

# Rede Neural Convolutacional e LSTM para Biometria Baseada em EEG no Modo de Identificação

Carlos Freitas<sup>1</sup>, Pedro Silva<sup>1</sup>, Gladston Moreira<sup>1</sup>, Eduardo Luz<sup>1</sup>

<sup>1</sup>Computing Department – Universidade Federal de Ouro Preto (UFOP)  
Ouro Preto – 35.400-000 – MG – Brazil

carlos.gabriel@aluno.ufop.edu.br, {silvap, gladston, eduluz}@ufop.edu.br

**Abstract.** *With the advancement of biometrics and the need for more robust security systems, different types of human traits have been taken into consideration for biometrics. One of these traits (modalities) is the electroencephalogram (brain signals). This paper evaluates a neural network model, whose architecture combines layers of Convolutional Neural Networks (CNN) and layers of Long Short-Term Memory (LSTM) for biometrics tasks from electroencephalograms. The experimental discussion has been performed on the Physionet Motor Movement/Imagery Dataset, with data from 109 individuals. By using a window size of 12 seconds, a state-of-the-art result of 99.7% accuracy is achieved, proving the efficiency of the applied methodology for biometric identification mode.*

**Resumo.** *Com o avanço da biometria e a necessidade de sistemas de segurança mais robustos, outros tipos de características humanas além das mais utilizadas foram levadas em consideração no desenvolvimento de sistemas biométricos. Uma destas características é o eletroencefalograma (sinais cerebrais). Este trabalho então avalia uma rede neural, cuja arquitetura combina camadas de Redes Neurais Convolutacionais e camadas de Long Short-Term Memory (LSTM), em um sistema biométrico no modo de identificação, e utiliza os dados dos 109 indivíduos presentes na base de dados EEG Motor Movement/Imagery Dataset. Ao utilizar um tamanho de janela de 12 seg., um resultado estado-da-arte de 99,7% de acurácia foi atingido, provando a eficiência da metodologia aplicada.*

## 1. Introdução

O desenvolvimento da biometria trouxe consigo o avanço tecnológico dos sistemas de identificação e de segurança. O acesso à caixas e urnas eletrônicas, por exemplo, passaram do uso de senhas, cartões e documentos impressos para o uso de digitais, as quais oferecem mais comodidade às pessoas ao eliminar o risco de esquecerem os elementos usados para sua identificação [Lumini and Nanni 2017], continuam permitindo o uso do serviço pelas pessoas em casos de roubo, e possuem menos chance de serem falsificadas [Obaidat et al. 2019].

Sistemas biométricos podem ser utilizados para combater fraude [Bobkowska et al. 2019] e aumentar o nível de segurança de serviços de saúde, transporte e sociais, dentre outros tipos, utilizados diariamente por um grande número de pessoas [Obaidat et al. 2019].

Uma das características humanas que vem sendo exploradas para construção de sistemas biométricos é o eletroencefalograma (EEG) [Carrión-Ojeda et al. 2021], que

consiste no registro dos sinais elétricos presentes na atividade cerebral através do posicionamento de eletrodos no couro cabeludo. Os sinais de um EEG variam de acordo com o modo como um indivíduo executa uma tarefa, seja ela motora ou imaginária, bem como o estado emocional em que ele se encontra [Lee and Hsieh 2014].

A fim de extrair as características que diferenciam um indivíduo do outro, e que “residem” dentro de um sinal EEG, podemos utilizar diferentes técnicas, como calcular a distância de Hamming entre dois sinais codificados [Damaševičius et al. 2018], similaridade por cosseno [Das et al. 2016], ou um tipo de rede neural explorado na área de *Deep Learning* chamado *Convolutional Neural Network* (CNN) [Schons et al. 2017], que é focado em realizar operações de convolução.

Outra arquitetura, presente em redes neurais, envolve a utilização de camadas de *Long Short-Term Memory* (LSTM). Este tipo de arquitetura apresenta bons resultados em experimentos que possuem dados sequenciais [Sherstinsky 2020], e além de serem utilizados em tradutores de texto [Wu and Qin 2022] e modelagem linguística [Lin and Tegmark 2017]. Também podem ser utilizados na construção de sistemas biométricos [Sun et al. 2019].

As contribuições deste trabalho podem ser sumarizadas em: (i) Fusionamento de camadas de *Long Short-Term Memory* (LSTM) com camadas de *Convolutional Neural Network* (CNN); e (ii) Resultados estado-da-arte para o contexto de identificação biométrica para a base *PhysioNet EEG Motor Movement/Imagery Dataset*.

O restante do trabalho é organizado da seguinte maneira: A Seção 2 destaca outros trabalhos relacionados com o proposto, bem como alguns resultados e/ou conclusões relatados neles. A base de dados é apresentada na Seção 3. Na Seção 4 é apresentada a metodologia adotada. A Seção 5 demonstra os resultados obtidos nos diferentes experimentos realizados, e discussões acerca deles. Por último, a Seção 6 apresentam as conclusões tomadas a partir do trabalho apresentado.

## 2. Trabalhos Relacionados

Nesta seção, são apresentados trabalhos presentes na literatura que estão relacionados com o tema proposto. Os critérios para seleção dos trabalhos relacionados foram os que utilizavam CNN na construção de um sistema biométrico baseado em EEG, sendo a única exceção o trabalho de [Yang et al. 2018], já que este contém os protocolos que foram avaliados pelo trabalho proposto.

Uma CNN e 108 dos 109 indivíduos presentes na base de dados *EEG Motor Movement/Imagery Dataset* [Schalk et al. 2004] foram utilizados em [Schons et al. 2017] para construir um sistema biométrico com foco no modo de verificação. Em [Schons et al. 2017], utilizou-se 90% dos sinais gravados em estado de repouso com olhos abertos (REO - *Rest Eyes Open*) para treinamento, os 10% remanescentes para validação, e os sinais gravados em estado de repouso com olhos fechados (REC - *Rest Eyes Closed*) para teste. A métrica de *Equal Error Rate* (EER), gerada por meio de comparações entre pares genuínos (intra-classe) e impostores (inter-classe) de vetores de características, foi utilizada para medir a eficiência da rede neural. Ao utilizar um tamanho de janela de 12 segundos e um filtro passa-banda de 30-50Hz, um EER de 0,19% foi obtido.

O trabalho proposto em [Mota et al. 2021] é uma extensão do trabalho

[Schons et al. 2017] e explora a utilização de camadas de *Squeeze-and-Excitation*. Ao compor os dados de treinamento e validação com a primeira sessão de uma tarefa da base de dados EEG *Motor Movement/Imagery Dataset* [Schalk et al. 2004], onde o indivíduo abre e fecha o punho correspondente ao lado da tela em que um alvo aparece, e os dados de teste com a segunda sessão desta mesma tarefa, um EER de 0,1% foi obtido.

Em [Das et al. 2017], sinais EEG de 50 indivíduos foram gravados pelos próprios autores utilizando 17 eletrodos. Uma tela mostrava aos indivíduos formas geométricas selecionadas aleatoriamente, e estes foram instruídos a se concentrar apenas quando um círculo aparecesse, sendo este o alvo, e ignorar a aparição de outras formas geométricas, sendo estas não-alvo. Uma CNN foi então utilizada para classificar os indivíduos no modo biométrico de identificação, e foi reportado 98,8% de acurácia em comparações não-alvo vs não-alvo (quando tanto a amostra do sinal EEG quanto a saída da CNN indicaram que o indivíduo estava ignorando um não-alvo) e 80,65% de acurácia em comparações alvo vs alvo.

Um sistema baseado em coeficientes de *wavelets* foi utilizado em [Yang et al. 2018] para a extração de características dos 109 indivíduos presentes na base de dados *PhysioNet*. Foram utilizados 9 dos 64 eletrodos disponíveis, julgados como os que podem apresentar informações de maior capacidade de diferenciação entre indivíduos. O trabalho também estabeleceu três protocolos para avaliação:

- Protocolo P1: qual(is) região(ões) cerebral(is) apresenta(m) uma maior performance na rede;
- Protocolo P2: qual combinação entre dados de treinamento e dados de teste apresenta uma maior performance na rede;
- Protocolo P3: o quanto a performance pode melhorar ao manter os mesmos dados de teste e adicionar mais dados de treinamento, provenientes de tarefas diferentes.

Os autores de [Yang et al. 2018] concluíram que quanto mais regiões cerebrais forem utilizadas para treinamento e teste do sistema, maior será a performance. Também foi observado que dados de treinamento e teste provenientes de diferentes tarefas realizadas pelos indivíduos não implicam na redução da performance, e que quanto mais dados de treinamento forem utilizados, maior será a performance da rede, obedecendo a Lei dos Rendimentos Decrescentes. Os protocolos P1, P2 e P3 mencionados também foram abordados no presente trabalho, e os resultados estão presentes na Subseção 5.3.

Já em [Jijomon and Vinod 2021] utilizou-se *auditory evoked potentials* (estímulos auditivos) em 20 indivíduos, 2 eletrodos localizados na parte frontal do cérebro para capturar os sinais de EEG e um tamanho de janela de 1 segundo. A coleta dos dados foi feita pelos próprios autores. Para o cenário de identificação de um sistema biométrico, foi utilizada uma rede neural que combina camadas de CNN com camadas de LSTM, e um resultado de 99,53% de acurácia foi atingido.

### 3. Base de Dados PhysioNet

A base de dados utilizada nesse trabalho é pertencente à *PhysioNet* [Goldberger et al. 2000] e se chama EEG *Motor Movement/Imagery Dataset* [Schalk et al. 2004]. Ela conta com 14 sessões de aquisição de dados, efetuadas por 109 indivíduos e registradas através do uso de 64 eletrodos, onde cada um é amostrado à uma frequência de 160Hz. As sessões são numeradas de 1 a 14 e as tarefas realizadas pelos indivíduos consistem em:

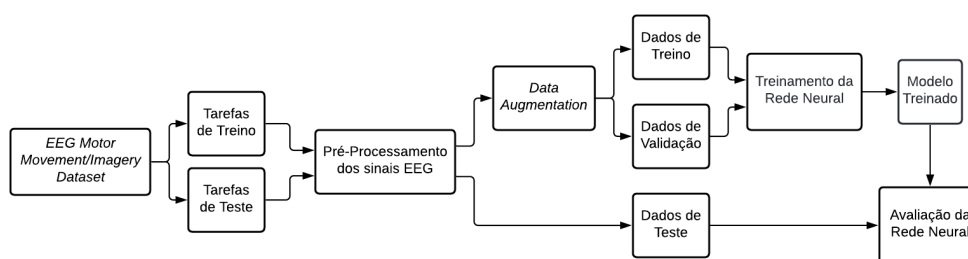
- *Resting Eyes Open* ou REO: O indivíduo permanece em estado de repouso com os olhos abertos (Sessão 1);
- *Resting Eyes Closed* ou REC: O indivíduo permanece em estado de repouso com os olhos fechados (Sessão 2);
- T1R1, T1R2 e T1R3: O indivíduo abre e fecha o punho correspondente ao lado da tela em que um alvo aparece (Sessões 3, 7 e 11);
- T2R1, T2R2 e T2R3: O indivíduo imagina abrir e fechar o punho correspondente ao lado da tela em que um alvo aparece (Sessões 4, 8 e 12);
- T3R1, T3R2 e T3R3: O indivíduo abre e fecha os punhos ou os pés, caso o alvo apareça em cima da tela ou embaixo, respectivamente (Sessões 5, 9 e 13);
- T4R1, T4R2 e T4R3: O indivíduo imagina abrir e fechar os punhos ou os pés, caso o alvo apareça em cima da tela ou embaixo, respectivamente (Sessões 6, 10 e 14);

As sessões 1 e 2 contam cada uma com 61 segundos de gravação ou 9760 amostras, enquanto cada uma das sessões de 3 a 14 contam com 2 minutos de gravação, ou 19200 amostras.

## 4. Metodologia

Nesta seção são apresentados os métodos utilizados para o processamento dos sinais de EEG e construção da rede neural utilizada. A metodologia do trabalho proposto seguiu o *workflow* apresentado na Figura 1.

Redes estado-da-arte são treinadas normalmente para imagens (2D). Como este trabalho utiliza dados 1D, não há uma variedade de modelos treinados o que dificulta o processo de transferência de aprendizado.



**Figura 1. Metodologia desenvolvida no trabalho proposto. Fonte: Próprios autores, 2022.**

### 4.1. Pré-Processamento

Como dito anteriormente, os sinais de EEG são sujeitos à muitos ruídos [Jamil et al. 2021], os quais podem atrapalhar na identificação de um indivíduo pela rede neural, e diminuir a acurácia da mesma. Deste modo, foi avaliado passar cada um dos sinais de um EEG em um filtro passa-banda, na tentativa de melhorar a qualidade dos sinais e dos dados obtidos por meio deles.

Testes foram realizados utilizando os filtros de 1-50 Hz, 10-30 Hz e 30-50 Hz, propostos em [Schons et al. 2017], e os resultados podem ser encontrados na Subseção 5.1.

## 4.2. Data Augmentation

Uma CNN necessita de uma grande quantidade de amostras para que ela possa convergir e aprender a tarefa designada à ela. O processo de *data augmentation* proposto em [Schons et al. 2017] para aumentar o número de dados que irão alimentar a rede, foi o de deslizar uma janela de 1920 amostras (equivalente à 12 segundos) pelo sinal completo. Essa técnica é importante, pois sem sobreposição das janelas, sinais de REO e REC terão somente 5 amostras, o que é insuficiente para o treinamento da rede.

Este processo foi implementado no trabalho proposto, com o tamanho da janela igual à 1920 amostras, e é efetuado para os sinais de cada indivíduo. O valor de *offset* escolhido foi o de 35 amostras, pois houve um ganho na acurácia em relação ao valor 40, proposto em [Mota et al. 2021].

## 4.3. Arquitetura Proposta

A arquitetura implementada inicialmente foi baseada na CNN utilizada em [Schons et al. 2017] e [Mota et al. 2021]. Vale lembrar que o formato da entrada da primeira camada da CNN (Conv1) é 1920 x 64 devido ao tamanho da janela utilizada nos trabalhos mencionados e ao número de eletrodos empregados na gravação dos sinais EEG, onde cada eletrodo corresponde a um canal do sinal.

A CNN é composta por três sequências de camadas de convolução, seguidas de *batch normalization* e *max pooling*, três camadas densas (4096, 4096 e 256 neurônios respectivamente), um *batch normalization*, uma camada de *dropout* (taxa de 10%) e finalmente outra camada densa que utiliza uma função de ativação *softmax* para classificar os indivíduos. A função de ativação utilizada nas camadas de convolução e em algumas das camadas densas foi a *Rectified Linear Unit*, ou ReLU.

Após as tarefas para compor os dados de treinamento e os dados de teste serem escolhidas, o pré-processamento ser aplicado aos sinais de EEG, e o processo de *data augmentation* previamente explicado for aplicado aos dados de treinamento, 10% desses últimos serão destinados a compor os dados de validação, divisão esta também utilizada por [Schons et al. 2017], usado como referência para o trabalho proposto.

A rede neural foi treinada utilizando *Stochastic Gradient Descent* com um coeficiente de *momentum* igual à 0.9 como o algoritmo de otimização, e o tamanho dos *mini-batches* foi definido como 100. Inicialmente, o número de épocas definido foi 60, mas foi mudado para 40 após uma melhora na acurácia para o contexto de identificação em relação à [Schons et al. 2017]. O *learning rate* empregado possui o valor 0,1 durante as primeiras 2 épocas, 0,01 das épocas de 3 a 37 e 0,001 nas épocas restantes. Desta forma, o *learning rate* é diminuído conforme o número de épocas na tentativa de obter um conjunto ótimo de pesos da rede neural [Bengio 2012].

Após resultados pouco expressivos, que podem ser encontrados na Subseção 5.1, foi levada em consideração a abordagem realizada em [Sun et al. 2019], que adicionou camadas de *Long Short-Term Memory* (LSTM) à rede neural que utilizou. Deste modo, diversos experimentos relacionados à utilização de LSTM foram realizados, visando aumentar a acurácia obtida no modo de identificação. Os resultados destes experimentos estão presentes na Subseção 5.2.

#### 4.4. Métrica utilizada

A métrica utilizada para medir a performance do modelo no modo de identificação foi a acurácia, que mede a frequência com que a rede neural acerta as previsões estabelecidas para um conjunto de dados de entrada. Considerando  $Y$  como o número de previsões que a rede neural acertou e  $Z$  como o número total de previsões feitas, a acurácia é obtida por  $Y/Z$ . Apenas a acurácia foi reportada pois o trabalho proposto em [Yang et al. 2018], que foi comparado com este, utilizou somente a acurácia para o modo de identificação.

#### 4.5. Protocolos Avaliados

Os protocolos P1, P2 e P3 estabelecidos em [Yang et al. 2018] foram avaliados com os 9 eletrodos utilizados em seu trabalho e com todos os 64 eletrodos disponíveis e os resultados destes experimentos estão presentes na Subseção 5.3.

O protocolo P1 utiliza duas tarefas de treino, sendo a primeira referente à primeira sessão de uma tarefa  $X$  e a segunda referente à terceira sessão de  $X$ , e a tarefa de teste é a segunda sessão de  $X$ . Enquanto os experimentos do protocolo P2 e P3 sempre utilizam todos os eletrodos, alguns experimentos deste protocolo consistem em utilizar eletrodos de uma região cerebral específica (frontal, motora ou occipital). Já o protocolo P2, apesar de compor as tarefas de treino do mesmo modo que o protocolo P1, utiliza sempre T1R2, T2R2, T3R2, T4R2, REO ou REC como tarefa de teste, independente do par de tarefas de treino escolhido. Finalmente, o protocolo P3 utiliza T1R2 como tarefa de teste, e consiste em mudar as tarefas de treino e/ou aumentar seu número. Por exemplo, o primeiro experimento utiliza apenas T1R1 como tarefa de treino, enquanto o último utiliza T1R1, T1R3, T2R1, T2R2, T2R3, T3R1, T4R1 e T4R2.

### 5. Resultados e Discussões

Nesta seção são apresentados os resultados dos experimentos realizados e discussões acerca deles. Os experimentos foram conduzidos em uma máquina com um processador Intel i9-10900 @ 2.80GHz, 128GB de RAM DDR4 e placa de vídeo RTX 3090 com 24GB de memória GDDR6X. O código referente ao artigo pode ser acessado pelo link: <https://github.com/Carlosgd-freitas/cnn-lstm-biometric-identification>.

Os experimentos citados nas Seções 5.1 e 5.2 tiveram os dados de treinamento e validação compostos por REO e os dados de teste compostos por REC.

#### 5.1. Filtragem e Normalização

Na Tabela 1 encontra-se a acurácia para cada filtro passa-banda testado e mencionado na Subseção 4.1, o qual é aplicado normalmente no sinal, e reversamente em seguida. Nessa mesma tabela também se encontra a acurácia para a utilização de uma normalização dos sinais EEG após a filtragem, utilizada em [Schons et al. 2017] e [Mota et al. 2021], onde os dados presentes em cada canal foram subtraídos por sua média, depois somados com o absoluto do valor mínimo presente neles, então divididos pelo seu desvio padrão, e por fim, divididos pelo valor máximo presente neles.

Como pode ser observado na Tabela 1, este tipo de normalização dos dados causa uma deterioração da performance na rede no modo de identificação, independentemente do tipo do filtro passa-banda utilizado, e portanto, não foi realizada nos próximos cenários de teste.

**Tabela 1. Filtro e normalização no pré-processamento dos dados. Acc-F = Acurácia sem normalização; Acc-FN = Acurácia com normalização.**

Filtro do Pré-Processamento	Acc-F (%)	Acc-FN (%)
1-50Hz	50,64	5,50
10-30Hz	44,22	3,67
30-50Hz	64,04	5,69

## 5.2. Avaliação de Arquitetura

A primeira avaliação feita foi o uso de camadas bidirecionais de LSTM ou não. Após a avaliação empírica de 5 camadas de LSTM com 10 unidades por bloco, observou-se que camadas bidirecionais (53,0275%) resultam em resultados inferiores ao uso de camadas comuns de LSTM (57,2477%).

Considerando os resultados obtidos por usar camadas de LSTM contiguamente, foi explorado o posicionamento de 5 dessas camadas, contíguos, com 10 unidades cada, adicionados ao modelo utilizado em [Schons et al. 2017], na tentativa de melhorar a acurácia. Também foi observado os resultados da aplicação de uma normalização diferente da descrita anteriormente, presente em [Sun et al. 2019], onde os dados presentes em cada canal foram subtraídos por sua média e depois divididos pelo seu desvio padrão. Os resultados destes experimentos estão presentes na Tabela 2.

**Tabela 2. Posição das camadas de LSTM na arquitetura da rede convolucional.**

Posição	Acurácia (%)
Entre Pool3 e FC1	40,7339
Anterior a Conv1	73,3945

Podemos observar que a melhor performance é obtida ao posicionar os camadas de LSTM anteriormente à primeira camada de convolução Conv1. Aplicar a normalização dos dados descrita não deteriorou a acurácia.

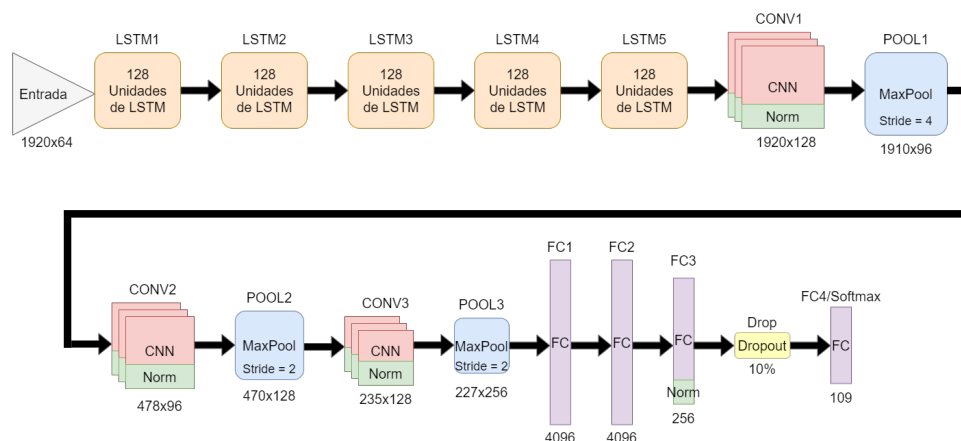
Uma vez definida a arquitetura, é necessário determinar a melhor quantidade de unidades por camada de LSTM. Deste modo, a Tabela 3 contém os resultados de variar o número de unidades por camada de LSTM, sendo cinco camadas posicionadas contiguamente anterior à camada Conv1. Podemos observar que a melhor acurácia foi obtida ao utilizar 128 unidades por bloco de LSTM.

**Tabela 3. Número de unidades por bloco da LSTM.**

# Unidades por bloco	10	16	32	64	128	256
Acurácia (%)	73,3945	90,6422	92,2936	94,3119	95,2294	94,6789

Para finalizar os experimentos relacionados à utilização de camadas de LSTM, avaliou-se a acurácia com o uso de quatro camadas de LSTM (94,6789%), cinco (95,2294%), seis (94,4954%) e dez (95,2294%), com 128 unidades cada. A maior acurácia é obtida ao posicionar cinco camadas antes da primeira camada de convolução Conv1.

Diante os resultados obtidos, a arquitetura final adotada foi a de cinco camadas de LSTM, com 128 unidades cada, seguidos da arquitetura usada em [Schons et al. 2017]. Esta arquitetura pode ser visualizada na Figura 2.



**Figura 2. Arquitetura proposta. "Norm" indica *Batch Normalization*. As camadas Conv1, Conv2 e Conv3 possuem um número de filtros igual a 96, 128 e 256, com tamanhos iguais a 11, 9 e 9, respectivamente. As camadas de *MaxPooling* possuem filtros de tamanho 2 sem *padding*. As camadas Conv1, Conv2, Conv3, FC1 e FC2 possuem funções de ativação ReLU. Fonte: Próprios autores, 2022.**

### 5.3. Protocolos P1, P2 e P3

Os experimentos dos protocolos P1, P2 e P3 presentes em [Yang et al. 2018] foram realizados utilizando o modelo descrito ao final da Subseção 5.2, com todos os 64 eletrodos disponíveis e com os 9 eletrodos utilizados em [Yang et al. 2018], assim tornando mais justa a comparação entre os dois trabalhos.

Os resultados referentes ao protocolo P1 estão presentes na Tabela 4. Nela, podemos observar que ao utilizar os 64 eletrodos disponíveis, a melhor performance no modo de identificação é obtida ao compor os dados de treino com T3R1 e T3R3 e os dados de teste com T3R2, que apresenta uma acurácia de 99,4465%. Ao utilizar os 9 eletrodos presentes em [Yang et al. 2018], a melhor performance no modo de identificação é obtida ao compor os dados de treino com T4R1 e T4R3 e os dados de teste com T4R2, todos estes considerando a região occipital, que apresenta uma acurácia de 80,7196%.

O resultado obtido em [Yang et al. 2018] no protocolo P1 foi de 96% de acurácia, e portanto, ao utilizar os 64 eletrodos disponíveis, o trabalho proposto apresentou um melhor resultado, e ao utilizar os mesmos 9 eletrodos, apresentou um pior resultado.

Analogamente, a Tabela 5 apresenta os resultados ao abordar o protocolo P2. Ao utilizar os 64 eletrodos disponíveis, o melhor resultado foi obtido ao utilizar T1R1 e T1R3 para compor os dados de treinamento e validação, e T3R2 para compor os dados de teste, que apresenta uma acurácia de 99,6310%. Utilizando os 9 eletrodos presentes em [Yang et al. 2018], o melhor resultado é obtido ao compor os dados de treinamento com as tarefas T4R1 e T4R3 e os dados de teste com a tarefa T2R2, onde foi obtido 73,8447% de acurácia.

O resultado obtido em [Yang et al. 2018] no protocolo P2 foi de 96,15% de



**Tabela 4. Resultado para o protocolo P1: Acc-9 = Acurácia com 9 eletrodos; Acc-64 = Acurácia com 64 eletrodos. Modo de identificação.**

Tarefas de treino	Teste	Acc-9 (%)	Acc-64 (%)
T1R1F + T1R3F	T1R2F	54,6125	—
T1R1M + T1R3M	T1R2M	61,5314	—
T1R1O + T1R3O	T1R2O	78,5055	—
T1R1 + T1R3	T1R2	65,1291	98,3395
T2R1F + T2R3F	T2R2F	54,5287	—
T2R1M + T2R3M	T2R2M	64,7874	—
T2R1O + T2R3O	T2R2O	79,8521	—
T2R1 + T2R3	T2R2	66,2662	98,7985
T3R1F + T3R3F	T3R2F	41,3284	—
T3R1M + T3R3M	T3R2M	62,7306	—
T3R1O + T3R3O	T3R2O	79,5203	—
T3R1 + T3R3	T3R2	64,8524	99,4465
T4R1F + T4R3F	T4R2F	51,0148	—
T4R1M + T4R3M	T4R2M	65,6827	—
T4R1O + T4R3O	T4R2O	80,7196	—
T4R1 + T4R3	T4R2	70,1107	98,8930

acurácia, e novamente, o trabalho proposto apresentou um melhor resultado ao utilizar os 64 eletrodos disponíveis, e um pior resultado ao utilizar os mesmos 9 eletrodos.

**Tabela 5. Resultado para o protocolo P2: Acc-9 = Acurácia com 9 eletrodos; Acc-64 = Acurácia com 64 eletrodos. Modo de identificação.**

Treino	Teste	Acc-9 (%)	Acc-64 (%)	Treino	Teste	Acc-9 (%)	Acc-64 (%)
T1R1 + T1R3	T1R2	65,1291	98,3395	T2R1 + T2R3	T1R2	63,0074	98,0627
	T2R2	60,5360	98,8909		T2R2	66,2662	98,7985
	T3R2	65,0369	99,6310		T3R2	52,7675	98,3395
	T4R2	65,2214	99,2620		T4R2	67,1587	99,0775
	REO	54,1284	95,0459		REO	56,1468	94,4954
	REC	32,6605	91,9266		REC	34,4954	90,2752
T3R1 + T3R3	T1R2	62,4539	98,2472	T4R1 + T4R3	T1R2	65,4059	98,8930
	T2R2	54,0665	97,1349		T2R2	73,8447	98,7061
	T3R2	64,8524	99,4465		T3R2	58,3948	96,5867
	T4R2	59,9631	98,8007		T4R2	70,1107	98,8930
	REO	44,5872	92,8440		REO	46,9725	93,3945
	REC	29,1743	89,1743		REC	30,8257	87,5229

Finalmente, a Tabela 6 apresenta os resultados ao abordar o protocolo P3, com-  
pondo os dados de teste com a tarefa T1R2. Tanto ao utilizar os 64 eletrodos disponíveis,  
quanto os 9 eletrodos presentes em [Yang et al. 2018], os melhores resultados foram

obtidos ao compor os dados de treinamento e validação com as tarefas T1R1, T1R3, T2R1, T2R2, T2R3, T3R1 e T4R1, resultando em 99,7232% e 93,8192% de acurácia, respectivamente.

O resultado obtido em [Yang et al. 2018] no protocolo P3 foi de 99% de acurácia, e outra vez, o trabalho proposto apresentou um melhor resultado ao utilizar os 64 eletrodos disponíveis, e um pior resultado ao utilizar os mesmos 9 eletrodos.

**Tabela 6. Resultado para o protocolo P3: Acc-9 = Acurácia com 9 eletrodos; Acc-64 = Acurácia com 64 eletrodos. Modo de identificação.**

Tarefas de treino	Acc-9 (%)	Acc-64 (%)
T1R1	35,7934	89,4834
T1R3	41,5129	92,1587
T1R1 + T1R3	65,1291	98,3395
T2R1 + T2R2 + T2R3	79,4280	97,4170
T1R1 + T1R3 + T2R1	74,3542	98,8930
T1R1 + T2R1 + T2R2 + T2R3	84,9631	99,5387
T1R1 + T1R3 + T2R1 + T2R2	84,2251	99,2620
T1R1 + T1R3 + T2R1 + T2R2 + T2R3	84,8709	99,6310
T1R1 + T1R3 + T2R1 + T2R2 + T2R3 + T3R1	85,8856	99,6310
T1R1 + T1R3 + T2R1 + T2R2 + T2R3 + T3R1 + T4R1	93,8192	99,7232
T1R1 + T1R3 + T2R1 + T2R2 + T2R3 + T3R1 + T4R1 + T4R2	93,5424	99,5387

#### 5.4. Discussões

**Performance de uma arquitetura CNN, LSTM e LSTM+CNN:** Embora CNNs apresentem bons resultados para extração de características e classificação [Liu 2018, Schons et al. 2017, Mota et al. 2021], a utilização de uma CNN com uma arquitetura convencional (que apresenta camadas de convolução, *pooling*, normalização, densas e de *dropout*) não obteve uma alta acurácia no modo de identificação. De modo semelhante, a utilização de camadas de LSTM isolados, que apresentam bons resultados para atividades que variam ao longo do tempo [Graves et al. 2005], também não apresentou uma boa performance. Entretanto, a união entre uma CNN convencional e camadas de LSTM obteve uma boa performance no modo de identificação, atingindo uma acurácia superior à 99,7%, como pode ser observado na Tabela 6. A natureza de um sinal de EEG, que varia com o tempo, aliado à necessidade de extração de características para distinguir indivíduos em um sistema biométrico, pode ser o motivo de que apenas a união das duas arquiteturas mencionadas foi capaz de atingir resultados satisfatórios.

**Performance por região cerebral:** Ao observar a Tabela 4, podemos deduzir que, ao utilizar os eletrodos presentes na região occipital, uma melhor performance é obtida em relação à outras regiões cerebrais. Além desta informação poder implicar que a região occipital do cérebro tem uma maior capacidade de diferenciar indivíduos, ela também pode implicar que podemos obter um resultado similar ao que utiliza 64 eletrodos, e que necessita de um menor poder computacional para ser gerado, se uma quantidade maior de

eletrodos presentes na região occipital for selecionada.

**Composição das tarefas de treino:** Os resultados presentes na Tabela 5 confirmam uma das conclusões obtidas em [Yang et al. 2018], em que compor os dados de treino com diferentes tarefas não implica necessariamente na diminuição da performance. No entanto, haverá sim uma diminuição considerável na performance se utilizarmos tarefas motoras ou imaginárias para compor os dados de treinamento e validação e REO ou REC para compor os dados de teste. Isso pode significar que, quando uma pessoa está em repouso, ela não está focada em realizar alguma tarefa específica, e portanto, seus sinais de EEG não possuem tanto valor para diferenciá-la de outra.

Os resultados apresentados na Tabela 6 também confirmam outra conclusão de [Yang et al. 2018], em que adicionar mais tarefas para compor os dados de treinamento implicará, na maioria dos cenários, no aumento da performance. A diminuição da performance no último cenário do protocolo P3 (adição de T4R2 nos dados de treino e validação) foi constatada tanto em [Yang et al. 2018] quanto no trabalho proposto, e ressalta que pode haver uma relação entre as tarefas utilizadas para treino e teste ou que deve haver algum estudo ou cuidado para selecionar as tarefas de treino que serão adicionadas a partir de um ponto, para que uma diminuição da performance não seja ocasionada.

## 6. Conclusões

Neste trabalho, foi avaliado a performance do fusionamento de uma arquitetura típica de CNN com camadas de LSTM, em um sistema biométrico baseado em sinais de EEG, com foco no modo de identificação. Através dos protocolos P1, P2 e P3 propostos em [Yang et al. 2018], também foi avaliado como a composição das tarefas utilizadas durante as etapas de treino e teste da rede neural afetam a acurácia apresentada por ela.

A utilização de um tamanho de janela de 12 segundos é impraticável em sistemas biométricos do dia-a-dia, os quais iriam fazer os usuários esperarem por uma grande quantidade de tempo. Por este motivo, mudanças no pré-processamento do sinal e na arquitetura da rede neural proposta ainda podem ser exploradas.

Finalmente, um resultado estado-da-arte foi atingido, o qual apresentou uma acurácia de 99,7%. Deste modo, o fusionamento de camadas de CNN com camadas de LSTM provou ser apto no trabalho proposto, e que sistemas biométricos robustos podem ser implementados através das técnicas utilizadas.

## Agradecimentos

Os autores agradecem a FAPEMIG (APQ-01518-21), e Universidade Federal de Ouro Preto pelo suporte no desenvolvimento deste trabalho.

## Referências

- Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. *CoRR*, abs/1206.5533:437–478.
- Bobkowska, K., Nagaty, K., and Przyborski, M. (2019). Incorporating iris, fingerprint and face biometric for fraud prevention in e-passports using fuzzy vault. *IET Image Processing*, 13(13):2516–2528.
- Carrión-Ojeda, D., Fonseca-Delgado, R., and Pineda, I. (2021). Analysis of factors that influence the performance of biometric systems based on eeg signals. *Expert Systems with Applications*, 165:113967.

- Damaševičius, R., Maskeliūnas, R., Kazanavičius, E., and Woźniak, M. (2018). Combining cryptography with EEG biometrics. *Computational Intelligence and Neuroscience*, 2018:1867548:1–1867548:11.
- Das, R., Maiorana, E., and Campisi, P. (2016). Eeg biometrics using visual stimuli: A longitudinal study. *IEEE Signal Processing Letters*, 23(3):341–345.
- Das, R., Maiorana, E., and Campisi, P. (2017). Visually evoked potential for EEG biometrics using convolutional neural network. In *25th European Signal Processing Conference, EUSIPCO 2017, August 28 - September 2, 2017*, pages 951–955, Kos, Greece. IEEE.
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220.
- Graves, A., Fernández, S., and Schmidhuber, J. (2005). Bidirectional LSTM networks for improved phoneme classification and recognition. In *15th International Conference on Artificial Neural Networks*, pages 799–804.
- Jamil, Z., Jamil, A., and Majid, M. (2021). Artifact removal from eeg signals recorded in non-restricted environment. *Biocybernetics and Biomedical Engineering*, 41(2):503–515.
- Jijomon, C. and Vinod, A. (2021). Person-identification using familiar-name auditory evoked potentials from frontal eeg electrodes. *Biomedical Signal Processing and Control*, 68:102739.
- Lee, Y.-Y. and Hsieh, S. (2014). Classifying different emotional states by means of eeg-based functional connectivity patterns. *PloS one*, 9(4):e95415.
- Lin, H. W. and Tegmark, M. (2017). Critical behavior in physics and probabilistic formal languages. *Entropy*, 19(7):299.
- Liu, Y. H. (2018). Feature extraction and image recognition with convolutional neural networks. *Journal of Physics: Conference Series*, 1087(6):062032.
- Lumini, A. and Nanni, L. (2017). Overview of the combination of biometric matchers. *Information Fusion*, 33:71–85.
- Mota, M. R., Silva, P. H., Luz, E. J., Moreira, G. J., Schons, T., Moraes, L. A., and Menotti, D. (2021). A deep descriptor for cross-tasking eeg-based recognition. *PeerJ Computer Science*, 7:e549.
- Obaidat, M. S., Rana, S. P., Maitra, T., Giri, D., and Dutta, S. (2019). Biometric security and internet of things (iot). In *Biometric-based physical and cybersecurity systems*, pages 477–509. Springer.
- Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043.
- Schons, T., Moreira, G. J. P., Silva, P. H. L., Coelho, V. N., and da S. Luz, E. J. (2017). Convolutional network for eeg-based biometric. In *22nd Iberoamerican Congress on Pattern Recognition*, pages 601–608.
- Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404:132306.
- Sun, Y., Lo, F. P.-W., and Lo, B. (2019). Eeg-based user identification system using 1d-convolutional long short-term memory neural networks. *Expert Systems with Applications*, 125:259–267.
- Wu, Y. and Qin, Y. (2022). Machine translation of english speech: Comparison of multiple algorithms. *Journal of Intelligent Systems*, 31(1):159–167.
- Yang, S., Deravi, F., and Hoque, S. (2018). Task sensitivity in eeg biometric recognition. *Pattern Analysis and Applications*, 21(1):105–117.