

Configuração Automática de Hiperparâmetros em Sistemas MI-BCI de Sub-bandas

Vitor Mendes Vilas-Boas e Cleison Daniel Silva

Resumo— Este trabalho apresenta uma arquitetura baseada em otimização bayesiana para ajuste automático de hiperparâmetros na classificação de tarefas motoras em interfaces cérebro-máquina. A abordagem agrega múltiplas técnicas de processamento para compor uma estrutura capaz de modelar a acurácia do classificador em função dos hiperparâmetros e encontrar uma configuração ótima personalizada. Simulações a partir de sinais públicos de eletroencefalografia (EEG) resultaram em ganho de desempenho médio de até 7.9% em relação à modelos de configuração manual, indicando que a auto configuração pode minimizar o impacto da variabilidade do sinal na performance e contribuir para um sistema mais generalista.

Palavras-Chave— Interface Cérebro-Máquina, Imagética Motora, Configuração Automática, Otimização Bayesiana.

Abstract— This work presents a Bayesian optimization-based architecture for automatic adjustment of hyperparameters in the classification of motor tasks in brain-machine interfaces. The strategy aggregates multiple processing techniques to compose a structure capable of modeling the accuracy of the classifier according to the hyperparameters and finding an optimal personalized configuration. Simulations from electroencephalography (EEG) public signals resulted in an average performance gain of up to 7.9% compared to manual configuration models, indicating that auto configuration can minimize the impact of signal variability on performance and contribute to a more generalist system.

Keywords— Brain Computer-Interface, Motor Imagery, Auto setup, Bayesian Optimization.

I. INTRODUÇÃO

Interfaces Cérebro-Máquina baseadas em Imagética Motora (Motor Imagery-based Brain-Computer Interfaces, MI-BCI) propõem-se a traduzir a (De)Sincronização da atividade cortical associada à simulações motoras, comumente registrada por EEG, em sinais de comunicação e controle [2]. Ampliar a autonomia de pessoas com comprometimento motor é um dos focos de estudo [1].

No entanto, decodificar adequadamente a intenção motora no EEG para traduzi-la em comandos confiáveis não é uma tarefa trivial. A alta variabilidade da atividade cerebral constitui padrões de sinais particulares a cada sujeito e exige o ajuste individual dos múltiplos hiperparâmetros em um MI-BCI [3].

A sensibilidade desses sistemas à natureza ruidosa e não-estacionária do EEG sustenta uma forte correlação entre a calibração dos algoritmos e o desempenho de classificação, sendo improvável que uma configuração independente do sujeito possa ser generalizável em nível satisfatório [5].

Ainda assim, a configuração única e manual é comum, especialmente em sistemas com arquiteturas complexas e

espaços de parâmetros de alta dimensionalidade, em que o número elevado de combinações para se obter uma solução ótima conjunta estimula a parametrização fixa [6].

Dada a eficiência dos modelos bayesianos na otimização de hiperparâmetros [6], este trabalho apresenta uma estratégia estruturada a partir desses modelos para ajuste automático de hiperparâmetros em sistemas MI-BCI com cadeias de processamento de sub-bandas.

A abordagem, denominada *auto setup*, modela a performance de generalização do classificador em função dos hiperparâmetros correspondentes a fim de encontrar uma configuração ótima personalizada que maximize a acurácia de classificação a partir de padrões particulares dos indivíduos.

Além de incluir múltiplas técnicas de classificação para formar um amplo espaço de busca hiperparamétrica, a abordagem diferencia-se de outros estudos ao empregar a decomposição em frequência via Transformada Rápida de Fourier (Fast Fourier Transform, FFT) a fim de tornar o pré-processamento dos sinais e a busca bayesiana mais eficientes.

Os detalhes são discutidos a seguir, a começar pela seção II que descreve a cadeia de processamento adotada. Em III as especificidades da abordagem *auto setup* são definidas, enquanto que na seção IV, discute-se os experimentos realizados e seus resultados. Por fim, a seção V fornece algumas conclusões.

Neste artigo, letras maiúsculas e minúsculas em negrito indicam matrizes e vetores, respectivamente, ao passo que letras em itálico denotam escalares. O símbolo \mathbb{R} corresponde ao conjunto dos números reais e \mathbb{C} ao conjunto dos complexos.

II. PROCESSAMENTO DE EEG EM SISTEMAS MI-BCI

Os Padrões Espaciais Comuns de Sub-bandas (Sub-Band Common Spatial Patterns, SBCSP) destacam-se como uma das abordagens de decodificação e classificação de sinais de EEG mais eficazes [4]. Diferentemente de arquiteturas tradicionais, caracterizadas pela filtragem do sinal em banda de frequência única, o SBCSP utiliza um banco de filtros para decompor os sinais em múltiplas sub-bandas.

Para cada sub-banda, os métodos Padrões Espaciais Comuns (Common Spatial Patterns, CSP) e Análise de Discriminantes Lineares (Linear Discriminant Analysis, LDA) são utilizados em série. O resultado é interpretado como pontuações de sub-banda e então são agrupados para finalmente serem classificados conforme a intenção motora [4].

A Figura 1 descreve a cadeia de processamento SBCSP cujos hiperparâmetros deseja-se ajustar automaticamente neste trabalho. Os valores candidatos para os hiperparâmetros em cada etapa definem o espaço de busca do *auto setup* proposto.

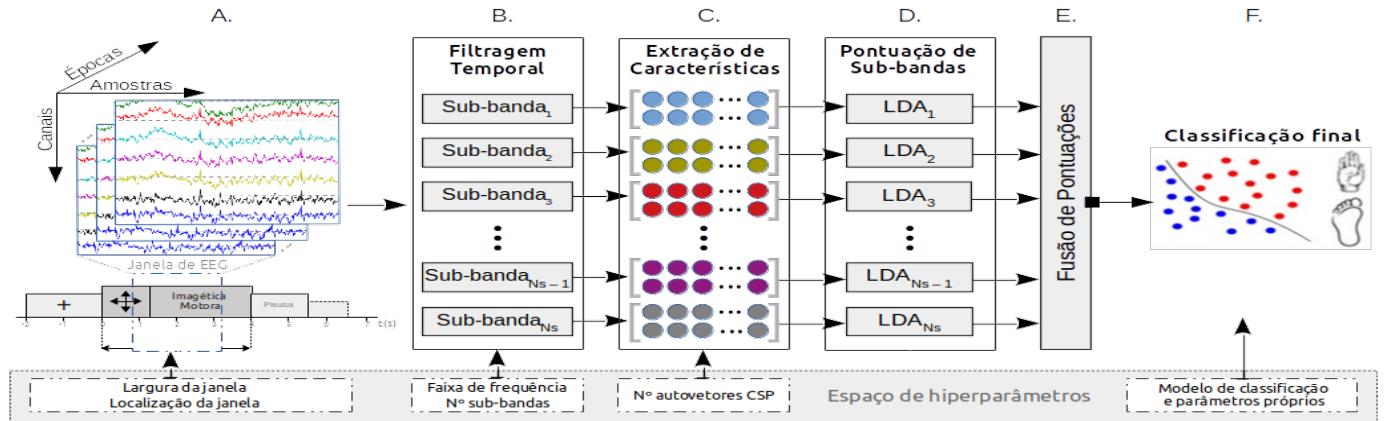


Fig. 1. Estrutura da abordagem SBCSP elaborada com base em [4] e [7]

A. Coleta e Segmentação dos Sinais de EEG

Em sistemas MI-BCI típicos, um protocolo baseado em dicas visuais orienta os indivíduos na modulação da atividade cerebral [8] enquanto os sinais de EEG são amostrados a uma dada frequência F_s .

Múltiplos segmentos de sinais de EEG associados à ocorrência de MI são obtidos a partir de N iterações do protocolo de coleta de sinais. Uma janela temporal, denominada época, é extraída de cada segmento no intervalo entre o momento da instrução (A_{min}) e o fim do tempo estipulado para MI (A_{max}).

Cada época, portanto, contempla as amostras no intervalo $\{A_-, A_+\}$, cujos extremos são determinados pelos hiperparâmetros α_i e α_f , respectivamente, sendo $A_- = F_s(A_{min} + \alpha_i)$ e $A_+ = F_s(A_{min} + \alpha_f)$. Juntos, os valores atribuídos a ambos os hiperparâmetros definem a localização e a largura de cada época em relação a A_{min} , respectivamente, e devem obedecer à relação $A_{max} \geq \alpha_f \geq \alpha_i + 1$.

A matriz $\mathbf{Z} \in \mathbb{R}^{P \times Q}$, cuja dimensão refere-se aos P sinais e suas Q amostras no intervalo mencionado, é utilizada para definir uma época na forma matricial, a qual é associada à tarefa de MI equivalente por um rótulo t . Assim, sejam Ω_1 e Ω_2 duas classes de tarefas motoras distintas, define-se $t = 1$ para épocas de classe Ω_1 e $t = 2$ para classe Ω_2 .

As épocas contribuem para que o processamento ocorra em duas fases sequenciais [1]. Primeiro, os algoritmos são treinados com base em um sub-conjunto com N_T épocas, denotado por \mathcal{E}_T , e seus respectivos rótulos \mathbf{t}_T . Na fase posterior, a validação do sistema devidamente calibrado ocorre a partir das N_V épocas sobressalentes, \mathcal{E}_V , e rótulos \mathbf{t}_V . Nota-se que $\mathcal{E}_T \cap \mathcal{E}_V = \emptyset$ e $N = N_T + N_V$.

B. Filtragem Temporal

Atenuar ruídos e delimitar o espectro de frequência nos sinais são tarefas importantes atribuídas à filtragem. Para tanto, o algoritmo FFT foi adotado neste trabalho com o objetivo de minimizar a elevação do custo de processamento inerente ao uso de sub-bandas em sistemas MI-BCI.

Ao calcular a FFT dos sinais em dada época \mathbf{Z} , obtém-se a informação espectral denotada pela matriz $\hat{\mathbf{X}} \in \mathbb{C}^{P \times Q}$ cujos elementos são definidos por: $\hat{x}_{pk} = \sum_{q=0}^{Q-1} z_{pq} e^{-j \frac{2\pi}{Q} kq}$, $\forall k =$

$0, 1, \dots, Q-1$. Em que z_{pq} representa a q -ésima amostra do p -ésimo sinal em \mathbf{Z} e os termos nos k -ésimos coeficientes complexos correspondem à decomposição na base de Fourier.

Seja \mathcal{F} um sub-conjunto com M bins de frequências discretas no intervalo $\Delta_f \triangleq \{f_l, f_u\}$, uma matriz de dimensão reduzida $\mathbf{X}^{P \times M} \in \mathbb{C}$ pode ser obtida extraído de cada linha de $\hat{\mathbf{X}}$, somente os M coeficientes \hat{x}_{pk} correspondentes aos bins contidos em \mathcal{F} . Define-se para tal, $M = \frac{(f_u - f_l)}{\delta_f}$, em que f_l e f_u representam, respectivamente, as frequências de corte inferior e superior no intervalo Δ_f e $\delta_f \triangleq \frac{F_s}{Q}$ é o espaçamento entre bins de frequências consecutivas.

Assim, dado um conjunto de N_s sub-bandas com sobreposição $\frac{S}{2}$ entre si (i.e. 50%), para ajusta-las aos M bins contidos em \mathcal{F} , a quantidade N_s deve satisfazer à relação $M = S(N_s + 1)$, de forma a obter o número de bins de frequências consecutivas em cada sub-banda definido por S .

C. Extração de Características

O CSP é um algoritmo popular em sistemas MI-BCI para extração de características espaciais relacionadas à atividade cortical [11]. O método busca filtros espaciais que maximizam a variância de uma classe e minimizam a de outra classe de forma a ressaltar a variação de energia que descreve a MI [9].

Seja N_κ épocas do conjunto \mathcal{E}_T associadas à classe Ω_κ , a matriz de correlação média na l -ésima sub-banda para Ω_κ é dada por $\mathbf{C}_{l\kappa} = \frac{1}{N_\kappa} \sum_{n=1}^{N_\kappa} \mathbf{X}_{l\kappa}^{(n)} \mathbf{X}_{l\kappa}^{(n)T}$, em que o sobrescrito T denota transposição matricial. Com isso, um filtro espacial $\mathbf{w}_l \in \mathbb{R}^{P \times 1}$, que maximiza a separabilidade entre as classes Ω_1 e Ω_2 na l -ésima sub-banda, pode ser obtido extremado:

$$\arg_{\mathbf{w}_l} \max \frac{\mathbf{w}_l^T \mathbf{C}_{l1} \mathbf{w}_l}{\mathbf{w}_l^T (\mathbf{C}_{l1} + \mathbf{C}_{l2}) \mathbf{w}_l}. \quad (1)$$

Uma possível solução para Eq. (1), é obtida ao resolver:

$$\mathbf{C}_{l1} \mathbf{w}_l = \lambda (\mathbf{C}_{l1} + \mathbf{C}_{l2}) \mathbf{w}_l, \quad (2)$$

em que λ e \mathbf{w}_l correspondem ao autovalor e autovetor generalizado, respectivamente. A Eq. (2) gera P filtros $\mathbf{w}_l^{(j)}$, para $j = 1, 2, \dots, P$, cada qual, ao ser aplicado na l -ésima matriz de coeficientes, produz o sinal espacialmente filtrado

$$\mathbf{y}_l^{(j)} = \mathbf{w}_l^{(j)T} \mathbf{X}_l. \quad (3)$$

O logaritmo da variância de $\mathbf{y}_\ell^{(j)}$ tem interpretação de energia e representa uma característica de sub-bandas, conforme:

$$\rho_\ell = \log \left(\mathbf{y}_\ell^{(j)} \mathbf{y}_\ell^{(j)\top} \right) = \log \left(\mathbf{w}_\ell^{(j)\top} \mathbf{X}_\ell \mathbf{X}_\ell^\top \mathbf{w}_\ell^{(j)} \right). \quad (4)$$

Logo, ao considerar todos os filtros espaciais, um vetor de características $\boldsymbol{\rho}_\ell \in \mathbb{R}^{P \times 1}$ pode ser definido por

$$\boldsymbol{\rho}_\ell = \log \left(\text{diag} \left(\mathbf{W}_\ell^\top \mathbf{X}_\ell \mathbf{X}_\ell^\top \mathbf{W}_\ell \right) \right), \quad (5)$$

em que $\text{diag}(\cdot)$ extrai os elementos da diagonal principal do argumento e \mathbf{W}_ℓ é a matriz de autovetores para a ℓ -ésima sub-banda cujas colunas são os filtros gerados na Eq. (2).

Uma vez que os filtros espaciais correspondentes aos maiores e menores autovalores são os que efetivamente maximizam, de modo recíproco, a variância de uma classe em detrimento de outra, enquanto que autovalores medianos tendem a contribuir menos para a separação, é comum a seleção de $R < P$ filtros para compor um vetor de características de dimensão reduzida, sendo R necessariamente natural, positivo e par [11].

D. Pontuação de Sub-bandas

Após definidos, os vetores de características $\boldsymbol{\rho}$ relativos à cada sub-banda são pontuados individualmente utilizando o LDA [9]. O método busca uma projeção ótima $\boldsymbol{\varpi}^*$ para cada vetor de entrada de forma a maximizar a separabilidade inter-classes e minimizar a dispersão intra-classe, conforme:

$$\arg_{\boldsymbol{\varpi}_\ell} \max \frac{\boldsymbol{\varpi}_\ell^\top \mathbf{S}_\beta^\ell \boldsymbol{\varpi}_\ell}{\boldsymbol{\varpi}_\ell^\top \mathbf{S}_\omega^\ell \boldsymbol{\varpi}_\ell}, \quad (6)$$

onde \mathbf{S}_β^ℓ e \mathbf{S}_ω^ℓ são as matrizes de espalhamento inter-classes e intra-classe, respectivamente, para a ℓ -ésima sub-banda. A solução ótima para Eq. (6) é dada por

$$\boldsymbol{\varpi}_\ell^* = (\mathbf{S}_\omega^\ell)^{-1} (\boldsymbol{\mu}_1^\ell - \boldsymbol{\mu}_2^\ell), \quad (7)$$

em que $\boldsymbol{\mu}_1^\ell$ e $\boldsymbol{\mu}_2^\ell$ contém as médias das classes Ω_1 e Ω_2 ao considerar os ℓ -ésimos vetores de características obtidos em Eq. (5) para cada época no conjunto \mathcal{E}_T .

A pontuação LDA para a ℓ -ésima sub-banda é dada por:

$$d_\ell = \boldsymbol{\varpi}_\ell^{\top} \boldsymbol{\rho}_\ell. \quad (8)$$

E. Fusão de Pontuações

Um meta-classificador bayesiano é usado para agrupar as pontuações LDA, como propõe [4]. Para tal, assume-se que as N_κ pontuações em cada sub-banda estão normalmente distribuídas, i. e.

$$\mathcal{L}(\Omega_\kappa | d_\ell) = \frac{1}{\sigma_\kappa^\ell \sqrt{2\pi}} \exp \left(-\frac{(d_\ell - \mu_\kappa^\ell)^2}{2(\sigma_\kappa^\ell)^2} \right), \quad (9)$$

com $\mu_\kappa^\ell = \frac{1}{N_\kappa} \sum_{n=1}^{N_\kappa} d_\ell^{(n)}$ e $(\sigma_\kappa^\ell)^2 = \frac{1}{N_\kappa} \sum_{n=1}^{N_\kappa} (\mu_\kappa^\ell - d_\ell^{(n)})^2$ representando a média e a variância das pontuações LDA atribuídas às N_κ épocas da classe Ω_κ , respectivamente.

Pode-se avaliar a densidade de probabilidade das distribuições em função da razão de verossimilhança entre as classes e obter uma meta-pontuação bayesiana,

$$v_\ell = \log \left(\frac{\mathcal{L}(\Omega_1 | d_\ell)}{\mathcal{L}(\Omega_2 | d_\ell)} \right), \quad (10)$$

e as meta-pontuações nas N_s sub-bandas podem ser agregadas por um vetor de características $\mathbf{v} = [v_1, v_2, \dots, v_{N_s}]^\top$.

F. Classificação Final

Conforme descrito em [4], como forma de compensar os erros produzidos pelos meta-classificadores individuais, um classificador final Φ é utilizado. Seu treinamento é realizado com base nos vetores \mathbf{v} obtidos a partir das épocas no conjunto \mathcal{E}_T , juntamente com seus respectivos rótulos \mathbf{t}_T .

Assume-se a forma genérica $\Phi = f(x)$ uma vez que o ajuste do método de classificação dar-se-á por *auto setup* e portanto não é definido previamente. Sua representação pode corresponder a um entre quatro modelos possíveis: LDA, Logistic Regression (LR), Support Vector Machines (SVM) e K -nearest neighbors (KNN) [1], [9].

Independentemente de sua forma, a saída s produzida por Φ para um dado \mathbf{v} , corresponde à classificação final do sinal quanto à tarefa de MI. De forma que, ao submeter todas as épocas do conjunto \mathcal{E}_V ao sistema devidamente treinado, obtém-se um vetor \mathbf{s} cujos elementos correspondem às N_V classificações $s^{(i)}$, $\forall i = 1, 2, \dots, N_V$, sendo $s^{(i)} = 1$ ou $s^{(i)} = 2$ para indicar as classes Ω_1 e Ω_2 , respectivamente.

III. AUTO SETUP DE SISTEMAS MI-BCI

Neste trabalho, propõe-se a otimização bayesiana (bayesian optimization, BO) como estratégia para automatizar e personalizar o ajuste do sistema apresentado na seção II. Por incorporar conhecimento prévio ao otimizar funções com expressões indefinidas e computacionalmente caras, a BO mostra-se eficiente para estimativa de hiperparâmetros [6].

A proposta é modelar a acurácia de generalização do classificador final em função dos hiperparâmetros correspondentes e encontrar um conjunto ótimo \mathbf{h}^* que maximiza a acurácia esperada. Para tal, $g : \mathcal{H} \rightarrow \mathbb{R}$ mapeia o espaço de hiperparâmetros \mathcal{H} para o conjunto de possíveis acurácias por *it* iterações e, com base em observações anteriores de $g(\mathbf{h})$, seleciona um novo grupo de valores candidatos $\mathbf{h} \in \mathcal{H}$ [12].

O problema de encontrar valores ótimos \mathbf{h}^* para os hiperparâmetros da cadeia SBCSP pode ser expressado como:

$$\mathbf{h}^* = \arg_{\mathbf{h} \in \mathcal{H}} \max g(\mathbf{h}), \quad (11)$$

em que $\mathcal{H} \subset \mathbb{R}^D$ cuja dimensão D é regida pela quantidade de hiperparâmetros no domínio.

A função $g(\mathbf{h})$, a qual deseja-se maximizar, representa a acurácia de classificação e é dada pela relação entre o número de épocas corretamente classificadas e o total de épocas no conjunto \mathcal{E}_V . Ao comparar as classificações em \mathbf{s} com os respectivos rótulos em \mathbf{t}_V , define-se o vetor \mathbf{a} para contabilizar as classificações corretas, em que cada elemento é dado por

$$a^{(i)} = \begin{cases} 1, & \text{se } s^{(i)} = t^{(i)} \\ 0, & \text{se } s^{(i)} \neq t^{(i)} \end{cases}, \forall i = 1, 2, \dots, N_V, \quad (12)$$

Portanto, acurácia do sistema cujos hiperparâmetros estão configurados com os valores \mathbf{h} em cada iteração *it* é definida por: $Acc_{it} = \frac{1}{N_V} \sum_{i=1}^{N_V} a^{(i)}$.

Tendo em vista que a efetividade da busca pelo conjunto ótimo \mathbf{h}^* depende de um número importante de iterações, a ponto de possibilitar a avaliação de um espaço amostral representativo, *it* = 100 iterações foram executadas.

A. Ferramenta e Espaço de hiperparâmetros

A biblioteca Python Hyperopt [13] foi utilizada para implementação do *auto setup* com base nos estimadores da Árvore de Parzen (Tree Parzen Estimators, TPE) [14]. Para compor o espaço de busca, considerou-se os hiperparâmetros associados a cada etapa do processamento, cada qual definido por um intervalo de busca específico conforme a Tabela I.

TABELA I
COMPOSIÇÃO DO ESPAÇO DE BUSCA DO AUTO SETUP

Hiperparâmetro	Intervalo de Busca
α_i	0, 0.5, ..., 2.5, 3
α_f	1, 1.5, ..., 3.5, 4
Δ_f	0, 1, ..., 50
N_s	0, 1, ..., 50
R	2, 4, ..., P
Φ	LDA, LR, SVM, k NN

Vale ressaltar, que o espaço detalhado na Tabela I também apresenta características implícitas de dependência interparâmetros. O intervalo de busca de N_s , por exemplo, pode ser reduzido conforme o intervalo definido em Δ_f . Assim como α_f relaciona-se com o valor assumido por α_i .

Além disso, o modelo de classificação Φ representa, por si só, um sub-espaço de busca com hiperparâmetros próprios condicionais, o qual é definido à posteriori à atribuição de Φ em uma mesma iteração, podendo ou não existir de acordo com o modelo atribuído. Para os modelos LDA e LR o sub-espaço de busca é inexistente, enquanto que para os demais modelos possíveis para Φ o sub-espaço pode ser formado pelos seguintes hiperparâmetros e intervalos condicionais:

- SVM
 - regularização $\Phi_c = \{10^{-8}, 10^{-7}, \dots, 10^4\}$
 - kernel $\Phi_k = \{\text{linear}, \text{sigmoide}, \text{gaussiana}, \text{poli } 3\}$
- k NN
 - nº vizinhos $\Phi_{nn} = \{2, 3, \dots, N_V\}$
 - métrica distância $\Phi_d = \{\text{euclidiana}, \text{manhattan}, \text{minkowski}, \text{chebyshev}\}$

IV. PROCEDIMENTOS EXPERIMENTAIS

O conjunto de dados 2a da Competição IV de BCI [15], o qual inclui sinais de 9 sujeitos que realizaram 4 tarefas de MI distintas (i.e. mão esquerda (ME), mão direita (MD), pés (PE) e língua (LG)), foi utilizado para validar a abordagem proposta. Os sinais foram gravados de $P = 22$ canais de EEG a $F_s = 250\text{Hz}$. Duas sessões de coleta foram disponibilizadas em subconjuntos distintos, cada qual contendo 72 tentativas por tarefa, totalizando 288 ensaios por sessão.

A. Avaliação de desempenho

O *auto setup* foi validado para cada sujeito considerando as 4 classes de MI disponíveis. Seis combinações foram formadas selecionando um par de classes por vez. A estratégia um-contrá-um adotada permite que a cada avaliação, $\{\Omega_1, \Omega_2\}$ corresponda a um sub-conjunto de $\{\{ME, MD\}, \{ME, PE\}, \{ME, LG\}, \{MD, PE\}, \{MD, LG\}, \{PE, LG\}\}$. Ao todo, $U = 54$ avaliações (9 sujeitos \times 6 pares de classes) foram realizadas.

A cada avaliação, as épocas associadas ao par $\{\Omega_1, \Omega_2\}$ são extraídas das sessões de coleta disponíveis. As 144 épocas

da sessão 1 são destinadas à busca pelo conjunto ótimo de hiperparâmetros. A busca por \mathbf{h}^* em cada avaliação se dá a partir do desempenho obtido nas $it = 100$ iterações, cada qual com um subconjunto de valores pertencentes ao espaço de busca definido em III-A. A cada iteração, uma validação cruzada de 5x5-pastas foi realizada para composição dos subconjuntos \mathcal{E}_T e \mathcal{E}_V e obtenção dos desempenhos individuais.

A acurácia por iteração Acc_{it} e conjunto \mathbf{h} correspondente é dada pela média entre as acurácias obtidas durante a validação cruzada. Concluídas as 100 iterações, o conjunto de hiperparâmetros ótimo \mathbf{h}^* representa os hiperparâmetros com maior acurácia de generalização (conjunto \mathcal{E}_V).

Uma vez obtido \mathbf{h}^* , a sessão 2 é utilizada na formação de um terceiro sub-conjunto de 144 épocas inéditas (72 por classe), denotado por \mathcal{E}_F , o qual é utilizado para validação final do *auto setup*. A acurácia gerada pelos valores de \mathbf{h}^* no conjunto \mathcal{E}_F em cada avaliação, Acc_u^* , foi utilizada para cálculo da acurácia média final do *auto setup*, na forma:

$$Acc_\mu = \frac{1}{U} \sum_{u=1}^U Acc_u^*, \quad (13)$$

B. Comparação entre técnicas

O desempenho do *auto setup* para o método SBCSP foi comparado à duas abordagens MI-BCI consolidadas.

1) *CSP-LDA*: um filtro butterworth de 5ª ordem foi aplicado nos sinais de EEG considerando um intervalo único de frequências entre 8 e 30Hz. $R = 8$ filtros CSP são usados para extração das características para classificação via LDA.

2) *SBCSP de configuração única*: um banco de filtros butterworth IIR de 5ª ordem filtra os sinais em $N_s = 9$ sub-bandas com intervalos de frequências definidos uniformemente para $\Delta_f = \{4, 40\}$. Em cada sub-banda, $R = 8$ filtros são utilizados na extração de características. O classificador final Φ assume um modelo SVM linear com regularização $\Phi_c = 10^{-4}$.

Em ambas as abordagens, as épocas foram extraídas do intervalo determinado por $\alpha_i = 0.5$ e $\alpha_f = 2.5$ e a Eq. (13) foi usada para obtenção do desempenho médio.

C. Resultados e Discussão

Ao se comparar as acurácias médias de classificação apresentadas na Tabela II e na Figura 2, observa-se um ganho de desempenho expressivo do *auto setup* em relação às abordagens de configuração única usadas como referência.

TABELA II
MÉDIA E DESVIO PADRÃO DAS ACURÁCIAS POR ABORDAGEM

Abordagem	Configuração	$Acc_\mu(\%)$
SBCSP	<i>auto setup</i>	88.57 ± 9.11
SBCSP	única/fixa	83.36 ± 12.65
CSP-LDA	única/fixa	80.67 ± 14.23

O auto ajuste dos hiperparâmetros SBCSP representou uma melhora significativa na generalização do sistema, dados os ganhos médios de 5.21% e 7.9% no desempenho em relação ao SBCSP e CSP-LDA de configuração única, respectivamente.

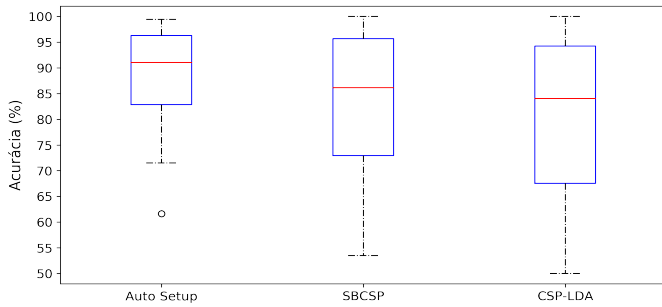


Fig. 2. Variância do desempenho obtido na abordagem proposta e abordagens de configuração única ao considerar todas as 54 avaliações

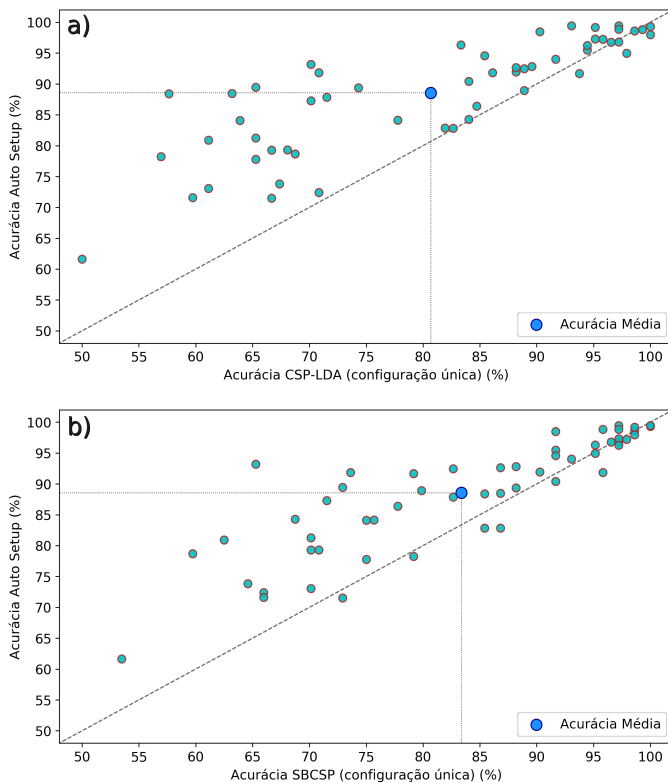


Fig. 3. Comparativo entre dispersões de acurácias: abordagens de configuração única versus abordagem proposta (*auto setup*)

O diagrama de caixas da Figura 2 descreve as variações de performance entre todas as 54 avaliações realizadas. O aumento da mediana e a redução do espalhamento entre as acurácias do *auto setup* evidenciam o ganho geral de desempenho em relação às abordagens de referência.

Os gráficos na Figura 3, por sua vez, apresentam a dispersão das U acurácias, de modo a correlacionar a performance individual do *auto setup* às abordagens CSP-LDA(a) e SBCSP(b).

Nota-se que, em ambas as comparações, o ganho médio supracitado refletiu-se em cada uma das 54 avaliações, denotando a eficácia do ajuste automático do sistema conforme o sujeito e tarefas de MI. Isso é evidenciado especialmente através das avaliações com piores desempenhos nas abordagens de referência (pontos próximos à extremidade esquerda) cujas acurácias foram notoriamente aumentadas pelo *auto setup*.

Além disso, o número de avaliações com alta generalização ($Acc_u^* > 90\%$) revela que, para um grupo significativo de sujeitos e pares de classes, a personalização do sistema pelo conjunto h^* é expressiva, sugerindo robustez da estratégia proposta independentemente do indivíduo e tarefas envolvidas.

V. CONCLUSÕES

Este trabalho propôs uma estratégia bayesiana para automatizar e personalizar o ajuste de hiperparâmetros em sistemas MI-BCI. O ganho significativo de desempenho em relação à configuração única, demonstra a eficácia do *auto setup* em reconhecer os padrões individuais que traduzem a intenção motora. A forte correlação entre o ajuste do sistema e seu desempenho, indicam que a configuração conforme o sujeito tende a minimizar o impacto da variabilidade do sinal de EEG no processo de classificação e a estabelecer um sistema mais generalista. Além disso, minimizar a imprecisão do ajuste manual e reduzir a dependência à expertise do projetista, especialmente para arquiteturas mais sofisticadas, o *auto setup* pode contribuir para a ampliação da usabilidade de sistemas MI-BCI. Por fim, a adoção da FFT na etapa de filtragem, como alternativa à elevação do tempo de processamento, também mostrou-se viável e sua relação custo-benefício pode ser melhor investigada em trabalhos futuros. Da mesma forma, o teste a partir de outros conjuntos de dados é recomendado para verificar a generalização da abordagem.

AGRADECIMENTOS

Agradecemos a Universidade Federal do Pará pelo apoio.

REFERÊNCIAS

- [1] L. F. Nicolas-Alonso e J. Gomez-Gil, "Brain computer interfaces, a review", *Sensors*, v. 12, n. 2, p. 1211-1279, 2012.
- [2] J. Wolpaw e E. W. Wolpaw (Ed), "Brain-computer interfaces: principles and practice", *OUN USA*, 2012.
- [3] M. Clerc, L. Bougrain, F. Lotte (Ed), *Brain-Computer Interfaces 1: Methods and Perspectives*, John Wiley & Sons, 2016.
- [4] Q. Novi et al., "Sub-band common spatial pattern (SBCSP) for brain-computer interface", *In: 2007 3rd International IEEE/EMBS Conference on Neural Engineering*, IEEE, 2007. p. 204-207.
- [5] H. Bashashati, R. K. Ward, A. Bashashati, "User-customized brain computer interfaces using Bayesian optimization", *Journal of neural engineering*, v. 13, n. 2, p. 026001, 2016.
- [6] M. Feurer e F. Hutter, "Hyperparameter optimization", *In: Automated Machine Learning*, Springer, p. 3-33, 2019.
- [7] S. Kumar, A. Sharma e T. Tsunoda, "An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information", *BMC bioinformatics*, 18(16), p.545, 2017.
- [8] G. Pfurtscheller e C. Neuper, "Motor imagery and direct brain-computer communication", *Proceedings of the IEEE*, v.89, n.7, p.1123-1134, 2001.
- [9] F. Lotte, "A tutorial on EEG signal-processing techniques for mental-state recognition in brain-computer interfaces", *In: Guide to Brain-Computer Music Interfacing*, Springer, London, p. 133-161, 2014.
- [10] K. K. Ang et al, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b", *Frontiers in neuroscience*, v.6, 2012.
- [11] B. Blankertz et al, "Optimizing spatial filters for robust EEG single-trial analysis", *IEEE Signal Processing Magazine*, v. 25, n.1, p. 41-56, 2008.
- [12] E. Brochu, V. M. Cora, N. De Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning", *arXiv:1012.2599*, 2010.
- [13] J. Bergstra, D. Yamins, D. Cox, "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms", *In: Proceedings of the 12th Python in science conference*, Citeseer, 2013.
- [14] Bergstra et al, "Algorithms for hyper-parameter optimization", *In: Advances in neural information processing systems*, p. 2546-2554, 2011.
- [15] C. Brunner et al, "BCI Competition 2008 – Graz data set A", Disponível em: http://www.bbci.de/competition/iv/desc_2a.pdf.