



PROJETO DE APRENDIZADO DE MÁQUINA

Primeira entrega na modalidade de relatório científico para a disciplina
Aprendizado de Máquina (SCC0276).

Docente: Prof. Fernando Pereira dos Santos

Equipe:

Calvin Suzuki NUSP: 11232420

Guilherme Soares Silvestre NUSP: 11299832

Vitor Fernando Rinaldini NUSP: 11232305

Sumário

1	Introdução	1
1.1	Inspiração na literatura	2
2	<i>Datasets</i>	3
2.1	Obtenção e pré-processamento dos dados	3
2.2	Datasets existentes	4
2.2.1	CHB-MIT Scalp EEG <i>Database</i>	4
2.2.2	Outros <i>datasets</i>	4
3	<i>Benchmarks</i>	6
3.0.1	Plataformas	6
3.0.2	Métricas	6
4	<i>Proposta</i>	7
	Referências bibliográficas	8

1 Introdução

A epilepsia é uma doença crônica comum que atinge aproximadamente 50 milhões de pessoas no mundo todo e, além disso, observa-se uma taxa de mortalidade prematura entre duas a três vezes maior em relação em indivíduos de até 5 anos de idade que apresentam a doença. (NUTRICIA, s.d.). Até o diagnóstico completo do paciente, as crises têm manifestações imprevisíveis no comportamento, no controle muscular, na consciência ou na sensibilidade do indivíduo, dificultando muito a percepção de uma crise (AMATO, s.d.).

O EEG (eletroencefalograma) é um exame que mede a atividade elétrica do cérebro, denominadas ondas cerebrais. Por meio da colagem de eletrodos na escalpe do paciente, podemos fazer o registro das ondas e coletar informações necessárias para o diagnóstico particular daquele paciente. No entanto, em casos particulares, não basta ter somente as informações coletadas do EEG para a produção do diagnóstico, para isso, pode-se prolongar a sessão para até 6 horas ou então optar pelo vídeo-EEG (*CHILDRENS HOSPITAL OF PITTSBURGH*, s.d.).

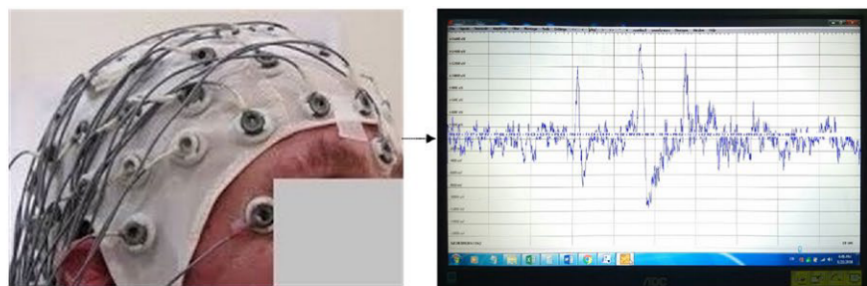


Figura 1.1: Eletrodos posicionados para o exame EEG e dados obtidos. Fonte: (SIDDQUI et al., 2020)

O vídeo-EEG é uma técnica que consiste em criar uma longa sessão de monitoramento em tempo real, visando detectar uma crise epilética com os dados das atividades cerebrais e mudanças de comportamento com o auxílio de câmeras (*CHILDRENS HOSPITAL OF PITTSBURGH*, s.d.). Nesses procedimentos a detecção é feita de forma manual, com uma equipe propondo atividades diversas para o paciente e uma outra equipe analisando seu comportamento e seus sinais cerebrais. Porém, esse processo pode apresentar falhas e atrasos da detecção de uma crise, impactando negativamente no tratamento do paciente(*CHILDRENS HOSPITAL OF PITTSBURGH*, s.d.).

1.1 Inspiração na literatura

A aplicação de métodos em *machine learning* para a realização de objetivos como detecção e predição de crises epiléticas são recorrentes na literatura, como foi apresentado no artigo "*A review of epileptic seizure detection using machine learning classifiers*" (SIDDIQUI et al., 2020). Esse trabalho apresentou sistemas que utilizam aprendizado de máquina para a predição e detecção das crises como auxílio para as sessões de vídeo-EEG.

A predição consiste em encontrar um momento "pré-crise" durante o monitoramento dos sinais do EEG do paciente. Isso permite que a equipe médica receba um alerta prévio para realizar os preparos para o diagnóstico, como inserção de contraste e ressonância magnética. Já na detecção de crises, o modelo objetiva inferir o momento de início da crise, informando o momento exato da disfunção neurológica.

2 *Datasets*

2.1 Obtenção e pré-processamento dos dados

Os *datasets* presentes na literatura são obtidos a partir da colagem de eletrodos em diversos locais na escalpe do indivíduo para a obtenção dos dados, resultando em um registro de vários sinais de tensão cerebral em um intervalo de tempo. Cada *dataset* possui seus próprios parâmetros como: taxa de amostragem; tamanho do intervalo; quantidade de eletrodos; e etc. Vale salientar que esses valores requerem pré-processamento antes de serem utilizados e validados por um método de aprendizado, por exemplo, para remoção de ruídos oriundos da imprecisão técnica do aparelho ou oscilações na rede elétrica.

No processo de limpeza são utilizados filtros de passa-alta, removendo componentes CC dos sinais e também os desvios, e de passa-baixa, eliminando as peças de alta frequência dos dados. Além dessas, também podem ser aplicadas outras formas, dependendo das impreviões ocorridas durante a etapa de monitoramento. Um exemplo é a aplicação da técnica de correção de artefatos de Eletrooculograma (EOG), que pode ser necessário se o sujeito sob gravação estiver com os olhos aberto. (RIERA, s.d.)

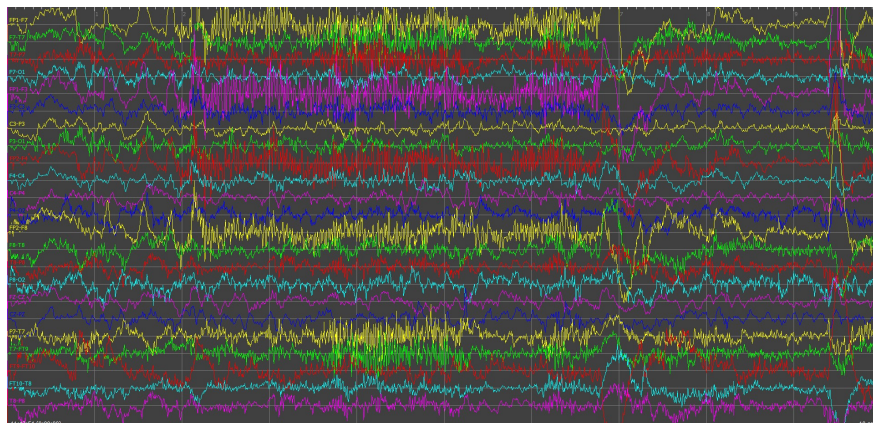


Figura 2.1: Exemplo de um Eletroencefalograma (EEG).

O próximo procedimento é a interpretação dos dados e, consequentemente, a extração de recursos. Dado um EEG, essa tarefa é complicada para ser realizada a olho nu (figura 2.1).

Exatamente por isto, foram desenvolvidos vários algoritmos complexos de processamento que fazem esse procedimento e separam as características no domínio do tempo e da frequência.

Como resultado da etapa anterior, deve-se obter uma alta gama de propriedades. Devido a isto, se faz necessário um procedimento de seleção de recursos, visando buscar apenas as características necessárias para a modelagem do problema.

Somente após a execução dos roteiros anteriores, os dados estarão prontos para alimentarem nossos modelos de aprendizado de máquina. Vale ressaltar que os processos de obtenção requerem anos ou até mesmo décadas para que acumulem uma quantidade suficiente de valores e, com isso, possibilitar a alimentação dos algoritmos. Entretanto, isso é algo que vem sendo armazenado há muito tempo e a melhor parte é que está disponível ao público. Na próxima seção, falaremos dos *datasets open-sources* existentes na internet.

2.2 Datasets existentes

2.2.1 CHB-MIT Scalp EEG Database

Esse banco de dados, consiste em registros de EEG de pacientes pediátricos com convulsões intratáveis. Os indivíduos foram monitorados por vários dias e, após a retirada da medicação anticonvulsivante, caracterizamos suas convulsões e avaliamos sua candidatura à intervenção cirúrgica (*CHB-MIT SCALP EEG DATABASE...*, s.d.).

Todos os sinais foram coletados em 256 amostras por segundo com resolução de 16 bits. A maioria dos arquivos contém 23 sinais de EEG (24 ou 26 em alguns casos). O sistema *International 10-20* de posições e nomenclatura dos eletrodos de EEG foi usado para esses registros.

2.2.2 Outros datasets

Existe um repositório no *GitHub* com uma ampla quantidade de *datasets* sobre EEG's (AGARWAL, s.d.) que estão separados por categorias como Motor-Imagem; Emoção-Reconhecimento; Potenciais Relacionados a Erros; e entre outros. Destacamos a possibilidade de utilizarmos o *dataset* de convulsões, visto que ela é um dos principais sintomas da crise epiléptica. Obviamente,

devemos lembrar que uma pessoa pode não apresentar esse sintoma durante um ataque, portanto esse banco de dados sozinho é ineficiente para o treinamento, contudo, ainda podemos utilizá-lo para testar a acurácia do modelo, após treinado.

3 *Benchmarks*

3.0.1 Plataformas

Infelizmente não se apresentam na comunidade e na literatura grandes quantidades de *benchmarks* contendo os trabalhos que foram desenvolvidos e os resultados alcançados. Entretanto, existem algumas plataformas que estão crescendo e apresentam a tarefa de detecção de crises epilépticas com alguns trabalhos catalogados. Entre elas o *Papers With Code* (CODE, s.d.), ou também o *Kaggle* (KAGGLE, s.d.).

3.0.2 Métricas

Para avaliar os modelos, a literatura utiliza métricas como a taxa de acertos, os falsos positivos e falsos negativos (GODOY, s.d.). Já que a tarefa é uma classificação binária, os meios de avaliação da resposta são mais simples em relação a outras tarefas de aprendizado de máquina. Além disso, métodos de avaliação dos modelos como a complexidade, a curva de treinamento e a avaliação de hiper-parâmetros permitem a análise mais aprofundada do modelo e promove consequências diretas nos resultados.

4 *Proposta*

Portanto, pretendemos aplicar os diferentes métodos de aprendizado de máquina para detectar momentos de uma crise epilética. Ao final, iremos comparar a acurácia e a complexidade de cada método para esse determinado problema e discutir dos resultados obtidos.

Por fim, Vale salientar que o objetivo é fazer uma classificação binária em positivo ou negativo para um processo epilético dado diversos intervalos diferentes de tempo. O resultado esperado está representado na figura 4.1.

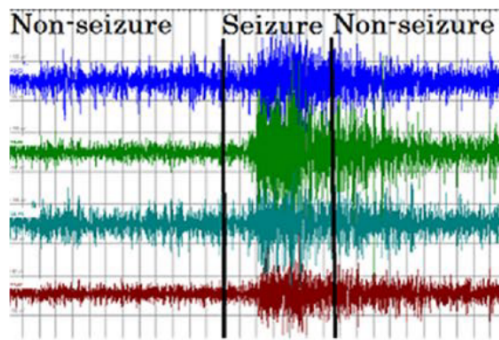


Figura 4.1: Classificação binária em uma amostra EEG

Referências bibliográficas

AGARWAL, M. **EEG-Datasets**. [S.l.: s.n.]. <https://github.com/meagmohit/EEG-Datasets>. Acessado no dia 21 de abril de 2022.

AMATO. **O que é epilepsia e convulsão. Epilepsia tem cura**. [S.l.: s.n.]. <https://www.amato.com.br/o-que-e-epilepsia-e-convulsao-epilepsia-tem-cura/>. Acessado no dia 21 de abril de 2022.

CHB-MIT SCALP EEG DATABASE. [S.l.: s.n.]. <https://physionet.org/content/chbmit/1.0.0/>. Acessado no dia 21 de abril de 2022.

CHILDRENS HOSPITAL OF PITTSBURGH. **Electroencephalogram (EEG) and Video EEG**. [S.l.: s.n.]. <https://www.chp.edu/our-services/brain/neurology/patient-procedures/eeg>. Acessado no dia 21 de abril de 2022.

CODE, P. W. **Seizure Detection**. [S.l.: s.n.]. <https://paperswithcode.com/task/seizure-detection>. Accessed 25 April 2022.

GODOY, R. V. d. **Predição de crises epiléticas utilizando técnicas de Deep Learning**. Tese (Doutorado) – Universidade de São Paulo.

KAGGLE. **EEG Seizure Analysis Dataset**. [S.l.: s.n.]. <https://www.kaggle.com/datasets/adibadea/chbmitseizuredataset>. Accessed 25 April 2022.

NUTRICIA, E. D. **Epilepsia infantil: tipos, causas, sintomas e o que fazer durante uma crise**. [S.l.: s.n.]. <https://www.danonenutricia.com.br/infantil/primeiros-meses/saude/epilepsia-infantil-tipos-causas-sintomas-e-o-que-fazer-durante-um-ataque-epiletico>. Acessado no dia 21 de abril de 2022.

RIERA, A. **EEG Signal Processing for Dummies**. [S.l.: s.n.]. <https://www.neuroelectrics.com/blog/2014/12/18/eeg-signal-processing-for-dummies/>. Acessado no dia 21 de abril de 2022.

SIDDIQUI, M. K.; MORALES-MENENDEZ, R.; HUANG, X.; HUSSAIN, N. A review of epileptic seizure detection using machine learning classifiers. **Brain Informatics**, Springer Berlin Heidelberg, v. 7, n. 1, p. 5, dez. 2020. ISSN 2198-4018. DOI: 10.1186/s40708-020-00105-1. Disponível em: <<https://doi.org/10.1186/s40708-020-00105-1>%20<https://braininformatics.springeropen.com/articles/10.1186/s40708-020-00105-1>>.