

# MRMR Aplicado a Sinais de EEG para Classificação de Emoções

Leandro A. Serafim, Marcelo Grilo Jr., Caroline P. A. Moraes, Denis Fantinato, Rodrigo Ramos, Aline Neves

**Resumo**— A classificação de emoções através do uso de sinais de EEG é um assunto que tem atraído a atenção de pesquisadores de diversas áreas, inclusive relacionado ao desenvolvimento de interfaces cérebro-máquina. Neste trabalho, focamos na etapa de redução de dimensionalidade, necessária antes da etapa de classificação. Propomos uma alteração no método de Máxima Relevância Mínima Redundância (MRMR) de forma a trabalhar com o sinal de EEG sem a necessidade de uma etapa de discretização. Mostramos como o método proposto melhora significativamente o desempenho dos classificadores quando comparado ao clássico PCA ou mesmo ao uso de todas as informações disponíveis.

**Palavras-Chave**— Classificação de emoções, Redução de dimensionalidade, PCA, MRMR, sinais de EEG.

**Abstract**— Emotion classification through EEG signals has attracted the attention of researchers in several fields, including those studying brain-computer interfaces. This work focus on feature selection necessary to achieve a good classification performance. We propose a modification on the Maximum Relevance Minimum Redundance method, applying it to EEG signals without the need of a discretization step. We show how the proposed method significantly improves the classifiers performance when compared to the use of classical PCA.

**Keywords**— Emotion classification, feature selection, PCA, MRMR, EEG signals

## I. INTRODUÇÃO

O entendimento e o conhecimento das emoções é assunto de interesse de pesquisadores em diversas áreas, buscando entender desde as influências que as emoções geram no cotidiano e nas interações sociais do indivíduo, até como essas emoções são geradas do ponto de vista fisiológico. Estudos que propõem entender como as emoções são geradas e interpretadas no corpo do indivíduo são realizados levando em conta diferentes aspectos, tais como expressões faciais [1][2], atividade cardíaca [3][4] ou atividades cerebrais [5][6].

Um dos métodos que pode ser usado para estudos de emoções baseados em atividades cerebrais é o uso de sinais de Eletroencefalograma (EEG) [7], um método de monitoramento eletrofisiológico que é utilizado para registrar as atividades elétricas do cérebro. O uso dessa metodologia consiste em coletar os sinais elétricos das atividades do cérebro, com uso de eletrodos, e assim fazer o tratamento e a análise sobre esses dados.

Leandro Abidias da Silva Serafim, Marcelo Grilo, Caroline P. A. Moraes, Denis Fantinato, Aline Neves, Universidade Federal do ABC, Santo André-SP, {leandro.abidias, marcelo.grilo, caroline.moraes, denis.fantinato, aline.neves}@ufabc.edu.br.; Rodrigo Ramos, Universidade Federal do Vale do São Francisco-PE, rodrigo.ramos@univasf.edu.br. Este trabalho foi parcialmente financiado pela CAPES e pela FAPESP (Processo no 2018/17678-3)

Neste estudo, procuraremos elicitare e classificar dois tipos de emoção: alegria e nojo. Para a análise dos dados coletados, uma série de etapas são necessárias, incluindo filtragem, janelamento e remoção de artefatos. Os movimentos involuntários que geram interferência no sinal coletado, como piscadas e movimentos musculares, são chamados de artefatos. Adicionalmente, muitas vezes é necessário efetuar uma redução na dimensionalidade dos dados antes da etapa de classificação. O foco deste trabalho está justamente nesta etapa.

Um dos métodos mais utilizados na literatura para redução de dimensionalidade em estudos de sinais de EEG voltados para classificação de emoções é a Análise de Componentes Principais (PCA - *Principal Component Analysis*) [8]. Outro método existente na literatura é aquele de Mínima Redundância Máxima Relevância (MRMR) [9][10], que utiliza o cálculo da informação mútua para buscar componentes que tenham a maior relevância e a menor redundância, quando comparados a uma classe pré-definida. O MRMR já foi usado dentro dos contextos de classificação de doenças [11][9][10], classificação de emoção [12][13] e predição de alergia [14].

Nas aplicações citadas, o MRMR sempre foi aplicado num contexto no qual os sinais são primeiramente discretizados. No contexto deste trabalho, aplicaremos o método MRMR diretamente ao sinal de EEG coletado e pré-processado (após a remoção dos artefatos). Assim, não discretizaremos os dados antes da aplicação do algoritmo. O desempenho da técnica será comparado com o resultado obtido com o PCA.

Por fim, após a etapa de redução de dimensionalidade, o sinal será finalmente classificado. Os classificadores mais comuns, utilizados dentro do contexto de classificação de emoções, são o *Linear Discriminant Analysis* (LDA) [15] e o *Support Vector Machine* (SVM) [16]. Utilizaremos os dois no contexto deste trabalho para compararmos os resultados.

Este artigo está organizado da seguinte forma: a Seção II detalha os métodos para a redução de dimensionalidade. A Seção III especifica o processo de elicitare das emoções e a aquisição de sinais de EEG, incluindo o pré-processamento dos sinais coletados. A Seção IV apresenta as simulações e os resultados obtidos e na Seção V são apresentadas as conclusões.

## II. MÉTODOS PARA REDUÇÃO DE DIMENSIONALIDADE

Iniciaremos esta seção com uma breve descrição do clássico método de Análise por Componentes Principais (PCA). Na sequência, abordaremos o método MRMR que é o foco deste trabalho.

### A. Análise de Componentes Principais

A análise de componentes principais é um método bastante utilizado em redução de dimensionalidade. Ele consiste em reduzir a dimensionalidade de um conjunto de dados que contém um grande número de variáveis inter-relacionadas, buscando um conjunto menor que tenha a menor redundância possível e que ainda assim represente, da melhor forma, o conjunto inicial. Isso é obtido gerando um novo conjunto de variáveis, que são denominadas de componentes principais. O método acaba sendo bastante utilizado devido à sua simplicidade: somente estatísticas de segunda ordem são utilizadas, ou seja, todo o processo é baseado somente no uso de correlações entre os sinais.

As componentes extraídas pelo PCA são descorrelacionadas e ortogonais entre si, e são ordenadas de forma que a primeira componente principal extraída é a que representa a maior variação dentro do conjunto de dados [8].

### B. Mínima Redundância Máxima Relevância

O MRMR (Mínima Redundância Máxima Relevância) é um algoritmo que vem ganhando popularidade na redução de dimensionalidade. Foi proposto pela primeira vez em 2005 [9], demonstrando que, com base em um conjunto de dezenas de milhares de genes, é possível selecionar um subconjunto que apresenta uma melhor precisão para detecção de uma doença.

O MRMR seleciona um subconjunto de características (*features*) que apresenta um ganho de precisão, levando em consideração a importância das características e as dependências estatísticas entre elas. Mesmo que duas características sejam altamente relevantes, a seleção das duas pode adicionar mais complexidade ao modelo aumentando a possibilidade de *overfitting* no momento da classificação. Devido ao fato de serem altamente correlacionadas (alta redundância) utilizar as duas características, não acrescenta informações significativas ao modelo.

Neste artigo, medimos a relevância das características a partir da informação mútua com relação a uma classe, conforme proposto em [9][10]. O objetivo é que a característica escolhida tenha a maior relevância no que se refere a representação da classe em questão, maximizando a informação mútua entre as duas, mas ao mesmo tempo, a característica escolhida não pode ser redundante face a outras já escolhidas, minimizando assim a redundância.

A informação mútua entre duas variáveis aleatórias  $X$  e  $Y$  é definida com base em sua densidade de probabilidade conjunta  $p(x, y)$  e as respectivas densidades de probabilidade marginais  $p(x)$  e  $p(y)$ :

$$I(x, y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \quad (1)$$

Calcular a informação mútua de variáveis contínuas não é uma tarefa simples, devido à necessidade de se estimar as densidades de probabilidades (pdf - *probability density function*) envolvidas, e principalmente a densidade de probabilidade

conjunta. Na grande maioria dos trabalhos existentes na literatura, os dados são discretizados [9][10], [11]. Neste trabalho, no entanto, utilizaremos o método de janela de Parzen [17] para estimar as densidades de probabilidades necessárias a partir das amostras disponíveis dos sinais em questão. Tal método já vem sendo utilizado, com bom desempenho, na estimativa de informação mútua dentro de outros contextos [18][19][20].

Dadas  $N$  amostras de um processo estocástico  $x$ , podemos estimar sua densidade de probabilidade fazendo:

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^N \kappa_{\sigma} \left( \frac{x - x_i}{\sigma} \right), \quad (2)$$

onde  $\kappa_{\sigma}$  é uma função kernel,  $x_i$  é a  $i$ -ésima amostra de  $x$  e  $\sigma$  é a largura da janela. Parzen comprovou que, com o kernel e a largura escolhidos corretamente, é possível estimar  $\hat{p}(x)$  de forma que ela pode convergir para a densidade verdadeira  $p(x)$ , quando  $N$  tende ao infinito [21]. Na literatura, a função mais utilizada como kernel é a Gaussiana. Para o caso de uma única variável, como na estimativa de  $p(x)$ , podemos usar  $\kappa_{\sigma} = G_{\sigma}$ , definida como:

$$G_{\sigma}(x_i - x_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{|x_i - x_j|^2}{2\sigma^2} \right). \quad (3)$$

No entanto, para o caso da probabilidade conjunta  $p(x, y)$ , é necessário usarmos a estimação multivariada. Considerando um vetor  $\mathbf{z}_{ij} = [(x_i - x_j), (y_i - y_j)]$  com amostras de  $x$  e  $y$ , a estimativa será dada por

$$\hat{p}(\mathbf{z}) = \frac{1}{N} \sum_{i,j} K \left( \frac{\mathbf{z} - \mathbf{z}_{ij}}{\sigma} \right), \quad (4)$$

onde  $K(\cdot)$  é uma função kernel multivariada já que opera sobre vetores e não mais sobre escalares. Novamente, a função kernel mais utilizada na literatura é a função Gaussiana multivariada dada por:

$$K_{\Sigma}(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp \left( -\frac{\mathbf{z}^T \Sigma^{-1} \mathbf{z}}{2} \right), \quad (5)$$

onde  $d$  é a dimensão de  $\mathbf{z}$ ,  $\Sigma$  é a matriz de covariância de  $\mathbf{z}$  e  $|\Sigma|$  é o seu determinante. Neste caso,  $\Sigma$  será uma matriz de dimensões  $2 \times 2$ .

Tendo estimado as densidades de probabilidades necessárias, podemos voltar ao algoritmo MRMR propriamente dito. O primeiro passo é definir um conjunto de classes  $c = \{c_1, c_2, \dots, c_M\}$ , levando em conta a tarefa de classificação dos sinais em questão.

Depois de definido o conjunto  $c$ , é necessário determinar, dentre as características disponíveis (no caso deste artigo, os sinais dos eletrodos do EEG), quais são mais relevantes tendo em vista o conjunto  $c$ . Para isto, calculamos as informações mútuas entre os sinais disponíveis  $x_i \in \mathcal{C}$ , onde  $\mathcal{C}$  representa o conjunto com todos os sinais (*features*) disponíveis, e a classe  $c$ ,  $I(x_i, c)$ . Após este processo, passamos então para a fase que busca a minimização da redundância. Para isto, são calculadas as informações mútuas entre o sinal  $x_i$  e todos os outros sinais que já fazem parte do subconjunto  $\mathcal{S}$  que estamos procurando,

$I(x_i, x_j)$  onde  $x_j \in \mathcal{S}$  e  $\mathcal{S}$  é o subconjunto das *features* já escolhidas.

Como precisamos otimizar as medidas de informação mútua diferentes, e uma precisa ser maximizada enquanto que a outra é minimizada, existem duas possibilidades para compormos estes dois critérios: considerar a diferença entre as duas informações mútuas (DIM):

$$\max_{x_i \in \mathcal{C}} \left[ I(x_i, c) - \frac{1}{|\mathcal{S}|} \sum_{x_j \in \mathcal{S}} I(x_i, x_j) \right], \quad (6)$$

onde  $|\mathcal{S}|$  é a cardinalidade de  $\mathcal{S}$ , ou então considerar o quociente entre as duas (QIM):

$$\max_{x_i \in \mathcal{C}} \left[ \frac{I(x_i, c)}{\frac{1}{|\mathcal{S}|} \sum_{x_j \in \mathcal{S}} I(x_i, x_j)} \right]. \quad (7)$$

Neste trabalho, escolhemos utilizar o critério QIM dado por (7), visto que o mesmo obteve o melhor desempenho quando testado em [10], ainda que em outro contexto. A cada iteração, a nova característica escolhida é inserida em  $\mathcal{S}$ , até obtermos o número de características (*features*) desejado. Em Algoritmo 1 é exibido um pseudocódigo do mRMR, no qual, para facilitar a notação,  $|\cdot|$  é a cardinalidade do conjunto.

---

#### Algoritmo 1 Mínima Redundância Máxima Relevância

---

**Entradas:** Dataset  $x$ , classe  $c$ , números de features a serem selecionadas  $n$

**Saída:** Conjunto de features selecionadas  $\mathcal{S}$

- 1: **para**  $i \leftarrow 1$  **até**  $|x|$  **faça**
  - 2:      $relevancia(i) = I(x_i, c)$
  - 3: **fim para**
  - 4:  $\mathcal{S}[1] = \max(relevancia)$
  - 5: exclui-se  $\mathcal{S}[1]$  de  $x$  e também de  $relevancia$  já que esta já é a primeira feature selecionada.
  - 6: **para**  $i \leftarrow 2$  **até**  $n$  **faça**
  - 7:     **para**  $j \leftarrow 1$  **até**  $|x|$  **faça**
  - 8:          $s = \sum_{k=1}^{|\mathcal{S}|} I(x_j, \mathcal{S}[k])$
  - 9:          $redundancia(j) = s / |\mathcal{S}|$
  - 10:     **fim para**
  - 11:      $\mathcal{S}[i] = \max(relevancia(redundancia+0,0001))$
  - 12:     exclui-se  $\mathcal{S}[i]$  de  $x$  e de  $relevancia$
  - 13: **fim para**
- 

### III. AQUISIÇÃO DOS DADOS E PRÉ-PROCESSAMENTO

Os sinais de eletroencefalograma (EEG) foram coletados na Universidade Federal do Vale de São Francisco (UNIVASF), utilizando o Eletroencefalógrafo Brain Wave II - Neurovirtual, com a configuração de 19 eletrodos, posicionados conforme o sistema internacional 10-20. Os registros foram gravados continuamente a uma taxa de amostragem de 200Hz.

Foram selecionados 23 voluntários, dentre eles 9 do sexo masculino e 14 do sexo feminino, com uma média de idade de  $24,43 \pm 4,38$  anos.

Durante a coleta dos sinais de EEG, os participantes foram submetidos a estímulos audiovisuais, assistindo uma sequência de 6 vídeos, conforme exibido na Figura 1, cada vídeo com

duração de 20 segundos. Esse procedimento assemelha-se ao que foi feito em [6]. Os vídeos 1, 3 e 6 se referem a emoções positivas e os vídeos 2, 4 e 5 são relacionados a emoções negativas. Houve um intervalo de 5 segundos entre a exibição de vídeos sucessivos no qual os participantes não foram submetidos a um estímulo audiovisual, e nesses períodos foram extraídos os sinais considerados como linha base.

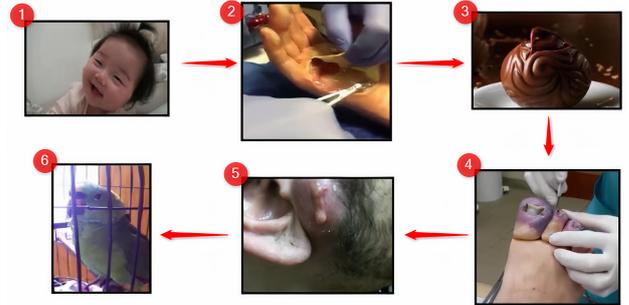


Fig. 1. Imagens que representam os vídeos exibidos para despertar as emoções, na ordem de apresentação.

Os vídeos foram utilizados de forma a induzir emoções nos participantes, considerando três parâmetros: classificação da expressão emocional sentida, compatibilidade e intensidade. Após a coleta dos sinais, cada participante respondeu um questionário on-line (utilizando Google Forms), de acordo com os procedimentos de *Self-Assessment Manikin* (SAM) [22], o qual avalia três aspectos associados às respostas afetivas: prazer (compatibilidade), excitação (intensidade) e dominância (controle). Esse procedimento foi o mesmo adotado por [23].

Os sinais foram filtrados na faixa de frequência entre 1 - 45 Hz, e pré-processados fazendo-se a remoção de artefatos através do algoritmo SOBI (*Second Order Blind Identification*) [24], algoritmo muito usado para este fim. A análise tempo-frequência foi implementada utilizando a função `newtimef` do EEGLAB, Matlab [25]. O sinal foi dividido em 3 bandas de frequência, Alfa (8 a 13 Hz), Beta (14 a 30 Hz), e Gama (31 a 45 Hz), com uma variação de 1 Hz entre cada faixa. Logo após foram extraídas as potências médias dos sinais em cada faixa de frequência, assim como para o sinal incluindo todas as faixas. Em [26] temos mais detalhes referente ao pré-processamento deste dataset.

### IV. RESULTADOS DE SIMULAÇÃO

Estudos anteriores [26][27][28][29][30] mostram que, abordando o problema de classificação de emoções através de sinais de EEG, a banda Alfa é a que apresenta maior atividade. Adicionalmente, tomando por base os dados coletados (descritos na Seção III), análises prévias realizadas em [26] mostraram que os sinais coletados durante a exibição dos vídeos 1 e 4 possuem uma qualidade melhor, e com menos artefatos. Adicionalmente, as respostas do questionário SAM foram mais representativas nestes vídeos, indicando que houve realmente a elicitación das emoções pretendidas. Com base nessas informações, os sinais coletados durante a exibição

destes dois vídeos foram utilizados como base para este estudo.

Para a análise e implementação dos algoritmos, utilizamos a plataforma Google Colaboratory, um ambiente na nuvem de notebooks Jupyter, voltado à criação e execução de códigos em Python. Os modelos de classificadores LDA e SVM foram baseados na biblioteca Sklearn (Version 0.22.2) [31], e o modelo PCA foi baseado na biblioteca Statsmodels (Version 0.10.2) [32]. O MRMR foi implementado conforme discutido na Seção II.

Para fins de comparação, três abordagens foram utilizadas. Primeiramente, realizamos a classificação utilizando as potências médias obtidas após o pré-processamento descrito na Seção III, tanto para os sinais coletados durante a exibição do vídeo 1 (emoção positiva) quanto para o vídeo 4 (emoção negativa), especificamente considerando a banda Alfa do sinal. Neste primeiro momento, não houve aplicação de nenhum método para redução de dimensionalidade. Como resultado, obteve-se uma acurácia média de 66,5% aplicando-se o LDA e 87,6% aplicando-se o SVM. Para este processo de classificação, os dados das duas classes (emoção positiva e negativa) foram embaralhados e separados de forma randômica, 2/3 dos dados foram usados para treinamento e 1/3 foi usado para teste. A acurácia média foi obtida a partir de 50 iterações (*bootstrapping validation*).

Na sequência, aplicou-se o PCA para redução de dimensionalidade. Dos sinais obtidos a partir dos 19 eletrodos utilizados para coleta, após o pré-processamento descrito na Seção III, foram extraídos até 14 sinais (*features*) mais representativos para cada sujeito. Aplicando-se novamente a etapa de classificação e variando a quantidade de *features* escolhida, chegou-se ao gráfico mostrado na Figura 2. Podemos ver que o SVM fornece uma acurácia média melhor na grande maioria dos casos. O melhor desempenho foi obtido usando-se 8 *features* e o classificador SVM, atingindo uma acurácia média de 85,2%. Por outro lado, o melhor desempenho obtido com o LDA foi utilizando 6 *features*, alcançando uma acurácia média de 82,9%, na Tabela I podemos ver a matriz de confusão dos melhores desempenhos.

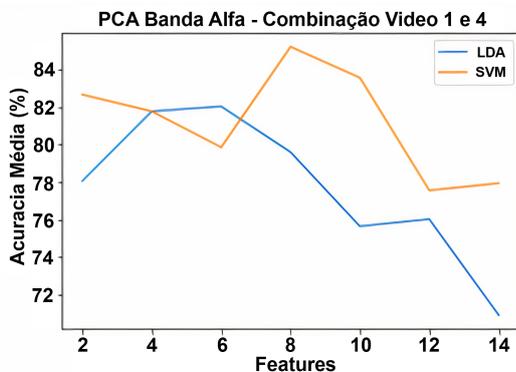


Fig. 2. Acurácia média utilizando modelo PCA.

Por fim, aplicamos o método que foi desenvolvido neste estudo: o MRMR com base no uso do sinal contínuo. O primeiro passo para podermos aplicar este método é definir

TABELA I  
MATRIZ DE CONFUSÃO DO PCA

		LDA - 6 Features		SVM - 8 Features	
		Positivo	Negativo	Positivo	Negativo
Positivo	Positivo	0,4094	0,1007	0,4183	0,0714
	Negativo	0,0841	0,4056	0,0752	0,4349

o valor de  $\sigma$  a ser utilizado em eq. (2) e o valor de  $\Sigma$  em eq. (5). Através de simulações, escolheu-se o valor de  $\sigma = 0,5$  e  $\Sigma$  é uma estimativa da matriz de correlação entre os sinais utilizados, obtida por meio de uma média temporal de todas as amostras disponíveis dos sinais.

Além disso, para calcular (7), é necessária uma classe (sinal de referência) tanto para a emoção positiva como para a emoção negativa. Para isto, usamos novamente análises prévias da base de dados em questão, realizadas em [26], e assim foi escolhido o sinal do eletrodo F4, do sujeito H2, vídeo 4 como classe para a emoção negativa e o sinal do eletrodo F3, mesmo sujeito, vídeo 1 como classe para a emoção positiva [33][34]. Adicionalmente, como a estimação das densidades de probabilidade necessárias através do método da janela de Parzen possuem um custo computacional relativamente alto, os sinais foram previamente decimados para diminuir pela metade o número de amostras disponíveis.

Assim como na simulação com o PCA, aplicamos o MRMR a fim de escolher até 14 *features* mais relevantes para cada sujeito. Utilizando os dados obtidos para classificação, mantendo a proporção de 2/3 para treinamento e 1/3 para testes, obteve-se o resultado mostrado na Figura 3. Claramente, neste caso, temos um melhor desempenho do classificador SVM. Comparando com os resultados obtidos pelo PCA mostrados na Figura 2, verificamos que neste caso o melhor resultado também foi obtido com o uso de 8 *features*, mas agora a acurácia média, com o SVM, ficou em 89,6%, superior ao melhor resultado obtido com o PCA (acurácia de 85,2%). Podemos também observar que, utilizando-se até 4 *features*, o PCA apresenta um desempenho superior. A partir daí, o uso do MRMR apresenta uma melhora significativa no desempenho, apresentando uma acurácia média superior. Adicionalmente, é interessante verificar que o LDA não apresentou um bom desempenho quando aplicado em conjunto com o MRMR. Na Tabela II podemos ver as matrizes de confusão obtidas para os dois classificadores com os parâmetros que resultaram na melhor acurácia.

TABELA II  
MATRIZ DE CONFUSÃO DO MRMR

		LDA - 8 Features		SVM - 8 Features	
		Positivo	Negativo	Positivo	Negativo
Positivo	Positivo	0,3954	0,1147	0,4719	0,0573
	Negativo	0,1058	0,3839	0,0408	0,4298

## V. CONCLUSÃO

Neste artigo, abordamos o uso do método de redução de dimensionalidade MRMR, utilizando diretamente os sinais contínuos, algo que ainda não havia sido abordado na literatura. Como contexto de aplicação, escolhemos o problema

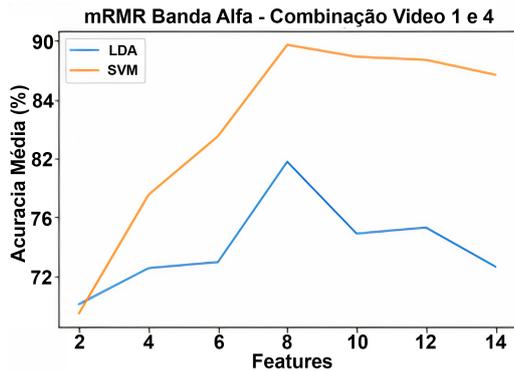


Fig. 3. Acurácia média utilizando modelo mRMR.

de classificação de emoções (positiva/neutra e negativa/nojo) através do uso de sinais de EEG.

Utilizando uma base de dados com 23 sujeitos e o sinal de EEG coletado através de 19 eletrodos, mostramos como a redução de dimensionalidade é uma etapa importante do pré-processamento, visto que podemos chegar a um ganho de desempenho de 13% em termos de acurácia, quando comparado com o caso em que esta etapa não é considerada. Comparando o PCA com o método MRMR proposto, vimos como este último apresenta um ganho de desempenho importante, podendo melhorar a acurácia média final da classificação em 4%.

#### REFERÊNCIAS

- [1] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *International Journal of Computer Vision*, vol. 25, pp. 23–48, 1997.
- [2] K. Anderson and P. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 1, pp. 96–105, 2006.
- [3] J. Wagner, J. Kim, and E. Andre, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," in *2005 IEEE International Conference on Multimedia and Expo*, pp. 940–943, 2005.
- [4] K. H. Kim, S. W. Bang, and S. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and Biological Engineering and Computing*, vol. 42, pp. 419–427, 2006.
- [5] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from eeg data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.
- [6] R. J. Davidson, P. Ekman, C. Saron, J. Senulis, and W. V. Friesen, "Approach/withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I.," *Journal of Personality and Social Psychology*, vol. 58, pp. 330–341, 1990.
- [7] L. Haas, "Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography," *Journal of neurology, neurosurgery, and psychiatry*, vol. 74, p. 9, 02 2003.
- [8] I. Jolliffe, "Principal component analysis: A beginner's guide - i. introduction and application," *Weather*, vol. 45, pp. 375–382, 10 1990.
- [9] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, pp. 523–528, 2003.
- [10] M. Mandal and A. Mukhopadhyay, "An improved minimum redundancy maximum relevance approach for feature selection in gene expression data," *Procedia Technology*, vol. 10, pp. 20–27, 2013. First International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013.
- [11] M. Radovic, M. Ghalwash, N. Filipovic, and Z. Obradovic, "Minimum redundancy maximum relevance feature selection approach for temporal gene expression data," *BMC Bioinformatics*, vol. 18, 01 2017.
- [12] J. Atkinson and D. Campos, "Improving bci-based emotion recognition by combining eeg feature selection and kernel classifiers," *Expert Systems with Applications*, vol. 47, pp. 35–41, 2016.
- [13] P. Ackermann, C. Kohlschein, J. A. Bitsch, K. Wehrle, and S. Jeschke, "Eeg-based automatic emotion recognition: Feature extraction, selection and classification methods," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–6, 2016.
- [14] J. Wang, D. Zhang, and J. Li, "Preal: Prediction of allergenic protein by maximum relevance minimum redundancy (mrmr) feature selection," *BMC systems biology*, vol. 7 Suppl 5, p. S9, 12 2013.
- [15] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [16] A. Ben-Hur, D. Horn, H. Siegelmann, and V. Vapnik, "Support vector clustering," *Journal of Machine Learning Research*, vol. 2, pp. 125–137, 11 2001.
- [17] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.
- [18] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2807–2820, 1999.
- [19] C. Moraes, D. Fantinato, and A. Neves, "An epanechnikov kernel based method for source separation in post-nonlinear mixtures," *XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 01 2019.
- [20] J. Saldanha, C. Moraes, A. Neves, and D. Fantinato, "Algoritmo baseado em estatísticas de segunda ordem para separação cega de fontes com misturas não lineares," *2019 Brazilian Technology Symposium*, 10 2019.
- [21] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [22] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [23] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Technical manual and affective ratings," *NIMH Center for the Study of Emotion and Attention*, p. 39–58, 1997.
- [24] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.
- [25] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics," *Journal of Neuroscience Methods*, vol. 134, pp. 9–12, 01 2004.
- [26] M. Grilo, L. Ribeiro, C. Moraes, C. Melo, D. Fantinato, L. Sampaio, A. Neves, and R. Ramos, "Artifact removal in eeg based emotional signals through linear and nonlinear methods," in *2019 E-Health and Bioengineering Conference (EHB)*, pp. 1–4, 2019.
- [27] M. Grilo, "Análise de padrões em eletroencefalogramas para diferentes situações mentais," Master's thesis, Universidade Federal do Vale do São Francisco, 2018.
- [28] Y. Y. Lee and S. Hsieh, "Classifying different emotional states by means of eeg-based functional connectivity patterns," *PloS one*, vol. 9, p. e95415, 04 2014.
- [29] M. Mikhail, K. El-Ayat, J. Coan, and J. Allen, "Using minimal number of electrodes for emotion detection using brain signals produced from a new elicitation technique," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 6, pp. 80–97, 12 2013.
- [30] M. Kostyunina and M. Kulikov, "Frequency characteristics of eeg spectra in the emotions," *Neuroscience and Behavioral Physiology*, vol. 26, no. 4, pp. 340–343, 1996.
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.
- [32] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," *Proceedings of the 9th Python in Science Conference*, vol. 2010, 01 2010.
- [33] W. Wittling and R. Roschmann, "Emotion-related hemisphere asymmetry: Subjective emotional responses to laterally presented films," *Cortex*, vol. 29, no. 3, pp. 431–448, 1993.
- [34] J. B. Hellige, *Hemispheric asymmetry: What's right and what's left?* Harvard University Press, 1993.