



# A B A K Ó S

Instituto de Ciências Exatas e Informática



Licença Creative Commons Attribution 4.0 International

## Aplicação da Máquina de Aprendizagem Mínima na Análise Emocional de Sinais Cerebrais: Comparação com Redes Neurais Artificiais e Máquinas de Vetores\*

Application of Minimal Learning Machine in the Emotional Analysis of Brain Signals: A Comparison with Artificial Neural Networks and Support Vector Machines

João Mariano Ipolito<sup>1</sup>

Roney Nogueira de Sousa<sup>2</sup>

Adonias Caetano de Oliveira<sup>3</sup>

Anderson Passos de Aragão<sup>4</sup>

Romulo Lopes Frutuoso<sup>5</sup>

### Resumo

Sistemas *Brain-Computer Interface* (BCI) utilizam sinais de eletroencefalografia (EEG) para traduzir padrões cerebrais em comandos computacionais. Contudo, a classificação desses sinais depende do ajuste de hiperparâmetros. Neste estudo, avaliou-se o desempenho da *Minimal Learning Machine* (MLM) no reconhecimento de emoções humanas a partir de sinais de EEG, comparando-a aos modelos *Multilayer Perceptron* (MLP) e *Support Vector Machine* (SVM). Foram conduzidos três experimentos computacionais, dos quais os resultados mostraram que a MLM superou MLP e SVM em dois experimentos, alcançando acurácia máxima de 96,7%. Conclui-se que a MLM é uma alternativa eficaz e flexível para aplicações em sistemas BCI.

**Palavras-chave:** Emoções humanas. Sinais EEG. Máquina de Aprendizagem Mínima. Reconhecimento de padrões.

\*Submetido em 16/01/2023 - Aceito em 27/08/2025

<sup>1</sup>Instituto Federal do Ceará (IFCE), Bacharelado em Ciência da Computação Brasil- ipolito.pesquisa@gmail.com.

<sup>2</sup>IFCE, Programa de Pós-Graduação em Computação (PPGCC) Brasil- roney.nogueira.sousa08@aluno.ifce.edu.br.

<sup>3</sup>Docente EBTT do IFCE, Bacharelado em Ciência da Computação Brasil- adonias.oliveira@ifce.edu.br.

<sup>4</sup>Docente EBTT do IFCE, Bacharelado em Ciência da Computação Brasil- anderson.aragao@ifce.edu.br.

<sup>5</sup>Docente EBTT do IFCE, Bacharelado em Engenharia Civil Brasil- romulo.lobes@ifce.edu.br.

### **Abstract**

Brain-Computer Interface (BCI) systems use electroencephalography (EEG) signals to translate brain patterns into computational commands. However, the classification of these signals depends on hyperparameter tuning. In this study, the performance of the Minimal Learning Machine (MLM) was evaluated for human emotion recognition using EEG signals, comparing it with Multilayer Perceptron (MLP) and Support Vector Machine (SVM) models. Three computational experiments were conducted, and the results showed that MLM outperformed MLP and SVM in two experiments, achieving a maximum accuracy of 96.7%. It is concluded that MLM is an effective and flexible alternative for applications in BCI systems.

**Keywords:** Human emotions. EEG signals. Minimal Learning Machine. Pattern recognition.

## 1 INTRODUÇÃO

Os sistemas de Interface Cérebro-Computador (BCI, *Brain-Computer Interface*) permitem a comunicação e o controle de dispositivos externos por meio da tradução de sinais gerados pela atividade cerebral em sinais de controle sem o envolvimento de nervos e músculos periféricos. Dentre as técnicas que podem registrar a atividade neural, as abordagens atuais concentram-se em sinais de Eletroencefalograma (EEG, *Electroencephalogram*) (Ouyang *et al.*, 2010; Alzahab *et al.*, 2021).

Os sinais de EEG correspondem a medidas da atividade elétrica agregada de milhões de neurônios situados nas proximidades do eletrodo de registro (Li *et al.*, 2014). Eles são usados em sistemas BCI não invasivos porque apresentam diversas vantagens por ser portátil, relativamente menos custosa financeiramente e fácil de usar com alta resolução temporal (Kayikcioglu; Aydemir, 2010; Kumar *et al.*, 2020; Alzahab *et al.*, 2021). Geralmente, o sistema BCI pode ser utilizado para restaurar as funções motoras por meio de um dispositivo controlado pela BCI, tais como cadeiras de rodas motorizadas (Velasco-alvarez *et al.*, 2013).

De acordo com Abiri *et al.* (2019), BCI se tornou uma das principais áreas de pesquisa em ciência aplicada devido aos avanços na ciência do cérebro e na tecnologia da computação na última década. Os autores ressaltam que esses avanços tecnológicos, como gravação sem fio, amplificadores de baixo custo, análise de aprendizado de máquina e resolução temporal em tempo real, aumentaram o interesse em abordagens BCI baseadas em eletroencefalografia (EEG). Abiri *et al.* (2019) afirmam que ainda há a necessidade de compreender os vários paradigmas experimentais utilizados em sistemas BCI baseados em EEG.

Este trabalho tem como objetivo realizar uma análise comparativa de diferentes abordagens de classificação aplicadas a conjuntos de dados de EEG. Mais especificamente, utiliza dados de emoções humanas relacionados com aplicação de Interface Cérebro-Computador. É avaliado o desempenho de três modelos variantes da Máquina de Aprendizagem Mínima (MLM, *Minimal Learning Machine*) em relação aos classificadores tradicionais de Aprendizado de Máquina (AM), a saber, Perceptron Multicamadas (MLP, *Multilayer Perceptron*) e Máquina de Vetores de Suporte (SVM, *Support Vector Machine*).

A motivação para o desenvolvimento do estudo se dá por conta da importância incluir a avaliação da capacidade do MLM em conjunto de dados de sistemas BCI. Conforme demonstram os trabalhos de Bashashati *et al.* (2015), Lotte *et al.* (2018), Abiri *et al.* (2019), Rashid *et al.* (2020), Gu *et al.* (2021) e Alzahab *et al.* (2021), numerosos algoritmos de classificação com abordagens e valores de hiperparâmetros foram propostos e/ou avaliados na literatura BCI baseada em EEG publicada, tais como a máquina de vetor de suporte, rede neural artificial, análise discriminante linear, classificador Bayesiano, *k*-vizinho mais próximo e até algoritmos de aprendizagem profunda (do inglês, *deep learning*).

Apesar do crescimento recente a passos largos da área de aprendizado profundo, muitas aplicações cotidianas de algoritmos de classificação não precisam de algoritmos muito com-

plexos (Fernández-delgado *et al.*, 2014). E mesmo que a necessidade de se usar algoritmos mais sofisticados seja bem caracterizada em alguns cenários, apenas deveria ser evidente após uma ampla investigação do desempenho de classificadores mais simples, tal com este estudo se propõe a fazer em relação ao MLM.

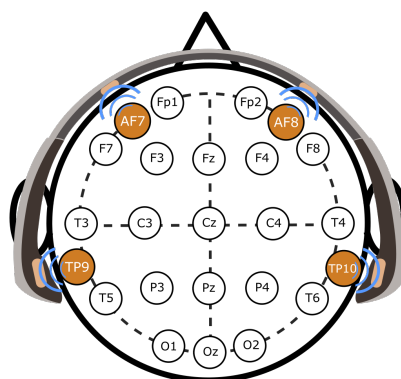
Este trabalho está dividido em sete seções. A seção 2 contém a fundamentação teórica necessária para a melhor compreensão sobre EEG, dados de emoções e algoritmos de classificação utilizadas neste trabalho, sobretudo a MLM e suas variantes. Os trabalhos relacionados que avaliam algoritmos de classificação aplicados em dados EEG de emoção são descritos na seção 3. A seção 4 apresenta a metodologia adotada, a saber, o conjunto de dados, técnicas de pré-processamento de dados, parâmetros de classificadores utilizados e as métricas de avaliação de desempenho. A discussão dos resultados e a análise comparativa com os trabalhos relacionados são abordados na seção 5. Por fim, a seção 7 expõe as considerações finais e trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Esta seção fornece os fundamentos teóricos necessários para a compreensão deste artigo, a saber, eletroencefalograma, Interface Cérebro-Computador e dados emocionais baseados em sinais EEG, bem como, uma descrição sucinta dos classificadores de padrões aplicados neste trabalho.

### 2.1 Eletroencefalograma e Dados emocionais

Figura 1 – Ponto de coleta de dados



Fonte: Adaptado de (Lelis; Filho, 2014).

O EEG é o registro gráfico da atividade elétrica cerebral. Ele representa o resultado do exame, ou seja, o traçado obtido a partir da captação dos sinais elétricos do cérebro. Por meio da técnica chamada eletroencefalografia, esses sinais são captados de forma não invasiva, com eletrodos posicionados na superfície do couro cabeludo, sem a introdução de agulhas ou aparatos na pele. Isso torna o procedimento prático, seguro e amplamente utilizado na investigação de distúrbios neurológicos (Caparelli *et al.*, 2007).

Na Figura 1 pode ser observado os pontos onde são colocados registradores dos eletrodos responsáveis pela captação dos sinais EEG através dos canais cerebrais. As ondas cerebrais consistem em intensidades eletromagnéticas originada das atividades elétrica do cérebro. São medidas em ciclos de tempo em Hz (Hertz). Esses sinais variam sua forma conforme a atividade elétrica dos neurônios (Lelis; Filho, 2014).

Os estados de emoções positivas são alcançados principalmente nas áreas da frente no lado esquerdo do cérebro, por outro lado, a emoções negativas são geradas principalmente na parte direita do cérebro (Huang *et al.*, 2021). Tendo em vista a aquisição dos registros EEG dessas regiões, com o apoio de classificadores adequados, é possível acionar eventos em sistema BCI com base no estado emocional do ser humano. Isso possibilita o suporte aos médicos, por exemplo, na identificação de emoções em pacientes com distúrbio de consciência, tais como o estado vegetativo, o coma, o estado de consciência mínima e entre outras situações em que os pacientes, geralmente, não conseguem expressar emoções adequadas (Huang *et al.*, 2021).

## 2.2 Máquina de Aprendizagem Mínima

A Máquina de Aprendizagem Mínima (MLM) se baseia em um modelo de regressão linear de múltiplas saídas entre as matrizes de distância de espaço de entrada e saída. As matrizes de distância são calculadas em relação a um subconjunto de pontos de dados referidos como pontos de referência. Embora o mapeamento entre as matrizes de distância seja conduzido linearmente, a MLM é baseada na construção de *kernel* usando distâncias de pares para os pontos de referência de tal forma que aprenda as relações de dados não lineares para problemas de classificação e regressão (Junior *et al.*, 2015; Linja *et al.*, 2020).

Uma vantagem significativa da MLM sobre outros métodos de aprendizagem supervisionados é a existência de apenas um hiperparâmetro. Logo, aplica-se ajuste fino para um único parâmetro a fim de obter o melhor modelo durante o processo de aprendizado. A complexidade computacional para o procedimento de treinamento na MLM é muito baixa, competindo com as mais rápidas abordagens de aprendizado de máquina (Junior *et al.*, 2015).

Resultados promissores foram obtidos com a MLM e suas variantes em diversos problemas de classificação e regressão, conforme são demonstrados em alguns trabalhos, tais como Júnior *et al.* (2013), Júnior (2014), Junior *et al.* (2015), Mesquita, Gomes e Junior (2015), Mesquita, Gomes e Junior (2017), Florêncio *et al.* (2018), Dias *et al.* (2018) e Linja *et al.* (2020).

### 2.2.1 Formulação Básica do MLM

Considere um conjunto de dados de entrada rotulados com  $N$  amostras representado por  $X = \{\mathbf{x}_i\}_{i=1}^N$ , com  $\mathbf{x}_i \in R^D$ , em que cada amostra há um padrão representado como um vetor  $\mathbf{x} = (a_1, a_2, \dots, a_D)^T$ , em que  $T$  é a operação matricial de transposta e  $a$  representa um

valor de atributo do conjunto de  $D$  características, isto é, um ponto no espaço  $D$ -dimensional. Considere também a matriz de saídas correspondente  $Y = \{y_i\}_{i=1}^N$ , com  $y_i \in R^S$ , em que as colunas da matriz  $Y$  correspondem às  $S$  saídas e as linhas às  $N$  observações.

Adicionalmente, considera-se a existência de um mapeamento  $f : X \rightarrow Y$  entre o espaço de entrada e saída, de tal forma que  $f$  seja estimado a partir dos dados com o modelo:

$$Y = f(X) + \mathcal{R}. \quad (1)$$

em que  $\mathcal{R}$  correspondem aos vetores de resíduo.

Dada as considerações apresentadas, pode-se descrever o funcionamento da MLM com base em duas etapas, discutidas nas próximas subseções, a saber:

- a) estimar o mapeamento existente entre as distâncias de entrada e saída;
- b) estimar a resposta a partir da configuração dos pontos de saída.

### 2.2.2 Regressão entre Distâncias

Primeiramente são selecionados pontos de entrada de referência  $R = \{m_k\}_{k=1}^K$  com  $R \subseteq X$  com seus correspondentes pontos de referência do espaço de saída  $T = \{t_k\}_{k=1}^K$  com  $T \subseteq Y$ . É definido a matriz de distâncias entre os pontos de referência do espaço de entrada em relação aos pontos de entrada  $D_x \in R^{N \times K}$  de tal forma que sua  $k$ -ésima coluna  $d(X, m_k)$  contenha as distâncias  $d(x_i, m_k)$  entre os  $N$  pontos de entrada e o  $k$ -ésimo ponto de referência  $m_k$ .

Analogamente, é definida a matriz de distâncias entre os pontos de referência do espaço de saída em relação aos pontos de saída  $\Delta_y \in R^{N \times K}$  de forma que sua  $k$ -ésima coluna  $\delta(Y, t_k)$  contém as distâncias  $\delta(y_i, t_k)$  entre os  $N$  pontos de saída  $y_i$  e a saída  $t_k$  do  $k$ -ésimo ponto de referência. O mapeamento  $g$  entre a matriz de distância de entrada  $D_x$  e a matriz de distância de saída correspondente  $\Delta_y$  pode ser reconstruído usando o modelo de regressão multi respostas.

$$\Delta_y = g(D_x) + E, \quad (2)$$

em que as colunas da matriz  $D_x$  correspondem aos  $K$  vetores de entrada e as colunas da matriz  $\Delta_y$  correspondem aos  $K$  vetores de resposta. As colunas da matriz  $E$  correspondem aos  $K$  resíduos. Assume-se que o mapeamento  $g$  entre distâncias de entrada e saída é linear, então o modelo de regressão multiresposta entre as matrizes de distância torna-se

$$\Delta_y = D_x B + E. \quad (3)$$

A coluna  $b_k$  da matriz de regressão  $B$  de dimensão  $K \times K$  corresponde aos coeficientes para as  $K$  respostas. A matriz  $B$  pode ser estimada a partir de dados através da minimização da soma dos quadrados dos resíduos com a função de custo:

$$\text{RSS}(b_k) = (\delta(Y, t_k) - D_x b_k)^T (\delta(Y, t_k) - D_x b_k). \quad (4)$$

Sob as condições normais em que o número de equações na Equação (3) é maior do que o número de incógnitas, o problema em sistemas lineares é sobredeterminado e, geralmente, sem solução. Isto corresponde ao caso em que o número de pontos de referência selecionada é menor que o número de pontos disponíveis para resolver o modelo (*i.e.*,  $K < N$ ) sendo necessário que a solução aproximada seja fornecida pela estimativa dos mínimos quadrados:

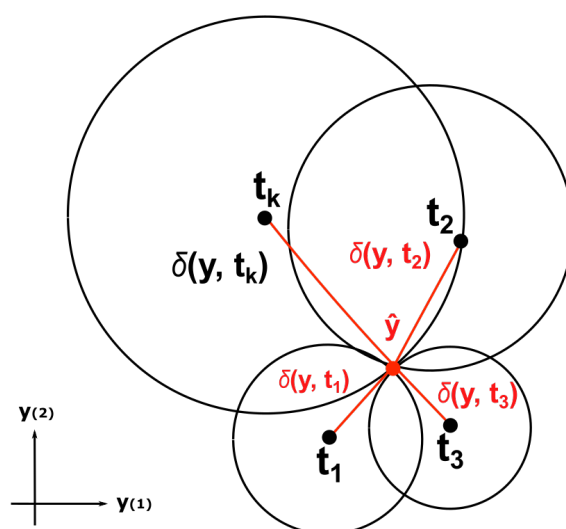
$$\hat{B} = (D_x^T D_x)^{-1} D_x^T \Delta_y. \quad (5)$$

Por outro lado, se na Equação (3) o número de equações é igual ao número de incógnitas (*i.e.*, todos os pontos de aprendizagem são também selecionados como ponto de referência e  $K = N$ ), o problema é unicamente determinado e, normalmente, com uma única solução  $\hat{B} = D_x^{-1} \Delta_y$ . Claramente, menos interessante é o caso em que a Equação (3) contém o número de equações menor do que o número de incógnitas (*i.e.*, para  $K > N$ ), sendo correspondente na qual depois de selecionar os pontos de referência, apenas um pequeno número de pontos de aprendizagem é usado, por isso leva a um problema subdeterminado com, normalmente, infinitas soluções.

Dada a possibilidade de  $\hat{B}$  ser exclusivamente resolvida ou estimada (Equação 5), para um ponto de treinamento  $x_i \in R^D$  cujas distâncias de  $K$  pontos de referência  $\{m_k\}_{k=1}^K$  são coletadas no vetor  $d(x, R) = [d(x, m_1) \dots d(x, m_K)]$ , as distâncias correspondentes entre a saída desconhecida  $y$  e as saídas conhecidas  $\{t_k\}_{k=1}^K$  dos pontos de referências são estimadas por

$$\hat{\delta}(y, T) = d(x, R) \hat{B}. \quad (6)$$

**Figura 2 – Procedimento para obter estimativa**



Fonte: Adaptado de (Junior *et al.*, 2015).

O vetor  $\hat{\delta}(y, T) = [\hat{\delta}(y, t_1) \dots \hat{\delta}(y, t_K)]$  fornece uma estimativa da configuração geométrica entre  $y$  e o conjunto de pontos de referência  $T$ , no espaço- $Y$  ilustrada pela Figura 2.

### 2.2.3 Estimativa da saída

O problema de estimar a saída  $y$ , dada as saídas  $\{t_k\}_{k=1}^K$  para todos os pontos de referências e uma estimativa  $\hat{\delta}(y, T)$  das suas distâncias mútuas, pode ser entendido como um problema de multilateração para estimar sua localização em  $Y$ . O problema de localizar  $y$  é equivalente a resolver um conjunto sobredeterminado de equações não-lineares correspondentes a  $K$  hiper-esferas centradas em  $t_k$  e passando por todo  $y$ , isto é, com um raio igual a  $\hat{\delta}(y, t_k)$ :

$$(y - t_k)^T (y - t_k) = \hat{\delta}^2(y, t_k), \forall k = 1, \dots, K. \quad (7)$$

O problema na Equação (7) pode ser formulado como um problema de otimização na qual uma estimativa  $\hat{y}$  pode ser obtida seguindo a minimização:

$$J(y) = \sum_{k=1}^K \left( (y - t_k)^T (y - t_k) - \hat{\delta}^2(y, t_k) \right)^2. \quad (8)$$

O objetivo tem um mínimo igual a 0 que pode ser alcançado se, e somente se,  $\hat{y}$  é a solução da Equação (7). Se existe, essa solução é global e única.

Devido à incerteza introduzida pelas estimativas entre os pontos de saída  $\hat{\delta}(y, t_k)$ , uma solução ótima para Equação (8) pode ainda ser alcançada usando o método gradiente descendente ou *Levenberg-Marquardt*.

## 2.3 MLM para Classificação

Uma importante classe de problemas é a classificação, na qual o objetivo é prever categorias, normalmente denotadas por saídas qualitativas (ou rótulos de classe). Novamente, tem-se disponível um conjunto de  $N$  pontos de entrada  $X = \{x_i\}_{i=1}^N$ , com  $x_i \in R^D$ , e um conjunto de rótulos correspondentes  $L = \{l_q\}_{q=1}^N$ , com  $l_q \in \{C_1, \dots, C_S\}$ , em que  $C_j$  denota a  $j$ -ésima classe; para  $S = 2$ , o problema é dito ser binário, enquanto para  $S > 2$  tem-se um problema multiclasse.

A Máquina de Aprendizagem Mínima pode ser especializada para problemas de classificação de uma maneira direta representando as  $S$  classes de forma vetorial através de um esquema binário de codificação (1-de- $S$ ). Nessa abordagem, uma variável qualitativa com  $S$  símbolos é representada por um vetor binário de dimensão  $S$ , dos quais somente uma componente é "ativa". Matematicamente, o conjunto de saídas  $Y = \{y_i\}_{i=1}^N$ , com  $y_i \in R^S$ , em correspondência aos pontos de entrada  $X$ , é definida de tal forma que a  $j$ -ésima componente de  $y_i$  é igual a  $\alpha$  se  $l_q = C_j$  e  $\beta$  caso contrário, onde  $\alpha$  e  $\beta$  são inteiros com  $\alpha > \beta$ . Uma escolha usual é  $\alpha = 1$  e  $\beta = -1$ .

Na etapa de classificação de uma observação de teste  $x$  com rótulo de classe desconhecido  $l \in \{C_1, \dots, C_S\}$ , a classe estimada  $\hat{l}$  associada com o vetor de saída estimado  $\hat{y}$  pelo método básico da MLM é dado por  $\hat{l} = C_{s^*}$ , em que

$$s^* = \arg_{s=1, \dots, S} \max \hat{y}^{(s)}. \quad (9)$$

## 2.4 Máquina de Aprendizagem Mínima com Vizinhos Mais Próximos

A Máquina de Aprendizagem Mínima com Vizinhos Mais Próximos (MLM-NN) consiste em uma MLM modificada para a classificação na qual uma busca do vizinho mais próximo é usada em vez do passo de otimização na fase de teste, reduzindo assim sua complexidade computacional, uma vez que tal passo é extensivamente calculado para cada membro do conjunto. Esse método é descrito formalmente por Mesquita, Gomes e Junior (2015).

Mesquita, Gomes e Junior (2015) demonstrou que usando a MLM-NN para tarefas de classificação é possível obter o mesmo resultado da formulação básica da MLM original. Isso significa que o passo de classificação pode ser realizado com base no rótulo do ponto de referência associado com a menor estimativa de distância de saída. Na prática, são utilizadas as distâncias estimadas  $\hat{\delta}(y, t_k)$  dadas a partir da Equação (10).

$$\hat{\delta}(y, T) = d(x, R)\hat{B}. \quad (10)$$

Dessa forma, a classe estimada  $\hat{l}$  para obter um padrão de entrada  $x$  é dada por  $\hat{l} = l_{k^*}$ , onde

$$k^* = \arg \min_{k=1, \dots, K} \{\hat{\delta}^2(y, t_k)\}, \quad (11)$$

e  $l_{k^*}$  representa o rótulo do  $k^*$ -ésimo ponto de referência.

## 2.5 Máquina de Aprendizagem Mínima Ponderada

A variante Máquina de Aprendizagem Mínima Ponderada (*wMLM*, *Weighted Minimal Learning Machine*) foi proposta por Gomes *et al.* (2015). A ideia principal da *wMLM* é ponderar as instâncias do conjunto de treinamento e modificar a contribuição de cada amostra na definição do modelo final da MLM. Ele consiste em uma generalização dos mínimos quadrados entre a matriz de distância de entrada  $D_x$  e a matriz de distância de saída  $\Delta_y$  correspondente.

A *wMLM* é obtida através da extensão do critério dos mínimos quadrados para a forma generalizada. Com a definição de uma matriz de peso  $W$ , então a função de custo da MLM pode ser reescrita como:

$$RSS(B) = \text{tr}\left(W(\Delta_y - D_x B)^T W(\Delta_y - D_x B)\right). \quad (12)$$

Onde  $W$  é uma matriz diagonal e cada elemento  $W_{ii}$  da sua diagonal representa o peso de cada

padrão de treinamento  $\mathbf{x}_i$ .

A otimização pode agora ser definida como um problema de mínimos quadrados ponderado e a matriz  $\mathbf{B}$  pode ser estimada por:

$$\hat{\mathbf{B}} = (\mathbf{D}'_x \mathbf{W} \mathbf{D}_x)^{-1} \mathbf{D}'_x \mathbf{W} \Delta_y. \quad (13)$$

O cálculo modificado de  $\hat{\mathbf{B}}$  define a  $w$ MLM uma vez que todos outros passos podem ser realizados conforme a formulação original.

## 2.6 Máquina de Vetores de Suporte

A Máquina de Vetores de Suporte (SVM) é baseada no princípio estatístico de aprendizado, esse raciocínio, determina uma série de convicções que devem ser seguidas, para o alcance de classificadores com generalização de boa capacidade (Faceli *et al.*, 2021).

O *kernel* é responsável por transformar os dados de entrada no formato necessário. Ele é um importante hiperparâmetro para se configurar no modelo SVM. O parâmetro de grau controla a flexibilidade do limite de decisão. Os núcleos de grau mais alto produzem um limite de decisão mais flexível, os parâmetros de *kernel* e a escolha do parâmetro de regularização,  $C$ , responsável pela penalidade do treinamento (Webb, 2003).

De acordo com Faceli *et al.* (2021), na prática, os *kernels* mais utilizados são os polinomiais, os de *Radial Basis Function Network* (RBF) e os sigmoidais. A função polinomial é expressa como  $(\gamma(\mathbf{x}_i \cdot \mathbf{x}_j) + \kappa)^\rho$ , sendo parametrizada por  $\gamma, \kappa, \rho$ . Já a função RBF é definida por  $\exp(-\sigma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$ , tendo  $\sigma$  como parâmetro principal. Por fim, a função sigmoide é representada por  $\tanh(\gamma(\mathbf{x}_i \cdot \mathbf{x}_j) + \kappa)$ , e depende dos parâmetros  $\gamma$  e  $\kappa$ . Os parâmetros são estipulados pelo usuário.

## 2.7 Perceptron multicamadas

Para solucionar problemas não separáveis linearmente utilizando as Redes Neurais Artificiais (RNAs), é adicionado uma ou mais camadas intermediárias. As redes do tipo Perceptron multicamadas apresentam uma ou mais camadas de neurônios intermediários e uma camada de saída.

Na existência de uma camada intermediária como primeira, cada um de seus neurônios utilizam nas camadas intermediárias funções não lineares de ativação, tais como a função sigmoide e hiperbólica. A função de cada neurônio na MLP é específica, cada neurônio de uma dada camada tem uma combinação de funções realizadas pelos neurônios da camada anterior conectados a ele (Faceli *et al.*, 2021).

A arquitetura do MLP consiste em um sistema de neurônios simples interligados. Segundo Gardner e Dorling (1998), esses neurônios são conectados por sinais e pesos de saída, ativados por funções de ativação. A capacidade da rede de representar funções não lineares

decorre da superposição de diversas funções de ativação não lineares simples. Em uma configuração típica de três camadas, sendo uma de entrada, duas ocultas e uma de saída, o vetor de entrada é transmitido à rede por meio da camada inicial. Tanto os vetores de entrada quanto os de saída podem ser representados como vetores únicos, enquanto as camadas ocultas estabelecem conexões entre a entrada e a saída, permitindo o processamento progressivo e hierárquico das informações.

O papel dos neurônios em cada camada de uma rede neural pode ser compreendido observando como o processamento se torna progressivamente mais complexo à medida que uma amostra avança de uma camada intermediária para a seguinte. Cada neurônio da nova camada realiza combinações dinâmicas de hiperplanos gerados pelos neurônios da camada anterior. Essa composição de funções desempenhadas por cada unidade da rede é o que define a função global associada à RNA, permitindo a classificação entre a **classe A** e a **classe B**, com base em uma nova entrada fornecida à rede previamente treinada.

### 3 TRABALHOS RELACIONADOS

Esta seção apresenta cinco estudos relacionados de aplicação para sistemas BCI que utilizaram dados emocionais baseados em sinais EEG. Eles obtiveram bons resultados (> 70%) na aplicação dos algoritmos MLP e SVM. A subseção 3.1 descreve a metodologia de seleção desses cinco trabalhos, enquanto a subseção 3.2 analisa os objetivos, principais métodos e resultados dos trabalhos, bem como, o diferencial para este estudo.

#### 3.1 Seleção dos trabalhos relacionados

Os cinco estudos selecionados serviram como sustentação teórica, referencial de metodologias adequadas preparação de dados e comparação de desempenho de modelos. As *strings* de busca utilizadas para localizar os trabalhos de Wang, Nie e Lu (2014), Bhatti *et al.* (2016) e Bhatti *et al.* (2016) foram adaptadas conforme o repositório consultado, considerando a aplicação de sinônimos relacionados ao tema com o intuito de ampliar a abrangência da busca e garantir uma varredura mais eficaz. O Quadro 1 apresenta a *string* de busca de cada um dos três trabalhos.

**Quadro 1 – Metodologia de seleção dos três trabalhos relacionados**

<b>Repositório</b>	<b>String de busca</b>	<b>Trabalho selecionado</b>
<i>ScienceDirect</i> <sup>6</sup>	"emotional state"AND "EEG"AND "machine learning"	Wang, Nie e Lu (2014)
<i>Google Scholar</i> <sup>7</sup>	"emotion human"AND "brain signals"AND "recognition"	Bhatti <i>et al.</i> (2016)
<i>IEEE Xplore</i> <sup>8</sup>	"emotion classification"AND "EEG in scalp"	Jalilifard, Pizzolato e Islam (2016)

**Fonte: Dados da pesquisa.**

Os critérios de inclusão dos três trabalhos foram: uso de dados emocionais baseados em sinais EEG, estudo de aplicação para sistemas BCI, avaliação de algoritmos MLP e/ou SVM e apresentar bons resultados (> 70%). Os critérios de exclusão foram: avaliação de outros algoritmos de aprendizagem sem SVM e MLP, não aplicação para sistemas BCI e dados emocionais não baseados em EEG.

Por sua vez, os trabalhos de Bird *et al.* (2019b) e Bird *et al.* (2019a) foram incluídos diretamente na análise, uma vez que os autores são responsáveis pela criação do conjunto de dados utilizado neste estudo.

### 3.2 Análise dos estudos selecionados

Wang, Nie e Lu (2014) investigaram as características dos atributos do EEG para classificação de emoções e técnica para rastrear a trajetória das mudanças emocionais. Uma série de experimentos de uso de clipes de filme foram projetados para despertar a emoção dos voluntários e um conjunto de dados de EEG de 6 pessoas foram coletados. Após aplicação de técnicas de pré-processamento de sinais e métodos de redução de dimensionalidade, a acurácia média dos experimentos de classificação de emoções pelo SVM (Kernel Linear) foi 87,53%. Eles também avaliaram o classificador Análise Discriminante Linear (LDA, *Linear Discriminant Analysis*) com 91,77% como melhor média de acurácia.

Bhatti *et al.* (2016) conduziram um estudo sobre reconhecimento de emoções felizes, tristes, de amor e raiva por sinais EEG em resposta a faixas de áudio de gêneros eletrônicos, *rap*, metal, *rock* e *hip-hop*. Os participantes ouviram faixas de música de áudio de 1 minuto para cada gênero em um ambiente livre de ruído. 30 pessoas de três diferentes faixas etárias (15 e 25 anos; 26 e 35 anos; 36 e 50 anos) passaram pelo experimento. Os experimentos computacionais incluíram a classificação das emoções humanas, nas quais foi evidenciado que o MLP forneceu a melhor taxa de acurácia de 78,11% em comparação com os classificadores SVM e *k*-Vizinhos Mais Próximos (*k*-NN, *k*-*Nearest Neighbor*) que apresentaram 75,62% e 72,80% de acurácia, respectivamente.

## Quadro 2 – Resumo dos estudos selecionados

Referência	Contexto do Estudo	Principais Achados
Wang, Nie e Lu (2014)	Classificação de emoções por EEG com estímulos de clipes de filmes (6 participantes)	SVM (Kernel Linear) obteve acurácia média de 87,53%; LDA apresentou melhor desempenho com 91,77%. Técnicas de pré-processamento e redução de dimensionalidade foram aplicadas.
Bhatti <i>et al.</i> (2016)	Reconhecimento de emoções (felicidade, tristeza, amor, raiva) por EEG em resposta a faixas de áudio de diferentes gêneros musicais (30 participantes, 3 faixas etárias)	MLP obteve melhor acurácia (78,11%) em comparação com SVM (75,62%) e <i>k</i> -NN (72,80%). Ambiente controlado e estímulos auditivos de 1 minuto.
Jalilifard, Pizzolato e Islam (2016)	Classificação de estados emocionais (neutro, relaxante, assustador) por EEG após visualização de videoclipes (19 participantes)	SVM obteve acurácia média de 92%; <i>k</i> -NN superou com 94%. Estudo focado em dois estados emocionais distintos.
Bird <i>et al.</i> (2019b)	Projeto do conjunto de dados utilizado neste estudo; avaliação de modelos com abordagem 10- <i>fold</i>	Random Forest obteve 97,89% de acurácia; MLP alcançou 94,89%. Dados utilizados como base para este trabalho.
Bird <i>et al.</i> (2019a)	Otimização evolutiva de MLP e comparação com LSTM para classificação de emoções por EEG	MLP otimizado obteve 96,23% de acurácia; LSTM apresentou desempenho superior com 97,06%. Estudo focado em ajuste de hiperparâmetros e aprendizado profundo.

Fonte: Dados da pesquisa.

Jalilifard, Pizzolato e Islam (2016) realizaram um estudo de classificação de dois estados emocionais diferentes por meio de dados de EEG. 19 pessoas participaram do experimento assistindo a um videoclipe que estimulava três estados emocionais: neutro, relaxante e assustador. Após técnicas de pré-processamento, os dados de EEG foram classificados, sendo obtida a taxa média de acurácia de 92% usando o SVM, enquanto o *k*-NN obteve uma acurácia média de 94%.

Bird *et al.* (2019b) projetaram o conjunto de dados utilizado neste trabalho (ver seção 4.1). Por meio de uma abordagem 10-*fold*, os modelos *Random Forest* e MLP alcançaram uma acurácia de 97,89% e 94,89%, respectivamente. Em um segundo artigo, os autores avaliaram a otimização de um MLP realizada com uma abordagem evolutiva antes da classificação para estimar os melhores hiperparâmetros da rede. Uma análise comparativa foi realizada com algoritmos de aprendizagem profunda *Long Short-Term Memory* (LSTM). O MLP otimizado

obteve uma acurácia de 96,23 %, enquanto LSTM obteve 97,06 % (Bird *et al.*, 2019a).

Conforme pode ser visto no resumo do Quadro 2, os trabalhos tratam da avaliação de algoritmos de aprendizado de máquina e aprendizagem profunda aplicadas na classificação de emoções baseadas em sinais EEG. Este trabalho avalia o desempenho da MLM aplicada em dados EEG de emoções com resultados semelhantes ou superiores aos algoritmos tradicionais, conforme discutido na seção 5.

## 4 MATERIAIS E MÉTODOS

A metodologia utilizada nesta pesquisa caracteriza-se como uma pesquisa aplicada, voltada para o uso do conhecimento científico na abordagem das variantes da MLM sobre dados de EEG. Quanto aos objetivos, trata-se de uma pesquisa descritiva e analítica, uma vez que busca caracterizar e analisar o desempenho dos modelos por meio de métricas quantitativas. Além disso, a pesquisa possui caráter de replicabilidade científica, ao utilizar o conjunto de dados projetado por Bird *et al.* (2019b) para executar experimentos adaptados neste estudo, permitindo a validação e comparação de resultados. O método adotado é o experimento comparativo, com manipulação de variáveis e aplicação de diferentes modelos de aprendizado de máquina, visando sustentar os achados do estudo por meio de análises quantitativas e interpretações técnicas (Prodanov; Freitas, 2018).

### 4.1 Conjunto de Dados

Pelo motivo em que os conjuntos de dados dos trabalhos relacionados não estarem disponíveis de forma pública, foi utilizado do conjunto de dados *Feeling Emotions*<sup>9</sup> do repositório público *kaggle* (Bird *et al.*, 2019b; Bird *et al.*, 2019a).

*Feeling Emotions* é um conjunto de dados de ondas cerebrais EEG composto por 2132 linhas e 2549 colunas, sendo uma delas a coluna de atributo alvo intitulada *label* e as outras 2548 colunas representam os canais. Os dados foram coletados de duas pessoas (1 homem, 1 mulher) durante 3 minutos por estado: positiva, neutra, negativa. Utilizaram uma bandana Muse EEG que registrou as colocações **EEG TP9**, **AF7**, **AF8** e **TP10** via eletrodos secos, como está apresentado na Figura 1. Seis minutos de dados neutros de descanso também são registrados. Os estímulos usados para evocar as emoções negativas foram exibidas cenas de morte e funerária. Para estimular emoções positivas, foi tocado musical de abertura de *La La Land*, incluindo lapso de tempo da natureza e clipes engraçados de cachorro.

*Feeling Emotions*, disponível publicamente, foi escolhida por sua acessibilidade e adequação ao objetivo de reconhecer emoções humanas por meio de sinais EEG. Apesar de ser apropriada para estudos exploratórios, apresenta limitações como o número reduzido de participantes e a curta duração das sessões, o que impacta a generalização dos resultados. Ainda assim, os estímulos emocionais utilizados foram cuidadosamente definidos por Bird *et al.*

---

<sup>9</sup>Disponível em <https://www.kaggle.com/birdy654/eeg-brainwave-dataset-feeling-emotions>

(2019b), Bird *et al.* (2019a), garantindo coerência na rotulação e viabilidade para análise preliminar dos modelos propostos.

## 4.2 Pré-processamento de dados

Para a etapa de pré-processamento dos dados contidos no conjunto de dados utilizado, não foi necessário a utilização de métodos de balanceamento de dados, pois a distribuição das amostras entre as classes são quase semelhantes: 716 amostras da classe Neutro, 708 registros da classe Positivo e 708 amostras de emoção Negativa.

Para manipular o atributo alvo foi necessário realizar a troca dos dados categóricos em texto para números, já que as arquiteturas de classificação necessitam operar com valores numéricos. Por conseguinte, a classe **NEUTRAL** foi substituída por 0, a classe **POSITIVE** substituída por 1 e **NEGATIVE** substituído por 2.

Em seguida, foram divididos os dados em dois subconjuntos  $X$  e  $Y$ : atributos preditores e atributo alvo, respectivamente. Por fim, os dados do subconjunto  $X$  foram normalizados para evitar que um atributo predomine sobre outro. Para tal, a normalização dimensiona cada variável de entrada separadamente para a faixa entre zero (0) e um (1) (Faceli *et al.*, 2021).

## 4.3 Experimentos computacionais

Para os experimentos foram implementados na Linguagem de Programação Python o ambiente laboratorial *Google Colaboratory*. Em relação aos classificadores convencionais foram utilizadas as bibliotecas *sklearn svm*<sup>1</sup> e *MLPClassifier*<sup>2</sup>. Para a família da *MLM* foram utilizadas a biblioteca *skmlm*<sup>3</sup>. Para as métricas de avaliação foi utilizado a biblioteca *sklearn metrics*<sup>4</sup>, sendo elas *accuracy score*, *precision score*, *recall score* e *F1-score*.

Foi realizado ajuste fino dos parâmetros dos modelos para obter o melhor modelo de classificação durante os experimentos. Os parâmetros do SVM foram inspirados no trabalho relacionado de Wang, Nie e Lu (2014), já que ele alternou entre os *Kernels* polinomial, linear, função de base radial e suas penalidades  $C$ . Os parâmetros do MLP foram alternados na taxa de aprendizagem e função de ativação, já que, a taxa de aprendizado inicial se permaneceu constante, mas diferente do valor padrão de 0.001.

Foi utilizada validação cruzada *holdout* com 20% dos dados para teste e os outros 80% para treinamento dos classificadores, visto que essa proporção obteve melhores resultados em comparação com 70% treinamento e 30% teste. As quantidades de execuções (rodadas para cada treinamento e validação) foram 30 vezes. A cada nova rodada os dados de entrada são embaralhados e divididos em 80% para treinamento e 20% para teste, garantindo assim, que os classificadores não tenham um sobreajuste (*overfitting*). O Quadro 3 apresenta a configuração

---

<sup>1</sup>Disponível em: <<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>>

<sup>2</sup>Disponível em: <[https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html)>

<sup>3</sup>Disponível em: <<https://github.com/omadson/scikit-mlm>>

<sup>4</sup>Disponível em: <[https://scikit-learn.org/stable/modules/model\\_evaluation.html](https://scikit-learn.org/stable/modules/model_evaluation.html)>

final dos hiperparâmetros dos modelos de cada experimento.

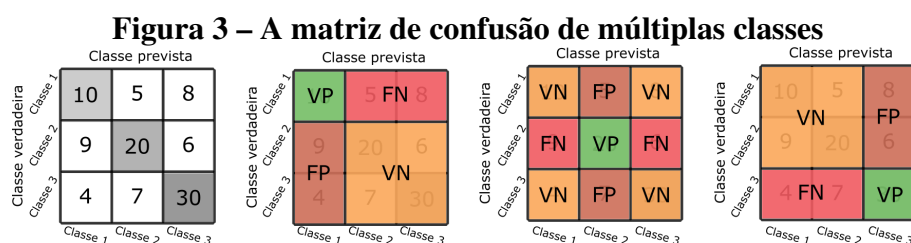
**Quadro 3 – Hiperparâmetros dos modelos de classificação dos experimentos realizados**

Experimento	MLP	SVM	MLM wMLM MLM-NN
1º	learning_rate = invscaling activation = 'logistic' learning_rate_init = 0.00001	Kernel = 'poly' C= 0.0625	rp_number = 275
2º	learning_rate = adaptive activation = 'tanh' learning_rate_init = 0.00001	Kernel = 'linear' C= 512	rp_number = 412
3º	learning_rate = constant activation = 'relu' learning_rate_init = 0.00001	Kernel = 'rbf' C= 1024	rp_number = 550

Fonte: Dados da pesquisa.

#### 4.4 Métricas de avaliação

As avaliações dos algoritmos de classificação utilizaram as métricas de acurácia, precisão, sensibilidade, medida-F1 e especificidade durante as iterações. Todas as referências desta seção foram baseadas nos trabalhos de Sokolova e Lapalme (2009), Faceli *et al.* (2021) e Markoulidakis *et al.* (2021).



Fonte: Adaptado de (Markoulidakis *et al.*, 2021).

As métricas de desempenho são calculadas a partir da matriz de confusão, ilustrada na Figura 3, que permite visualizar a eficácia do classificador. Para interpretar essas métricas corretamente, é essencial compreender a disposição dos valores na matriz: os verdadeiros positivos (VP) representam as amostras corretamente classificadas como positivas; os verdadeiros negativos (VN) indicam os padrões corretamente identificados como negativos; os falsos positivos (FP) correspondem às previsões incorretas em que padrões negativos são classificados como positivos; e os falsos negativos (FN) referem-se aos exemplos positivos que o classificador identifica erroneamente como negativos.

As métricas contribuem na avaliação do classificador, por exemplo, a acurácia releva a taxa em que as classes foram rotuladas de maneira esperada. O percentual de acerto é obtida

pela razão entre o total de entradas e a quantidade de acertos. Na matriz de confusão, ela é obtida pelo somatório dos valores da diagonal principal da matriz dividida pelo somatório total dos elementos. Para calcular a acurácia, usa-se a Equação (14):

$$acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (14)$$

A precisão aponta a proporção do modelo em prever corretamente positivo entre todas os dados rotulados como positivo. Essa métrica é calculada pela quantidade de verdadeiros positivos sobre a somatório dos valores verdadeiros positivos mais os falsos positivos, conforme Equação (15).

$$precisão = \frac{VP}{VP + FP} \quad (15)$$

A sensibilidade avalia a capacidade do sistema em prever corretamente a classe como positiva, em relação ao somatório dos valores positivos com o falsos negativos. Ela é calculada pela Equação (16).

$$sensibilidade = \frac{VP}{VP + FN} \quad (16)$$

A medida-F1, lida com a média harmônica que considera precisão e sensibilidade de acordo com a Equação (17).

$$medida-F1 = \frac{2 \times precisão \times sensibilidade}{precisão + sensibilidade} \quad (17)$$

A especificidade avalia o método em detectar resultados negativos pela divisão da quantidade de verdadeiros positivos pela soma de todas as amostras de classe negativa conforme a Equação (18).

$$especificidade = \frac{VN}{VN + FP} \quad (18)$$

## 5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesta seção são apresentados os resultados obtidos pelos classificadores avaliados no contexto da aplicação proposta. Para cada experimento computacional, são analisadas as métricas de desempenho dos modelos de classificação, com o objetivo de verificar o comportamento da abordagem baseada em MLM em comparação ao MLP e ao SVM, considerando o número reduzido de hiperparâmetros envolvidos na configuração

**Tabela 1 – Resultados do 1º experimento computacional: média das métricas**

<b>Classificador</b>	<b>Acurácia</b>	<b>Precisão</b>	<b>Sensibilidade</b>	<b>F1-Score</b>	<b>Especificidade</b>
SVM	93,15%	93,58%	93,10%	93,05%	<b>98,63%</b>
MLM	<b>94,61%</b>	<b>94,81%</b>	<b>94,63%</b>	<b>94,61%</b>	98,62%
wMLM	<b>94,97%</b>	<b>95,16%</b>	<b>95,07%</b>	<b>94,96%</b>	98,54%
MLM-NN	<b>95,15%</b>	<b>95,27%</b>	<b>95,18%</b>	<b>95,11%</b>	98,43%
MLP	91,56%	91,86%	91,45%	91,34%	97,10%

Fonte: Dados da pesquisa.

Na Tabela 1 obtivemos as médias de cada métrica das 30 execuções do primeiro experimento computacional. Podemos observar pelas métricas acurácia, precisão, sensibilidade, medida-F1 e especificidade que todas as médias ficaram acima de 91%. Com exceção da especificidade, a família da MLM obteve desempenho superior em relação aos demais classificadores. A especificidade do SVM foi acima com 98,63%, enquanto a MLM ficou bem próxima com 98,62% (desempenho semelhante).

Na Tabela 2, referente ao segundo experimento computacional, observa-se que os classificadores da família MLM apresentaram valores próximos de 96% nas métricas de acurácia, precisão e medida F1. Na métrica de especificidade, os resultados ficaram acima de 98%. O classificador MLP apresentou valores entre 90,3% e 96,3% em todas as métricas avaliadas. Já o SVM obteve valores intermediários entre os demais, com destaque nas métricas de acurácia e precisão. Esses resultados indicam variações de desempenho entre os modelos testados, conforme evidenciado pelas métricas apresentadas.

**Tabela 2 – Resultados do 2º experimento computacional: média das métricas**

<b>Classificador</b>	<b>Acurácia</b>	<b>Precisão</b>	<b>Sensibilidade</b>	<b>F1-Score</b>	<b>Especificidade</b>
SVM	95,52%	95,52%	95,55%	95,51%	97,72%
MLM	<b>96,07%</b>	<b>96,16%</b>	<b>96,10%</b>	<b>96,05%</b>	<b>98,76%</b>
wMLM	<b>95,61%</b>	<b>95,73%</b>	<b>95,63%</b>	<b>95,58%</b>	<b>98,25%</b>
MLM-NN	<b>95,71%</b>	<b>95,81%</b>	<b>95,65%</b>	<b>95,66%</b>	<b>98,95%</b>
MLP	90,55%	90,71%	90,48%	90,27%	96,31%

Fonte: Dados da pesquisa.

Neste segundo experimento todos os classificadores com exceção do MLP se saíram melhores que no primeiro experimento.

**Tabela 3 – Resultados do 3º experimento computacional: média das métricas**

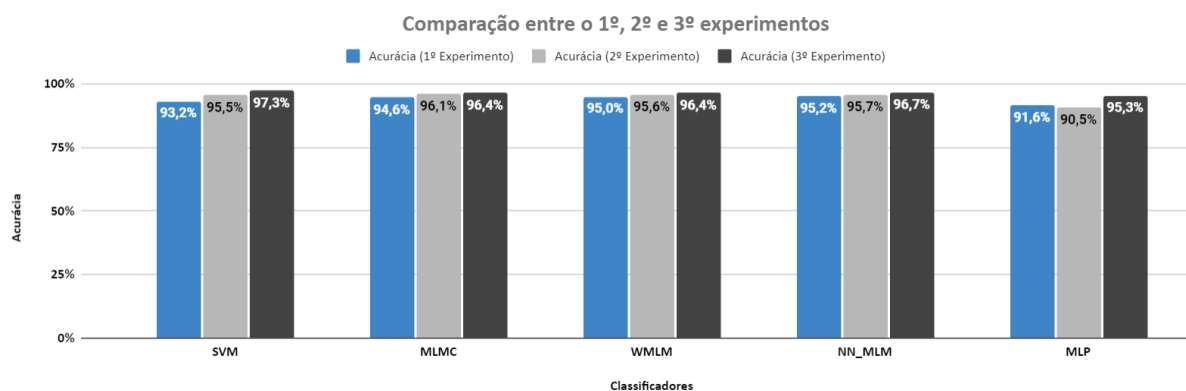
<b>Classificador</b>	<b>Acurácia</b>	<b>Precisão</b>	<b>Sensibilidade</b>	<b>F1-Score</b>	<b>Especificidade</b>
SVM	<b>97,28%</b>	<b>97,28%</b>	<b>97,25%</b>	<b>97,24%</b>	98,50%
MLM	96,42%	96,46%	96,41%	96,38%	<b>98,82%</b>
wMLM	96,37%	96,35%	95,63%	96,35%	<b>98,91%</b>
MLM-NN	96,67%	96,71%	96,64%	96,62%	<b>98,62%</b>
MLP	95,27%	95,43%	95,29%	95,24%	98,01%

Fonte: Dados da pesquisa.

Em relação ao terceiro experimento computacional, conforme apresentado na Tabela 3, observa-se que o classificador SVM obteve médias acima de 95% nas métricas avaliadas, com exceção da especificidade. A família MLM apresentou resultados próximos ao SVM, com destaque na métrica de especificidade. O MLP registrou desempenho em torno de 95% nas métricas analisadas. Nesse experimento, os três modelos apresentaram resultados mais elevados em comparação aos experimentos anteriores, conforme evidenciado pelas médias das métricas de acurácia, precisão, sensibilidade, medida F1 e especificidade.

Com base nos resultados dos três experimentos, nos permite evidenciar a semelhança ao nível de capacidade de assertividade da família da MLM com os outros classificadores SVM e MLP para o problema de classificação deste trabalho. A Figura 4 permite visualizar melhor como cada modelo se comportou nos diferentes experimentos. Para cada experimento há uma alteração nos resultados de classificação, nos levando a entender que os hiperparâmetros modificados tiveram sua parcela na forma como os dados se apresentaram no resultado. Portanto, é importante encontrar boas configurações para obtenção dos melhores resultados. Além disso, mostra que em relação aos modelos tradicionais, as variantes da MLM propostas possuem ótimos resultados, sendo assim, são alternativas eficazes para uso em sistemas BCI.

**Figura 4 – Histograma de resultados agregados dos experimentos computacionais**



Fonte: Dados da pesquisa.

## 6 COMPARATIVO INDIRETO COM TRABALHOS SIMILARES

Neste tipo de pesquisa envolvendo dados EEG e métodos de Inteligência Computacional, é uma prática científica aceitável comparar trabalhos moderadamente similares. Isso pode ser observado nos trabalhos de Bhatti *et al.* (2016) e Bird *et al.* (2019b) que realizaram uma comparação com trabalhos similares mesmo utilizando conjuntos de dados EEG de emoções distintos. Portanto, o nosso intuito na comparação com trabalhos moderadamente relacionados visa contextualizar o esquema proposto com os métodos de reconhecimento de emoções humanas usando sinais de EEG de última geração.

**Tabela 4 – Análise indireta da nossa proposta com trabalhos relacionados**

<b>Estudo</b>	<b>Método</b>	<b>Acurácia</b>
<b>Este estudo</b>	MLP	95.27%
<b>Este estudo</b>	SVM	97.28%
<b>Este estudo</b>	MLM	96.42%
<b>Este estudo</b>	wMLM	96.37%
<b>Este estudo</b>	MLM-NN	96.67%
Bird <i>et al.</i> (2019b)	MLP	94.89%
Bird <i>et al.</i> (2019a)	MLP	96.23%
Wang, Nie e Lu (2014)	SVM	87.53%
Bhatti <i>et al.</i> (2016)	MLP	78.11%
Bhatti <i>et al.</i> (2016)	SVM	75.62%
Jalilifard, Pizzolato e Islam (2016)	SVM	92.00%

**Fonte: Dados da pesquisa.**

A Tabela 4 compara as melhores acurácias de cada classificador abordado neste trabalho em comparação com os classificadores SVM e MLP dos trabalhos relacionados, apesar do conjunto de dados serem distintos na maioria das obras. A meta é mostrar que os classificadores da família da MLM apresentam bons ou ótimos resultados em comparação com outros algoritmos tradicionais encontradas na literatura sobre reconhecimento de emoções humanas.

Os projetistas Bird *et al.* (2019b), Bird *et al.* (2019a) do conjunto de dados utilizado nesta pesquisa obtiveram as acurácias de 94.89% e 96.23% ao aplicar o MLP com hiperparâmetros distintos (Bird *et al.*, 2019b; Bird *et al.*, 2019a). Os demais trabalhos apresentaram uma acurácia entre 75% e 92%. Destacam-se os trabalhos de Bhatti *et al.* (2016) com 75.62% (menor acurácia) e Jalilifard, Pizzolato e Islam (2016) com 92.00% (maior acurácia) ao aplicarem o SVM. Tendo uma semelhança nos resultados com os outros classificadores do estado da arte, pode-se concluir que a MLM é um classificador com resultados promissores para a aplicação em problemas com reconhecimento de padrões emocionais humanas a partir de dados EEG, bem como no desenvolvimento de sistemas BCI.

## **6.1 Limitações e forças do estudo**

Embora este estudo tenha alcançado seu objetivo de realizar uma análise comparativa entre diferentes abordagens de classificação aplicadas a dados de EEG para reconhecimento de emoções humanas, algumas limitações devem ser consideradas. A primeira limitação refere-se ao conjunto de dados utilizado, que contém registros de apenas duas pessoas (ver Seção 4.1). Essa amostra reduzida pode comprometer a capacidade de generalização dos resultados, uma vez que não representa a diversidade interindividual presente em sinais cerebrais relacionados às emoções. Estudos com maior número de participantes são essenciais para validar a robustez dos modelos propostos. Além disso, não foram avaliados outros algoritmos de apren-

dizado relevantes na literatura, tais como *Random Forest*, *Extra Trees*, *Gradient Boosting* e *Deep Learning*, que poderiam oferecer diferentes perspectivas de desempenho e complexidade computacional.

O estudo apresenta diversas contribuições relevantes, destacando-se pela atualidade do tema e pela aplicação da MLM no reconhecimento de emoções humanas a partir de sinais EEG. A estrutura metodológica adotada, aliada à realização de três experimentos com variações de hiperparâmetros e múltiplas métricas de avaliação, demonstra rigor e consistência na análise dos resultados. Além disso, a comparação indireta com trabalhos anteriores reforça a competitividade da MLM frente aos algoritmos consagrados como SVM e MLP. Os achados indicam que a MLM possui desempenho semelhante ou superior, com potencial de aplicação prática em sistemas BCI, beneficiando-se de sua flexibilidade na construção de modelos de aprendizado.

## 7 CONSIDERAÇÕES FINAIS

Este trabalho contribuiu com um estudo da área de Inteligência Computacional Aplicada que avaliou o desempenho da MLM na classificação de dados EEG de emoções humanas. Uma análise comparativa com os modelos SVM e MLP mostra a semelhança da MLM, que possui a vantagem de realizar um processo de ajuste fino em relação ao único hiperparâmetro: quantidade de pontos de referência. Ademais, a complexidade computacional para o procedimento de treinamento na MLM é muito baixa, competindo com as mais rápidas abordagens de aprendizado de máquina. Com isso, o objetivo do estudo foi alcançado, pois foi possível realizar uma análise comparativa entre duas abordagens tradicionais de classificação aplicadas a dados EEG, evidenciando o desempenho promissor da MLM frente aos modelos MLP e SVM e sua viabilidade para aplicações em Interface Cérebro-Computador.

Observou-se que a MLM se mostrou como um classificador mais flexível de construir o modelo adequado por ter menor quantidade de hiperparâmetros para ser configurado, com ocasiões havendo a necessidade da configuração de um único hiperparâmetro. A MLM assemelha-se ao nível de resultados com os outros classificadores sendo uma alternativa eficaz para classificar dados EEG relacionados com sistemas BCI. Dentro dos seus melhores resultados a família da MLM podemos destacar a acurácia de 96,7%, o SVM de 97,3% e o MLP 95,3%.

Como trabalhos futuros, sugere-se a ampliação da análise por meio da inclusão de algoritmos adicionais de aprendizado supervisionado, tais como *Random Forest*, *Extra Trees*, *Gradient Boosting* e técnicas de *Deep Learning*, que podem oferecer diferentes perspectivas quanto ao desempenho e à complexidade computacional. Recomenda-se também a aplicação de métodos Inteligência Artificial Explicável, a fim de interpretar as decisões dos classificadores e compreender melhor os padrões extraídos dos sinais EEG. Além disso, propõe-se investigar o impacto de diferentes técnicas de redução de dimensionalidade, como *Principal Component Analysis* (PCA), *Locally-Linear Embedding* (LLE), *t-Distributed Stochastic Neighbor Embedding* (t-SNE) e *Uniform Manifold Approximation and Projection* (UMAP), bem como o uso

de métodos de seleção de atributos, como *Sequential Forward Selection* (SFS) e *Sequential Backward Selection* (SBS). A adoção de validação cruzada com o método k-fold pode contribuir para uma avaliação mais robusta dos modelos. Por fim, recomenda-se explorar variantes adicionais da MLM, como a versão com opção de rejeição e abordagens baseadas em Lógica *Fuzzy*, ampliando o potencial de adaptação da técnica a diferentes contextos de classificação.

## REFERÊNCIAS

ABIRI, Reza *et al.* A comprehensive review of EEG-based brain–computer interface paradigms. **Journal of neural engineering**, IOP Publishing, v. 16, n. 1, p. 011001, 2019.

ALZAHAB, Nibras Abo *et al.* Hybrid Deep Learning (hDL)-Based Brain-Computer Interface (BCI) Systems: A Systematic Review. **Brain Sciences**, v. 11, n. 1, 2021. ISSN 2076-3425. Disponível em: <<https://www.mdpi.com/2076-3425/11/1/75>>.

BASHASHATI, Hossein *et al.* Comparing different classifiers in sensory motor brain computer interfaces. **PLoS one**, Public Library of Science San Francisco, CA USA, v. 10, n. 6, p. e0129435, 2015.

BHATTI, Adnan Mehmood *et al.* Human emotion recognition and analysis in response to audio music using brain signals. **Computers in Human Behavior**, Elsevier, v. 65, p. 267–275, 2016. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0747563216305945?via%3Dihub>>.

BIRD, Jordan J *et al.* A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction. **Complexity**, Hindawi, v. 2019, 2019.

BIRD, Jordan James *et al.* Mental emotional sentiment classification with an EEG-based brain-machine interface. In: **Proceedings of the International Conference on Digital Image and Signal Processing (DISP'19)**. United Kingdom: St Hugh's College, Oxford University, 2019. p. 75–78. ISBN 978-1-912532-05-6. **Anais:**<[https://imm.dtu.dk/~alan/DISP2019\\_proceedings.pdf](https://imm.dtu.dk/~alan/DISP2019_proceedings.pdf)>. Disponível em: <<https://jordanjamesbird.com/publications/Mental-Emotional-Sentiment-Classification-with-an-EEG-based-Brain-machine-Interface.pdf>>.

CAPARELLI, Thiago Bruno *et al.* Projeto e desenvolvimento de um sistema multicanal de biotelemetria para detecção de sinais ECG, EEG e EMG. Universidade Federal de Uberlândia, 2007. Dissertação de Mestrado do Programa de Pós-graduação em Engenharia Elétrica.

DIAS, Madson Luiz Dantas *et al.* Opposite neighborhood: a new method to select reference points of minimal learning machines. In: **European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning 2018**. Bruges (Belgium): ESANN, 2018. p. 201–206. ISBN 978-287587047-6. **Anais:**<<https://www.esann.org/proceedings/2018>>.

FACELI, Katti *et al.* **Inteligência artificial: uma abordagem de aprendizado de máquina**. Barueri (SP): LTC, 2021. ISBN 9788521637349.

FERNÁNDEZ-DELGADO, Manuel *et al.* Do we need hundreds of classifiers to solve real world classification problems? **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 3133–3181, 2014. Disponível em: <<http://jmlr.org/papers/v15/delgado14a.html>>.

FLORÊNCIO, José AV *et al.* A fuzzy c-means-based approach for selecting reference points in minimal learning machines. In: SPRINGER. **North American Fuzzy Information Processing Society Annual Conference**. Fortaleza (CE): Springer, Cham, 2018. p. 398–407.

GARDNER, Matt W; DORLING, Stephen R. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. **Atmospheric environment**, Elsevier, v. 32, n. 14-15, p. 2627–2636, 1998.

GOMES, João Paulo Pordeus *et al.* A cost sensitive minimal learning machine for pattern classification. In: SPRINGER. **International Conference on Neural Information Processing**. Montréal (Canadá): Springer, Cham, 2015. p. 557–564. **Anais:**<<https://link.springer.com/book/10.1007/978-3-319-26532-2>>. Disponível em: <[https://doi.org/10.1007/978-3-319-26532-2\\_61](https://doi.org/10.1007/978-3-319-26532-2_61)>.

GU, Xiaotong *et al.* EEG-based brain-computer interfaces (BCIs): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. **IEEE/ACM transactions on computational biology and bioinformatics**, IEEE, 2021.

HUANG, H. *et al.* An EEG-based brain computer interface for emotion recognition and its application in patients with Disorder of Consciousness. **IEEE Transactions on Affective Computing**, IEEE, 2021.

JALILIFARD, Amir; PIZZOLATO, Ednaldo Brigante; ISLAM, Md Kafiul. Emotion classification using single-channel scalp-EEG recording. In: IEEE. **2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. Florida (USA): IEEE, 2016. p. 845–849. **Anais:**<<https://ieeexplore.ieee.org/xpl/conhome/7580725/proceeding>>. Disponível em: <<https://ieeexplore.ieee.org/document/7590833>>.

JÚNIOR, Amauri Holanda de Souza. Regional models and minimal learning machines for nonlinear dynamical system identification. Universidade Federal do Ceará, 2014. Tese de doutorado do Programa de Pós-Graduação em Engenharia de Teleinformática.

JUNIOR, Amauri Holanda de Souza *et al.* Minimal learning machine: a novel supervised distance-based approach for regression and classification. **Neurocomputing**, Elsevier, v. 164, p. 34–44, 2015.

JÚNIOR, Amauri Holanda de Souza *et al.* Extending the Minimal Learning Machine for Pattern Classification. In: IEEE. **2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence**. Ipojuca (PE): IEEE, 2013. p. 236–241.

KAYIKCIOGLU, Temel; AYDEMIR, Onder. A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data. **Pattern Recognition Letters**, Elsevier, v. 31, n. 11, p. 1207–1215, 2010.

KUMAR, M Keerthi *et al.* Comparative analysis to identify efficient technique for interfacing bci system. **IOP Conference Series: Materials Science and Engineering**, IOP Publishing, v. 925, n. 1, p. 012062, sep 2020. Disponível em: <<https://dx.doi.org/10.1088/1757-899X/925/1/012062>>.

LELIS, Atahualpa Bastos; FILHO, Jarbas Cabral. Utilização de Ondas Cerebrais para Controle de Componentes Eletrônicos. **Semana Acadêmica, Fortaleza**, v. 1, n. 52, 2014.

LI, Jing *et al.* Using permutation entropy to measure the changes in EEG signals during absence seizures. **Entropy**, Multidisciplinary Digital Publishing Institute, v. 16, n. 6, p. 3049–3061, 2014.

LINJA, Joakim *et al.* Do Randomized Algorithms Improve the Efficiency of Minimal Learning Machine? **Machine Learning and Knowledge Extraction**, Multidisciplinary Digital Publishing Institute, v. 2, n. 4, p. 533–557, 2020.

LOTTE, F. *et al.* A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. **Journal of neural engineering**, IOP Publishing, v. 15, n. 3, p. 031005, 2018.

MARKOULIDAKIS, I. *et al.* Multiclass confusion matrix reduction method and its application on net promoter score classification problem. **Technologies**, Multidisciplinary Digital Publishing Institute, v. 9, n. 4, p. 81, 2021.

MESQUITA, Diego Parente Paiva; GOMES, Joao Paulo Pordeus; JUNIOR, Amauri H Souza. Ensemble of efficient minimal learning machines for classification and regression. **Neural Processing Letters**, Springer, v. 46, n. 3, p. 751–766, 2017. Disponível em: <<https://doi.org/10.1007/s11063-017-9587-5>>.

MESQUITA, Diego Parente Paiva; GOMES, Joao Paulo Pordeus; JUNIOR, Amauri Holanda Souza. Ensemble of minimal learning machines for pattern classification. In: **International Work-Conference on Artificial Neural Networks**. Palma de Mallorca (Spain): Springer, 2015. p. 142–152. **Anais**:<<https://link.springer.com/book/10.1007/978-3-319-19222-2>>. Disponível em: <[https://doi.org/10.1007/978-3-319-19222-2\\_12](https://doi.org/10.1007/978-3-319-19222-2_12)>.

OUYANG, Gaoxiang *et al.* Ordinal pattern based similarity analysis for EEG recordings. **Clinical Neurophysiology**, Elsevier, v. 121, n. 5, p. 694–703, 2010.

PRODANOV, C. C.; FREITAS, E. C. Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico–2. ed.–novo hamburgo: Feevale, 2013 qedu. **Censo Escolar EEB Mun. Prof. Nair Alves Bratti**, Novo Hamburgo, v. 15, 2018.

RASHID, Mamunur *et al.* Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review. **Frontiers in neurobotics**, Frontiers, v. 14, p. 25, 2020.

SOKOLOVA, Marina; LAPALME, Guy. A systematic analysis of performance measures for classification tasks. **Information Processing & Management**, Elsevier, v. 45, n. 4, p. 427–437, 2009.

VELASCO-ALVAREZ, Francisco *et al.* Audio-cued motor imagery-based brain–computer interface: Navigation through virtual and real environments. **Neurocomputing**, Elsevier, v. 121, p. 89–98, 2013.

WANG, Xiao-Wei; NIE, Dan; LU, Bao-Liang. Emotional state classification from EEG data using machine learning approach. **Neurocomputing**, Elsevier, v. 129, p. 94–106, 2014.

WEBB, Andrew. R. **Statistical pattern recognition**. San Francisco, USA: John Wiley & Sons, 2003.