

# Reconhecimento de estados dos olhos utilizando máquinas de aprendizado profundo a partir de ondas cerebrais

Gabriel Anísio S. Soares, Bruno Otávio P. Prado, Gilton José Ferreira da Silva,  
Hendrik Teixeira Macedo, Leonardo Nogueira Matos

Departamento de Computação  
Universidade Federal de Sergipe – São Cristóvão – SE – Brasil

{gabriel.soares, bruno, gilton, hendrik, leonardo}@dcomp.ufs.br

**Abstract.** *The brain-computer interface is one of the emerging fields of human-computer interaction due to its broad spectrum of applications, especially those that deal with human cognition. In this work, electroencephalography (EEG) is used as base data for classifying the state of the eyes (open or closed) by applying Long Short-Term Memory (LSTM) networks and variants. For benchmarking purposes, the EEG data set with the eye state record was used, available in the Machine Learning repository at UCI. The results obtained indicated that the LSTM and GRU bidirectional cells models are applicable to the classification of the data, presenting an accuracy greater than 95%, and that its performance is good compared to the more expensive models computationally.*

**Resumo.** *A interface cérebro-computador é um dos campos emergentes da interação homem-computador devido ao seu amplo espectro de aplicações, especialmente as que lidam com a cognição humana. Neste trabalho, a eletroencefalografia (EEG) é usada como dado base para classificar o estado dos olhos (abertos ou fechados) aplicando redes Long Short-Term Memory (LSTM) e variantes. Para fins de benchmarking, foi utilizado o conjunto de dados de EEG com registro do estado do olho, disponível no repositório de Aprendizado de Máquina da UCI. Os resultados obtidos indicaram que os modelos bidirecionais das células LSTM e GRU são aplicáveis na classificação dos dados, apresentando acurácia superior a 95%, e que seu desempenho é bom comparado aos modelos mais caros computacionalmente.*

## 1. Introdução

A interface cérebro-computador é um dos campos em ascensão da interação homem-computador, possui um amplo espectro de aplicações, sendo as mais proeminentes as aplicações industriais e médicas. Entre essas aplicações, destacam-se as focadas no mapeamento ou reparação da cognição ou funções motoras de pessoas que possuem seus respectivos sistemas comprometimentos [Narejo et al. 2016].

Também em destaque devido ao seu avanço abrupto nas últimas décadas, a área de Aprendizado Profundo se tornou o estado da arte em vários campos, seu conjunto de algoritmos, que tentam aprender hierarquicamente representações não lineares de dados, obteve sucesso em diversas aplicações práticas, inclusive no campo da bioinformática [Goodfellow et al. 2016, Längkvist et al. 2014].

Muitos estudos sobre sinais biomédicos, como os obtidos numa eletroencefalografia (EEG), que é o registro da atividade elétrica espontânea do cérebro durante um período de tempo, trouxeram descobertas importantes e úteis para a classificação do estado cognitivo humano [Narejo et al. 2016, Wang et al. 2014, Sabancı e Koklu 2015]. Entre eles estão trabalhos bem sucedidos que utilizaram os sinais do EEG como dado base para classificação do estado dos olhos (abertos ou fechados) afim de detectar crises epiléticas [Polat e Günes 2007], identificação de característica de estresse [Sulaiman et al. 2011], detecção de sonolência ao dirigir [Yeo et al. 2009], entre outros.

Como os sinais biomédicos representam dados naturalmente sequenciais, diversas técnicas de Aprendizado de Máquina e abordagens estatísticas podem ser empregadas para resolver problemas como os de classificação desses dados [Långkvist et al. 2014]. Entre essas técnicas, destacam-se as Redes Neurais Recorrentes, que possuem uma arquitetura de aprendizado apropriada para analisar dados e que, geralmente, produz resultados promissores.

Este trabalho tem como objetivo propor uma abordagem para a identificação do estado do olho a partir dos sinais cerebrais do EEG usando Redes Neurais Recorrentes *vanilla* (doravante RNN *vanilla*), redes *Long Short-Term Memory* (doravante LSTM), redes *Gated Recurrent Unit* (doravante GRU) e suas variantes bidirecionais (doravante BRNN, BLSTM e BGRU, respectivamente). Para facilitar a identificação das vantagens e desvantagens da metodologia de aprendizado de máquina utilizada neste trabalho, será utilizado o conjunto de dados de *benchmark EEG Eye State Data Set*<sup>1</sup> que está disponível publicamente.

O restante deste artigo está organizado como segue: na Seção 2 são apresentados os conceitos teóricos utilizados, além de uma análise da literatura acerca da tarefa em questão; na Seção 3 são detalhados os procedimentos, recursos e arranjos utilizados no desenvolvimento do trabalho; na Seção 4 são expostos e analisados os resultados obtidos a partir dos experimentos desenvolvidos; e por fim, na Seção 5 são apresentadas as considerações finais a respeito dos resultados obtidos, além de suas limitações e panoramas de continuidade futura.

## 2. Revisão da Literatura

### 2.1 Eletroencefalografia

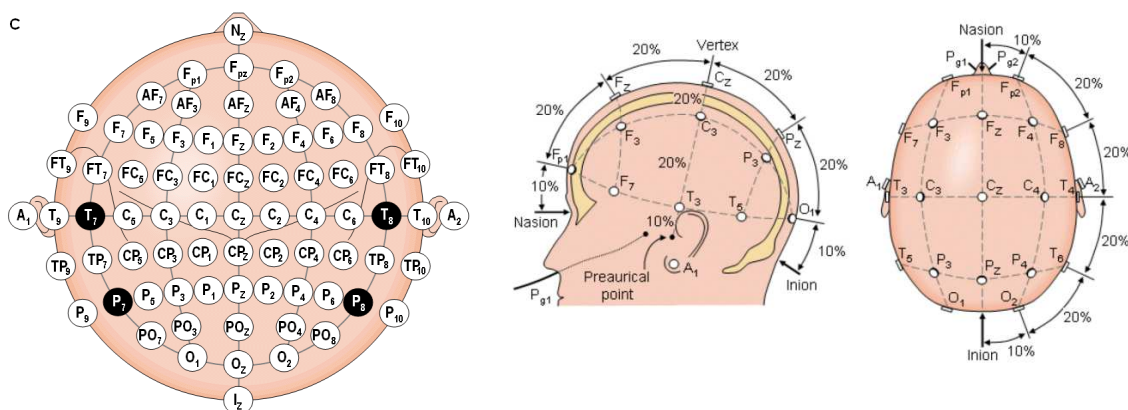
Eletroencefalografia (EEG) é a medida da atividade elétrica produzida pelo cérebro, registrada por eletrodos colocados no couro cabeludo. Esta atividade é registrada, condicionada e exibida na tela do computador. O EEG é utilizado para diagnosticar lesões cerebrais, tumores e outras anomalias de forma não invasiva, além de ser utilizado em pesquisas de neurociência, ciência cognitiva, psicologia cognitiva, neurolinguística e pesquisa psicofisiológica [Northrop 2012].

As localizações topográficas dos eletrodos são padronizadas e geralmente aderem ao sistema internacional 10-20 [Acharya et al. 2016], vide Figura 1. O “10” e o “20” referem-se ao fato de que as distâncias reais entre os eletrodos adjacentes são 10% ou 20% da distância total da frente ou da direita e esquerda do crânio. Cada posição de

---

<sup>1</sup> Disponível em <<http://archive.ics.uci.edu/ml/datasets/EEG+Eye+State>>.

gravação é identificada por uma combinação de letras e números. Por exemplo, F = Frontal, Fp = Pós frontal, C = Central, P = Parietal, O = Occipital, T = Temporal, A = Lóbulo da orelha, Pg = Faríngea, Z = Zero (linha média). Os números ímpares estão no lado esquerdo da cabeça, e os números pares estão à direita.

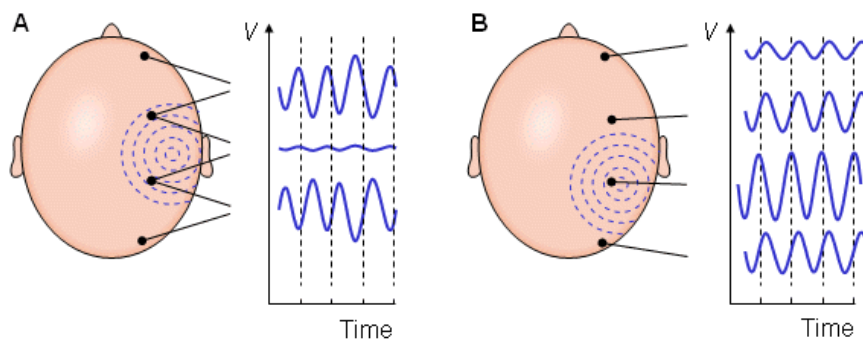


**Figura 1. (Esquerda) Localização e nomenclatura dos eletrodos, de acordo com o padrão da American Electroencephalographic Society. (Centro e Direita) Vista lateral esquerda e de topo do posicionamento dos eletrodos de acordo com o padrão.**

Fonte: [Malmivuo and Plonsey 1995].

Segundo [Northrop 2012], por ser geralmente registrada a partir do couro cabeludo, a atividade elétrica do córtex cerebral deve passar pelas membranas pia e dura-máter, líquido cefalorraquidiano, crânio e couro cabeludo. Por causa dessas estruturas anatômicas, há uma atenuação considerável em relação à atividade elétrica captada pelos eletrodos. Os maiores potenciais de EEG registrados no couro cabeludo são de aproximadamente 150  $\mu$ V-pico.

Na aferição EEG podem ser usadas configurações bipolares ou unipolares de eletrodos. Na primeira configuração (ver Figura 2, A), a diferença de potencial entre um par de eletrodos é medida. Na segunda (ver Figura 2, B), o potencial de cada eletrodo é comparado a um eletrodo neutro ou à média de todos os eletrodos (média espacial).



**Figura 2. (A) Configuração bipolar e (B) Configuração unipolar.**

Fonte: [Malmivuo and Plonsey 1995].

Os ritmos cerebrais estão, tradicionalmente, agrupados em bandas de frequência, que incluem [Cohen 2014, Northrop 2012]:

- As ondas delta ( $\delta$ ) são definidas para frequências de 2 a 4 Hz, ocorrem em adultos em sono profundo;
- As ondas teta ( $\theta$ ) são definidas para frequências de 4 a 8 Hz, ocorrem durante o sono em adultos e em crianças pré-púberes, também pode ser vista durante a meditação;
- As ondas alfa ( $\alpha$ ) são definidas para frequências de 8 a 14 Hz, ocorrem em adultos conscientes, mas relaxados, com os olhos fechados, desaparece quando os olhos estão abertos e o indivíduo se concentra em alguma tarefa;
- As ondas beta ( $\beta$ ) são definidas para frequências de 14 a 30 Hz. As ondas beta são subdivididas nos tipos I e II: as ondas beta tipo I desaparecem durante a atividade mental intensa, enquanto as tipo II aparecem.

Vale ressaltar que não existem limites precisos que definem as bandas de frequência. Além disso, as diferenças individuais nas frequências de pico estão ligadas a uma série de características individuais, incluindo estrutura do cérebro, idade, capacidade de memória de trabalho e composição química do cérebro [Cohen 2014].

Diferentes padrões de EEG em áreas distintas do cérebro estão intimamente associados com o nível de consciência do indivíduo. À medida que a atividade aumenta, o EEG muda para maior frequência dominante e menor amplitude. Estados de excitação, falta de atenção, sonolência, diferentes estágios do sono, presença de diferentes tipos de drogas e toda uma série de estados clínicos se manifestam distintamente no EEG. Quando os olhos estão fechados, as ondas alfa tendem a dominar o EEG. Quando a pessoa adormece, a frequência de EEG dominante diminui. No sono profundo, o EEG possui amplitudes grandes e frequência baixa (ondas delta) [Malmivuo e Plonsey 1995]. Durante a meditação, geralmente ocorrem aumentos de amplitude nas ondas teta e alfa e diminuição da frequência geral [Cahn e Polich 2006].

Existem diversas formas de exibir os dados do EEG, entre elas estão a visualização bruta e a espectral [Cohen 2014]. A primeira, chamada de não processada, ou EEG bruto, oferece uma excelente resolução temporal, no entanto, tendências ao longo do tempo são difíceis de avaliar em uma tela de computador, particularmente porque elas se relacionam com mudanças na frequência. A segunda maneira de visualizar o EEG é chamada processada, ou EEG espectral. É chamado processado porque o computador analisa a forma de onda afim de fornecer mais informações, como, por exemplo, mostrar como as frequências do EEG mudam ao longo do tempo.

## 2.2 Redes Neurais Recorrentes

Uma Rede Neural Recorrente, do inglês *Recurrent Neural Network* (RNN), é uma rede neural artificial em que as conexões entre as unidades ocultas formam ciclos. As Redes Neurais Recorrentes são um tipo específico de rede neural projetada para problemas que envolvem sequências [Graves 2012].

Dada uma rede *perceptron feedforward*, uma rede neural recorrente pode ser pensada como a adição de *loops* à arquitetura. Desse modo, a entrada da RNN inclui também a entrada da etapa anterior. A rede processa a entrada atual, e usa um *loop* de

*feedback* para considerar as entradas do passo anterior, também chamado de passado recente, para o contexto. As conexões recorrentes adicionam estado, ou memória, à rede e permitem que ela lide com abstrações mais amplas das sequências de entrada.

Formalmente, uma RNN *vanilla* é definida como segue [Goodfellow et al. 2016]:

$$h_t = \varphi_h(W_x x_t + W_h h_{t-1} + b_h) \quad (1)$$

$$y_t = \varphi_y(W_y h_t + b_y) \quad (2)$$

Onde  $x_t$ ,  $h_t$  e  $y_t$  são *arrays* das camadas de entrada, oculta e de saída, respectivamente;  $W_h$ ,  $W_y$  e  $W_x$  são matrizes de pesos (conexões entre neurônios);  $\varphi_h$  e  $\varphi_y$  são as funções de ativação da camada oculta e da camada de saída, respectivamente; e,  $b_x$  e  $b_y$  são o *bias* para a camada oculta e de saída, respectivamente. Os subíndices  $t$  referem-se ao período de tempo.

Ao lidar com abstrações mais amplas das sequências de entrada, as RNNs *vanilla* podem ficar suscetíveis a dificuldades de treinamento para resolver problemas que exigem o aprendizado de dependências temporais longas. Isso ocorre porque o gradiente da função de perda decai exponencialmente com o tempo (chamado de problema do desaparecimento do gradiente) [Graves 2012]. Para tratar esse problema, a literatura relacionada sugere a utilização de um conjunto de "portas" para controlar o fluxo de informações entre as unidades.

### 2.3 Long Short-Term Memory

As redes *Long Short-Term Memory* (LSTM) são um tipo de RNN com arquitetura apropriada para lidar com sequências e séries temporais, uma rede LSTM pode aprender dependências de longo prazo entre as etapas de tempo destas entradas [Hochreiter e Schmidhuber 1997, Graves 2013]. Para isso, essas redes mantêm dinamicamente em suas unidades uma célula de memória, além dos portões de entrada, esquecimento e saída que modulam o fluxo de informações dentro da unidade [Greff et al. 2017].

Estas redes abordam o problema do desaparecimento do gradiente, comumente encontrado nas RNNs, incorporando funções de ativação nas dinâmicas de estado de seus neurônios. Em cada instante de tempo, uma unidade LSTM mantém um vetor oculto  $h$  e um *array* de memória  $c$  responsável pelo controle de atualizações e saídas de estados [Zaremba et al. 2014].

Formalmente, uma rede LSTM é definida como segue [Yao et al. 2015]:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1}) \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1}) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1}) \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t) \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

Onde  $i_t$ ,  $f_t$  e  $o_t$  são chamados de portas de entrada, esquecimento e saída, respectivamente;  $c_t$  é a memória interna da unidade;  $h_t$  é o estado oculto de saída obtido pela multiplicação da memória com o portão de saída;  $\sigma$  é a função de ativação;  $W_{xi}$ ,  $W_{xf}$ ,  $W_{xc}$  e  $W_{xo}$  são matrizes de pesos da entrada;  $W_{hi}$ ,  $W_{hf}$ ,  $W_{hc}$  e  $W_{ho}$  são matrizes de pesos recorrentes;  $W_{ci}$ ,  $W_{cf}$  e  $W_{co}$  são os pesos *peephole*.

## 2.4 Gated Recurrent Unit

As redes *Gated Recurrent Unit* (GRU) são uma versão mais simples das redes LSTM (não possuem células de memória dentro de suas unidades), foram propostas por [Cho et al. 2014] para fazer com que cada unidade recorrente capture de forma adaptativa as dependências de diferentes escalas de tempo. Essa arquitetura também utiliza portões que controlam o fluxo de informações dentro do neurônio, no entanto, diferentemente da unidade LSTM, possui apenas dois portões, o de atualização e o de redefinição [Chung et al. 2014, Yao et al. 2015].

Formalmente, em relação ao instante de tempo  $t$ , uma rede GRU é definida como segue [Yao et al. 2015]:

$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t \quad (8)$$

$$z_t = \sigma(W_zx_t + U_zh_{t-1}) \quad (9)$$

$$\tilde{h}_t = \tanh(W_hx_t + U(r_t \odot h_{t-1})) \quad (10)$$

$$r_t = \sigma(W_rx_t + U_rh_{t-1}) \quad (11)$$

Onde  $z_t$  e  $r_t$  são chamados de portas de atualização e redefinição, respectivamente;  $\tilde{h}_t$  é a saída candidata;  $W_z$ ,  $W_r$ ,  $W_h$ ,  $U_z$  e  $U_r$  são matrizes de pesos.

## 2.5 Redes Neurais Recorrentes Bidirecionais

As arquiteturas mencionadas nas Seções 2.2, 2.3 e 2.4 aprendem representações das sequências a partir de etapas anteriores de tempo. No entanto, dependendo da aplicação, pode ser útil que a rede aprenda com a série temporal completa em cada etapa de tempo, como, por exemplo, em aplicações de reconhecimento de fala e de escrita [Goodfellow et al. 2016].



Uma Rede Neural Recorrente Bidirecional pode ter suas unidades baseadas nas arquiteturas apresentadas anteriormente (Seções 2.2, 2.3 e 2.4), mantendo sempre duas conexões com direções opostas: uma avançando no tempo (lidando com representações anteriores da sequência), e uma retrocedendo no tempo (lidando com representações futuras da sequência) [Goodfellow et al. 2016].

## 2.6 Métricas de avaliação

As métricas de avaliação dos modelos de aprendizado de máquina são muito importantes, tendo em vista que a escolha delas influencia em como o desempenho dos algoritmos é medido e comparado [Raschka 2015]. Este trabalho explora uma tarefa de classificação, portanto, as métricas abordadas são aplicáveis a tarefas de classificação.

### 2.6.1 Matriz de confusão

A matriz de Confusão em si não é uma medida de desempenho como tal, mas quase todas as métricas de desempenho são baseadas nela. Essa matriz é uma forma intuitiva de visualizar a exatidão e acurácia do modelo (detalhamento das classificações corretas e incorretas para cada classe) [Raschka 2015]. Os termos associados a uma matriz de confusão são os seguintes (vide Tabela 1):

- Verdadeiros Positivos (TP): são os casos em que a classe real do ponto era P (Verdadeiro) e o previsto também era P (Verdadeiro);
- Verdadeiros Negativos (TN): são os casos em que a classe real do ponto era N (Falso) e o previsto também era N (Falso);
- Falsos Positivos (FP): são os casos em que a classe real do ponto era N (Falso) e o previsto era P (Verdadeiro);
- Falsos Negativos (FN): são os casos em que a classe real do ponto era P (Verdadeiro) e o previsto era N (Falso).

**Tabela 1. Matriz de confusão**

		P	N	
P	Classe real	Verdadeiro Positivo (TP)	Falso Negativo (FN)	
N		Falso Positivo (FP)	Verdadeiro Negativo (TN)	
		Classe prevista		

Fonte: Adaptado de Raschka (2015).

### 2.6.2 Acurácia

Em problemas de classificação, acurácia é o número de previsões corretas feitas pelo modelo ( $TP + TN$ ) sobre todas as previsões efetuadas ( $TP + FP + FN + TN$ ), ou seja, o quão frequente o classificador faz a previsão correta. A acurácia é uma boa medida

quando as classes nos dados são aproximadamente balanceadas. A métrica é definida como segue [Raschka 2015]:

$$\text{Acurácia} = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

### 2.6.3 Precisão

Precisão é o número de resultados positivos corretamente classificados ( $TP$ ) dividido pelo número de todos os resultados positivos retornados pelo classificador ( $TP + FP$ ). Intuitivamente, a métrica de precisão é a capacidade do classificador de não rotular como positiva uma amostra que é negativa. A métrica é definida como segue [Raschka 2015]:

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (13)$$

### 2.6.4 Recall

*Recall* é o número de resultados positivos corretamente classificados ( $TP$ ) dividido pelo número de todas as amostras que deveriam ter sido identificadas como positivas retornadas pelo classificador ( $TP + FN$ ). Intuitivamente, a métrica de *Recall* é a capacidade do classificador de encontrar todas as amostras positivas. A métrica é definida como segue [Raschka 2015]:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (14)$$

### 2.6.5 F1-score

*F1-score* é a média harmônica entre as métricas *Precisão* e *Recall*. A métrica é definida como segue [Raschka 2015]:

$$\text{F1score} = 2 \frac{\text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (15)$$

## 2.7 Trabalhos Relacionados

A classificação do estado dos olhos é um tipo de problema de série temporal comum para detectar o estado cognitivo humano. Estudos anteriores apresentam várias abordagens aplicadas na identificação do estado ocular do EEG.

Na literatura, o *corpus* construído em [Rösler e Suendermann 2013], que registrou o estado ocular e o EEG de um indivíduo por 117 segundos, é tido como um problema de referência, e é disponibilizado pelo Repositório de Aprendizagem de Máquina, Universidade da Califórnia, Irvine (UCI) [Dheeru e Karra Taniskidou 2017]. No estudo [Rösler e Suendermann 2013], foram empregados 42 métodos diferentes de aprendizados de máquina e abordagens estatísticas utilizando o Weka [Holmes et al.



1994] para tentar prever se os olhos de um indivíduo estavam abertos ou fechados. Foi obtida uma taxa de classificação correta de 97,3% com o algoritmo K\* [Cleary e Trigg 1995].

Os pesquisadores de [Wang et al. 2014] abordaram o Aprendizado de Atributos Incrementais (IAL), no qual as *features* são gradualmente importadas para o sistema uma por vez para prever a rotulagem de classe. O estudo [Hamilton et al. 2015] desenvolveu três diferentes modelos de aprendizagem, o modelo mais preciso foi o construído a partir da combinação *Rotational Forests* (RRF) e K\* com uma acurácia de 97.4%. Já o trabalho [Kim et al. 2016] propôs um sistema baseado em um pequeno número de regras *Neuro-Fuzzy* para um problema similar de classificação. O melhor desempenho para o *corpus* em questão foi uma taxa de erro média de 4,0%.

No trabalho [Narejo et al. 2016], foi proposta a utilização de duas arquiteturas de aprendizado profundo para a tarefa de classificação do estado do olho do EEG. Especificamente *Deep Belief Network* (DBN) e *stacked AutoEncoder* (SAE). O melhor desempenho para o *corpus* em questão foi obtido pelo modelo SAE, com 98,1% de acurácia.

Já o trabalho [Bashivan et al. 2015] propôs a utilização de redes híbridas (redes neurais convolucionais e recorrentes) no intuito de absorver as características espacial, espectral e temporal do EEG. Para isso, em vez de representar recursos de EEG bruto como um vetor, transformou os dados em uma sequência de imagens multiespectrais que preservam a topologia e usa esses dados para treinar redes recorrentes e convolucionais profundas para aprender representações robustas a partir da sequência de imagens.

Todos esses trabalhos mostraram que diversas técnicas de Aprendizado de Máquina e abordagens estatísticas são viáveis na resolução do problema da classificação dos sinais cerebrais que compõem um EEG para a identificação do estado do olho de um indivíduo.

### 3. Metodologia

A metodologia empregada neste projeto contempla duas fases: a primeira envolve a classificação de padrões no EEG utilizando redes profundas LSTM, e a segunda envolve a comparação dos resultados obtidos com os da literatura.

Na primeira etapa serão implementados<sup>2</sup> modelos de redes profundas LSTM em linguagem Python, utilizando a API de redes neurais de alto nível Keras [Chollet et al. 2015] com o *backend* do *framework* TensorFlow [Abadi et al. 2015]. Os testes iniciais serão realizados com bases de dados disponíveis gratuitamente, como a base EEG disponível no repositório UCI, intitulada *EEG Eye State Data Set*, na qual todos os dados são de uma medição de EEG contínua com o dispositivo *Emotiv*<sup>3</sup> *EEG Neuroheadset*. Foram utilizados 14 eletrodos nas posições AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. O estado do olho foi detectado através de uma câmera durante a medição do EEG e adicionado manualmente ao arquivo depois de analisar os quadros de vídeo. '1' indica o olho fechado e '0' o estado do olho aberto.

---

<sup>2</sup> Códigos referentes a esta implementação disponíveis em <<https://git.dcomp.ufs.br/golf-sierra/eeg-lstm>>.

<sup>3</sup> Mais informações disponíveis em <<https://www.emotiv.com/epoc>>.

Todos os valores estão em ordem cronológica com o primeiro valor medido na parte superior dos dados.

Durante os experimentos, a base foi particionada em conjuntos de treinamento e teste com as respectivas divisões de 90% e 10%, com ordem aleatória de seleção, afim de utilizar a técnica de validação de modelo *k-fold cross-validation*, com  $k = 10$ , para avaliar como os resultados serão generalizados para um conjunto de dados independente.

Nos experimentos foram utilizados modelos baseados nas arquiteturas RNN *vanilla*, LSTM e GRU, bem como suas variantes bidirecionais. Por fim, os resultados do modelo proposto (apresentando métricas que descrevem o desempenho completo do modelo) são comparados com os de trabalhos relacionados que utilizaram a mesma base de dados.

#### 4. Resultados

Ao efetuar uma análise exploratória da base de dados notou-se a presença de valores fora do normal, foi realizado o tratamento da base, removendo estes *outliers*<sup>4</sup>, a base resultante tratada tem 8172 instâncias válidas de "olho aberto" ('0') e 6720 instâncias válidas de "olho fechado" ('1').

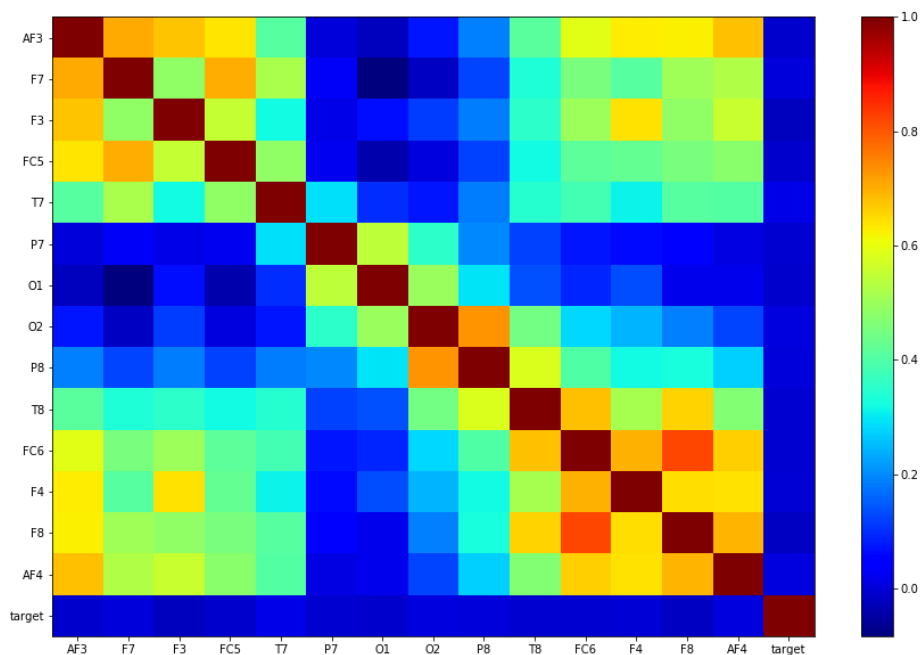
Em seguida verificou-se os coeficientes de correlação (matriz de correlação), ver Figura 3, para tentar identificar as *features* (colunas do conjunto de dados) que têm uma alta correlação com a variável de *target*<sup>5</sup>. No caso desta base de dados, as correlações se mantém próximas a zero, ou seja, provavelmente não se encontrará uma base ótima para representar os dados de forma compacta, isso pode ser observado verificando as cores frias na última linha (e na última coluna, já que se trata de uma matriz simétrica), tendo em vista que a classe ou variável de decisão (*target*) está associada à última posição nesta matriz.

Também se quantificou a relação de dependência linear (matriz de dispersão) entre as *features*, que fornece um resumo gráfico das relações no conjunto de dados, verificou-se que, os histogramas das variáveis parecem estar normalmente distribuídos. Além disso, as relações entre alguns pares sugerem um padrão linear (pontos mais quentes no gráfico de calor da matriz de correlação).

---

<sup>4</sup> *Outliers* são valores extremos muito além das outras observações do conjunto de dados.

<sup>5</sup> Variável de *target*, no contexto deste trabalho, é a variável que é ou deve ser a saída. Por exemplo, pode ser 0 ou 1 binário (tarefa de classificação) ou pode ser uma variável contínua (tarefa de regressão).



**Figura 3. Matriz de correlação da base de dados representada como um mapa de calor.**

Fonte: Produzido pelos autores.

Antes de utilizar os dados nos modelos, em uma etapa complementar à remoção de *outliers* do conjunto de dados, foi realizado um escalonamento das instâncias para que permanecessem entre um valor mínimo e máximo, neste caso entre zero e um. Isto porquê, em geral, os algoritmos de aprendizado se beneficiam da padronização do conjunto de dados.

Todos os modelos utilizados, possuíam três camadas ocultas, com 28 unidades cada, com *dropout* de 50% entre as camadas, é importante destacar também a utilização da regularização L2 ( $\lambda = 0,0001$ ) nas matrizes recorrentes das unidades. A utilização do *dropout* e da regularização foram medidas preventivas contra *overfitting* dos modelos.

A rede RNN *vanilla* obteve acurácia de  $81,53 \pm 3,20\%$  10-*fold*, as demais métricas são apresentadas nas Tabelas 2 e 3. Já sua variante bidirecional obteve acurácia de  $88,59 \pm 1,47\%$  10-*fold*, as demais métricas são apresentadas nas Tabelas 4 e 5.

**Tabela 2. Resumo da precisão, recall, F1-score para cada classe do modelo RNN**

	<i>Precisão</i>	<i>Recall</i>	<i>F1 score</i>	<i>Support</i>
Olho aberto	0,897	0,876	0,886	1635
Olho fechado	0,854	0,877	0,865	1344
<b>Média/Total</b>	<b>0,877</b>	<b>0,877</b>	<b>0,877</b>	<b>2979</b>

**Tabela 3. Matriz de confusão do modelo RNN**

		Olho aberto	Olho fechado	
Estado real	Olho aberto	1179	165	
	Olho fechado	202	1433	
		Estado previsto		

**Tabela 4. Resumo da precisão, recall, F1-score para cada classe do modelo BRNN**

	<i>Precisão</i>	<i>Recall</i>	<i>F1 score</i>	<i>Support</i>
Olho aberto	0,898	0,943	0,920	1635
Olho fechado	0,926	0,970	0,897	1344
<b>Média/Total</b>	<b>0,911</b>	<b>0,910</b>	<b>0,910</b>	<b>2979</b>

**Tabela 5. Matriz de confusão do modelo BRNN**

		Olho aberto	Olho fechado	
Estado real	Olho aberto	1169	175	
	Olho fechado	93	1542	
		Estado previsto		

A rede LSTM obteve acurácia de  $94,35 \pm 0,91\%$  10-fold, as demais métricas são apresentadas nas Tabelas 6 e 7. Já sua variante bidirecional obteve acurácia de  $96,13 \pm 0,54\%$  10-fold, as demais métricas são apresentadas nas Tabelas 8 e 9.

**Tabela 6. Resumo da precisão, recall, F1-score para cada classe do modelo LSTM**

	<i>Precisão</i>	<i>Recall</i>	<i>F1 score</i>	<i>Support</i>
Olho aberto	0,935	0,956	0,946	1635
Olho fechado	0,945	0,920	0,932	1344
<b>Média/Total</b>	<b>0,940</b>	<b>0,940</b>	<b>0,940</b>	<b>2979</b>

**Tabela 7. Matriz de confusão do modelo LSTM**

		Olho aberto	Olho fechado	
Estado real	Olho aberto	1236	108	
	Olho fechado	72	1563	
		Estado previsto		

**Tabela 8. Resumo da precisão, recall, F1-score para cada classe do modelo BLSTM**

	<i>Precisão</i>	<i>Recall</i>	<i>F1 score</i>	<i>Support</i>
Olho aberto	0,967	0,967	0,961	1635
Olho fechado	0,946	0,960	0,953	1344
<b>Média/Total</b>	<b>0,957</b>	<b>0,957</b>	<b>0,957</b>	<b>2979</b>

**Tabela 9. Matriz de confusão do modelo BLSTM**

		Olho aberto	Olho fechado		
Estado real	Olho aberto	1290	54	Estado previsto	
	Olho fechado	74	1561		

Por fim, a rede GRU obteve acurácia de  $93,73 \pm 1,00\%$  10-fold, as demais métricas são apresentadas nas Tabelas 10 e 11. Já sua variante bidirecional obteve acurácia de  $96,13 \pm 0,76\%$  10-fold, as demais métricas são apresentadas nas Tabelas 12 e 13.

**Tabela 10. Resumo da precisão, recall, F1-score para cada classe do modelo GRU**

	<i>Precisão</i>	<i>Recall</i>	<i>F1 score</i>	<i>Support</i>
Olho aberto	0,931	0,956	0,944	1635
Olho fechado	0,945	0,914	0,929	1344
<b>Média/Total</b>	<b>0,937</b>	<b>0,937</b>	<b>0,937</b>	<b>2979</b>

**Tabela 11. Matriz de confusão do modelo GRU**

		Olho aberto	Olho fechado		
Estado real	Olho aberto	1229	115	Estado previsto	
	Olho fechado	72	1563		

**Tabela 12. Resumo da precisão, recall, F1-score para cada classe do modelo BGRU**

	<i>Precisão</i>	<i>Recall</i>	<i>F1 score</i>	<i>Support</i>
Olho aberto	0,961	0,960	0,961	1635
Olho fechado	0,952	0,952	0,952	1344
<b>Média/Total</b>	<b>0,957</b>	<b>0,957</b>	<b>0,957</b>	<b>2979</b>

**Tabela 13. Matriz de confusão do modelo BGRU**

		Olho aberto	Olho fechado		
Estado real	Olho aberto	1280	64	Estado previsto	
	Olho fechado	65	1570		

#### 4.1 Discussão

Baseado nestes resultados, constata-se que os modelos de redes recorrentes bidirecionais com unidades LSTM e GRU obtiveram o melhor desempenho na tarefa de

classificação do *corpus*, evidenciando a utilidade de se considerar as informações do passado e do futuro da sequência de entrada (melhor entendimento do contexto). As variantes unidirecionais destas arquiteturas apresentaram desempenho inferior, permanecendo próximos aos índices das redes bidirecionais, resultado que evidencia a eficácia dos mecanismos de portões das unidades das redes. Vale destacar também, que as redes LSTM e GRU apresentam desempenho comparável [Chung et al. 2014], resultado que se observa ainda nas variantes bidirecionais destas. Já o modelo baseado em células RNN *vanilla* obteve desempenho inferior as demais, atribui-se a isto sua arquitetura mais simples e limitada, como discutido anteriormente na Seção 2.2. Sua variante bidirecional (BRNN), no entanto, apresentou desempenho superior comparada a RNN *vanilla*.

O estado da arte para classificação deste conjunto de dados obteve resultados superiores aos obtidos neste trabalho ao custo de maior esforço computacional e consequentemente maior complexidade (vide Tabela 14) [Rösler e Suendermann 2013, Narejo et al. 2016]. Os classificadores K\* [Rösler e Suendermann 2013] e K\*+RRF [Hamilton et al. 2015] obtêm resultados superiores a 90% a um custo computacional mais elevado, como aponta [Rösler e Suendermann 2013]. Também são destacados os classificadores *Neuro-Fuzzy* [Kim et al. 2016], que apresentaram resultados similares aos de [Rösler e Suendermann 2013, Hamilton et al. 2015]. Por fim, o modelo proposto por [Narejo et al. 2016], utilizando *Stacked AutoEncoders* (SAE), apresenta o melhor desempenho. Os modelos propostos neste trabalho possuem arquitetura simplificada e visam atender minimamente ao compromisso de melhor desempenho e melhor classificação possível.

**Tabela 14. Comparação com os modelos da literatura**

<b>Modelo</b>	<b>Sucesso na classificação (%)</b>
MLP [Sabaci e Koklu 2015]	56,4
RNN	81,5
kNN [Sabaci e Koklu 2015]	84,1
BRNN	88,6
GRU	93,7
LSTM	94,3
<i>Neuro-Fuzzy</i> [Kim et al. 2016]	96,0
BGRU	96,1
BLSTM	96,1
K* [Rösler e Suendermann 2013]	97,3
K*+RRF [Hamilton et al. 2013]	97,4
SAE [Narejo et al. 2016]	98,9

## 5. Considerações Finais

Neste artigo, uma abordagem de classificação de séries temporais baseada em redes neurais recorrentes é proposta para a identificação do estado ocular do paciente utilizando seus dados de EEG. A abordagem utilizada focou na utilização de modelos baseados em redes LSTM e suas variantes para efetuar tal classificação. Os resultados obtidos pelos modelos são próximos ou equivalentes aos apresentados pela literatura.

Os resultados finais indicaram que os modelos bidirecionais BLSTM e BGRU (que utilizam informações do passado e futuro do sinal de entrada) são aplicáveis na tarefa de classificação de séries temporais de EEG utilizadas, e que o seu desempenho é razoável se comparado aos modelos com métricas superiores e mais caros computacionalmente.

Trabalhos futuros podem investigar novas abordagens, como arquiteturas híbridas, a exemplo das redes CNN/LSTM, entre outras. Sugere-se também a experimentação de técnicas para melhoria as métricas dos modelos apresentados, como a de *data augmentation*, já que a quantidade de dados do *corpus* utilizado é limitada e essa classe de algoritmos tendem a se beneficiar com o aumento do conjunto de dados, como é sabido na literatura relacionada. Além disso, pode ser explorada a extração de características do sinal como novas *features*, por exemplo. Sempre tendo em vista o compromisso entre custo computacional e desempenho do modelo.

## Referências

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Man e, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- Acharya, J. N., Hani, A. J., Cheek, J., Thirumala, P., and Tsuchida, T. N. (2016). American clinical neurophysiology society guideline 2: guidelines for standard electrode position nomenclature. *The Neurodiagnostic Journal*, 56(4):245–252.
- Bashivan, P., Rish, I., Yeasin, M., and Codella, N. (2015). Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448.
- Cahn, B. R. and Polich, J. (2006). Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychological bulletin*, 132(2):180.
- Cho, K., Van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.



- Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press.
- Cleary, J. G.; Trigg, L. E. K\*: An instance-based learner using an entropic distance measure. In: 12th International Conference on Machine Learning. [S.l.: s.n.], 1995. p. 108–114.
- Dheeru, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- Graves, A. (2012). *Supervised Sequence Labelling with Recurrent Neural Networks*. Studies in Computational Intelligence 385. Springer-Verlag Berlin Heidelberg, 1st edition.
- Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2017). Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*.
- Hamilton, C. R., Shahryari, S., and Rasheed, K. M. (2015). Eye state prediction from EEG data using boosted rotational forests. In *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on*, pages 429–432. IEEE.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Holmes, G., Donkin, A., and Witten, I. H. (1994). Weka: A machine learning workbench. In *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*, pages 357–361. IEEE.
- Kim, Y., Lee, C., and Lim, C. (2016). Computing intelligence approach for an eye state classification with EEG signal in BCI. In *Software Engineering and Information Technology: Proceedings of the 2015 International Conference on Software Engineering and Information Technology (SEIT2015)*, pages 265–270. World Scientific.
- Långkvist, M., Karlsson, L., and Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11–24.
- Malmivuo, J. and Plonsey, R. (1995). *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA.
- Narejo, S., Pasero, E., and Kulsoom, F. (2016). EEG based eye state classification using deep belief network and stacked autoencoder. *International Journal of Electrical and Computer Engineering (IJECE)*, 6(6):3131–3141.
- Northrop, R. B. (2012). *Analysis and Application of Analog Electronic Circuits to Biomedical Instrumentation, Second Edition*. Biomedical engineering series (Boca Raton Fla.). CRC Press, 2nd ed edition.
- Polat, K. and Günes, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast fourier transform. *Applied Mathematics and Computation*, 187(2):1017–1026.

- Raschka, S. (2015). Python Machine Learning. Packt Publishing Ltd.
- Rösler, O. (2013). EEG Eye State Data Set.  
<https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State>.
- Rösler, O. and Suendermann, D. (2013). A first step towards eye state prediction using EEG. Proc. of the AIHLS.
- Sabancı, K. and Koklu, M. (2015). The classification of eye state by using KNN and MLP classification models according to the EEG signals. International Journal of Intelligent Systems and Applications in Engineering, 3(4):127–130.
- Sulaiman, N., Taib, M. N., Lias, S., Murat, Z. H., Aris, S. A., and Hamid, N. H. A. (2011). Novel methods for stress features identification using EEG signals. International Journal of Simulation: Systems, Science and Technology, 12(1):27–33.
- Wang, T., Guan, S.-U., Man, K. L., and Ting, T. (2014). EEG eye state identification using incremental attribute learning with time-series classification. Mathematical Problems in Engineering, 2014.
- Yao, K., Cohn, T., Vylomova, K., Duh, K., and Dyer, C. (2015). Depth-gated recurrent neural networks. arxiv preprint. arXiv preprint arXiv:1508.03790, 9.
- Yeo, M. V., Li, X., Shen, K., and Wilder-Smith, E. P. (2009). Can SVM be used for automatic EEG detection of drowsiness during car driving? Safety Science, 47(1):115–124.
- Zaremba, W., Sutskever, I., and Vinyals, O. (2014). Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.