as classes de cada hipótese, a transformação a ser realizada sobre esses dados e a montagem dos vetores de características. Além disso, as hipóteses serão apresentadas.

A Figura 4 representa a modelagem completa do experimento a ser realizado, mostrando todas as etapas e a conexão entre elas. A seguir será explicado como será realizada a leitura dos arquivos dos pacientes.

Figura 4 - Modelagem do experimento.



Fonte: O autor.

4.1 Conjunto de dados

O conjunto de dados utilizados no presente trabalho pertence ao projeto *CHB-MIT Scalp EEG Dataset*, que pode ser acessado através do site: <https://physionet.org/content/chbmit/1.0.0/>. Consiste em gravações contínuas de EEG de Escalpo, realizados em 23 pacientes pediátricos do Hospital Infantil de Boston, após a retirada de medicação. Dos 23 pacientes, 6 são do gênero masculino com idades entre 3 e 22 anos, e 17 são do gênero feminino com idades entre 1,5 e 19 anos. O conjunto de dados completo descompactado corresponde a um total de 42,6 GB de dados.

Cada paciente contém entre 9 e 42 arquivos com extensão EDF gravados de forma contínua. A maioria dos arquivos contém gravações de uma hora de sinal EEG digitalizado, embora os arquivos pertencentes ao paciente chb10 tenham duas horas de duração e aqueles

pertencentes aos pacientes chb04, chb06, chb07, chb09 e chb23 tenham quatro horas de duração. A privacidade dos pacientes foi protegida através da substituição de algumas informações.

Todos os sinais foram gravados utilizando 256 amostras por segundo com 16 bits de resolução. A maioria dos arquivos contem 23 sinais (canais) de EEG que foram gravados utilizando o sistema de posicionamento internacional de eletrodos 10-20.

Em alguns arquivos, foram adicionados sinais extras, tais como um sinal de ECG nos 36 últimos arquivos pertencentes ao paciente chb04 e um sinal de estímulo de nervo vago (VNS) nos últimos 18 arquivos pertencentes ao paciente chb09.

De todos os 664 arquivos, 129 deles apresentam crises, sendo um total de 198 crises já que alguns deles apresentam mais de uma crise por arquivo. Para cada paciente existe um arquivo chamado *chbnn-summary.txt* contendo informações sobre todos os arquivos do paciente e suas crises, quando existentes. São essas informações que nos mostram qual arquivo contém crise e em que instante de tempo elas iniciam e terminam.

O conjunto de dados mencionado serviu de referência para os trabalhos de Ali Shoeb, entre eles:

- Ali Shoeb, John Guttag. [Application of Machine Learning to Epileptic Seizure Onset Detection](#). *27th International Conference on Machine Learning (ICML)*, June 21-24, 2010, Haifa, Israel.
- Ali Shoeb, Herman Edwards, Jack Connolly, Blaise Bourgeois, S. Ted Treves, John Guttag. [Patient-Specific Seizure Onset Detection](#). *Epilepsy and Behavior*. August 2004, 5(4): 483-498. [doi:10.1016/j.yebeh.2004.05.005].

4.1.1 Subconjunto de dados

Para efetuar os testes das sete hipóteses deste trabalho, optou-se por trabalhar apenas com um subconjunto do conjunto de dados disponível. No entanto, os testes relativos à hipótese 1 já tinham sido executados para todos os 23 pacientes em momento anterior a criação das outras hipóteses. Portanto, essa hipótese servirá como ponto de partida para avaliação e comparação das outras hipóteses.

A execução da primeira hipótese forneceu uma primeira impressão de desempenho do conjunto de dados de cada paciente. Conseguiu-se obter boa acurácia com alguns pacientes (em torno de 98%), enquanto para outros, a acurácia ficou em torno de 60%. A diferença de

desempenho se dá devido à grande variedade de dados entre pacientes causada por diversos motivos tais como presença de artefatos e a própria particularidade de atividade cerebral de cada paciente.

Visando a execução dos testes das hipóteses com a maior variedade de dados possível, procurou-se selecionar pacientes que apresentaram bons resultados (pacientes chb01, chb03, chb05 e chb18) e outros que apresentaram resultados inferiores (pacientes chb07 e chb14).

4.2 Leitura dos arquivos

O conjunto de dados de um paciente, conforme já explicado, é composto por diversos arquivos com extensão EDF (European Data Format), geralmente com uma hora de duração. Alguns desses arquivos contêm uma ou mais convulsões, sendo que o instante de início e fim destas convulsões foi previamente sinalizado por um especialista. Neste trabalho utilizou-se apenas os arquivos que apresentam uma ou mais crises (convulsão).

Para realizar a leitura de cada arquivo EEG (extensão .EDF) utilizou-se a função *edfread*, que foi escrita originalmente para ser utilizada no *Matlab* e posteriormente incorporada à ferramenta *Signal Processing Toolbox* do mesmo. Nesse trabalho, no entanto, utilizou-se a função *edfread* no ambiente *Octave*², ferramenta que apresenta funcionalidades parecidas (compatíveis) com o *Matlab*³. Optou-se pelo uso dessa ferramenta por se tratar de ferramenta de software livre.

4.3 Balanceamento das classes

Conforme explicado anteriormente, o intervalo de duração de uma convulsão (período ictal) é muito pequeno quando comparado com o de não convulsão (período inter-ictal ou não-ictal). Esse grande desbalanceamento entre as classes (ictal e não-ictal) tende a provocar mau funcionamento em grande parte dos classificadores.

Portanto, é necessário que haja um balanceamento entre as classes (ictal e não-ictal) antes de realizar a extração de características.

² www.gnu.org/software/octave/

³ www.mathworks.com

Os períodos ictais são fixos, já que foram previamente rotulados por um especialista. No entanto, os períodos não-ictais, por representarem a maioria das amostras do paciente, precisam ser extraídos em quantidade proporcional às amostras ictais. Não se sabe, a priori, qual seria a proporção ictal/não-ictal ideal. Isso, inclusive, é um dos itens testados nas diversas hipóteses explicadas mais à frente.

Nos experimentos realizados neste trabalho optou-se por um balanceamento entre as classes ictal/não-ictal em proporções 1:1 ou 1:3, dependendo da hipótese.

Tabela 1 – Proporção de amostras não-ictal/ictal das hipóteses.

HIPÓTESE	NÃO-ICTAL/ICTAL
1	1:1
2	1:1
3	1:1
4	4:1
5	4:1
6	4:1
7	4:1

Fonte: O autor.

A Tabela 1 mostra a proporção entre as amostras não-ictais e ictais de todas as hipóteses levantadas neste trabalho.

4.4 Construção dos Vetores de Características

Após a extração, de forma balanceada, das amostras ictais e não-ictais de cada arquivo, serão listadas a seguir as tarefas necessárias para realizar a montagem dos vetores de características. São elas:

1. Concatenação das amostras ictais extraídas de todos os arquivos do paciente, formando uma matriz contendo todas as amostras ictais do paciente;
2. Concatenação das amostras não-ictais (de acordo com a hipótese) extraídas de todos os arquivos do paciente, formando uma matriz contendo todas as amostras não-ictais do paciente;
3. Concatenação das matrizes de amostras não-ictais e ictais formando assim outra matriz, contendo todos os dados não-ictais e ictais do paciente;

4. Divisão da matriz formada no passo 3 em intervalos de dois segundos, chamados de épocas. Cada época é uma matriz de dimensão 23 x 512, considerando um arquivo de EEG composto por 23 canais;
5. Rotulação de cada época do passo 4 em ictal (+1) e não-ictal (-1). Essa informação será guardada no vetor **Y**, que será a classe de saída ou target;
6. Cálculo da Matriz de Covariância de cada época, seguido da vetorização dos elementos da parte triangular superior dessa matriz. Além disso, cada elemento não pertencente a diagonal principal será multiplicado pelo coeficiente $\sqrt{2}$. O vetor formado será salvo como vetor coluna da matriz **X**;
7. Aplicação do passo anterior a todas as épocas restantes e concatenação dos vetores formados a Matriz **X**. Ao final, a matriz **X** representará os dados de entrada a serem usados juntamente com o vetor **Y** para treinamento de um classificador binário, em nosso caso, o classificador OLS.

Dessa forma, após o processamento de todos os arquivos de um paciente, teremos duas matrizes, uma representando as amostras de entrada (**X**) e outra representando a classe de saída (**Y**). A Figura 5 mostra a estrutura das matrizes **X** e **Y**, onde **E** representa o número de épocas e consequentemente, de vetores de características.

Figura 5 - Matrizes **X** e **Y** após Extração dos vetores de características.

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,E} \\ x_{2,1} & x_{2,2} & \dots & x_{2,E} \\ \dots & \dots & \dots & \dots \\ x_{276,1} & x_{276,2} & \dots & x_{276,E} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_{1,1} & y_{1,2} & \dots & y_{1,E} \end{bmatrix}$$

Fonte: O autor.

Segundo Shoeb (2010), devido ao fato do EEG ser não estacionário e não poder ser dividido em unidades fisiológicas muito pequenas, épocas de 2 segundos são normalmente usadas.

No passo 4 a mesma abordagem será utilizada formando épocas compostas de 512 amostras, já que o EEG foi gravado utilizando frequência de 256 Hz.

Considerando que a maioria dos arquivos de EEG tem 23 canais, no passo 5, quando aplicada a matriz de covariância a cada época, ou seja, a cada matriz de dimensão 23 x 512 amostras, é retornada uma matriz de covariância C de dimensão 23 x 23 amostras. Como a matriz de covariância é simétrica em relação a sua diagonal principal, é necessário extrair apenas os elementos de sua parte triangular superior.

Conforme Barachant (2013), o total de elementos extraídos da matriz C , incluindo os elementos da diagonal principal, poderá ser obtido pela expressão abaixo, onde N representa o número de canais do EEG.

$$(N + 1) \frac{N}{2} \quad (2)$$

Para o exemplo fornecido, a matriz de covariância gerará um vetor coluna de dimensão 276 x 1, como pode ser visto na Figura 5.

Esse procedimento será repetido para cada época do arquivo que está sendo processado. No final, teremos um arquivo de dimensão 276 x E , sendo E o número de épocas do arquivo original. Esse arquivo representará a matriz de dados X a ser utilizada no treinamento do classificador.

O passo 5 associa a cada época do arquivo X um valor no vetor de saída Y , ou seja, para cada época, ele informa se a mesma representa uma crise (ictal) ou não-crise (não-ictal). Para isso, cada época ictal será rotulada em Y como +1 e cada época não-ictal será rotulada em Y como -1.

Ao final do passo 7, teremos a matriz X e o vetor Y , que poderão ser utilizados para treinar uma máquina preditiva através do uso de um ou mais classificadores (neste caso, o classificador OLS) e assim poder realizar previsões quando novos dados forem a ela submetidos.

O treinamento e teste do classificador costuma ser realizado dividindo-se os dados contidos em X e Y em dados de treino e teste. No entanto, realizaremos a fase de teste através de dados não contidos em X e Y . Para isso, separaremos um arquivo completo do paciente (não contido em X e Y), e executaremos apenas os passos de 4 a 7. Considerando um arquivo com uma hora ou 3600 segundos, no final teremos um arquivo de teste de dimensão 276 x 1800 amostras, já que cada época tem 2 segundos de duração.

Essa diferença na forma de realização do teste será feita visando termos um teste mais próximo da realidade, já que nenhuma alteração (balanceamento das classes) será realizada nos dados originais.

Além da proporção (balanceamento) entre as classes, testaremos a posição e a distância das amostras não-ictais em relação às amostras ictais.

Na próxima seção listaremos todas as hipóteses que serão exploradas neste trabalho, fornecendo as principais diferenças entre elas.

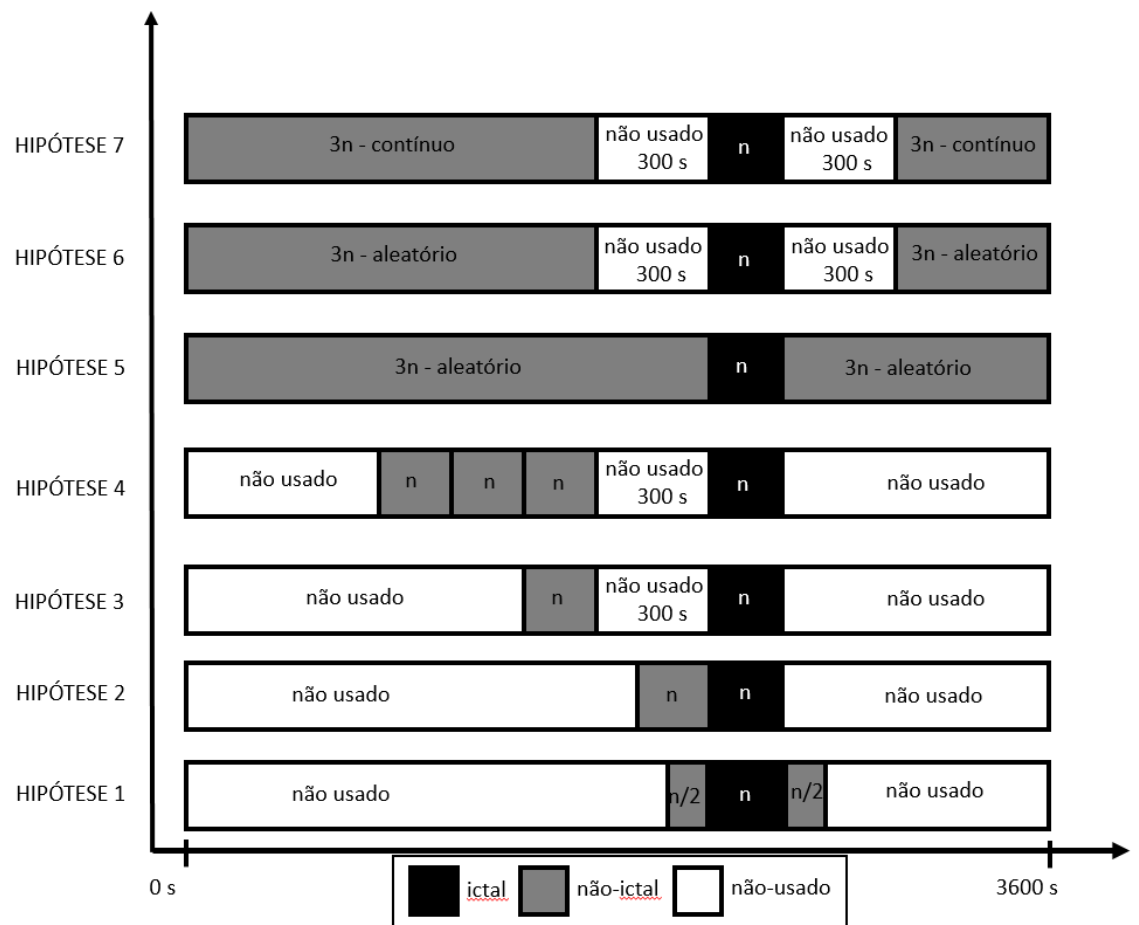
4.5 Hipóteses desenvolvidas para seleção de amostras de treinamento

Conforme falado anteriormente, o foco principal desse trabalho consiste no teste de várias hipóteses de extração de dados não convulsivos a serem utilizados durante a fase de treinamento de um algoritmo de classificação.

Trabalharemos com sete hipóteses, onde cada uma delas representa uma forma diferente de extração de dados não convulsivos (não-ictais) tendo como ponto de partida os dados convulsivos (ictais). A título de explicação, considerando uma crise eletrográfica de 10 segundos em determinado arquivo de um paciente com 3600 segundos de duração, realizamos a extração de dados não convulsivos de mesma duração, ou seja, 10 segundos, sendo 5 segundos extraídos antes e 5 segundos extraídos depois do intervalo convulsivo.

A Figura 6 resume todas as sete hipóteses que serão trabalhadas. A seguir descreveremos as principais características de cada hipótese.

Figura 6 - Hipóteses mostrando a proporção ictal/não-ictal



Fonte: O autor.

4.5.1 Hipótese 1

Dado determinado intervalo convulsivo (ictal) de tamanho n , extrair de forma contínua, intervalo não-convulsivo (não-ictal) do mesmo tamanho de n , sendo este extraído metade antes e metade depois do período convulsivo, conforme mostrado na Figura 6. Esta hipótese é aquela que servirá como parâmetro para as demais, pois a partir dela, verificaremos se haverá melhora nos resultados do classificador.

4.5.2 Hipótese 2

Dado determinado intervalo convulsivo (ictal) de tamanho n , extrair de forma contínua, intervalo não-convulsivo (não-ictal) do mesmo tamanho de n , sendo este extraído antes do período convulsivo conforme mostrado na Figura 6.

4.5.3 Hipótese 3

Dado determinado intervalo convulsivo (ictal) de tamanho n , extrair de forma contínua, intervalo não-convulsivo (não-ictal) do mesmo tamanho de n , sendo este extraído antes do período convulsivo, mas distante 300 segundos deste, conforme mostrado na Figura 6.

4.5.4 Hipótese 4

Dado determinado intervalo convulsivo (ictal) de tamanho n , extrair de forma contínua, intervalo não-convulsivo (não-ictal) na proporção de três vezes o tamanho de n , sendo este extraído antes do período convulsivo, mas distante 300 segundos deste, conforme Figura 6.

4.5.5 Hipótese 5

Dado determinado intervalo convulsivo (ictal) de tamanho n , extrair de forma aleatória, intervalo não-convulsivo (não-ictal) na proporção de três vezes o tamanho de n , sendo este extraído antes ou depois do período convulsivo, conforme Figura 6.

4.5.6 Hipótese 6

Dado determinado intervalo convulsivo (ictal) de tamanho n , extrair de forma aleatória, intervalo não-convulsivo (não-ictal) na proporção de três vezes o tamanho de n , sendo este extraído antes ou depois do período convulsivo, mas distante 300 segundos deste, conforme Figura 6. Esta hipótese extrai aleatoriamente cada segundo que compõe o intervalo.

4.5.7 Hipótese 7

Dado determinado intervalo convulsivo (ictal) de tamanho n , extrair de forma aleatória e contínua, intervalo não-convulsivo (não-ictal) na proporção de três vezes o tamanho de n , sendo este extraído antes ou depois do período convulsivo, mas distante 300 segundos deste, conforme Figura 6. Diferente da Hipótese 6, esta hipótese extrai um intervalo contínuo.

4.6 Classificador para avaliação das hipóteses

Para realizar o teste das sete hipóteses será utilizado o classificador linear dos mínimos quadrados (*ordinary least squares* - OLS) para classificar cada vetor de características do conjunto de dados de treinamento (Matrix \mathbf{X}) em ictal ou não-ictal. Nos classificadores lineares, o objetivo é obter um mapeamento linear entre os dados de entrada e os dados de saída (vetor \mathbf{Y}) correspondentes.

Conforme já mencionado, cada EEG foi previamente analisado por um especialista e o período ictal foi sinalizado. Desta forma, é possível identificar dois tipos de dados em cada arquivo EEG de um paciente: dados ictais e dados não-ictais. Tem-se então um problema de característica binária.

Assim, é possível criar uma máquina preditiva que fará uso de um classificador (OLS) para classificar os dados em ictais e não-ictais. Para isso, a classe de saída \mathbf{Y} (ou target) será rotulada com os valores +1 para uma época ictal e -1 para uma época não-ictal.

Sendo \mathbf{X} a matriz das amostras de entrada e \mathbf{Y} a matriz de rótulos de saída, o mapeamento entre os dados de entrada e de saída pode ser obtido através da relação

$$\mathbf{Y} = \mathbf{W}\mathbf{X}.$$

Caso a matriz \mathbf{X} seja quadrada, a matriz \mathbf{W} poderá ser obtida da seguinte forma:

$$\mathbf{W} = \mathbf{Y}\mathbf{X}^{-1}.$$

Entretanto, nos problemas práticos de classificação, tem-se uma quantidade bem superior de amostras comparada ao número de atributos. Portanto, a matriz \mathbf{X} não pode ser invertida diretamente. Faz-se necessário o uso da matriz pseudoinversa conhecida como técnica

dos mínimos quadrados ordinários (*ordinary least square*), que visa minimizar a soma quadrática dos erros na aproximação do mapeamento. Assim:

$$\mathbf{W} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}.$$

Após o cálculo da matriz \mathbf{W} , pode-se testar o desempenho do classificador OLS para as amostras que não foram utilizadas para treinamento, comparando a saída estimada pelo classificador.

A matriz $\mathbf{X}\mathbf{X}^T$ a ser invertida pode estar mal condicionada ou ser singular, o que pode levar a resultados numericamente instáveis. Este problema pode ser evitado usando o comando *pinv* presente no *Matlab/Octave*, abaixo:

```
>> W=Y*pinv(X);
```

De posse da matriz \mathbf{W} , as previsões Y_tst para o conjunto de teste X_tst podem ser obtidas pelo comando abaixo:

```
>> Y_tst_predict=W*X_tst;
```

4.7 Resumo do capítulo

Neste capítulo explicamos toda a metodologia utilizada no trabalho. Apresentamos as diferentes hipóteses desenvolvidas para seleção de amostras de treinamento, explicamos a construção dos vetores de características e o classificador linear dos mínimos quadrados.

No próximo capítulo serão apresentados os resultados obtidos ao aplicar a metodologia aqui exposta aos arquivos EEG de seis pacientes.

5 RESULTADOS

Neste capítulo serão apresentados os resultados relativos aos testes de todas as hipóteses para cada paciente. Conforme mencionado anteriormente, dos 23 pacientes disponíveis, apenas seis foram escolhidos para realizar a análise. A Tabela 2 mostra algumas informações sobre os pacientes que foram selecionados para os testes das hipóteses.

Tabela 2 - Pacientes selecionados para teste das hipóteses.

PACIENTE	Nº ARQUIVOS	GÊNERO	IDADES	Nº DE CRISES
chb01	7	F	11	7
chb03	7	F	14	7
chb05	5	F	7	5
chb07	3	F	14,5	3
chb14	7	F	9	8
chb18	6	F	18	6

Fonte: O autor

Vale ressaltar novamente que os testes das hipóteses são do tipo paciente específico, onde apenas os dados do mesmo paciente são utilizados para treino e teste.

5.1 Paciente chb01

A Tabela 3 apresenta um resumo dos arquivos do paciente chb01 que foram utilizados para treino e teste do classificador OLS. É importante lembrar que um dos arquivos é separado para teste (inicialmente aquele arquivo que apresenta crise de maior duração) e os demais são processados para formar os dados de treino.

Entretanto, houve a curiosidade de se verificar o desempenho do classificador utilizando-se cada um dos arquivos como arquivo de teste. Por isso, tanto para o paciente chb01 como para os demais, serão apresentados os resultados através de validação cruzada.

Nesta seção serão apresentados mais detalhes do que nas seções referentes aos demais pacientes. Isso se faz necessário para que o leitor entenda o significado das colunas das tabelas apresentadas. Para os demais pacientes, os procedimentos se repetem, não sendo necessário, portanto, repetir o mesmo detalhamento.

Tabela 3 - Arquivos do paciente chb01 contendo crises.

ARQUIVO	INICIO (s)	FIM (s)	DURAÇÃO (s)
chb01_03	2996	3036	41
chb01_04	1467	1494	28
chb01_15	1732	1772	41
chb01_16	1015	1066	52
chb01_18	1720	1810	91
chb01_21	327	420	94
chb01_26	1862	1963	102

Fonte: O autor

A Tabela 4 mostra os resultados para as Hipóteses 1 e 2 do paciente chb01. A coluna “arq_treino” mostra quais arquivos desse paciente foram combinados para formar o arquivo de treino e a coluna “arq_teste” mostra o respectivo arquivo de teste.

Cada linha da tabela representa os resultados do classificador OLS para uma combinação de arquivos de treino (coluna “arq_treino”) e um arquivo de teste (“coluna arq_teste”). Para cada hipótese são apresentadas a sua acurácia (ACUR), especificidade (ESPECIF) e sensibilidade (SENSIB).

Foi realizada a validação cruzada para cada hipótese, seguida dos cálculos de suas médias aritmética e geométrica. A média geométrica é utilizada para que as três grandezas (ACUR, ESPECIF, SENSIB) sejam maximizadas simultaneamente. É calculada da seguinte forma:

$$MG = \sqrt[3]{(ACUR \times ESPECIF \times SENSIB)}$$

Tabela 4 - Resultados para as hipóteses 1 e 2 do paciente chb01.

PACIENTE 01		HIPOTESE 1			HIPOTESE 2		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb01_04_15_16_18_21_26	chb01_03	94,444	94,716	71,429	95,833	96,346	52,381
chb01_03_15_16_18_21_26	chb01_04	96,667	96,695	93,333	94	94,174	73,333
chb01_03_04_16_18_21_26	chb01_15	96,556	97,021	57,143	95,667	96,121	57,143
chb01_03_04_15_18_21_26	chb01_16	90,833	91,032	77,778	89,889	90,073	77,778
chb01_03_04_15_16_21_26	chb01_18	98,556	98,917	84,783	97,5	98,233	69,565
chb01_03_04_15_16_18_26	chb01_21	93,556	94,406	62,500	86,5	87,272	58,333
chb01_03_04_15_16_18_21	chb01_26	89,157	90,459	60,784	93,373	94,959	58,824
MÉDIA ARITMÉTICA		94,253	94,749	72,536	93,252	93,883	63,908
MÉDIA GEOMÉTRICA		86,523			82,401		

Fonte: O autor

Percebe-se pela Tabela 4, que não houve melhora no desempenho da Hipótese 2 com relação a Hipótese 1.

A Tabela 5 mostra os resultados para as Hipóteses 3 e 4. Nesse caso, percebe-se que a Hipótese 4 apresentou melhora com relação a Hipótese 1. Já a Hipótese 3 apresenta uma ligeira perda de desempenho quando comparada com a Hipótese 1.

Tabela 5 - Resultados para as Hipóteses 3 e 4 do paciente chb01.

PACIENTE 01		HIPOTESE 3			HIPOTESE 4		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb01_04_15_16_18_21_26	chb01_03	93,389	93,704	66,667	97,5	97,864	66,667
chb01_03_15_16_18_21_26	chb01_04	93,778	93,725	100	96,889	96,919	93,333
chb01_03_04_16_18_21_26	chb01_15	85,333	85,497	71,429	92,611	93,03	57,143
chb01_03_04_15_18_21_26	chb01_16	88,056	88,268	74,074	96,056	96,334	77,778
chb01_03_04_15_16_21_26	chb01_18	97	97,434	80,435	98,389	98,689	86,957
chb01_03_04_15_16_18_26	chb01_21	84,667	85,845	41,667	87,944	88,813	56,25
chb01_03_04_15_16_18_21	chb01_26	93,201	94,869	56,863	95,439	96,67	68,627
MÉDIA ARITMÉTICA		90,775	91,335	70,162	94,975	95,474	72,394
MÉDIA GEOMÉTRICA		83,477			86,909		

Fonte: O autor

A Tabela 6 mostra que as Hipóteses 5 e 6 apresentaram melhora quando comparadas com a Hipótese 1, enquanto na Tabela 7 percebe-se que quase não houve alteração na Hipótese 7 quando comparada com a Hipótese 1.

Tabela 6 - Resultados para as Hipóteses 5 e 6 do paciente chb01.

PACIENTE 01		HIPOTESE 5			HIPOTESE 6		
arq_treino	arq_teste	Pacerto	ESPECIF	SENSIB	Pacerto	ESPECIF	SENSIB
chb01_04_15_16_18_21_26	chb01_03	98,056	98,426	66,667	96,944	97,358	61,905
chb01_03_15_16_18_21_26	chb01_04	98,944	99,048	86,667	99,056	99,216	80
chb01_03_04_16_18_21_26	chb01_15	94,944	95,166	76,19	96,556	96,965	61,905
chb01_03_04_15_18_21_26	chb01_16	96,222	96,277	92,593	96,667	96,842	85,185
chb01_03_04_15_16_21_26	chb01_18	98,889	99,373	80,435	99,111	99,829	71,739
chb01_03_04_15_16_18_26	chb01_21	94,167	94,92	66,667	97,111	97,831	70,833
chb01_03_04_15_16_18_21	chb01_26	91,997	92,169	88,235	94,923	96,31	64,706
MÉDIA ARITMÉTICA		96,174	96,483	79,636	97,195	97,764	70,896
MÉDIA GEOMÉTRICA		90,408			87,663		

Fonte: O autor

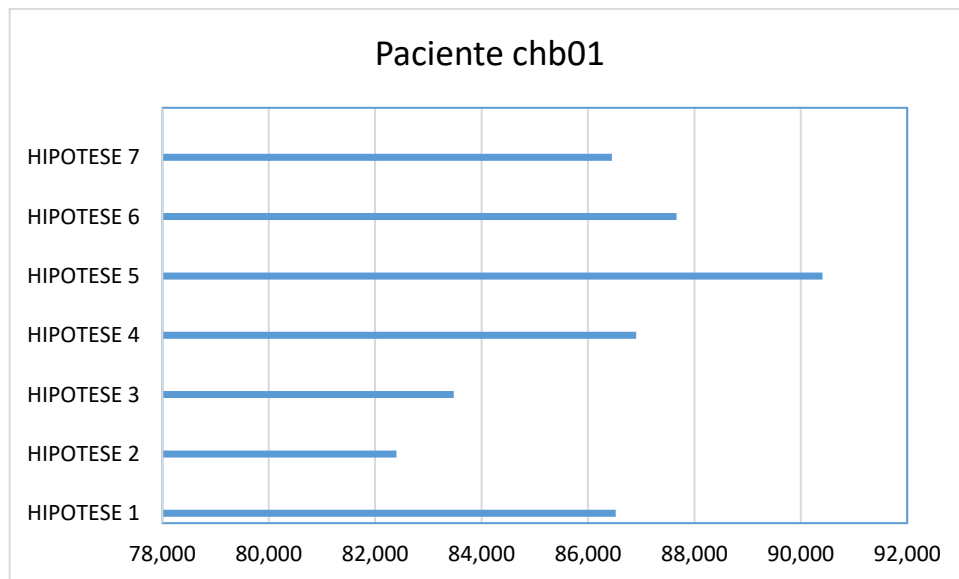
Tabela 7 - Resultados para a Hipótese 7 do paciente chb01.

PACIENTE 01		HIPOTESE 7		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB
chb01_04_15_16_18_21_26	chb01_03	97,056	97,358	71,429
chb01_03_15_16_18_21_26	chb01_04	98,944	99,048	86,667
chb01_03_04_16_18_21_26	chb01_15	96,611	97,021	61,905
chb01_03_04_15_18_21_26	chb01_16	86,5	86,576	81,481
chb01_03_04_15_16_21_26	chb01_18	98,5	99,544	58,696
chb01_03_04_15_16_18_26	chb01_21	93,111	94,064	58,333
chb01_03_04_15_16_18_21	chb01_26	92,771	93,249	82,353
MÉDIA ARITMÉTICA		94,785	95,266	71,552
MÉDIA GEOMÉTRICA		86,450		

Fonte: O autor

O Gráfico 1 e a Tabela 8 apresentam o resumo dos resultados do paciente chb01.

Gráfico 1 - Resultados do paciente chb01.



Fonte: O autor

Tabela 8 – Resumo dos resultados do paciente chb01.

chb01	ACUR	ESPECIF	SENSIB	MÉD. GEOMÉTRICA
Hipótese 1	94,253	94,749	72,536	86,525
Hipótese 2	93,252	93,883	63,908	82,401
Hipótese 3	90,775	91,335	70,162	83,477
Hipótese 4	94,975	95,474	72,394	86,909
Hipótese 5	96,174	96,483	79,636	90,408
Hipótese 6	97,195	97,764	70,896	87,663
Hipótese 7	94,785	95,266	71,552	86,450

Fonte: O autor

A seguir mostraremos os resultados para os demais pacientes. Lembrando que, por se tratarem das mesmas informações do paciente chb01 (os mesmos procedimentos foram realizados), comentaremos apenas os resultados.

5.2 Paciente chb03

A Tabela 9 apresenta um resumo dos arquivos do paciente chb03. As Tabelas 10, 11 12 e 13 mostram os resultados para todas as hipóteses quando aplicadas aos dados do paciente chb03.

Tabela 9 - Arquivos do paciente chb03 contendo crises.

ARQUIVO	INICIO (s)	FIM (s)	DURAÇÃO (s)
chb03_01	362	414	53
chb03_02	731	796	66
chb03_03	432	501	70
chb03_04	2162	2214	53
chb03_34	1982	2029	48
chb03_35	2592	2656	65
chb03_36	1725	1778	54

Fonte: O autor

Tabela 10 - Resultados para as Hipóteses 1 e 2 do paciente chb03.

PACIENTE 03		HIPOTESE 1			HIPOTESE 2		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb03_02_03_04_34_35_36	chb03_01	98,389	98,703	77,778	96,778	97,349	59,259
chb03_01_03_04_34_35_36	chb03_02	95,444	95,753	79,412	96,667	97,225	67,647
chb03_01_02_04_34_35_36	chb03_03	97,167	98,074	51,429	83,389	83,343	85,714
chb03_01_02_03_34_35_36	chb03_04	99,778	99,944	88,889	99,222	99,38	88,889
chb03_01_02_03_04_35_36	chb03_34	94,500	94,876	66,667	93,222	94,032	33,333
chb03_01_02_03_04_34_36	chb03_35	92,889	93,492	60,606	87,444	87,719	72,727
chb03_01_02_03_04_34_35	chb03_36	93,556	94,018	64,286	87,056	87,246	75
MÉDIA ARITMÉTICA		95,960	96,409	69,867	91,968	92,328	68,938
MÉDIA GEOMÉTRICA		86,462			83,652		

Fonte: O autor

Tabela 11 - Resultados para as Hipóteses 3 e 4 do paciente chb03.

PACIENTE 03		HIPOTESE 3			HIPOTESE 4		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb03_02_03_04_34_35_36	chb03_01	96,722	97,293	59,259	98,167	98,421	81,481
chb03_01_03_04_34_35_36	chb03_02	92,722	93,262	64,706	96,833	97,169	79,412
chb03_01_02_04_34_35_36	chb03_03	91,667	92,521	48,571	94,167	94,561	74,286
chb03_01_02_03_34_35_36	chb03_04	99,389	99,718	77,778	99,611	99,887	81,481
chb03_01_02_03_04_35_36	chb03_34	93,222	93,525	70,833	95,333	95,664	70,833
chb03_01_02_03_04_34_36	chb03_35	92,611	93,605	39,394	96,556	97,17	63,636
chb03_01_02_03_04_34_35	chb03_36	94,611	95,203	57,143	95,556	96,219	53,571
MÉDIA ARITMÉTICA		94,421	95,018	59,669	96,603	97,013	72,100
MÉDIA GEOMÉTRICA		81,197			87,751		

Fonte: O autor

Tabela 12 - Resultados para as Hipóteses 5 e 6 do paciente chb03.

PACIENTE 03		HIPOTESE 5			HIPOTESE 6		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb03_02_03_04_34_35_36	chb03_01	98,889	99,098	85,185	98,444	98,872	70,370
chb03_01_03_04_34_35_36	chb03_02	97,778	97,961	88,235	98	98,188	88,235
chb03_01_02_04_34_35_36	chb03_03	95,667	96,034	77,143	95	95,581	65,714
chb03_01_02_03_34_35_36	chb03_04	99,500	99,887	74,074	99,667	99,944	81,481
chb03_01_02_03_04_35_36	chb03_34	95,000	95,383	66,667	95,778	96,059	75
chb03_01_02_03_04_34_36	chb03_35	94,889	95,246	75,758	94,556	95,19	60,606
chb03_01_02_03_04_34_35	chb03_36	95,889	96,670	46,429	91,556	91,574	57,143
MÉDIA ARITMÉTICA		96,802	97,183	73,356	96,143	96,487	71,221
MÉDIA GEOMÉTRICA		88,369			87,096		

Fonte: O autor

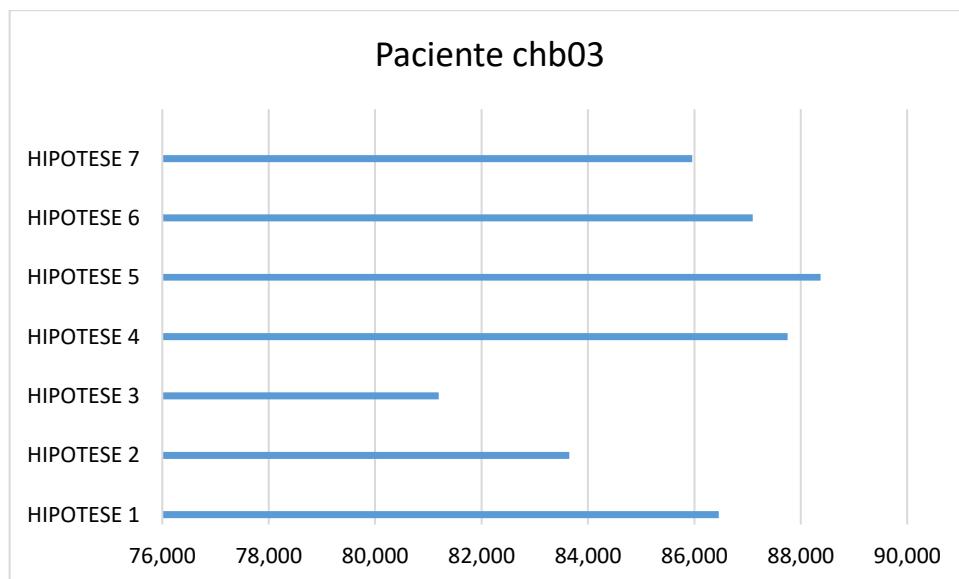
Tabela 13 - Resultados para a Hipótese 7 do paciente chb03.

PACIENTE 03		HIPOTESE 7		
arq_treino	arq_teste	Pacerto	ESPECIF	SENSIB
chb03_02_03_04_34_35_36	chb03_01	98,833	99,154	77,778
chb03_01_03_04_34_35_36	chb03_02	98	98,358	79,412
chb03_01_02_04_34_35_36	chb03_03	99,056	99,773	62,857
chb03_01_02_03_34_35_36	chb03_04	99,611	99,831	85,185
chb03_01_02_03_04_35_36	chb03_34	94,778	95,158	66,667
chb03_01_02_03_04_34_36	chb03_35	91,722	92,756	36,364
chb03_01_02_03_04_34_35	chb03_36	95,222	95,711	64,286
MÉDIA ARITMÉTICA		96,746	97,249	67,507
MÉDIA GEOMÉTRICA		85,958		

Fonte: O autor

O Gráfico 2 e a Tabela 14 apresentam o resumo dos resultados do paciente chb03.

Gráfico 2 - Resultados do paciente chb03.



Fonte: O autor

Tabela 14 – Resumo dos resultados do paciente chb03.

chb03	ACUR	ESPECIF	SENSIB	MÉD. GEOMÉTRICA
Hipótese 1	95,960	96,409	69,867	86,462
Hipótese 2	91,968	92,328	68,938	83,652
Hipótese 3	94,421	95,018	59,669	81,197
Hipótese 4	96,603	97,013	72,100	87,751
Hipótese 5	96,802	97,183	73,356	88,369
Hipótese 6	96,143	96,487	71,221	87,096
Hipótese 7	96,746	97,249	67,507	85,958

Fonte: O autor

De acordo com a Tabela 10, não houve melhora nos resultados da Hipótese 2 com relação a Hipótese 1. A Tabela 11 mostra que não houve melhora nos resultados da Hipótese 3, mas a Hipótese 4 apresentou melhora. A Tabela 12 apresenta melhoras tanto para a Hipótese 5 quanto para a Hipótese 6. A Tabela 13 mostra uma ligeira piora nos resultados.

5.3 Paciente chb05

A Tabela 15 apresenta um resumo dos arquivos do paciente chb05.

Tabela 15 - Arquivos do paciente chb05 contendo crises.

ARQUIVO	INICIO (s)	FIM (s)	DURAÇÃO (s)
chb05_06	417	532	116
chb05_13	1086	1196	111
chb05_16	2317	2413	97
chb05_17	2451	2571	121
chb05_22	2348	2465	118

Fonte: O autor

As Tabelas 16, 17, 18, e 19 mostram todos os resultados para este paciente.

Tabela 16 - Resultados para as Hipóteses 1 e 2 do paciente ch05.

PACIENTE 05		HIPOTESE 1			HIPOTESE 2		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb05_13_16_17_22	chb05_06	97,333	98,392	66,102	96,556	97,99	54,237
chb05_06_16_17_22	chb05_13	97,167	98,394	58,929	96,444	97,42	66,071
chb05_06_13_17_22	chb05_16	97,667	98,401	71,429	97,389	98,287	65,306
chb05_06_13_16_22	chb05_17	96,389	97,642	60,656	96,111	97,47	57,377
chb05_06_13_16_17	chb05_22	91,722	93,337	44,068	89	90,408	47,458
MÉDIA ARITMÉTICA		96,056	97,233	60,237	95,100	96,315	58,090
MÉDIA GEOMÉTRICA		82,553			81,032		

Fonte: O autor

Tabela 17 - Resultados para as Hipóteses 3 e 4 do paciente ch05.

PACIENTE 05		HIPOTESE 3			HIPOTESE 4		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb05_13_16_17_22	chb05_06	97,444	98,679	61,017	98,167	99,426	61,017
chb05_06_16_17_22	chb05_13	97,111	98,108	66,071	98,667	99,484	73,214
chb05_06_13_17_22	chb05_16	96,722	97,601	65,306	98,611	99,429	69,388
chb05_06_13_16_22	chb05_17	96,333	97,601	65,306	98,278	99,367	67,213
chb05_06_13_16_17	chb05_22	89	90,178	54,237	94,333	95,979	45,763
MÉDIA ARITMÉTICA		95,322	96,433	62,387	97,611	98,737	63,319
MÉDIA GEOMÉTRICA		83,087			84,821		

Fonte: O autor

Tabela 18 - Resultados para as Hipóteses 5 e 6 do paciente ch05.

PACIENTE 05		HIPOTESE 5			HIPOTESE 6		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb05_13_16_17_22	chb05_06	98	99,655	49,153	98,722	99,77	67,797
chb05_06_16_17_22	chb05_13	98,556	99,484	69,643	98,556	99,025	83,929
chb05_06_13_17_22	chb05_16	98,389	99,143	71,429	98,5	99,086	77,551
chb05_06_13_16_22	chb05_17	97,889	98,965	67,213	97,667	98,907	62,295
chb05_06_13_16_17	chb05_22	94,778	96,611	40,678	91,833	93,567	40,678
MÉDIA ARITMÉTICA		97,522	98,772	59,623	97,056	98,071	66,450
MÉDIA GEOMÉTRICA		83,122			85,839		

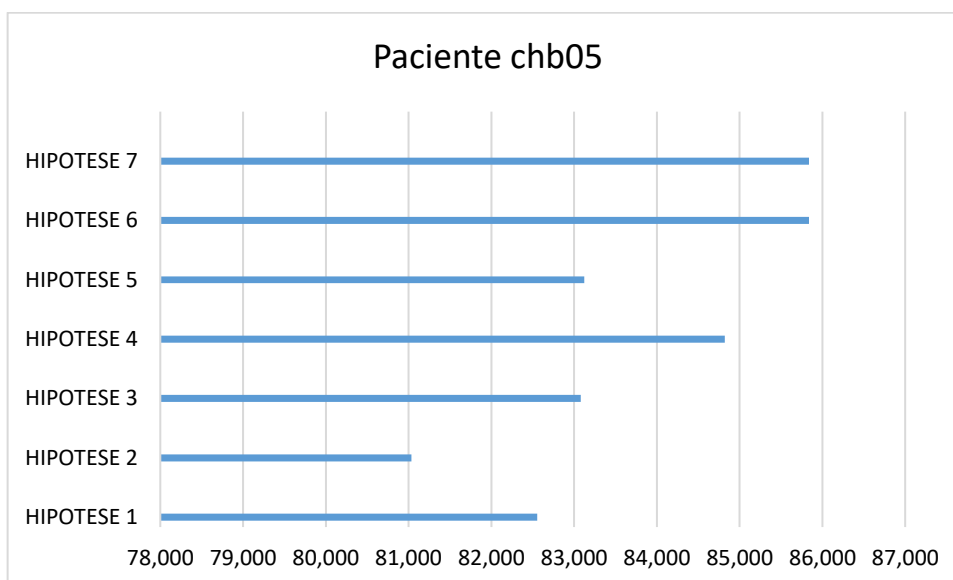
Fonte: O autor

Tabela 19 - Resultados para a Hipótese 7 do paciente ch05.

PACIENTE 05		HIPOTESE 7		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB
chb05_13_16_17_22	chb05_06	98,722	99,77	67,797
chb05_06_16_17_22	chb05_13	98,556	99,025	83,929
chb05_06_13_17_22	chb05_16	98,5	99,086	77,551
chb05_06_13_16_22	chb05_17	97,667	98,907	62,295
chb05_06_13_16_17	chb05_22	91,833	93,567	40,678
MÉDIA ARITMÉTICA		97,056	98,071	66,450
MÉDIA GEOMÉTRICA		85,83914751		

Fonte: O autor

Gráfico 3 - Resultados do paciente chb05.



Fonte: O autor

Tabela 20 – Resumo dos resultados do paciente chb05.

chb05	ACUR	ESPECIF	SENSIB	MÉD. GEOMÉTRICA
Hipótese 1	96,056	97,233	60,237	82,553
Hipótese 2	95,100	96,315	58,090	81,032
Hipótese 3	95,322	96,433	62,387	83,082
Hipótese 4	97,611	98,737	63,319	84,821
Hipótese 5	97,522	98,772	59,623	83,122
Hipótese 6	97,056	98,071	66,450	85,839
Hipótese 7	97,071	98,071	66,450	85,839

Fonte: O autor

Com exceção da Hipótese 2, houve melhora nos resultados de todas as hipóteses deste paciente. O Gráfico 3 e a Tabela 20 resumem os resultados deste paciente.

5.4 Pacientes chb07, chb14 e chb18

Para os pacientes chb07, chb14 e chb18 serão apresentados apenas os gráficos e tabelas de resumo dos resultados deixando assim o texto mais claro. Os resultados completos para estes pacientes podem ser verificados no **APÊNDICE A**.

5.4.1 Paciente chb07

A Tabela 21 apresenta um resumo dos arquivos do paciente chb07.

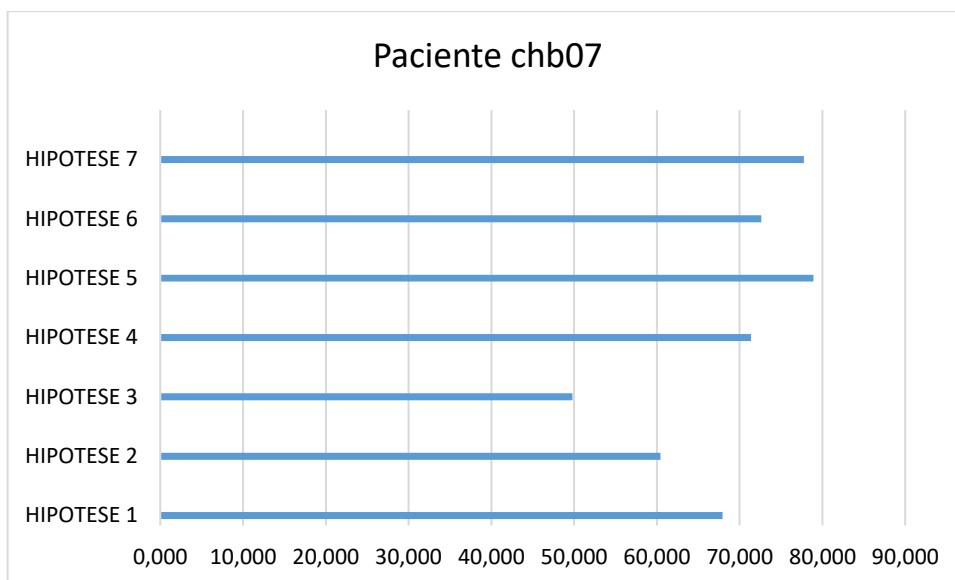
Tabela 21 - Arquivos do paciente chb07 contendo crises.

ARQUIVO	INICIO (s)	FIM (s)	DURAÇÃO (s)
chb07_12	4920	5006	87
chb07_13	3285	3381	97
chb07_19	13688	13831	144

Fonte: O autor

A Tabela 22 e o Gráfico 4 apresentam os resultados para todas as hipóteses do paciente chb07.

Gráfico 4 - Resultados do paciente chb07.



Fonte: O autor

Tabela 22 – Resumo dos resultados do paciente chb07.

chb07	ACUR	ESPECIF	SENSIB	MÉD. GEOMÉTRICA
Hipótese 1	72,544	72,67	59,437	67,921
Hipótese 2	84,069	84,906	30,898	60,418
Hipótese 3	56,806	56,598	38,428	49,806
Hipótese 4	93,396	94,111	41,390	71,387
Hipótese 5	95,802	96,428	53,224	78,927
Hipótese 6	95,326	95,966	41,852	72,613
Hipótese 7	94,087	94,739	52,754	77,762

Fonte: O autor

Os arquivos deste paciente, diferente dos demais, apresentam arquivos contendo quatro horas de gravação. Os resultados deste paciente não apresentaram melhoras para as Hipóteses 2 e 3, conforme Tabela 22. No entanto, houve melhoras significativas para as Hipóteses 4, 5, 6 e 7.

5.4.2 Paciente chb14

A Tabela 23 apresenta um resumo dos arquivos do paciente chb14. O arquivo chb14_04 aparece repetido porque o mesmo apresenta duas crises.

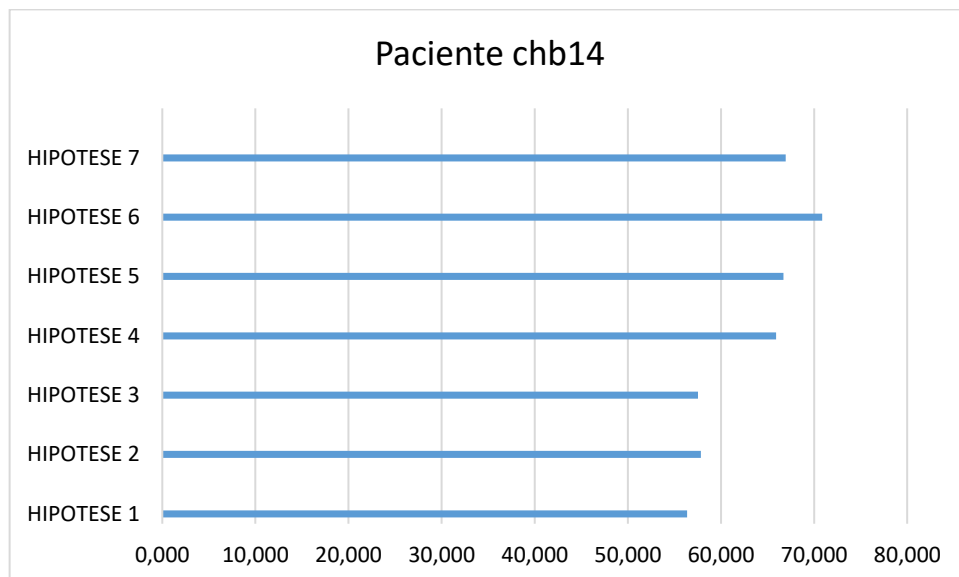
Tabela 23 - Arquivos do paciente chb14 contendo crises.

ARQUIVO	INICIO (s)	FIM (s)	DURAÇÃO (s)
chb14_03	1986	2000	15
chb14_04	1372	1392	21
chb14_04	2817	2839	23
chb14_06	1911	1925	15
chb14_11	1838	1879	42
chb14_17	3239	3259	21
chb14_18	1039	1061	23
chb14_27	2833	2849	17

Fonte: O autor

A Tabela 24 e o Gráfico 5 apresentam todos os resultados para este paciente.

Gráfico 5 - Resultados do paciente chb14.



Fonte: O autor

Tabela 24 – Resumo dos resultados do paciente chb14.

chb14	ACUR	ESPECIF	SENSIB	MÉD. GEOMÉTRICA
Hipótese 1	55,762	55,754	57,578	56,358
Hipótese 2	54,262	54,142	65,927	57,858
Hipótese 3	52,468	52,372	69,260	57,521
Hipótese 4	72,651	72,735	54,181	65,909
Hipótese 5	76,540	76,692	50,558	66,703
Hipótese 6	75,206	75,263	62,915	70,881
Hipótese 7	69,040	69,083	62,911	66,947

Fonte: O autor

Todas as hipóteses deste paciente apresentaram melhoras nos resultados com relação a Hipótese 1.

5.4.3 Paciente chb18

A Tabela 25 mostra todos os arquivos do paciente chb18 e a duração das crises de cada arquivo.

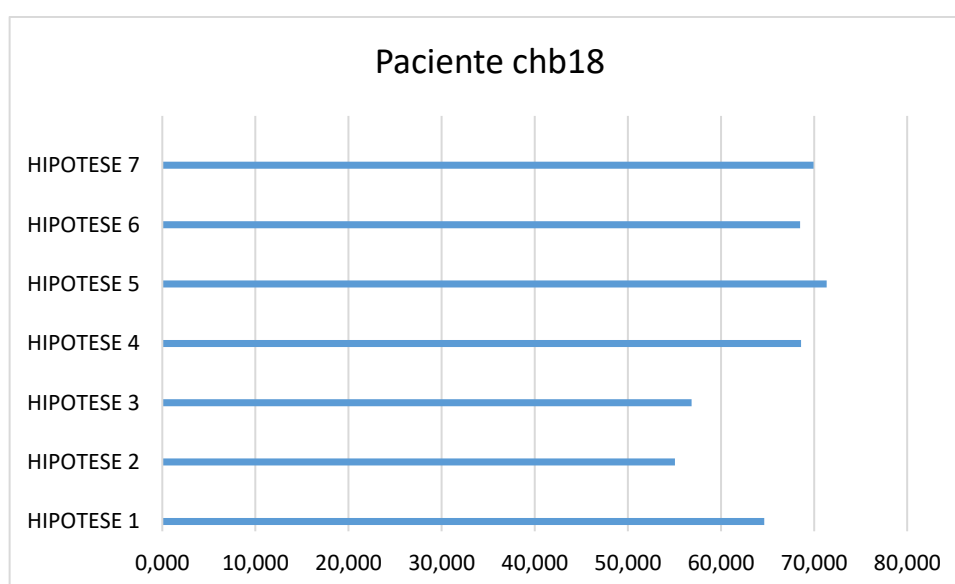
Tabela 25 - Arquivos do paciente chb18 contendo crises.

ARQUIVO	INICIO (s)	FIM (s)	DURAÇÃO (s)
chb18_29	1986	2000	15
chb18_30	1372	1392	21
chb18_31	2817	2839	23
chb18_32	1911	1925	15
chb18_35	1039	1061	23
chb18_36	2833	2849	17

Fonte: O autor

A Tabela 26 e o Gráfico 6 apresentam os resultados de todas as hipóteses deste paciente.

Gráfico 6 - Resultados do paciente chb18.



Fonte: O autor

Tabela 26 – Resumo dos resultados do paciente chb18.

chb18	ACUR	ESPECIF	SENSIB	MÉD. GEOMÉTRICA
Hipótese 1	82,740	83,487	39,091	64,635
Hipótese 2	68,055	68,708	35,689	55,056
Hipótese 3	67,864	68,301	39,623	56,842
Hipótese 4	89,955	90,818	39,550	68,619
Hipótese 5	77,444	77,477	41,777	71,361
Hipótese 6	85,204	85,943	43,872	68,489
Hipótese 7	86,476	87,286	45,362	69,959

Fonte: O autor

Pela análise do Gráfico 6 e da Tabela 26, pode-se verificar que as Hipóteses 2 e 3 não apresentaram melhora nos resultados apresentados pela Hipótese 1. Enquanto isso, as Hipóteses 4, 5, 6 e 7 apresentaram melhoras significativas.

5.7 Resumo do capítulo

Neste capítulo foram apresentados os resultados da classificação utilizando o algoritmo OLS aplicado a todas as hipóteses levantadas.

No próximo capítulo serão apresentadas as conclusões relativas às hipóteses levantadas assim como a discussão acerca de alguns pontos relevantes.

6 CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo serão apresentadas as conclusões e observações acerca das diversas hipóteses levantadas, as discussões do que poderia ser feito visando melhorar a análise dos dados e propostas de trabalhos futuros.

6.1 Conclusões

Neste trabalho, avaliou-se como a extração de intervalos não-ictais de exames de EEG, a partir da localização dos intervalos ictais, pode impactar os resultados de um classificador.

Foram levantadas sete diferentes hipóteses para seis pacientes, onde os resultados de cada hipótese foram comparados com a primeira hipótese, que representa uma referência para as demais hipóteses.

Após os testes de todas as hipóteses para cada paciente, chegou-se as seguintes conclusões:

1. De modo geral, as Hipóteses 2 e 3 não ofereceram melhora no desempenho em relação a Hipótese 1. Isso já era esperado, já que as amostras de EEG vizinhas a uma crise costumam apresentar comportamentos próximos aos apresentados por uma crise. A exceção fica por conta da Hipótese 3 do paciente chb05 e das Hipóteses 2 e 3 do paciente chb14;
2. As Hipóteses 4, 5, 6 e 7 apresentaram melhoras significativas se comparadas com a Hipótese 1. Estas hipóteses têm em comum a seleção de amostras não-ictais distantes das amostras ictais e em proporção 3 vezes maior. Isso diminuem as chances de se obter amostras com características de crise e também fornecem mais dados para treinamento do modelo.

6.2 Trabalhos Futuros

A seguir são listadas algumas possibilidades de continuidade para este trabalho:

1. Alguns arquivos do conjunto de dados utilizados apresentam muitos artefatos e não receberam qualquer tratamento (filtragem) neste trabalho. Como o desempenho de

- qualquer classificador está diretamente relacionado à qualidade dos dados de entrada, a exclusão de artefatos e outros ruídos poderia apresentar melhores resultados;
2. Desenvolvimento de uma máquina preditiva multiclases com o objetivo de classificar, além dos rótulos ictais e não-ictais, também os rótulos pré-ictais e pós-ictais;
 3. Analisar como se sairia um classificador usando uma abordagem diferente do tipo paciente específico usada neste trabalho. A ideia seria realizar a etapa de treinamento utilizando arquivos de todos os pacientes. Para a etapa de teste poderia ser reservado um arquivo de cada paciente;
 4. Desenvolvimento de aplicação capaz de identificar e excluir artefatos conhecidos, antes da aplicação da técnica utilizada neste trabalho;
 5. Utilização de uma plataforma de testes baseada em Python, que permitirá melhor exploração dos dados e criação de ambientes de teste automatizados;

REFERÊNCIAS

AGUIRRE, L. **Introdução à identificação de sistemas – Técnicas Lineares e Não Lineares: Teoria e Prática**. Belo Horizonte. UFMG, 4. Ed., p. 274-275, 2015.

BARACHANT, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). **Classification of covariance matrices using a Riemannian-based kernel for BCI applications**. *Neurocomputing* 112, 172–178. doi:10.1016/j.neucom.2012.12.039.

G. A. Barreto (2020). Estimação da matriz de covariância. Notas de Aulas. Disciplina de Introdução ao Reconhecimento de Padrões. Curso de Engenharia de Computação da UFC.

MARCUSE, LARA V; FIELDS, MADELINE C; YOO, JIYEOUN. *Rowan's PRIMER of EEG*. ELSEVIER, 2 Ed., 2016.

MCEWEN, J. A.; ANDERSON, G. B. **Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity**. *IEEE Transactions on Biomedical Engineering*, BME-22, n. 5, p. 361–369, Sep. 1975. ISSN 0018-9294.

MISRA, UK; KALITA, J. **Clinical eletroencephalography**, Elsevier. 2 Ed. p. 12, 2018. ISBN: 978-81-312-5296-3.

Russell, Stuart J.; Norvig, Peter. **Inteligência artificial**. p.822. Elsevier, 3 Ed., 2013, p. 82, ISBN: 978-85-352-3701-6.

SANEI, S.; CHAMBERS, J. A. *EEG signal processing*. [S.l.]: John Wiley & Sons, 2007. ISBN 0470025816,9780470025819.

SHOEB, A. CHB-MIT Scalp EEG Database. [S.l.]: physionet.org, 2010.

SHOEB, A. H. Application of machine learning to epileptic seizure onset detection and treatment, PhD Theses. [S.l.]: Massachusetts Institute of Technology, 2009.

SILVA JÚNIOR, J. P. Métodos de extração de atributos para detecção de crises epiléticas: uma abordagem comparativa. 2019. 155 f. Dissertação (Mestrado em Engenharia de Teleinformática)-Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2019.

RENCER, ALVIN C. **Methods of Multivariate Analysis**. John Wiley & Sons, Inc., 2002. ISBN 0-471-41889-7.

WORLD HEALTH ORGANIZATION. WHO. Epilepsy. 2018. Disponível em: <<http://www.who.int/en/news-room/fact-sheets/detail/epilepsy>>. Acesso em: 15/02/2021.

WORRELL, G. A.; PARISH, L.; CRANSTOUN, S. D.; JONAS, R.; BALTUCH, G.; LITT, B. High-frequency oscillations and seizure generation in neocortical epilepsy. *Brain*, v. 127, n. 7, p. 1496–1506, 2004. Disponível em: <<http://dx.doi.org/10.1093/brain/awh149>>.

UNIVERSIDADE FEDERAL DO CEARÁ. Biblioteca Universitária. **Guia de normalização de trabalhos acadêmicos da Universidade Federal do Ceará**. Fortaleza, 2013.

APÊNDICE A – RESULTADOS POR PACIENTE

Figura 4 – Resultados dos pacientes chb01, chb03 e chb05.

PACIENTE 01		HIPOTESE 1			HIPOTESE 2			HIPOTESE 3			HIPOTESE 4			HIPOTESE 5			HIPOTESE 6			HIPOTESE 7		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ACUR	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb01_04_15_16_18_21_26	chb01_03	94,444	94,716	71,429	95,833	96,346	52,381	93,389	93,704	66,667	97,5	97,864	66,667	98,056	98,426	66,667	96,944	97,358	61,905	97,056	97,358	71,429
chb01_03_15_16_18_21_26	chb01_04	96,667	96,695	93,333	94	94,174	73,333	93,778	93,725	100	96,889	96,919	93,333	98,944	99,048	86,667	99,056	99,216	80	98,944	99,048	86,667
chb01_03_04_16_18_21_26	chb01_15	96,556	97,021	57,143	95,667	96,121	57,143	85,333	85,497	71,429	92,611	93,03	57,143	94,944	95,166	76,19	96,556	96,965	61,905	96,611	97,021	61,905
chb01_03_04_15_18_21_26	chb01_16	90,833	91,032	77,778	89,889	90,073	77,778	88,056	88,268	74,074	96,056	96,334	77,778	96,222	96,277	92,593	96,667	96,842	85,185	86,5	86,576	81,481
chb01_03_04_15_16_21_26	chb01_18	98,556	98,917	84,783	97,5	98,233	69,565	97	97,434	80,435	98,389	98,689	86,957	98,889	99,373	80,435	99,111	99,829	71,739	98,5	99,544	58,696
chb01_03_04_15_16_18_26	chb01_21	93,556	94,406	62,500	86,5	87,272	58,333	84,667	85,845	41,667	87,944	88,813	56,25	94,167	94,92	66,667	97,111	97,831	70,833	93,111	94,064	58,333
chb01_03_04_15_16_18_21	chb01_26	89,157	90,459	60,784	93,373	94,959	58,824	93,201	94,869	56,863	95,439	96,67	68,627	91,997	92,169	88,235	94,923	96,31	64,706	92,771	93,249	82,353
MÉDIA ARITMÉTICA		94,253	94,749	72,536	93,252	93,883	63,908	90,775	91,335	70,162	94,975	95,474	72,394	96,174	96,483	79,636	97,195	97,764	70,896	94,785	95,266	71,552
MÉDIA GEOMÉTRICA		86,5248362			82,40103487			83,47719958			86,90918504			90,40787839			87,66302194			86,450114		

PACIENTE 03		HIPOTESE 1			HIPOTESE 2			HIPOTESE 3			HIPOTESE 4			HIPOTESE 5			HIPOTESE 6			HIPOTESE 7		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb03_02_03_04_34_35_36	chb03_01	98,389	98,703	77,778	96,778	97,349	59,259	96,722	97,293	59,259	98,167	98,421	81,481	98,889	99,098	85,185	98,444	98,872	70,370	98,833	99,154	77,778
chb03_01_03_04_34_35_36	chb03_02	95,444	95,753	79,412	96,667	97,225	67,647	92,722	93,262	64,706	96,833	97,169	79,412	97,778	97,961	88,235	98	98,188	88,235	98	98,358	79,412
chb03_01_02_04_34_35_36	chb03_03	97,167	98,074	51,429	83,389	83,343	85,714	91,667	92,521	48,571	94,167	94,561	74,286	95,667	96,034	77,143	95	95,581	65,714	99,056	99,773	62,857
chb03_01_02_03_34_35_36	chb03_04	99,778	99,944	88,889	99,222	99,38	88,889	99,389	99,718	77,778	99,611	99,887	81,481	99,500	99,887	74,074	99,667	99,944	81,481	99,611	99,831	85,185
chb03_01_02_03_04_35_36	chb03_34	94,500	94,876	66,667	93,222	94,032	33,333	93,222	93,525	70,833	95,333	95,664	70,833	95,000	95,383	66,667	95,778	96,059	75	94,778	95,158	66,667
chb03_01_02_03_04_34_36	chb03_35	92,889	93,492	60,606	87,444	87,719	72,727	92,611	93,605	39,394	96,556	97,17	63,636	94,889	95,246	75,758	94,556	95,19	60,606	91,722	92,756	36,364
chb03_01_02_03_04_34_35	chb03_36	93,556	94,018	64,286	87,056	87,246	75	94,611	95,203	57,143	95,556	96,219	53,571	95,889	96,670	46,429	91,556	91,574	57,143	95,222	95,711	64,286
MÉDIA ARITMÉTICA		95,960	96,409	69,867	91,968	92,328	68,938	94,421	95,018	59,669	96,603	97,013	72,100	96,802	97,183	73,356	96,143	96,487	71,221	96,746	97,249	67,507
MÉDIA GEOMÉTRICA		86,46215359			83,65216492			81,19718764			87,7510589			88,36947802			87,09606749			85,95844786		

PACIENTE 05		HIPOTESE 1			HIPOTESE 2			HIPOTESE 3			HIPOTESE 4			HIPOTESE 5			HIPOTESE 6			HIPOTESE 7		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb05_13_16_17_22	chb05_06	97,333	98,392	66,102	96,556	97,99	54,237	97,444	98,679	61,017	98,167	99,426	61,017	98	99,655	49,153	98,722	99,77	67,797	98,722	99,77	67,797
chb05_06_16_17_22	chb05_13	97,167	98,394	58,929	96,444	97,42	66,071	97,111	98,108	66,071	98,667	99,484	73,214	98,556	99,484	69,643	98,556	99,025	83,929	98,556	99,025	83,929
chb05_06_13_17_22	chb05_16	97,667	98,401	71,429	97,389	98,287	65,306	96,722	97,601	65,306	98,611	99,429	69,388	98,389	99,143	71,429	98,5	99,086	77,551	98,5	99,086	77,551
chb05_06_13_16_22	chb05_17	96,389	97,642	60,656	96,111	97,47	57,377	96,333	97,601	65,306	98,278	99,367	67,213	97,889	98,965	67,213	97,667	98,907	62,295	97,667	98,907	62,295
chb05_06_13_16_17	chb05_22	91,722	93,337	44,068	89	90,408	47,458	89	90,178	54,237	94,333	95,979	45,763	94,778	96,611	40,678	91,833	93,567	40,678	91,833	93,567	40,678
MÉDIA ARITMÉTICA		96,056	97,233	60,237	95,100	96,315	58,090	95,322	96,433	62,387	97,611	98,737	63,319	97,522	98,772	59,623	97,056	98,071	66,450	97,056	98,071	66,450
MÉDIA GEOMÉTRICA		82,55303673			81,03228938			83,08178855			84,82122539			83,1222423			85,83914751			85,83914751		

Fonte – O autor.

Figura 5 – Resultados dos pacientes chb07, chb14 e chb18.

PACIENTE 07		HIPOTESE 1			HIPOTESE 2			HIPOTESE 3			HIPOTESE 4			HIPOTESE 5			HIPOTESE 6			HIPOTESE 7		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb07_13_19	chb07_12	81,958	82,015	72,727	84,264	84,53	40,909	78,972	79,29	27,273	93,819	94,187	34,091	96,597	96,87	52,273	95,292	95,752	20,455	93,472	93,628	68,182
chb07_12_19	chb07_13	69,136	69,239	65,306	81,267	83,076	14,286	20,827	19,35	75,51	94,418	95,81	42,857	93,988	95,314	44,898	94,364	95,424	55,102	94,257	95,645	42,857
chb07_12_13	chb07_19	66,537	66,755	40,278	86,676	87,111	37,5	70,618	71,154	12,5	91,95	92,337	47,222	96,822	97,1	62,5	96,322	96,722	50	94,532	94,943	47,222
MÉDIA ARITMÉTICA		72,544	72,670	59,437	84,069	84,906	30,898	56,806	56,598	38,428	93,396	94,111	41,390	95,802	96,428	53,224	95,326	95,966	41,852	94,087	94,739	52,754
MÉDIA GEOMÉTRICA		67,921			60,418			49,806			71,387			78,927			72,613			77,762		

PACIENTE 14		HIPOTESE 1			HIPOTESE 2			HIPOTESE 3			HIPOTESE 4			HIPOTESE 5			HIPOTESE 6			HIPOTESE 7		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb14_04_06_11_17_18_27	chb14_03	56,056	56,194	25	57,5	57,701	12,5	46,444	46,373	62,5	70,167	70,145	75	80,833	81,027	37,5	75,889	75,949	62,5	69,222	69,196	75
chb14_03_06_11_17_18_27	chb14_04	56,556	56,387	69,565	68,667	68,486	82,609	47,333	47,271	52,174	66,167	66,123	69,565	75,667	75,746	69,565	72,5	72,369	82,609	58,389	58,582	43,478
chb14_03_04_11_17_18_27	chb14_06	38,167	38,002	75	42,889	42,746	75	47,222	47,042	87,5	67	66,694	75	74,5	74,554	62,5	74,222	74,163	87,5	65,778	65,848	50
chb14_03_04_06_17_18_27	chb14_11	71,389	71,669	47,619	23,278	22,709	71,429	54,389	54,356	57,143	75,111	75,379	52,381	77,222	77,347	66,667	69,389	69,365	71,429	73	72,962	76,19
chb14_03_04_06_11_18_27	chb14_17	47,611	47,513	63,636	71,111	71,101	72,727	40,444	40,246	72,727	79,389	79,542	54,545	72,389	72,554	45,455	71,167	71,381	36,364	65,222	65,12	81,818
chb14_03_04_06_11_17_27	chb14_18	46,222	46,085	66,667	67	67,058	58,333	62,778	62,696	75	68,167	68,345	41,667	69,833	69,966	50	77,278	77,349	66,667	69,333	69,407	58,333
chb14_03_04_06_11_17_18	chb14_27	74,333	74,428	55,556	49,389	49,19	88,889	68,667	68,621	77,778	82,556	82,915	11,111	85,333	85,65	22,222	86	86,265	33,333	82,333	82,468	55,556
MÉDIA ARITMÉTICA		55,762	55,754	57,578	54,262	54,142	65,927	52,468	52,372	69,260	72,651	72,735	54,181	76,540	76,692	50,558	75,206	75,263	62,915	69,040	69,083	62,911
MÉDIA GEOMÉTRICA		56,35804315			57,85790381			57,52104557			65,90892276			66,70267008			70,88090344			66,94711957		

PACIENTE 18		HIPOTESE 1			HIPOTESE 2			HIPOTESE 3			HIPOTESE 4			HIPOTESE 5			HIPOTESE 6			HIPOTESE 7		
arq_treino	arq_teste	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB	ACUR	ESPECIF	SENSIB
chb18_30_31_32_35_36	chb18_29	74,167	74,577	46,154	56,611	56,483	65,385	78,667	79,087	50	98,167	98,816	53,846	98,944	99,493	61,538	97,778	98,083	76,923	96,333	97,069	46,154
chb18_29_31_32_35_36	chb18_30	89	89,519	31,25	90,5	91,143	18,75	87	87,668	12,5	94,722	95,348	25	96,556	97,253	18,75	96,667	97,309	25	96,222	96,917	18,75
chb18_29_30_32_35_36	chb18_31	70,273	71,416	34,286	71,944	73,503	22,857	65,347	65,426	62,857	92,172	93,376	54,286	95,251	97,005	40	94,723	96,007	54,286	90,413	91,47	57,143
chb18_29_30_31_35_36	chb18_32	77,111	77,54	50	82,833	83,634	32,143	67,278	67,269	67,857	71,333	71,67	50	92,944	93,567	53,571	74,556	74,944	50	92,611	93,341	46,429
chb18_29_30_31_32_36	chb18_35	94,222	95,637	22,857	97,444	99,377	0	97,278	99,15	2,857	97,778	99,717	0	97,944	99,377	0	96,944	98,81	2,8571	96,167	97,96	5,7143
chb18_29_30_31_32_35	chb18_36	91,667	92,23	50	9	8,108	75	11,611	11,205	41,667	85,556	85,98	54,167	77,444	77,477	75	50,556	50,507	54,167	47,111	46,959	58,333
MÉDIA ARITMÉTICA		82,740	83,487	39,091	68,055	68,708	35,689	67,864	68,301	39,623	89,955	90,818	39,550	93,181	94,029	41,477	85,204	85,943	43,872	86,476	87,286	45,362
MÉDIA GEOMÉTRICA		64,63535439			55,05565831			56,84206807			68,61933385			71,3612538			68,48894481			69,95902804		

Fonte – O autor.