

16º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2025

ESTUDO DO CICLO SONO-VIGÍLIA A PARTIR DA ANÁLISE DE SINAIS ELETROENCEFALOGRÁFICOS DE RATOS COM EPILEPSIA GENÉTICA (GEAS)

MARINA M. REZENDE¹, CAROLINE S. BRITO², RICARDO PIRES³, GABRIELA B. BATISTA⁴,
ANGELA CRISTINA DO VALLE⁵, SARA D. DOS SANTOS⁶

¹ Graduanda em Engenharia de Controle e Automação, Bolsista do CNPq, IFSP, Campus São Paulo, marina.rezende@aluno.ifsp.edu.br

² Graduanda em Engenharia de Controle e Automação, Bolsista do CNPq, Campus São Paulo, caroline.brito@ifsp.edu.br

³ Doutor em Sistemas Automáticos e Microeletrônicos, U.M.II, Docente do IFSP, Campus São Paulo, ricardo.pires@ifsp.edu.br

⁴ Graduada em Ciências Biológicas pela FMU gabibenicio98@gmail.com

⁵ Doutora em Neurociências pela UNIFESP, Pesquisadora em Neurociências, FMUSP angela.valle@fm.usp.br

⁶ Doutora em Ciências, USP, Docente do Departamento de Elétrica, IFSP, Campus São Paulo, sarad@ifsp.edu.br
Área de conhecimento (Tabela CNPq): 3.13.01.01-0 - processamento de sinais biológicos

RESUMO: A epilepsia é uma condição neurológica caracterizada por crises recorrentes que afeta mais de 50 milhões de pessoas. A pesquisa com modelos animais permite que cientistas descrevam aspectos da fisiopatologia das epilepsias, contribuindo com dados para a clínica. Um modelo animal com crises epiléticas de ausência foi desenvolvido no Laboratório de Neurociências na Faculdade de Medicina da Universidade de São Paulo e pode ser o primeiro modelo de origem genética para o estudo das crises de ausência. Para controle das crises, desenvolvem-se tratamentos com técnicas de neuromodulação não-invasiva, baseadas na modulação de redes neurais específicas por correntes elétricas de baixa intensidade. Uma análise acurada dos eventos epileptiformes presentes nos sinais eletroencefalográficos (EEG) é fundamental para a compreensão da fenomenologia das crises e sua relação com o ciclo vigília-sono. Nesse contexto, um programa de computador foi desenvolvido, baseado em um sistema com aprendizado de máquina que, a partir dos sinais EEG dos animais estudados, classificou, com acurácia de 99,74%, os períodos de vigília-sono dessa amostra e quantificou as horas de cada estado. Esses dados serão usados pelo grupo de pesquisa XXXX (a ser revelado apenas na versão final deste artigo) na compreensão dos efeitos da neuromodulação nos animais estudados. do Laboratório de Neurociências Prof. Cesar Timo-Iaria, LIM 01, FMUSP.

PALAVRAS-CHAVE: máquina de vetor de suporte; aprendizado de máquina; epilepsia; sono-vigília

STUDY OF SLEEP-WAKE CYCLE BASED ON THE ANALYSIS OF ELECTROENCEPHALOGRAPHIC SIGNAL FROM RATS WITH GENETIC EPILEPSY (GEAS)

ABSTRACT: Epilepsy is a neurological condition characterized by recurrent spontaneous epileptic seizures that affects approximately 50 million people. Research using animal models allows scientists to better describe the pathophysiology of epilepsy. In Brazil, a new animal model with absence seizures has been established at the Faculty of Medicine of the University of São Paulo (FMUSP), and it may be the first genetically based model for studying the mechanisms of spike-and-wave complex (SWC) generation in absence seizures and for improving the understanding of the pathophysiology of this phenomenon. One of the most recent adjunctive treatment methods is neuromodulation, whose mechanism is based on the modulation of specific neural networks through low-intensity electrical currents. An accurate analysis of epileptiform events present in electroencephalographic (EEG) signals is essential for a better understanding of the phenomenology of seizures and their relationship with the sleep-wake cycle. In this context, a computer program was developed, based on a machine-learning system that, based on the EEG signals of the studied animals, classified, with 99.74% accuracy, the

wake-sleep periods of these individuals, in addition to quantifying the hours of each state. This data will be used by the XXXX research group (disclosed only in the final version) and will help understand the effects of neuromodulation on the studied animals.

KEYWORDS: support vector machine; machine learning; epilepsy, sleep-wake

INTRODUÇÃO

A epilepsia é uma condição neurológica caracterizada por crises epiléticas recorrentes que afeta aproximadamente 50 milhões de pessoas. As crises emergem de um aumento súbito na excitabilidade de uma determinada população neuronal, resultando em descargas elétricas hiper-sincronizadas. Elas são classificadas em focais ou generalizadas, acompanhadas, ou não, por abalos musculares e convulsões e são comumente registradas em aparelhos eletrencefalográficos (EEG) para efeitos de monitoramento (ILAE, 1989; 2024). Embora os mecanismos exatos das crises ainda não sejam completamente compreendidos, diversos estudos apontam para uma relação com a regulação neural do sono, seja do ponto de vista neuroanatômico e funcional, como comportamental. Do ponto de vista estrutural, inúmeros autores têm descrito uma correlação entre as redes tálamo-corticais, fundamentais para a geração das fases do sono não-REM e uma atividade disfuncional para geração das espículas-onda que culmina na eclosão das crises. Há a constatação de que a privação do sono leva a um aumento na frequência das crises e esse aumento piora significativamente a qualidade do sono dos pacientes. A pesquisa com modelos animais permite que cientistas possam descrever melhor a fisiopatologia das epilepsias (Jafarian et al., 2020).

A área biológica pode ser beneficiada com o uso de ferramentas baseadas em processamento de sinais e aprendizagem de máquina, a qual consiste na habilidade de os sistemas computadorizados aprenderem a partir de um conjunto de dados (IA) (Obeid; Selesnick; Picone, 2020; Aripin et al, 2024). O aprendizado supervisionado requer uma fase de treinamento do sistema, pela apresentação de um conjunto de dados de entrada e de sua respectiva saída esperada (Haykin, 2009). Pode ser realizado pelas máquinas de vetores de suporte (*Support Vector Machine*, SVM), desenvolvidas para resolver problemas de classificação binária usando um hiperplano separador ótimo entre as classes (Cortes; Vapnik, 1995).

Este trabalho visa ao desenvolvimento de um programa de computador que, a partir dos sinais EEG de animais estudados, classifique os seus períodos de vigília-sono e quantifique as horas de cada estado para uso do grupo de pesquisa do XXXXX Laboratório de Neurociências Prof. Cesar Timo-Iaria, LIM 01, FMUSP.

FUNDAMENTAÇÃO TEÓRICA

A atividade elétrica cerebral manifestada, captada e registrada no eletroencefalograma (EEG), é comumente caracterizada pelas frequências de suas oscilações, as quais são identificadas por meio da transformada discreta de Fourier (Sharma; Meena, 2024). Este é um modelo matemático que, em um intervalo de tempo, encontra a representação na forma de raias espectrais, no domínio da frequência, de um sinal amostrado no domínio de tempo. A Transformada Rápida de Fourier (*Fast Fourier Transform*, FFT) é uma implementação dessa transformada, executada com velocidade relativamente alta (Oppenheim; Schaffer, 2009). Uma das utilidades dessa transformada é extrair dados que caracterizam bem sinais de várias naturezas, para entregá-los a algoritmos de aprendizagem de máquina.

A Máquina de Vetores de Suporte, do inglês, *Support Vector Machine* (SVM) é um modelo de aprendizado de máquina para classificação, que busca a divisão de dados em duas classes por meio de um hiperplano linear. Para a análise não linear dos dados, é necessária a aplicação de uma função *kernel* (linear, sigmóide, polinomial ou RBF) que, por diferentes modelos matemáticos, aumenta a dimensão do conjunto de dados, aumentando-se a probabilidade de se encontrar um hiperplano separador entre as classes (Cortes; Vapnik, 1995).

Após o ajuste da SVM, considerando-se que vetores x a serem classificados sejam mapeados pelo *kernel* para um espaço com mais dimensões, onde são chamados z , a função que indica a classe de cada dado (vetor) z é dada na equação 1, onde α_i são multiplicadores de Lagrange, z_i são vetores de suporte naquele espaço, b_0 é o termo que descreve o deslocamento do hiperplano de separação em

relação à origem do sistema de coordenadas, o ponto representa produto escalar e *sign* é a função que retorna o sinal de seu argumento (Cortes; Vapnik, 1995). Ela retorna sinal positivo para vetores de uma classe e negativo para vetores da outra classe.

$$I(z) = \text{sign}\left(\sum_{\text{support vectors}} \alpha_i z_i \cdot z + b_0\right) \quad (1)$$

Os vetores de suporte são definidos, no espaço original, pelas condições de margem apresentadas nas equações 2 e 3. Para dados da classe positiva $y_i = +1$, enquanto para a classe negativa $y_i = -1$ (Cortes; Vapnik, 1995). Estas condições são unificadas na Equação 4

$$w \cdot x_i + b \geq 1 \quad \text{se } y_i = 1 \quad (2)$$

$$w \cdot x_i + b \leq -1 \quad \text{se } y_i = -1 \quad (3)$$

$$y_i(w \cdot x_i + b) \geq 1 \quad i = 1, \dots, l \quad (4)$$

Para a análise não linear dos dados, pode ser necessária a aplicação de um *kernel* (linear, sigmoide, polinomial ou RBF), que por diferentes modelos matemáticos, aumenta a dimensão do conjunto de dados.

Ao subtrair um $\xi_i \leq 0$ ao lado direito da desigualdade, como observado na equação 5, permite-se um grau de erro à equação 4 (Cortes; Vapnik, 1995). Esse recurso é considerado no método pela letra C, sendo muito utilizado para a tolerância para valores discrepantes.

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, l \quad (5)$$

A equação 6 representa o equilíbrio entre maximizar a margem e minimizar os erros de classificação. O parâmetro C controla esse compromisso: valores altos de C penalizam mais os erros, resultando em margens menores mas com menos violações, enquanto valores baixos permitem margens maiores com mais tolerância a erros.

$$\frac{1}{2}w^2 + C F\left(\sum_{i=1}^l \xi_i^\sigma\right) \quad (6)$$

Para o kernel RBF, existe também o parâmetro γ , que determina os raios de abrangência de cada classe em torno dos vetores de suporte. Matematicamente, é o valor de γ , na equação 7, que controla a largura da função gaussiana, sendo que valores mais altos geram fronteiras mais localizadas, enquanto valores mais baixos geram fronteiras mais largas (Cortes; Vapnik, 1995).

$$K(u, v) = \exp\left(-\gamma|u - v|^2\right) \quad (7)$$

Na prática, esses parâmetros podem ser alterados para que se avalie o impacto no desempenho da SVM no processo de classificação. Para se obter a melhor escolha, no que se refere à combinação desses parâmetros, é possível utilizar um recurso chamado *grid*, que testa muitas combinações automaticamente e fornece aquela com melhor acurácia. No pacote em linguagem Python scikit-learn (Pedregosa et al, 2011), o *grid* faz com que sejam percorridas e testadas automaticamente faixas de valores de C e de γ escolhidas pelo usuário.

Uma outra forma de se buscar o melhor desempenho de um sistema com aprendizagem de máquina é empregar a técnica de validação cruzada. Trata-se de um método que divide o conjunto de dados em subconjuntos, sendo que um deles é reservado para teste enquanto os demais participam do treinamento do sistema. Esse processo é realizado utilizando diferentes combinações dos subconjuntos, até todos serem usados para ambas as funções. Ao final, o desempenho obtido é a média de todas as iterações (Ertel, 2017).

Para avaliar o modelo de classificação, é feita uma matriz de confusão, que visa a comparar os valores previstos como positivos, separando-os naqueles que realmente eram positivos (verdadeiros positivos, VP) daqueles que eram falsos (falso positivo, FP), e a mesma coisa é feita para os valores previstos como negativos, dividindo-os em verdadeiros negativos (VN) e falsos negativos (FN) (Stehman, 1997). Com essa matriz é possível calcular vários parâmetros que descrevem o desempenho

de um sistema com aprendizagem de máquina. A acurácia é um deles, sendo definida pela equação 8. Trata-se de uma métrica que indica a proporção de todos os acertos em função de todos os dados. Além disso, é possível calcular o F1-score (equação 9), que equivale à média harmônica entre a acurácia e proporção de todos os verdadeiros positivos com todos os dados positivos (Tharwat, 2021).

$$Ac = \frac{VP+VN}{VP+VN+FP+FN} \quad (8)$$

$$F1 = \frac{2VP}{2VP+FP+FN} \quad (9)$$

MATERIAL E MÉTODOS

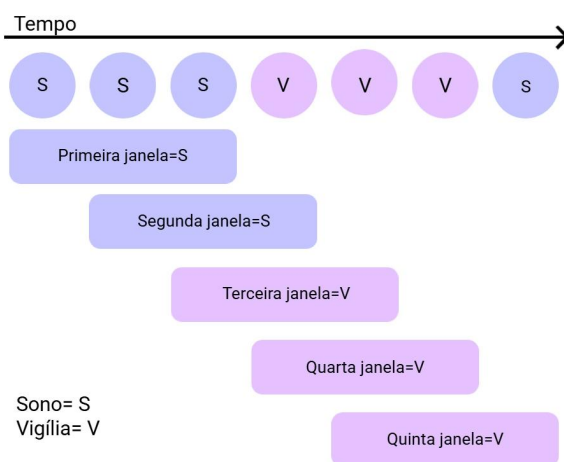
O laboratório parceiro forneceu um documento em formato de texto com medições, com frequência de 256 Hz, de três canais do EEG de um rato durante 24 horas. (A aprovação por um comitê de ética tinha ocorrido naquela instituição.) Ademais, foi fornecida uma tabela descrevendo os eventos de acordo com o intervalo de tempo, cujas classes são: crise, vigília, sono e evento de EPI. Para a leitura dos arquivos foi feita com a linguagem Python. Para o primeiro arquivo, foi necessário formatá-lo e transformá-lo em uma lista. Para o segundo, foi necessário atribuir um valor para cada uma das classes, gerando uma tabela apenas com valores de 0 a 3 de acordo com a classe e com o intervalo de tempo, a classe 1 sendo para os dados em crise, classe 2 para vigília, classe 3 para sono, e classe 0 para eventos de EPI e linhas sem marcação. Posteriormente, cada intervalo de tempo foi expandido, de modo a transformá-lo na quantidade de linhas com aquela rotulagem, como exemplo: um intervalo de 8 segundos em que a rotulagem era três, foi transformado em oito linhas de rotulagem três.

A FFT foi aplicada a cada janela de um segundo do arquivo com as medições do EEG para medir a média das raízes de frequência cerebral em cada trecho do sinal, tornando possível resumir e facilitar a análise dos dados.

Foram retiradas dos arquivos aquelas linhas equivalentes à crise e evento de EPI, já que o presente trabalho se debruçou apenas nos períodos de sono e vigília dos animais. Após isso, foi selecionada a mesma quantidade de linhas de cada classe, pois cerca de 80% dos dados eram de sono e, para maior acurácia da SVM treinada, é necessário que as classes estejam equilibradas.

Com estes dados, foi possível fazer o treino da SVM. O *grid* foi utilizado para encontrar a melhor combinação entre parâmetros. A fim de aumentar a acurácia, foi utilizada a validação cruzada e a votação temporal. Esta última consiste na análise de uma sequência de janelas previstas pela SVM, em número ímpar, de forma a se decidir as classes (sono ou vigília) com base na maioria dos votos dessa sequência. Esse processo pode ser melhor compreendido pela Figura 1. Foram testadas sequências de votação de 3 a 301. Pela análise dos dados anteriores e posteriores de um item, atribui-se ao elemento do centro da sequência a classe da maioria dos itens presentes nessa sequência.

Figura 1 - Classificação por votação temporal.



Fonte: Autor

Por fim, foi criada a interface gráfica, utilizando a biblioteca Tkinter, que organiza os dados já classificados, assim como informa ao usuário sobre o programa e o carregamento de dados, tornando-o acessível para aqueles que não possuem conhecimento técnico na área de programação.

RESULTADOS E DISCUSSÃO

Ao treinar a SVM com um dos três canais disponíveis, alterando apenas os *kernels*, a maior acurácia obtida foi para o polinomial, chegando a 51,27%. Visando a aumentá-la, foi aplicado o *grid*, sendo o melhor resultado com o kernel RBF, com os parâmetros $C=1$ e $\gamma=0.001$, alcançando a acurácia de 62,07%. Com a validação cruzada, a acurácia foi de 51,46% e após a votação temporal, subiu para 63,54%.

Em busca de melhores resultados, foi escolhido fazer a análise dos três canais EEG juntos. Nesse caso, após a votação temporal, a acurácia foi de 88,75%. Expandindo-se o tamanho das janelas de votação, o melhor desempenho foi para as sequências de 201, 213 e 225 janelas, com acurácia de 99,74% para todas essas três.

Após a definição dos parâmetros, foi feita a matriz de confusão, e calculado o F1-score que resultou, assim como a acurácia, em 99,74% (Tabela 1).

Tabela 1- Matriz de confusão para sequências contendo 201 janelas.

		Valor predito	
		Vigília	Sono
Real	Vigília	11299	59
	Sono	0	11709

Com a obtenção desses resultados, foi possível passar à criação da interface gráfica que possui um botão para a seleção do arquivo que será avaliado, uma tabela que possui uma coluna para o intervalo de tempo, em segundos, uma para o resultado vigília, e outra para o resultado sono, que serão marcadas pelo programa com um “X”, de acordo com a classe. Mais abaixo, é possível observar um botão para salvar a avaliação feita pela SVM. Há um botão com informações sobre a criação do programa. Como resultado é apresentado, além da tabela já citada, o tempo total avaliado, assim como a duração dos períodos de sono e de vigília. Esta interface pode ser observada na Figura 2.

Figura 2 - Interface gráfica

Intervalo de Tempo	Vigília	Sono
0-529		X
529-531	X	
531-534		X
534-582	X	
582-584		X
584-585	X	
585-593		X
593-600	X	
600-602		X
602-610	X	

Tempo total: 02:10:11.000
Tempo de sono: 01:37:50.000 (75.2%)
Tempo de vigília: 00:32:21.000 (24.8%)

Fonte: Autor

CONCLUSÕES

O presente trabalho alcançou o objetivo de criar um modelo de aprendizado de máquina para classificar trechos de sinais de EEG em sono ou vigília. Além disso, uma interface gráfica foi desenvolvida para o melhor manuseio da ferramenta pelo grupo de pesquisa parceiro. Naturalmente, o programa criado depende do modo como os sinais EEG são gravados e rotulados por aquele grupo, necessitando de adaptações para se ajustar a outros tipos de registros.

Apesar dos resultados promissores, com acurácia e F1-score em 99,74%, o sistema, que consiste em uma SVM com kernel RBF ($C=1$ e $\gamma=0.001$) e método de votação com sequências de 201 janelas, por exemplo, pode ser aprimorado para analisar um conjunto maior de exames e diferentes indivíduos, analisando se as previsões do modelo seguirão satisfatórias. No contexto atual, os sinais EEG dos três canais disponíveis dos animais foram usados. Como oportunidade futura, também é possível considerar a inclusão de outros sinais, tais como o eletromiográfico (EMG) no desempenho do sistema.

CONTRIBUIÇÕES DOS AUTORES

Todos os autores contribuíram com a revisão do trabalho e aprovaram a versão submetida.

AGRADECIMENTOS

Os autores agradecem ao Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação (PIBITI) do CNPq pelas bolsas concedidas a XXXXX M. M. Rezende e a C. S. Brito.

REFERÊNCIAS

ARIPIN, Ipin et al. Artificial intelligence in biology and learning biology: A literature review. *Jurnal Mangifera Edu*, v. 8, n. 2, p. 41-48, 2024.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, v. 20, n. 3, p. 273-297, 1995.

ERTEL, W. **Introduction to Artificial Intelligence**. Second Edition. Springer International Publishing, 2017.

JAFARIAN, M.; ALIPOUR, M. E.; KARIMZADEH, F. **Experimental models of absence epilepsy**. *Basic and Clinical Neuroscience*, v. 11, n. 6, p. 715, 2020.

HAYKIN, Simon. *Neural Networks and Learning Machines*. 3. ed. New Jersey: Prentice Hall, 2009.

OBEID, JP Iyad; SELESNICK, Ivan; PICONE, Joseph. **Signal processing in medicine and biology**. Springer, 2020.

OPPENHEIM, A.; SCHAFER, R. **Discrete-Time Signal Processing**. Prentice Hall Signal Processing. 3rd Edition. ISBN-13 : 978-0131988422. 2009.

PEDREGOSA *et al.* **Scikit-learn: Machine Learning in Python**, *JMLR* 12, pp. 2825-2830, 2011.

SHARMA, Ramnivas; MEENA, Hemant Kumar. **Emerging trends in EEG signal processing: A systematic review**. *SN Computer Science*, v. 5, n. 4, p. 415, 2024.

STEHMAN, Stephen V. **Selecting and interpreting measures of thematic classification accuracy**. *Remote sensing of Environment*, v. 62, n. 1, p. 77-89, 1997.

THARWAT, A. **Classification assessment methods**. *Applied Computing and Informatics* 4 January 2021; 17 (1): 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>