

INFLUÊNCIA DA REDUÇÃO DO NÚMERO DE CANAIS DE EEG NA PREDIÇÃO DE SURTOS EPILÉPTICOS BASEADA EM MÁQUINAS DE VETORES DE SUPORTE

THIAGO KODAMA¹, MIGUEL ANGELO DE ABREU DE SOUSA², SARA DERESTE DOS SANTOS³, RICARDO PIRES⁴

¹Graduando em Engenharia Eletrônica, Bolsista PIBITI, IFSP, *Campus* São Paulo, tk.thiago0.tk@gmail.com

²Doutor em Ciências, USP, Docente do Departamento de Elétrica, IFSP, *Campus* São Paulo, angelo@ifsp.edu.br

³Doutora em Ciências, USP, Docente do Departamento de Elétrica, IFSP, *Campus* São Paulo, sarad@ifsp.edu.br

⁴Doutor em Sistemas Automáticos e Microeletrônicos, Université Montpellier II, Docente do Departamento de Elétrica, IFSP, *Campus* São Paulo, ricardo_pires@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 3.13.01.01-0 Processamento de Sinais Biológicos

RESUMO: A epilepsia é uma das doenças neurológicas mais comuns no mundo. É caracterizada por crises causadas por distúrbios na atividade elétrica do cérebro, a qual é medida na forma do chamado eletroencefalograma (EEG), usando-se, normalmente, mais de 20 canais. Há grande interesse no desenvolvimento de sistemas para a previsão da ocorrência das crises, a serem usados pelos pacientes. Porém, um dispositivo apenas se tornaria viável se fosse pequeno, leve e pouco perceptível pelas demais pessoas. Em função disso, neste trabalho, é investigada uma forma de reduzir a quantidade de canais do EEG sem afetar a qualidade da previsão de um surto epilético. Através da técnica de Análise de Discriminante Linear (LDA) foi possível reduzir a quantidade de canais de 23 para apenas 4, mantendo a acurácia do sistema de classificação, que consiste em Máquinas de Vetores de Suporte (SVM), acima de 90% para a maioria dos casos avaliados. Também se verificou que pacientes podem apresentar crises incompatíveis entre si, o que dificulta, nesses casos, a previsão de um surto mesmo com o uso de todos os canais do EEG.

PALAVRAS-CHAVE: *Epilepsia; Predição; Redução de dimensionalidade; Análise de Discriminante Linear; Máquinas de Vetores de Suporte.*

INFLUENCE OF REDUCING THE NUMBER OF EEG CHANNELS ON THE PREDICTION OF EPILEPTIC OUTBREAKS BASED ON SUPPORT VECTOR MACHINES

ABSTRACT: Epilepsy is one of the most common neurological diseases in the world. It is characterized by crises caused by disturbances in the electrical activity of the brain, which is measured in the form of the so-called electroencephalogram (EEG), normally using more than 20 channels. There is great interest in the development of systems for predicting the occurrence of crises, to be used by patients. However, a device would only become viable if it were small, light and barely noticeable to others. As a result, this work investigates a way to reduce the amount of EEG channels without affecting the quality of the prediction of an epileptic outbreak. Through the technique of Linear Discriminant Analysis (LDA) it was possible to reduce the number of channels from 23 to just 4, maintaining the accuracy of the classification system, which consists of Support Vector Machines (SVM), above 90% for the most cases evaluated. It was also found that patients may present crises that are incompatible with each other, which makes it difficult, in these cases, to predict an outbreak even with the use of all EEG channels.

KEYWORDS: *Epilepsy; Prediction; Dimensionality reduction; Linear Discriminant Analysis; Support Vector Machines.*

INTRODUÇÃO

A epilepsia é uma das doenças neurológicas mais comuns no mundo, atingindo mais de 50 milhões de pessoas de todas as idades (EPILEPSY FOUNDATION, 2019; WORLD HEALTH ORGANIZATION, 2019).

¹ tk.thiago0.tk@gmail.com

² {angelo; ricardo_pires; sarad}@ifsp.edu.br

A atividade elétrica do cérebro é medida na forma do chamado eletroencefalograma (EEG), usando-se, normalmente, mais de 20 eletrodos metálicos (SEECK et al., 2017). Usualmente, é empregado o chamado sistema 10-20 de posicionamento de eletrodos na superfície do crânio, sistema este padronizado internacionalmente e descrito por Malmivuo e Plonsey (1995). O EEG, portanto, manifesta-se como ondas ao longo do tempo, uma para cada eletrodo. Cada um desses eletrodos e sua onda correspondente é chamado de canal. O EEG é usado nos principais testes de diagnósticos para a epilepsia, além de ser usado no diagnóstico de outros distúrbios cerebrais (MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH, 2019).

As opções para tratamento da epilepsia são medicamentos ou, em casos mais raros, cirurgia. Esses medicamentos não são eficazes para uma grande parcela dos pacientes (GADHOUMI, 2016). Por isso, há grande interesse no desenvolvimento de sistemas para a previsão da ocorrência das crises epiléticas, a serem usados pelos pacientes em suas vidas cotidianas. Se um tal sistema pudesse advertir seu usuário da iminência de uma crise epilética, com boa confiabilidade e boa antecedência, ao ser advertido, o usuário poderia se colocar em situação de segurança.

Nesse contexto, este projeto tem como objetivo o desenvolvimento de um sistema para previsão de crises que faça uso de poucos canais do EEG, a fim de tornar o protótipo mais atrativo para o usuário final.

MATERIAL E MÉTODOS

Os sinais EEG utilizados neste trabalho foram obtidos de vários indivíduos e estão disponíveis em PHYSIONET (2010) (SHOEB, 2009). Cada indivíduo é identificado pela nomenclatura ChbXX, com XX referindo-se ao seu número.

Foram selecionadas crises epiléticas no banco de dados para serem usadas na fase de treinamento do sistema e separadas das demais crises para teste, com o cuidado de se terem, para cada paciente, crises para treino e para teste. Respeitou-se a ideia de que o sistema deve ser testado, a cada vez, recebendo dados do mesmo paciente para o qual foi treinado. Mas, os dados de treino são sempre diferentes dos dados de teste.

As durações das janelas de EEG usadas (4 segundos) e a forma de montagem dos vetores de características a partir delas (usando a transformada Wavelet) foram aquelas apresentadas em Elgohary, Eldawlatly e Khalil (2016), por terem sido investigadas em detalhe por aqueles autores e por estarem em uso, com sucesso, em projetos do grupo de pesquisa onde este trabalho foi conduzido (KITANO, 2018).

Como os sinais EEG de PHYSIONET têm 23 canais e o método de Elgohary, Eldawlatly e Khalil (2016) faz com que se obtenha um número para cada canal, cada vetor de características original, a princípio, ficou com 23 componentes. Para cada crise epilética usada no treinamento do sistema, foi aplicado o método de Bandarabadi (2015) para a identificação da duração do período pré-ictal. Os vetores de características originais daquela crise foram, então, rotulados como pertencentes ao período interictal ou pré-ictal usando a duração encontrada pelo método. Para cada crise usada no treino, os vetores rotulados foram submetidos à Análise de Discriminante Linear (LDA). Para isso, foi usada a biblioteca computacional scikit-learn (PEDREGOSA, 2011), na linguagem Python. Como a aplicação da LDA requer um grande esforço computacional, o qual cresce rapidamente com o volume de dados envolvidos no cálculo, sua aplicação neste trabalho se deu por amostragem. Foram, então, escolhidos aleatoriamente alguns vetores, cujas quantidades são dadas para três pacientes na Tabela 1. Foi aplicada a LDA a eles, para a identificação das direções principais. As 23 direções originais foram classificadas em ordem de importância, da maior para a menor. Este processo foi repetido mil vezes para vetores de cada paciente, variando-se os vetores escolhidos aleatoriamente a cada vez, para se obter uma média da classificação de cada direção dentre as 23 direções.

Tabela 1: Quantidade de vetores escolhidos aleatoriamente comparado com o total de vetores. “NU” indica “não utilizado”, devido ao fato de a crise não ser compatível com crises usadas no treino da SVM (conforme explicado na seção Resultados).

	Chb01		Chb02		Chb05	
Crise	Escolhidos	Total	Escolhidos	Total	Escolhidos	Total

1	650	749	750	843	90	105
2	250	367	NU	NU	NU	NU
3	NU	NU	25	33	450	580
4	NU	NU	NU	NU	500	613
5	350	430	NU	NU	450	587

A partir desse ponto, testes de classificação de janelas de EEG que não foram usadas no treino foram feitos em duas versões: com vetores completos (23 componentes originais) e com vetores reduzidos. A classificação foi realizada usando-se Máquinas de Vetores de Suporte, comumente usadas nesse tipo de trabalho e implementadas na biblioteca LIBSVM (CHANG, LIN, 2011).

RESULTADOS E DISCUSSÃO

Os resultados dos cálculos para a identificação das direções principais usando LDA, para os três pacientes de Physionet, estão na Tabela 2. Ela mostra a média da classificação de cada canal (direção) dentre os 23 canais originais. Quanto mais baixo for o valor da classificação, maior é a importância do canal. Por exemplo, para o paciente Chb01, vê-se que o canal 7 é o mais importante para se detectar se o estado é o interictal ou o pré-ictal, já que sua média de classificação foi 1,0 (primeiro colocado).

Tabela 2: Média da classificação (importância) de cada canal dentre os 23 canais originais, obtida por meio de LDA.

canal	Média de classificação		
	Chb01	Chb02	Chb05
1	11,49	20,12	20,08
2	13,30	1,00	3,67
3	16,31	5,88	21,59
4	8,33	18,99	22,51
5	6,00	21,94	10,90
6	22,12	9,01	16,27
7	1,00	14,36	8,31
8	7,82	10,98	9,40
9	22,76	6,36	16,02
10	13,71	2,00	11,54
11	3,00	3,00	15,97
12	21,11	23,00	2,00
13	17,88	8,00	5,08
14	4,90	4,12	13,82

15	4,10	12,59	7,00
16	2,00	16,61	3,33
17	19,92	20,49	8,80
18	15,85	17,36	17,73
19	16,27	5,64	21,63
20	18,76	17,49	13,18
21	7,85	14,50	19,15
22	10,72	10,01	1,00
23	10,79	12,55	7,04

Nos experimentos realizados neste trabalho, verificou-se que, para alguns pacientes, uma SVM treinada usando-se vetores de períodos interictal e pré-ictal anteriores a uma crise obtinha alta acurácia na classificação de vetores anteriores a outras crises do mesmo paciente. Nestes casos, essas crises foram aqui chamadas de “crises compatíveis entre si”. Para outros pacientes, no entanto, não foram encontradas crises compatíveis entre si, ou seja, uma SVM treinada usando-se vetores anteriores a uma crise de um desses pacientes obtinha baixa acurácia ao tentar classificar vetores anteriores a outras crises dele. Estes casos foram marcados como NU (“não utilizado”) na Tabela 1. A Tabela 3 apresenta os quatro melhores canais (correspondentes às quatro primeiras direções principais) para cada paciente do Physionet para os quais foram encontradas crises compatíveis. Para aqueles pacientes, a Tabela 4 mostra as acurácias na classificação de vetores para crises compatíveis, em duas versões: “normal”, em que foram mantidos dados dos 23 canais originais nos vetores, e “reduzido”, em que foram mantidos somente os dados dos quatro melhores canais identificados pela LDA.

Tabela 3: Pacientes com crises compatíveis e melhores canais.

Paciente	Canal A	Canal B	Canal C	Canal D
Chb01	C3-P3	C4-P4	T8-P8	P8-O2
Chb02	F7-T7	F4-C4	C4-P4	F8-T8
Chb05	F7-T7	P4-O2	P8-O2	FT10-T8

Tabela 4: Comparação da média de acurácia entre vetores normais e reduzidos para três pacientes do banco de dados.

Crise	Chb01		Chb02		Chb05	
	Normal	Reduzido	Normal	Reduzido	Normal	Reduzido
1	99%	97%	90%	89%	97%	95%
2	97%	83%	NU	NU	NU	NU
3	NU	NU	95%	90%	94%	70%
4	NU	NU	NU	NU	99%	95%

5	99%	98%	NU	NU	99%	94%
---	-----	-----	----	----	-----	-----

CONCLUSÕES

Verificou-se experimentalmente que, com o método aqui utilizado, mesmo mantendo-se os 23 componentes originais dos vetores, uma SVM treinada usando-se vetores de períodos interictal e pré-ictal anteriores a uma crise de um paciente nem sempre obtém alta acurácia na classificação de vetores anteriores a outras crises do mesmo paciente. As crises para as quais se obteve alta acurácia no processo de treino e teste foram aqui chamadas de “crises compatíveis entre si”. Para alguns pacientes, não foram encontradas crises compatíveis entre si.

Para as crises compatíveis entre si, verificou-se que uma redução no número de canais de 23 para apenas 4 causou, no pior caso, uma redução de 94% para 70% na acurácia na classificação. Como era esperado, devido à grande redução no número de canais, houve perda de acurácia em todos os casos testados. Mas, na maioria dos casos, essa perda foi muito menor do que a do pior caso, conforme a Tabela 4. Em vários casos, mesmo após a redução de 23 para 4 canais, a acurácia se manteve acima de 90%.

Verificou-se, na Tabela 3, que o conjunto dos 4 melhores canais variou entre os pacientes. Isso dificulta a construção de um aparelho padrão a ser usado por todos os pacientes. Uma alternativa para a continuação deste trabalho e para obtenção desse aparelho padrão seria escolher-se um conjunto uniforme, dentre todos os conjuntos possíveis formados por 4 canais, de forma que a pior acurácia dentre as de todos os pacientes fosse a maior possível nessa condição. Outra possibilidade seria admitir-se o uso de mais do que 4 canais, escolhendo-se um subconjunto maior dentre os canais indicados na Tabela 3, contendo os melhores canais de vários pacientes.

REFERÊNCIAS

- BANDARABADI, M. et al. On the proper selection of preictal period for seizure prediction. **Epilepsy & Behavior**, v. 46, p. 158-166, 2015.
- CHANG, C-C.; LIN, C-J. LIBSVM: A library for support vector machines. **ACM transactions on intelligent systems and technology (TIST)**, v. 2, n. 3, p. 27, 2011.
- ELGOHARY, S.; ELDAWLATLY, S.; KHALIL, M. I. Epileptic seizure prediction using zero-crossings analysis of EEG wavelet detail coefficients. In: **2016 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)**. IEEE, 2016. p. 1-6.
- EPILEPSY FOUNDATION. Disponível em <https://www.epilepsy.com/>. Acesso em: 18 abr. 2019.
- GADHOUMI, K. et al. Seizure prediction for therapeutic devices: A review. **Journal of neuroscience methods**, v. 260, p. 270-282, 2016.
- KITANO, L. A. S. et al. Epileptic Seizure Prediction from EEG Signals Using Unsupervised Learning and a Polling-Based Decision Process. **Artificial Neural Networks**
- MALMIVUO, J.; PLONSEY, R.; **Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields**. Capítulo 13, **Electroencephalography**. Oxford University Press. 1995. Disponível em <http://www.bem.fi/book/13/13.htm#03>. Acesso em: 18 abr. 2019.
- MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. EEG (electroencephalogram). 2019. Disponível em <https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875>. Acesso em: 18 abr. 2019.
- PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python, **Journal of Machine Learning Research**, v. 12, pp. 2825-2830, 2011.
- SEECK, M. et al. The standardized EEG electrode array of the IFCN. **Clinical Neurophysiology**, v. 128, p. 2070-2077, 2017.
- SHOEB, Ali Hossam. **Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment**. PhD Thesis, Massachusetts Institute of Technology, Cambridge, 2009.