

Classificação de Janelas de Eletroencefalogramas com e sem Crises Epilépticas

Davi C. Nascimento* Gean Carlos L. Sousa**
Jonathan A. Queiroz*** Allan Kardec D. Barros****

* *Universidade Federal do Maranhão, MA, (e-mail:
davi.nascimento777@gmail.com).*

** *Universidade Federal do Maranhão, MA (e-mail:
gean_m@ yahoo.com.br)*

*** *Universidade Federal do Maranhão, MA, (e-mail:
prof.queiroz.jonathan@gmail.com)*

**** *Universidade Federal do Maranhão, MA, (e-mail:
akduailibe@gmail.com)*

Abstract: Epilepsy is a chronic neurological disease that increases an individual's predisposition to recurrent seizures. In aggressive cases, seizures can lead to injury, trauma or even sudden death. People suffering from epilepsy can live free from seizures since they have been diagnosed receiving adequate treatment. Therefore, methodologies that simplify and streamline the diagnosis and treatment of these individuals are valid and necessary. The objective of this work is the development of a new model of classification of electroencephalogram (EEG) windows with or without epileptic seizures. This methodology is based on the calculation of metrics that depend on the second, third and fourth order statistical moments. There were two stages of classification. The first with the feature vectors obtained and the second with the rotation of the same vectors, using principal component analysis (PCA). 1 second windows were extracted from the CHB-MIT (Children's Hospital Boston) databases. The proposed classification model obtained for the feature vectors and for the components from PCA, accuracy of 86.4 % and 94.6 %, respectively. It was believed that with appropriate adaptations and improvements, the model can be embedded in a device, for classification of windows in real time.

Resumo: A epilepsia é uma doença neurológica crônica, que aumenta a predisposição de um indivíduo sofrer convulsões recorrentes. Em casos mais agressivos as crises podem promover lesões, traumatismos ou mesmo morte súbita. Pessoas que sofrem de epilepsia poderiam viver livres de convulsões, desde que fossem diagnosticadas previamente ou recebessem um tratamento adequado. Portanto, metodologias que simplifiquem e agilizem o diagnóstico e o tratamento destes indivíduos são válidas e necessárias. Esse trabalho tem por objetivo a apresentação de uma metodologia de identificação de janelas de eletroencefalogramas (EEG's) com presença ou ausência de crises epiléticas. Essa metodologia é baseada no cálculo de métricas que dependem dos momentos estatísticos de segunda, terceira e quarta ordem. Foram realizadas duas etapas de classificação. A primeira com os vetores de características obtidos a partir dos momentos estatísticos, a segunda com a rotação dos mesmo vetores, utilizando a análise de componentes principais (PCA). Foram extraídas janelas de 1 segundo de arquivos da base de dados CHB-MIT (Children's Hospital Boston). O modelo de classificação proposto obteve para os vetores de características e para as componentes oriundas do PCA, acurácia de 86.4% e 94.6%, respectivamente. Acreditasse que com devidas adaptações melhorias o modelo pode ser embarcado em um dispositivo, para classificação de janelas em tempo real.

Keywords: Epilepsy; Electroencephalogram; Machine Learning; Statistics; Feature Extraction.

Palavras-chaves: Epilepsia; Eletroencefalograma; Aprendizado de Máquina; Estatística; Extração de Características.

1. INTRODUÇÃO

A epilepsia consiste em uma alteração temporária do funcionamento do cérebro devido, atividades neuronais síncronas e excessivas. Tais alterações no comportamento neural promovem eventos involuntários que vão desde alterações na consciência, eventos involuntários de natureza muscular, psíquica, sensitiva e autonômica, de acordo com Thurman et al. (2011). O diagnóstico de epilepsia é comumente definido pela ocorrência de *duas crises epiléticas* não provocadas, separadas por um intervalo de mais de 24 horas, Yacubian and Kochen (2014). A epilepsia, portanto é uma doença crônica, caracterizada pela predisposição persistente do cérebro de gerar crises epiléticas recorrentes, Fisher et al. (2014).

Estimasse que em todo mundo em torno de 65 milhões de pessoas possuem epilepsia, Ngugi et al. (2010). Cerca de 70% dessas pessoas poderiam viver livres de convulsões, se fossem diagnosticadas previamente ou recebessem o tratamento adequado, WHO (2019). Em casos mais extremos, a ocorrência de crises pode aumentar o risco de um indivíduo sofrer queimaduras, lacerações, fraturas cranianas e até mesmo morte súbita e inesperada, Ficker (2000).

Logo, verifica-se que o desenvolvimento de tecnologias que facilitem o diagnóstico da epilepsia, e que auxiliem no tratamento de pacientes faz-se necessário.

A base de dados CHB-MIT (Children's Hospital Boston - Massachusetts Institute of Technology), Shoeb and Gutttag (2010), tem sido objeto de estudo de vários pesquisadores, que já sugeriram soluções para o problema de identificação de padrões eletroencefalográficos.

Podemos destacar Khan et al. (2012), que ao extrair os coeficientes de variação, variação normalizada e de variação relativa, obteve 91,8% de acurácia ao classificar registros com e sem crises epiléticas. Para esse fim foi o utilizando o LDA. Fergus et al. (2015) ainda no objetivo de classificar registros com e sem crise, desenvolveu um modelo de classificação baseado em aprendizado de máquina supervisionado. Extraíndo características como frequência média e picos de frequência, e utilizando novamente o k-NN, foram alcançados 98% de acurácia. Mesclando características do domínio do tempo e da frequência, Nandy et al. (2019) obteve 97,05% de acurácia utilizando o SVM com kernel RBF.

Vários métodos para identificar crises epiléticas são propostos. Utilizando análises de eletroencefalogramas (EEG), técnicas de processamento de sinais, aplicações de algoritmos de aprendizado de máquina mais tradicionais, como Shoeb et al. (2011), e nos últimos anos abordando a problemática com uso de aprendizado profundo, semelhante a Acharya et al. (2018); Thodoroff et al. (2016). No entanto, verifica-se que estas técnicas computacionais, muitas vezes, não chega a prática médica, talvez pela complexidade dos algoritmos e a difícil compreensão dessas metodologias.

Dada a dificuldade de se estabelecer uma metodologia de fácil compreensão e que seja capaz de identificar, com precisão, padrões de crises epiléticas é que nesse artigo

propõe-se que conceitos estatísticos sejam aplicados a análise de sinais de EEG's para a análise e classificação de janelas com e sem episódios de epilepsia.

Outro fator que torna a identificação de episódios de epilepsia tarefa importante é a possibilidade de identificação de início de crises epiléticas tal como Ahammad et al. (2014); Shoeb and Gutttag (2010).

Diante do que foi exposto, objetivo deste trabalho é o desenvolvimento de um modelo de classificação, e de extração de características, de janelas de EEG com presença e ausência de crises epiléticas baseado no cálculo de momentos estatísticos.

2. BASE DE DADOS

Os exames utilizados nas simulações computacionais, a fim de testar a metodologia proposta, são sinais de eletroencefalogramas disponíveis na base de dados CHB-MIT (Children's Hospital Boston - Massachusetts Institute of Technology), Shoeb and Gutttag (2010). Essa base possui gravações de 23 pacientes, homens e mulheres, de 1,5 a 22 anos de idade. Foram realizadas entre 9 e 42 coletas de EEG para cada paciente. O conjunto de dados é composto por 664 arquivos, totalizando 916 horas de EEG's. Entre esses EEG's, 129 são rotulados com crise. Todos com frequência amostral de 256 Hz.

3. METODOLOGIA

3.1 Pré-processamento

Ao analisar a base de dados observou-se que em alguns EEG's existem canais repetidos ou gerados artificialmente. Também foram identificados sinais biológicos diferentes, como eletrocardiograma e estímulo do nervo vago. Tais sinais foram removidos dos EEG's.

Foram identificadas também diferentes montagens de EEG dentro da base de dados. Em alguns casos foram adotadas mais de uma montagem para o mesmo paciente. Foi identificada a montagem de EEG com maior ocorrência dentro da base (FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10 e FT10-T8). EEG's com montagens diferentes da Figura 1 não foram utilizados nas demais etapas desse trabalho.

A partir dos EEG's rotulados como positivos para epilepsia, foram retirados os seus respectivos períodos de crise. Concatenando os trechos dos EEG's com os episódios de crise foi possível então, formar uma matriz $\mathbf{A}$. De forma semelhante, foram retirados dos demais EEG's períodos com ausência de crises e os mesmos foram concatenados para gerar outra matriz $\mathbf{B}$.

$$\mathbf{A}_{m,n} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \quad e \quad \mathbf{B}_{m,n} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{bmatrix}$$

em que m corresponde ao número de canais do EEG e n o número de amostras.

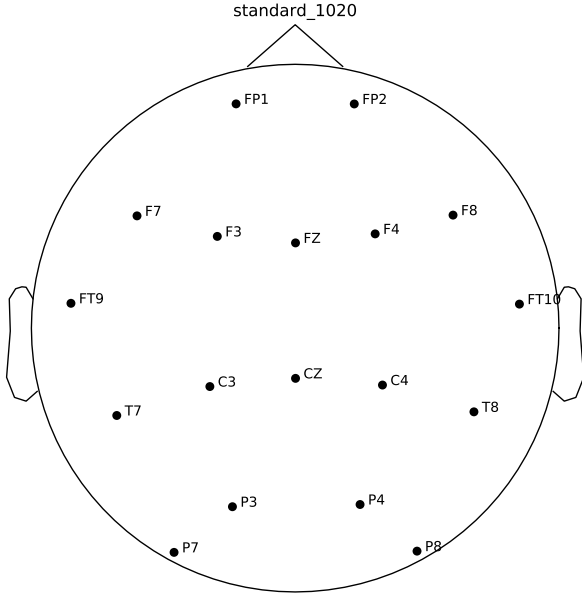


Figura 1. Montagem de EEG adotada.

3.2 Extração de Características e Momentos Estatísticos

O processo de extração de características proposto nesse artigo tem como base o cálculo de momentos estatísticos sobre janelas do EEG resultante da concatenação realizada na definição das matrizes **A** e **B**. A cada matriz citada foi aplicado o janelamento, tal como ilustrado na Figura 2.

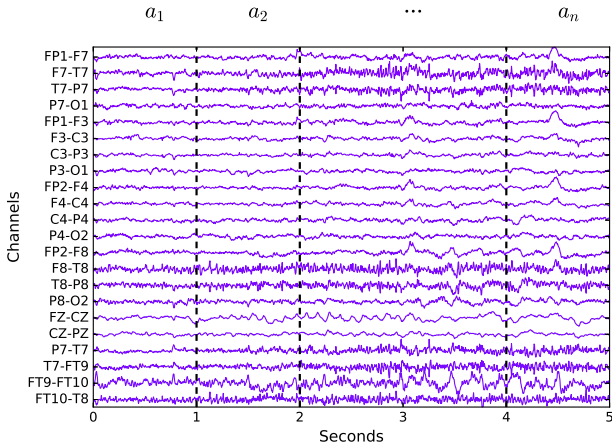


Figura 2. Janelamento aplicado, exemplificado na matriz **A**.

A proposta de metodologia desse trabalho, consiste em calcular a partir de cada janela J (Fig. 2), métricas que dependem dos momentos estatísticos de segunda, terceira e quarta ordem. Um momento estatístico de ordem n , de uma variável aleatória X , é definido de acordo com a seguinte integral, Papoulis and Pillai (2002):

$$m_n = E\{\mathbf{x}^n\} = \int_{-\infty}^{+\infty} x^n f(x) dx. \quad (1)$$

em que $f(x)$ é a função densidades de X e E o operador esperança. Nessa mesma perspectiva, é definido também o k -ésimo momento central ($k > 1$) (Eq. 2).

$$\mu_k = E\{(\mathbf{x} - m_1)^k\} \quad (2)$$

Considerando o segundo, terceiro e quarto momento algumas métricas podem ser definidas, a saber, variância - σ^2 , assimetria - γ e curtose - κ Hyvärinen and Oja (2000); Joanes and Gill (1998).

$$\sigma^2 = \mu_2 \quad (3)$$

$$\gamma = \frac{\mu_3}{\mu_2^{3/2}} \quad (4)$$

$$\kappa = \frac{\mu_4}{\mu_2^2} - 3. \quad (5)$$

As métricas definidas nas equações (3), (4) e (5), aplicadas sob as linhas da matriz **A** geram portanto, novas matrizes $\mathbf{A}_{\sigma^2}$, $\mathbf{A}_{\gamma}$ e $\mathbf{A}_{\kappa}$.

$$\mathbf{A}_{\sigma^2} = \begin{bmatrix} \sigma_{1,1}^2 & \cdots & \sigma_{1,k}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{m,1}^2 & \cdots & \sigma_{m,k}^2 \end{bmatrix} \quad (6)$$

$$\mathbf{A}_{\gamma} = \begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,k} \\ \vdots & \ddots & \vdots \\ \gamma_{m,1} & \cdots & \gamma_{m,k} \end{bmatrix} \quad (7)$$

$$\mathbf{A}_{\kappa} = \begin{bmatrix} \kappa_{1,1} & \cdots & \kappa_{1,k} \\ \vdots & \ddots & \vdots \\ \kappa_{m,1} & \cdots & \kappa_{m,k} \end{bmatrix} \quad (8)$$

Por fim, são obtidos vetores de características definidos pela média das colunas das matrizes $\mathbf{A}_{\sigma^2}$, $\mathbf{A}_{\gamma}$ e $\mathbf{A}_{\kappa}$:

$$v_{\sigma^2} = \left(\frac{\sum_{i=1}^m \sigma_{i,1}^2}{m}, \frac{\sum_{i=1}^m \sigma_{i,2}^2}{m}, \dots, \frac{\sum_{i=1}^m \sigma_{i,k}^2}{m} \right), \quad (9)$$

$$v_{\gamma} = \left(\frac{\sum_{i=1}^m \gamma_{i,1}}{m}, \frac{\sum_{i=1}^m \gamma_{i,2}}{m}, \dots, \frac{\sum_{i=1}^m \gamma_{i,k}}{m} \right), \quad (10)$$

$$v_{\kappa} = \left(\frac{\sum_{i=1}^m \kappa_{i,1}}{m}, \frac{\sum_{i=1}^m \kappa_{i,2}}{m}, \dots, \frac{\sum_{i=1}^m \kappa_{i,k}}{m} \right). \quad (11)$$

O mesmo processo é realizado considerando a matriz **B**, e de modo análogo se obtém três vetores de características. Os vetores de características das matrizes **A** e **B** são então utilizados como entrada nos algoritmos de classificação que serão apresentados posteriormente.

3.3 Análise de Componentes Principais

Além de utilizar os vetores de características como entradas de algoritmos de classificação, a metodologia proposta também sugere que esses vetores podem ser representados em um novo espaço cujos eixos representam a maximização

das variações dos mesmos. Para essa tarefa foi então aplicada a Análise de Componentes Principais - PCA, Duda et al. (2012). Essa técnica algébrica-estatística procura representar dados n-dimensionais em um novo espaço vetorial de menor ou igual dimensão, Webb (2003).

Desde que $\mathbf{X}$ seja uma matriz composta pelos vetores de características e $\mathbf{Y}$ a representação de $\mathbf{X}$ no novo espaço obtido pela PCA. Procura-se então, encontrar uma matriz de rotação $\mathbf{R}(n \times n)$ tal que $\mathbf{Y} = \mathbf{R}\mathbf{X}$ implica em $var(y_1) \geq var(y_2) \geq var(y_3)$ e $corr(y_i, y_j) = 0; \forall i \neq j; i, j = 1, 2, 3$ e ainda $var(y_1)$ seja máxima. É conhecido que $\mathbf{R}$ é composta pelos autovetores da matriz de autocovariância dos dados e que $var(y_i)$ é o respectivo i -ésimo autovalor.

3.4 Classificadores

A proposta de classificação dos vetores de características e das componentes do PCA fez uso de cinco algoritmos de classificação descritos a seguir.

Análise Discriminante Quadrática A Análise Discriminante Quadrática (QDA), trata-se de um algoritmo de classificação que pressupõe que cada uma das classe do *dataset* têm sua única matriz de covariância Backes and Junior (2016). De modo que sua fórmula é dada por

$$P(c_i|\vec{x}) = -\frac{1}{2} \ln \left| \sum_i \left| -\frac{1}{2} (\vec{x} - \vec{\mu}_i)^T \sum_i (\vec{x} - \vec{\mu}_i) + \ln P(c_i) \right. \right| \quad (12)$$

Naive Bayes O Naive Bayes é um algoritmo de aprendizado supervisionado o qual aplica a teoria de Bayes Theodoridis et al. (2010).

$$P(\omega_i|x)p(x) = p(x|\omega_i)P(\omega_i). \quad (13)$$

Onde $P(\omega_i)$ é a probabilidade a priori da classe ω_i e $P(\omega_i|x)$ a probabilidade a posteriori da classe ω_i ter o valor de x . Dentre as variações do classificador está o Naive Bayes Gaussiano. Seu uso assume que os vetores de característica são distribuições Gaussianas Chan et al. (1982):

$$P(x|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right). \quad (14)$$

A máquina de vetores de suporte (SVM) A máquina de vetores de suporte (SVM) tem o objetivo de produzir um modelo de classificação binária. Vetores de treinamento são mapeados em um espaço de determinada dimensão. Então o SVM define um hiperplano de separação e suas devidas margens. Hsu et al. (2003). Para casos de dados linearmente separáveis é adotado o seu kernel linear:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j. \quad (15)$$

Em casos onde não é possível a separação linear das amostras será necessário o uso do kernel de função de base radial (RBF). Dessa forma as amostras serão mapeadas não-linearmente em um espaço de dimensão superior:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0. \quad (16)$$

K-vizinho mais próximo (k-NN) No K-vizinho mais próximo (k-NN), de forma geral, ocorre uma memorização do conjunto de treinamento. Ao ser inserido uma nova amostra, a esta será atribuído o rótulo de seus k vizinhos mais próximos Shalev-Shwartz and Ben-David (2014). A proximidade de uma nova amostra em relação aos seus vizinhos é definida pela distância euclidiana de seus vetores de características Artificial (2011)

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^d (x_i^l - x_j^l)^2}. \quad (17)$$

Redes Neurais Artificiais As redes neurais Multi Layer Perceptron (MLP) têm como principal característica a presença de pelo menos uma camada neurônios oculta, entre a camada de entrada e a camada de saída Da Silva et al. (2010). Logo uma rede MLP possuirá no mínimo duas camadas de neurônios, distribuídos entre camadas intermediárias e a cama da de saída.

3.5 Métricas de Avaliação

Posteriormente foram obtidos os percentuais de acurácia, sensibilidade e especificidade de cada um dos classificadores

$$Accuracy = \frac{VP + VN}{VP + VN + FN + FP} \times 100. \quad (18)$$

$$Sensitivity = \frac{VP}{VP + FN} \times 100. \quad (19)$$

$$Specificity = \frac{VN}{VN + FP} \times 100. \quad (20)$$

Onde VP corresponde ao número de verdadeiros positivos para presença de crises epiléticas, VN os verdadeiros negativos para ausência de crise, FP para registros com crises classificados com ausência e FN para classificações positivas em registros com ausência. Como técnica de validação cruzada foi usado o método k-fold Friedman et al. (2001), tendo $k = 10$. Adotou-se a área sob a curva ROC (AUC) Prati et al. (2008) como critério geral para comparação de desempenho entre os classificadores.

Por fim, a sequência de etapas da metodologia proposta para extração de características e classificação é ilustrada pela Figura 3.

4. RESULTADOS

4.1 Extração de Características

Após a extração de características, foi feita uma análise visual em três dimensões dos dois grupos de janelas. E também os gráficos de dispersão de cada uma das características estatísticas. As Figuras 4 e 5 ilustram tais informações.

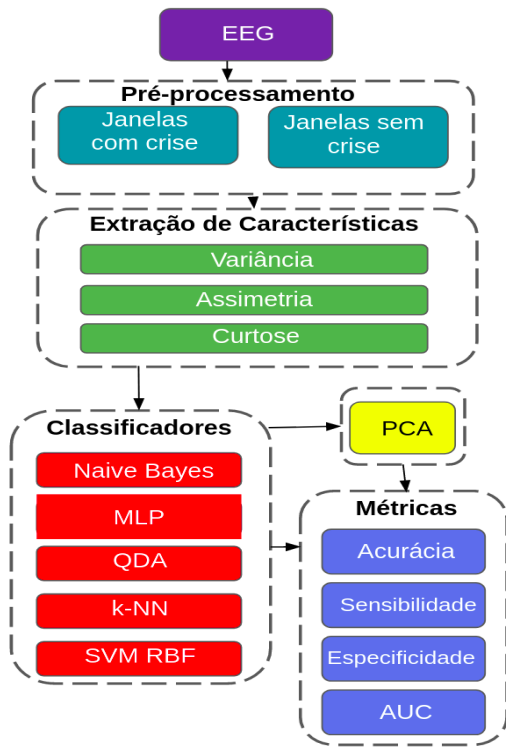


Figura 3. Fluxograma para ilustração da extração de características e classificação de janelas de EEG.

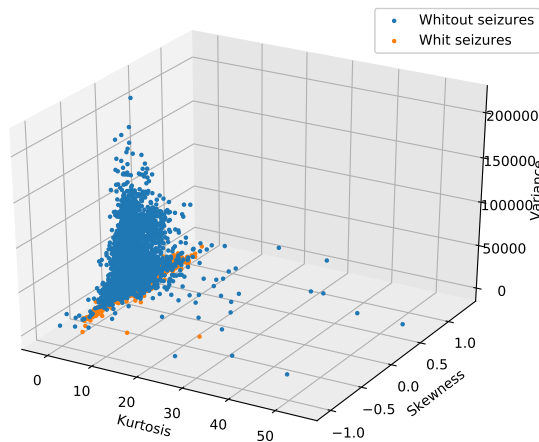


Figura 4. Características estatísticas

Nota-se aparente aglomeração os grupos na Figura 4. Mas após aplicação de transformação logarítmica percebe-se que há uma separação entre os grupos, ilustrada pela Figura 5.

Em seguida foi aplicado o PCA para redução em 3 componentes principais, afim de alterar a distribuição dos vetores de característica. As figuras a seguir ilustram gráficos de dispersão 3D das componentes obtidas pelo PCA. Observa-se um reposicionamento de janelas com crises na Figura 6. E uma expressiva separação dos grupos após aplicação de transformação logarítmica nas componentes do PCA, ilustrada pela Figura 7.

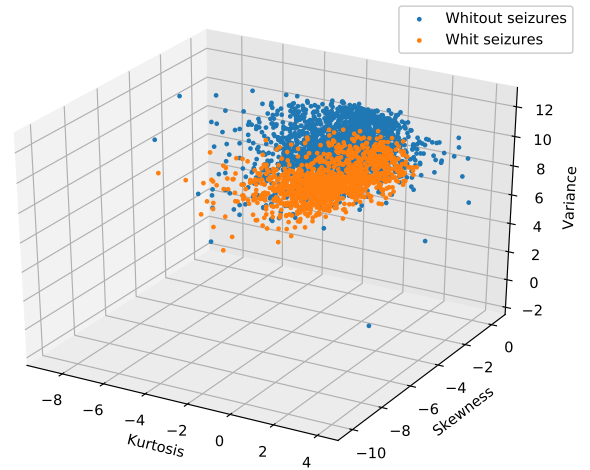


Figura 5. Log das características estatísticas

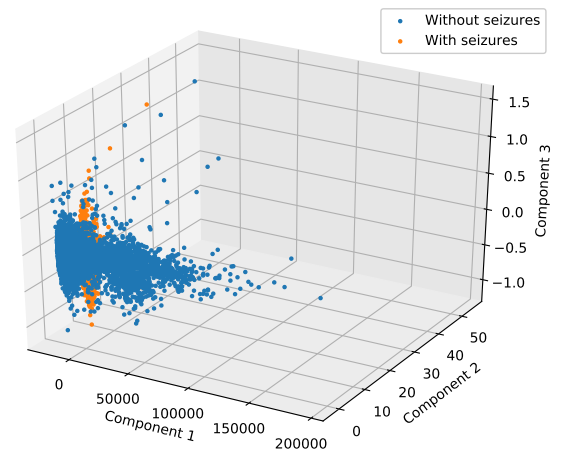


Figura 6. Componentes do PCA

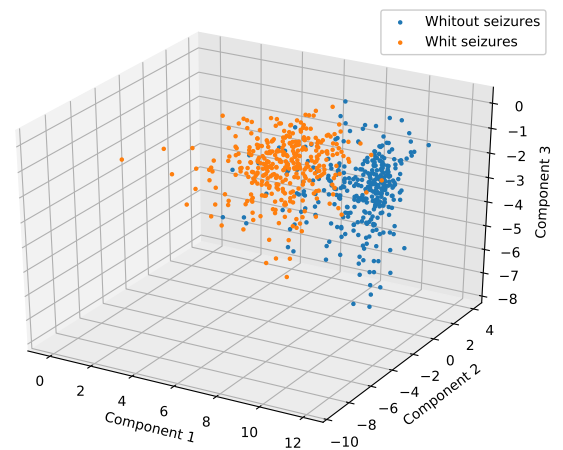


Figura 7. Log das componentes do PCA

4.2 Classificação

Foram realizadas duas etapas de classificação. Classificação utilizando os vetores de característica dos dois grupos de janelas e classificação utilizando as componentes do

PCA dos grupos. Foram estabelecidos 80% dos dados para treino e 20% para teste.

Ambas as etapas de classificações usaram os algoritmos: SVM com kernel RBF, $C = 10.0$ e $\gamma = \frac{1}{n \times \sigma^2(X)}$; k-NN com $k = 5$; uma MLP com duas camadas ocultas com 5 e 2 neurônios respectivamente, com função de ativação linear retificada e otimização de pesos pela função Adam; Naive Bayes Gaussiano e QDA. Vale ressaltar que não houve um método específico para otimização dos hiperparâmetros dos classificadores. Todos usaram parâmetros atribuídos por padrão na biblioteca *Scikit Learn*, Pedregosa et al. (2011)

Na Tabela 1 estão contidos os valores das métricas de classificação (acurácia, sensibilidade e especificidade) das classificações dos vetores de característica.

Tabela 1. Métricas de classificação dos vetores de característica.

Metrics	SVM-RBF	k-NN	QDA	MLP	Naive Bayes
Accuracy	85.5	86.4	82.6	50.0	83.2
Sensitivity	84.1	88.7	95.3	30.0	96.5
Specificity	86.9	84.2	69.88	70.0	70.0
AUC	0.86	0.91	0.9	0.5	0.91

A Figura 8 ilustra a curva ROC das classificações dos vetores de característica, afim de fazer um comparativo entre os desempenhos de cada classificador.

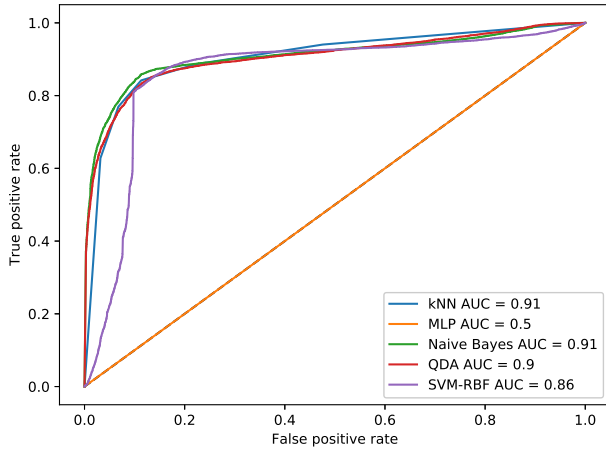


Figura 8. Curva ROC dos classificadores dos vetores de característica

A Tabela 2 mostra os valores das métricas de classificação das classificações das componentes oriundas do PCA.

Tabela 2. Métricas de classificação das componentes do PCA.

Metrics	SVM-RBF	k-NN	QDA	MLP	Naive Bayes
Accuracy	92.4	94.6	91.1	94.3	91.8
Sensitivity	96.3	93.0	87.2	90.5	87.3
Specificity	88.5	96.2	69.88	98.1	96.3
AUC	0.96	0.97	0.97	0.96	0.98

A Figura 10 ilustra o comparativo entre os desempenhos de cada classificador das componentes resultantes do PCA, através da curva ROC.

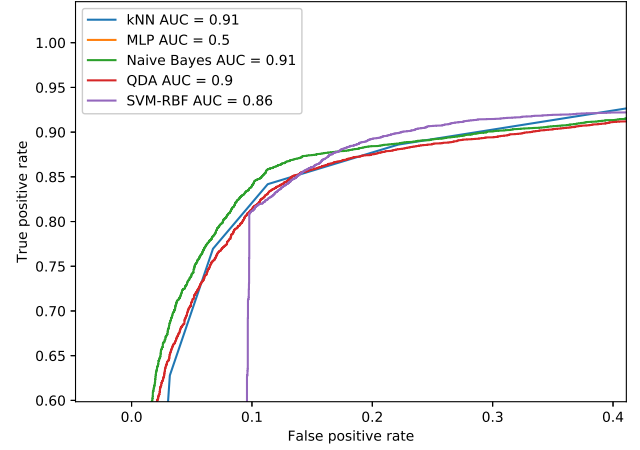


Figura 9. Zoom na curva ROC dos classificadores dos vetores de característica

5. DISCUSSÃO

Nota-se que após a extração de características ocorre maior agrupamento das janelas de EEG. Em especial a maior aglomeração de janelas com crise e alta dispersão das janelas sem crise no eixo da variância na Figura 4. A separação entre os dois grupos fica mais evidente após a transformação logarítmica, Figura 5.

Foi realizado um teste T com os vetores de características das janelas. No teste foi adotada a hipótese nula de que as amostras são de uma distribuição normal Pearson et al. (1977). Essa hipótese foi rejeitada. Foi realizado um novo teste estatístico paramétrico Mann-Whitney U-test Gibbons and Chakraborti (2011). Foi adotada a hipótese nula de que os grupos são de distribuições contínuas e com medianas iguais. Tal hipótese foi rejeitada para as três características extraídas das janelas de EEG. Os p -valores obtidos foram 0, $2,6^{-33}$ e $2,4^{-11}$, para variância, assimetria e curtose, respectivamente.

Após a aplicação do PCA houve uma mudança na disposição dos dados, tal como na Figura 6. A separação dos grupos fica ainda mais evidente após aplicação da transformação logarítmica, como mostra a Figura 7. Na classificação das componentes do PCA, os percentuais dos classificadores aumentaram, com destaque para os índices da MLP.

Na classificação dos vetores de característica todos os classificadores obtiveram valores de acurácia semelhantes, com exceção da MLP (acurácia de 50%). Durante as simulações das duas etapas de classificação o SVM e a MLP foram os mais custosos computacionalmente. Apesar dos bons índices obtidos pelos dois classificadores, o seu custo e tempo de convergência os impedi de serem aplicados em futuras aplicações para classificação de janelas em tempo real.

O Naive Bayes e QDA tiveram desempenho semelhante. Na classificação dos vetores de características o Naive Bayes obteve $AUC = 0.91$ e acurácia de 83.2%; e na classificação das componentes do PCA obteve $AUC = 0.98$ e acurácia de 91.8%. Já o QDA classificando vetores de características alcançou $AUC = 0.9$, acurácia de 82.6%;

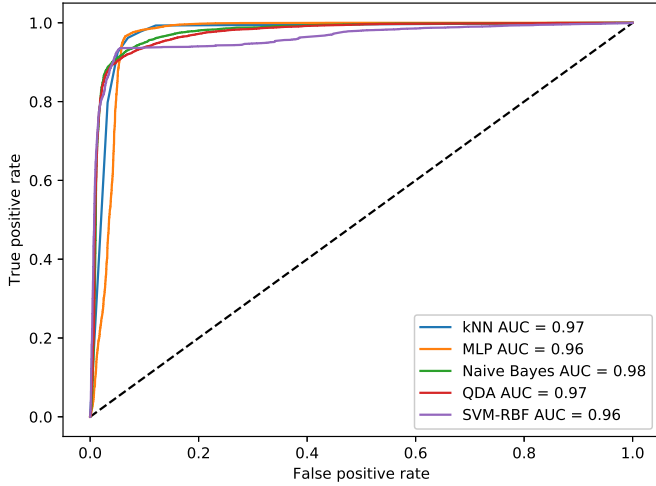


Figura 10. Curva ROC de classificação das componentes do PCA

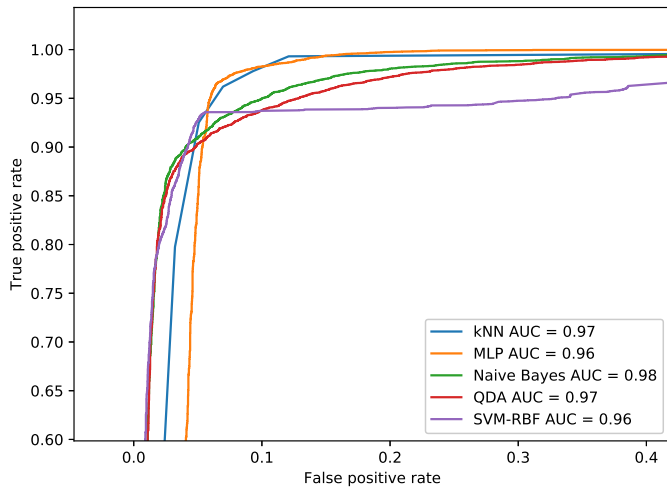


Figura 11. Zoom na curva ROC de classificação das componentes do PCA

e classificando componentes do PCA obteve $AUC = 0.97$ e acurácia de 91.1%. Com resultados tão semelhantes e assumindo que as matrizes de covariância do QDA são diagonais, temos que as entradas de cada classe são condicionalmente independentes. Logo o QDA será um classificador equivalente ao Naive Bayes Gaussiano. Porém ressalta-se que nos testes estatísticos anteriores a hipótese nula, de que os vetores de característica são de distribuições normais, foi rejeitada. Sabendo que o Naive Bayes Gaussiano admite que seus dados de entrada sejam de distribuições normais, podemos considerar que as métricas obtidas foram um acaso específico para os dados utilizados. Por fim podemos desconsiderar o uso dos dois classificadores.

Tanto nas classificações dos vetores de característica quanto nas dos componentes do PCA, o k-NN os melhores percentuais de acurácia, respectivamente 86.4% e 94%. Tendo em vista o uso da distância euclidiana para a classificação e seus bons índices de acurácia, podemos concluir que ambos os grupos foram bem separados. O que denota

eficiência da metodologia de extração de características adotada.

Em comparação a metodologias anteriores podemos notar semelhanças. Assim como Fergus et al. (2015) a melhor acurácia obtida foi através do k-NN. Porém, foram utilizadas apenas características no domínio do tempo, as quais podem ser computacionalmente melhor obtidas. Nandy et al. (2019) obteve acurácia mais elevada, porém ainda utilizando a frequência. E apesar disso usou o SVM-RBF, geralmente utilizado em classificações de dados não linearmente separáveis, o que denota que a sua extração de características não foi tão satisfatória. Khan et al. (2012) utilizou características igualmente fáceis de serem calculadas, porém a sua acurácia foi menor se comparada a do presente trabalho.

A Tabela 3 ilustra um comparativo entre os melhores resultados obtidos neste trabalho em relação as soluções citadas, tomando por base os percentuais de acurácia de cada um.

Tabela 3. Métricas de classificação das componentes do PCA.

Autor	Características	Classificador	Acurácia
Fergus, 2014	Frequência média e picos	k-NN	98%
Khan, 2012	COV, RCOV e NCOV	LDA	91.0%
Nandy, 2019	Tempo e frequência	SVM	97.05%
Este Trabalho	VAR, SKEW e KUR	k-NN	94.6%

A partir da metodologia de extração de característica proposta, bem como o uso de algoritmos de classificação, seria possível elaborar futuramente um modelo de identificação de início e fim de crises epiléticas, semelhante a Shoeb and Gutttag (2010); Shoeb et al. (2011). Tal metodologia poderia ser portada para um dispositivo embarcado. Para assim efetuar classificação de janelas em tempo real. Dessa forma seria possível auxiliar a triagem de pacientes com suspeita de epilepsia.

6. CONCLUSÃO

Neste trabalho foi proposto um modelo para classificação de janelas de eletroencefalogramas com e sem epilepsia. A extração de características gerou vetores estatisticamente distintos. E a aplicação do PCA proporcionou maior separação entre os grupos. Foram obtidos altos índices de acurácia, com destaque para o k-NN com 94.6%. A partir da metodologia proposta será possível implementar um modelo de identificação de início e término de crises epiléticas. E também uma metodologia para auxílio a triagem de pacientes com suspeita da doença.

REFERÊNCIAS

- Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., and Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Computers in biology and medicine*, 100, 270–278.
- Ahammad, N., Fathima, T., and Joseph, P. (2014). Detection of epileptic seizure event and onset using eeg. *BioMed research international*, 2014.

- Artificial, I. (2011). Uma abordagem de aprendizagem de máquina/katti facili...[et al.]. *Rio de Janeiro: LTC*.
- Backes, A.R. and Junior, J.J.d.M.S. (2016). *Introdução à visão computacional usando Matlab*. Alta Books Editora.
- Chan, T.F., Golub, G.H., and LeVeque, R.J. (1982). Updating formulae and a pairwise algorithm for computing sample variances. In *COMPSTAT 1982 5th Symposium held at Toulouse 1982*, 30–41. Springer.
- Da Silva, I., Spatti, D., and Flauzino, R. (2010). Redes neurais artificiais para engenharia e ciências aplicadas: curso prático, artliber editora ltda, são paulo, sp, brasil. Technical report, ISBN978-85-88098-53-4.
- Duda, R.O., Hart, P.E., and Stork, D.G. (2012). *Pattern classification*. John Wiley & Sons.
- Fergus, P., Hignett, D., Hussain, A., Al-Jumeily, D., and Abdel-Aziz, K. (2015). Automatic epileptic seizure detection using scalp eeg and advanced artificial intelligence techniques. *BioMed research international*, 2015.
- Ficker, D.M. (2000). Sudden unexplained death and injury in epilepsy. *Epilepsia*, 41, S7–S12.
- Fisher, R.S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J.H., Elger, C.E., Engel Jr, J., Forsgren, L., French, J.A., Glynn, M., et al. (2014). Ilae official report: a practical clinical definition of epilepsy. *Epilepsia*, 55(4), 475–482.
- Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.
- Gibbons, J.D. and Chakraborti, S. (2011). *Nonparametric statistical inference*. Springer.
- Hsu, C.W., Chang, C.C., Lin, C.J., et al. (2003). A practical guide to support vector classification.
- Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5), 411–430.
- Joanes, D. and Gill, C. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 183–189.
- Khan, Y.U., Farooq, O., and Sharma, P. (2012). Automatic detection of seizure onset in pediatric eeg. *Int. J. Emb. Syst. Appl*, 2, 81–89.
- Nandy, A., Alahe, M.A., Uddin, S.N., Alam, S., Nahid, A.A., and Awal, M.A. (2019). Feature extraction and classification of eeg signals for seizure detection. In *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 480–485. IEEE.
- Ngugi, A.K., Bottomley, C., Kleinschmidt, I., Sander, J.W., and Newton, C.R. (2010). Estimation of the burden of active and life-time epilepsy: a meta-analytic approach. *Epilepsia*, 51(5), 883–890.
- Papoulis, A. and Pillai, S.U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.
- Pearson, E.S., D “AGOSTINO, R.B., and Bowman, K.O. (1977). Tests for departure from normality: Comparison of powers. *Biometrika*, 64(2), 231–246.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12, 2825–2830.
- Prati, R.C., Batista, G., Monard, M.C., et al. (2008). Curvas roc para avaliação de classificadores. *Revista IEEE América Latina*, 6(2), 215–222.
- Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.
- Shoeb, A., Kharbouch, A., Soegaard, J., Schachter, S., and Gutttag, J. (2011). A machine-learning algorithm for detecting seizure termination in scalp eeg. *Epilepsy & Behavior*, 22, S36–S43.
- Shoeb, A.H. and Gutttag, J.V. (2010). Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 975–982.
- Theodoridis, S., Pikrakis, A., Koutroumbas, K., and Cavouras, D. (2010). *Introduction to pattern recognition: a matlab approach*. Academic Press.
- Thodoroff, P., Pineau, J., and Lim, A. (2016). Learning robust features using deep learning for automatic seizure detection. In *Machine learning for healthcare conference*, 178–190.
- Thurman, D.J., Beghi, E., Begley, C.E., Berg, A.T., Buchhalter, J.R., Ding, D., Hesdorffer, D.C., Hauser, W.A., Kazis, L., Kobau, R., et al. (2011). Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*, 52, 2–26.
- Webb, A.R. (2003). *Statistical pattern recognition*. John Wiley & Sons.
- WHO (2019). World health organization (who), epilepsy. <https://www.who.int/news-room/fact-sheets/detail/epilepsy>. Online; Acesso em: 30-maio-2020.
- Yacubian, E.M. and Kochen, S. (2014). Crises epiléticas. *São Paulo: Ed. Casa Leitura Médica*, 6.