Efficient Calibration in Motor Imagery BCIs Under Data Constraints via Subject Transfer

Tekin Gunasar[1]
[1] Brown University
Providence, RI 02912
tekin_gunasar@brown.edu

Virginia R. de Sa[2]
[2] Department of Cognitive Science, HDSI
University of California, San Diego, La Jolla, CA 92093
desa@ucsd.edu

Abstract

Motor imagery brain-computer interfaces (MI-BCIs) often require large amounts of subject-specific calibration data. We propose Riemannian Transfer CSP (RTCSP), a transfer learning method that aligns EEG covariance matrices from prior users to a new subject using Riemannian geometry. This alignment yields spatial filters that maintain high performance even with limited calibration data, enabling efficient and practical MI-BCI deployment.

1 Introduction

Motor imagery brain computer interfaces (MI-BCIs) translate neural activity into control signals, typically using linear classifiers trained on subject-specific calibration data. A standard method is Common Spatial Patterns (CSP), which learns spatial filters that projects bandpass filtered original EEG into a space where variance is maximized for one class while it is minimized for the other [4]. However, EEG is highly non-stationary and subject-specific, so spatial filters trained on one subject generalize poorly to others [10, 19], making per-user and per-session calibration an obstacle for practical BCIs.

Transfer learning addresses this by leveraging data from previous users (sources) to aid a new user (target) in training classifier that generalizes better to new data. Prior work includes transferring features generated by CSP directly [5], aligning covariance matrices on Riemannian manifolds for direct classification of the covariances[15], and embedding transfer learning into the CSP formulation itself [6, 9, 16]. Yet most studies assume full datasets, leaving their effectiveness in low-data settings unclear and often difficult to conduct because of reliance on extensive hyperparameter tuning.

We propose Riemannian Transfer CSP (RTCSP), which combines covariance alignment on a Riemannian manifold with CSP to compute more robust subject-specific filters. By aligning source covariances to the target's, RTCSP expands the training set and improves generalization of learned spatial filters. While gains over CSP are modest with full data, RTCSP yields substantial improvements in low-data regimes, making it a promising approach for faster, more practical MI-BCI calibration.

2 Related Work

2.1 Riemannian Manifolds

The cone of $n \times n$ Symmetric Positive Definite (SPD) matrices, $\mathcal{P}(n)$ [3], forms a Riemannian manifold which provides us a mathematical structure to manipulate the covariance matrices associated with blocks of EEG that represent a motor imagery execution. Briefly, manifolds are spaces that appear flat locally but may curve globally; with bijective mappings called charts, they can be given

*This work was supported by NSF IIS-1817226 and hardware funding from Sony.
39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Foundation Models for the Brain and Body NeurIPS 2025 Workshop.

a smooth differential structure, enabling tangent spaces $T_X\mathcal{P}(n)$ for local linearization within the neighborhood of X on $\mathcal{P}(n)$. A Riemannian manifold further equips each tangent space with an inner product to measure distances and angles.

We map a matrix $X \in \mathcal{P}(n)$ to the tangent space at A using the logarithm map and back via the exponential map:

$$\tilde{X} = \mathrm{Log}_A(X) = X^{1/2}\log(X^{-1/2}AX^{-1/2})X^{1/2}, \quad X = \mathrm{Exp}_A(\tilde{X}) = \tilde{X}^{1/2}\exp(\tilde{X}^{-1/2}A\tilde{X}^{-1/2})\tilde{X}^{1/2},$$

where $\log$ and $\exp$ denote the matrix logarithm and exponential. By definition, $\mathrm{Log}_A(A) = 0$, so A is the origin of $T_A\mathcal{P}(n)$. Equipping each tangent space with the Affine-Invariant Riemannian Metric (AIRM),

$$\forall \tilde{X}^{(1)}, \tilde{X}^{(2)} \in T_A\mathcal{P}(n),\ \langle \tilde{X}^{(1)}, \tilde{X}^{(2)} \rangle_A = \mathrm{tr}(A^{-1}\tilde{X}^{(1)}A^{-1}\tilde{X}^{(2)}),$$

yields the squared geodesic distance between $X^{(1)}$ and $X^{(2)}$:

$$\delta^2_{\text{AIRM}}(X^{(1)}, X^{(2)}) = \sum_{k=1}^{n} \log^2(\lambda_k),$$

where λ_k are eigenvalues of $(X^{(1)})^{-1}X^{(2)}$. The Riemannian mean of SPD matrices $X^{(1)}, ..., X^{(n)}$ is then

$$\mu = \arg\min_{\mu \in \mathcal{M}} \sum_{i=1}^{n} \delta^2_{\text{AIRM}}(X^{(i)}, \mu),$$

which can be solved via gradient descent [1].

2.2 Common Spatial Patterns

CSP computes class-wise average covariance matrices

$$\Sigma_{(c)} = \frac{1}{|I_c|} \sum_{i \in I_c} \frac{X^{(i)}X^{(i)\prime}}{\mathrm{tr}(X^{(i)}X^{(i)\prime})},$$

with $c \in \{-1, +1\}$ and I_c are the set of indices associated with class c and $X^{(i)} \in \mathbb{R}^{C \times T}$, where C refers to the number of channels, and T refers to the number of samples in each channel for a single motor imagery trial. Spatial filters W are obtained from the generalized eigenvector problem

$$\Sigma_{(-1)}W = \lambda\Sigma_{(+1)}W,$$

If we wish to select n spatial filters for each class, we can do so by using the first and last n columns of W. Features for a trial $X^{(i)}$ are

$$f_i = \log\Big(\mathrm{diag}\{W'X^{(i)}X^{(i)\prime}W\}\Big).$$

For pre-processing, we bandpass EEG (8–30 Hz) with a 5th-order Butterworth filter and use three pairs of spatial filters. For multiclass data, CSP is applied in a One versus Rest(OvR) manner to create features for multiple binary classifiers in an OvR set up. Each binary classifier yields a probability, and the label with highest probability is predicted [17].

2.3 Riemannian Geometry in BCIs

EEG trials can be represented by covariance matrices and classified directly on $\mathcal{P}(n)$ by using the Riemannian Minimum Distance to Mean (RMDM) [2]. Recent methods combine neural networks with Riemannian geometry to capture spectral/temporal features [8], and applications extend to error-related activity [11]. Reported gains over CSP are mixed [7], and a comprehensive comparison remains open.

Early work such as Composite CSP [9] used linear combinations of source and target covariances. More recent approaches, e.g. Riemannian Procrustes Analysis (RPA) [15], align source/target covariances in a shared subspace and train RMDM. Extensions further perform unsupervised tangent-space alignment and classification on lower-dimensional representations [18].

3 Riemannian Transfer CSP

3.1 Problem Setup

Let source subject S provide N labeled trials $\{(X_S^{(i)}, y_S^{(i)})\}_{i=1}^N$ and target subject T provide M labeled trials $\{(X_T^{(j)}, y_T^{(j)})\}_{j=1}^M$, where each trial $X \in \mathbb{R}^{C\times T}$ has C channels and T samples. Trials from S and T are denoted $\boldsymbol{X_S}$ and $\boldsymbol{X_T}$ and $X_S^{(i)}$ would denote the i^{th} motor imagery trial from source subject S. We can obtain a normalized covariance from each trial in $\mathbb{R}^{C\times C}$.

$$\boldsymbol{\Sigma_S} = \left\{ \frac{X_S^{(i)} X_S^{(i)\prime}}{\mathrm{tr}(X_S^{(i)} X_S^{(i)\prime})} \right\}_{i=1}^{N}, \quad \boldsymbol{\Sigma_T} = \left\{ \frac{X_T^{(j)} X_T^{(j)\prime}}{\mathrm{tr}(X_T^{(j)} X_T^{(j)\prime})} \right\}_{j=1}^{M}.$$

Our goal is to align the distribution of covariances in $\boldsymbol{\Sigma_S}$ with $\boldsymbol{\Sigma_T}$. The aligned source covariances $\boldsymbol{\Sigma_{S_{al}}}$ are concatenated with $\boldsymbol{\Sigma_T}$ and their labels, i.e.,

$$\boldsymbol{\Sigma} = \mathrm{Concat}[\boldsymbol{\Sigma_{S_{al}}}, \boldsymbol{\Sigma_T}] \quad y = \mathrm{Concat}[y_S, y_T],$$

on which CSP is applied to compute spatial filters for the target subject's train and test trials.

3.2 Single Subject Transfer

We describe alignment for one class with the understanding that it is simply repeated for the other class. First, compute the Riemannian means of $\boldsymbol{\Sigma_S}$ and $\boldsymbol{\Sigma_T}$, denoted $\mathbf{M_S}$ and $\mathbf{M_T}$, and map each covariance to the tangent spaces $T_{\mathbf{M_S}}\mathcal{P}(C)$ and $T_{\mathbf{M_T}}\mathcal{P}(C)$ with the logarithm map. Each matrix in these tangent spaces is then vectorized by flattening its upper-triangular part, giving vectors of length $\frac{C(C+1)}{2}$:

$$\tilde{\mathbf{\Sigma}}_\mathbf{S} = \{\mathrm{vec}(Log_{\mathbf{M_S}}(\mathbf{\Sigma}_S^{(i)}))\}_{i=1}^N, \quad \tilde{\mathbf{\Sigma}}_\mathbf{T} = \{\mathrm{vec}(Log_{\mathbf{M_T}}(\mathbf{\Sigma}_T^{(j)}))\}_{j=1}^M.$$

We extract the top two principal components of these list of vectors: $\mathbf{P_S}, \mathbf{P_T} \in \mathbb{R}^{2\times\frac{C(C+1)}{2}}$ and compute Cholesky decompositions in these reduced spaces:

$$\mathbf{P_S}\tilde{\mathbf{\Sigma}}_S\tilde{\mathbf{\Sigma}}_S'\mathbf{P_S}' = \mathbf{L_S}\mathbf{L_S}', \quad \mathbf{P_T}\tilde{\mathbf{\Sigma}}_T\tilde{\mathbf{\Sigma}}_T'\mathbf{P_T}' = \mathbf{L_T}\mathbf{L_T}'.$$

We do these Cholesky Decompositions in the principal component space because often $\frac{C(C+1)}{2}$ is larger than the number of available samples. This yields a transformation aligning source to target:

$$\tilde{\mathbf{\Sigma}}_{\mathbf{S_{al}}}' = \{\mathbf{P_T}'\mathbf{L_T}\mathbf{L_S^{-1}}\mathbf{P_S}\,\tilde{\mathbf{\Sigma}}_S^{(i)'}\}_{i=1}^N.$$

The aligned vectors are reshaped into symmetric matrices and mapped back to the manifold via $Exp_{\mathbf{M_T}}$:

$$\mathbf{\Sigma}_{\mathbf{S_{al}}} = \{Exp_{\mathbf{M_T}}(\mathrm{mat}(\tilde{\mathbf{\Sigma}}_{\mathbf{S_{al}}}^{(i)}))\}_{i=1}^N.$$

Finally, CSP is run on $\mathbf{\Sigma} = \mathbf{\Sigma}_{\mathbf{S_{al}}} \cup \boldsymbol{\Sigma_T}$, producing a filter $\mathbf{W_S}$. Applying $\mathbf{W_S}$ to target trials yields log-variance features

$$\mathbf{f_S} = \{\log(\mathrm{diag}(\mathbf{W_S}'\mathbf{X_T}^{(j)}\mathbf{X_T}^{(j)'}\mathbf{W_S}))\}_{j=1}^M,$$

paired with labels $\mathbf{y_T}$ to train an LDA classifier. The same filter is used at test time. This describes transfer from a single source; extension to multiple sources and multiclass tasks follows by repeating the procedure in a one-vs-rest manner.

3.3 Multiple Subject Transfer

We extend the single-subject alignment method to incorporate multiple source subjects using two approaches: RTCSP Single Spatial Filter (SSF) and RTCSP-Combine.

In RTCSP-SSF, K source subjects $\mathbf{S_1}, \ldots, \mathbf{S_K}$ with covariances $\boldsymbol{\Sigma_{S_1}}, \ldots, \boldsymbol{\Sigma_{S_K}}$ are aligned to the target $\mathbf{S_T}$ with $\boldsymbol{\Sigma_T}$, producing aligned covariances $\boldsymbol{\Sigma_{S_1^{al}}}, \ldots, \boldsymbol{\Sigma_{S_K^{al}}}$. These are merged with the target:

$$\mathbf{\Sigma} = \mathrm{Concat}[\boldsymbol{\Sigma_T}, \boldsymbol{\Sigma_{S_1^{al}}}, \ldots, \boldsymbol{\Sigma_{S_K^{al}}}], \quad \mathbf{y} = \mathrm{Concat}[\mathbf{y_T}, \mathbf{y_{S_1}}, \ldots, \mathbf{y_{S_K}}].$$

Running CSP on $\mathbf{\Sigma}, \mathbf{y}$ yields a single spatial filter $\mathbf{W_S}$, which is applied to target trials for log-variance features which are then used to train an LDA classifier. The same filter is used at test time.

RTCSP-Combine instead generates individual spatial filters $\mathbf{W_{S_1}}, \ldots, \mathbf{W_{S_K}}, \mathbf{W_T}$ from each aligned source and the target. Each filter is applied to $\mathbf{X_T}$ to produce separate log-variance feature sets, which are concatenated for LDA training. If the target has M trials, this produces $M(K+1)$ feature vectors. Filters are only used to generate additional data points to train LDA with, not to expand feature dimensions. At test time, the RTCSP-SSF filter $\mathbf{W_S}$ is used for prediction, and the pre-trained classifier is used to make predictions. One may refer to appendix A to see how the method is implemented concretely in terms of an algorithm.

4 Results

We evaluate our approach on three motor imagery datasets and compare against established baselines, with hyper-parameter tuning applied where appropriate.

Datasets. We used three benchmark datasets. *BCI Competition III Dataset IVa* is a two-class (left hand, right foot) dataset with five subjects (aa, al, aw, av, ay) having 168, 224, 56, 84, and 28 trials, respectively; channels were reduced from 118 to 68 to avoid rank deficiency in trial covariances following [16], and Ledoit-Wolf shrinkage was applied. *BCI Competition IV Dataset 2a* is a four-class (left hand, right hand, tongue, foot) dataset [14] with 22 channels and 72 trials per class per subject, excluding artifacts, with only 22 channels. Given this dataset only had 22 measurement channels, covariance shrinkage methods were not needed. *Dataset Three* [13, 12] is a two-class (left hand, right hand) dataset with 12 subjects, each with 64 channels and 136 trials, also with Ledoit-Wolf shrinkage applied.

Compared Methods. We benchmark against two transfer learning approaches. *Composite CSP (cCSP)* constructs spatial filters from a linear combination of average class covariances of target and source subjects, then applies the resulting filter to all trials of the target. *Riemannian Procrustes Analysis (RPA)* transforms both source and target covariances into a shared space and uses a Riemannian Minimum Distance to Mean (RMDM) classifier trained on both aligned source and target covariances.

Hyper-parameter Tuning. Tuning was performed only for cCSP using target subject training data with validation splits, while all source data was included. The parameter λ, that controls how much target versus source subject covariances are weighted, was chosen to minimize validation error: for BCIC IV 2a and Dataset Three, 10-fold cross validation was run with $\lambda \in [0.1, 0.9]$ [13, 11]; for BCIC III 4a, given its variable trial counts, Leave-One-Out cross validation was used. Results are presented in Table 1 and Figures 1 and 2.

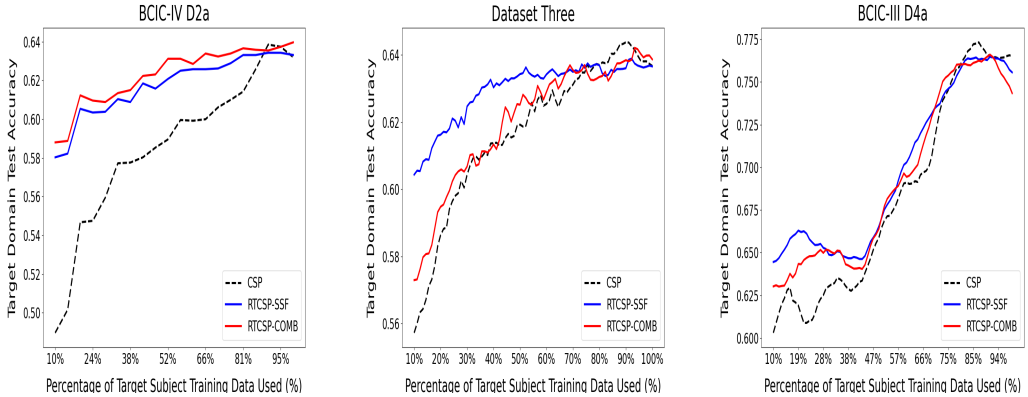


Figure 1: Performance of our methods averaged over all subjects in a dataset when different amounts of training data are available for the target subject. Our methods generalize to unseen data in low training data regimes more effectively than standard CSP. For more experimental details refer to appendix A, and to see what these curves look like for individual subjects refer to appendix B.

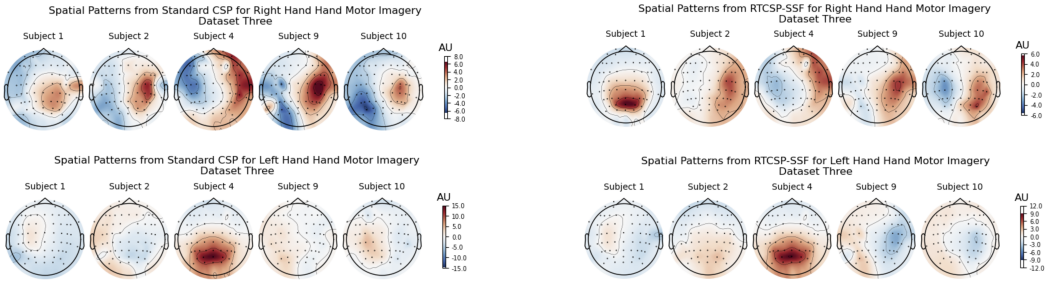


Figure 2: Spatial patterns for standard CSP and RTCSP-SSF, showing how presumed neural sources project onto the scalp at each electrode. Polarity reflects the relative contribution of each source to the measured signals in sensor space. Generalized eigenvalues from the CSP problem were ordered from highest to lowest, and we plot one spatial filter associated with the three largest eigenvalues, and one filter from the three smallest.

Subject	Dataset Three					BCIC-IV 2a					BCIC-III 4a				
	Ours		Baselines			Ours		Baselines			Ours		Baselines		
	SSF	COMB	CSP	cCSP	RPA	SSF	COMB	CSP	cCSP	RPA	SSF	COMB	CSP	cCSP	RPA
1	55.9	55.4	58.2	54.1	61.3	74.7	74.7	72.9	74.0	**77.4**	**69.6**	67.8	62.5	64.3	57.1
2	**52.6**	52.3	48.5	51.5	47.6	**45.5**	**45.5**	44.1	44.4	32.6	**65.2**	62.5	60.7	60.2	52.2
3	57.9	51.3	**61.8**	54.0	48.6	79.5	78.8	**81.6**	80.6	70.8	59.2	57.1	59.2	**60.2**	59.7
4	65.4	64.2	60.8	**66.2**	63.6	62.2	63.5	**64.2**	63.5	47.2	71.4	71.4	76.8	78.5	**89.3**
5	81.4	81.4	77.9	69.3	63.6	38.5	39.2	39.9	**41.0**	36.5	73.4	73.8	71.0	**74.2**	50.4
6	**70.8**	70.6	68.3	56.2	60.9	46.5	**46.9**	43.1	42.4	40.3					
7	57.3	57.6	57.6	51.7	**58.7**	78.5	**79.2**	73.6	75.0	67.4					
8	**62.9**	62.5	60.5	56.0	54.0	76.7	77.1	78.5	78.1	**77.8**					
9	69.0	68.5	65.5	**70.5**	61.0	72.6	73.3	70.8	73.2	**74.0**					
10	58.5	57.9	60.2	57.9	**63.8**										
11	46.4	**48.6**	42.9	47.7	50.0										
12	57.7	57.4	**59.8**	50.3	55.6										
Mean	**61.3**	60.7	60.1	57.1	57.2	63.9	**64.2**	63.2	63.6	58.0	**67.8**	66.5	66.0	67.5	61.8

Table 1: Performance across three datasets. Gray cells indicate subjects not present in that dataset. Bold numbers indicate the best-performing method per subject.

5 Discussion

Our results demonstrate that RTCSP achieves more robust calibration than standard CSP and other transfer learning baselines, especially when subject-specific data is severely limited. While performance gains are modest with full datasets, the key advantage of our approach is its ability to reach optimal accuracy with far fewer trials, as shown in Figure 1. This efficient calibration directly addresses a major bottleneck for the practical deployment of MI-BCIs.

It is important to note that our transfer paradigm compensates for data scarcity, not inherent signal quality issues. RTCSP aligns to the target subject's data distribution; therefore, its performance is contingent on the quality of the limited target data available. If performance is poor due to factors like low signal quality or user inattention, RTCSP cannot overcome these limitations. However, when performance is limited primarily by data quantity, our method provides substantial improvement by leveraging transferred data to learn more robust spatial filters.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] A. Barachant et al. Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2011.

[3] R. Bhatia. *Positive Definite Matrices*. Princeton University Press, 2009.

[4] B. Blankertz et al. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2007.

[5] Z. Chen, M. Mousavi, and V. R. de Sa. Multi-subject unsupervised transfer with weighted subspace alignment for common spatial patterns. In *2022 10th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2022.

[6] D. Devlaminck et al. Multisubject learning for common spatial patterns in motor-imagery bci. *Computational Intelligence and Neuroscience*, 2011(1):217987, 2011.

[7] M. Eder, J. Xu, and M. Grosse-Wentrup. Benchmarking brain–computer interface algorithms: Riemannian approaches vs convolutional neural networks. *Journal of Neural Engineering*, 21(4):044002, 2024.

[8] C. Ju and C. Guan. Tensor-cspnet: A novel geometric deep learning framework for motor imagery classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12):10955–10969, 2022.

[9] H. Kang, Y. Nam, and S. Choi. Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Processing Letters*, 16(8):683–686, 2009.

[10] F. Lotte et al. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 4(2):R1, 2007.

[11] M. Mousavi and V. R. de Sa. Spatio-temporal analysis of error-related brain activity in active and passive brain–computer interfaces. *Brain-Computer Interfaces*, 6(4):118–127, 2019.

[12] M. Mousavi and V. R. de Sa. Motor imagery performance from calibration to online control in eeg-based brain-computer interfaces. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2021.

[13] M. Mousavi, L. R. Krol, and V. R. de Sa. Hybrid brain-computer interface with motor imagery and error-related brain activity. *Journal of Neural Engineering*, 17(5):056041, 2020.

[14] M. Naeem et al. Seperability of four-class motor imagery data using independent components analysis. *Journal of Neural Engineering*, 3(3):208, 2006.

[15] P. L. C. Rodrigues, C. Jutten, and M. Congedo. Riemannian procrustes analysis: transfer learning for brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 66(8):2390–2401, 2018.

[16] W. Samek et al. Stationary common spatial patterns for brain–computer interfacing. *Journal of Neural Engineering*, 9(2):026013, 2012.

[17] W. Wu, X. Gao, and S. Gao. One-versus-the-rest (ovr) algorithm: An extension of common spatial patterns (csp) algorithm to multi-class case. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006.

[18] W. Zhang and D. Wu. Manifold embedded knowledge transfer for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(5):1117–1127, 2020.

[19] S.-M. Zhou, J. Q. Gan, and F. Sepulveda. Classifying mental tasks based on features of higher-order statistics from eeg signals in brain–computer interface. *Information Sciences*, 178(6):1629–1640, 2008.

A Experimental and Method Details

A.1 Generation of Figure 1

In this section we provide a detailed explanation of how the curves in Figure 1 were generated. Consider a dataset with k subjects, denoted $S_1, \ldots, S_k$. Let S_j represent the target subject, and let the remaining subjects $\{S_1, \ldots, S_k\} \setminus \{S_j\}$ serve as source subjects. Each subject's training set contains N motor imagery trials. For the target subject S_j, we consider the condition in which only $p\%$ of the available training data is used, corresponding to $\lfloor pN \rfloor$ trials.

Under this setting, standard CSP is trained exclusively on the $\lfloor pN \rfloor$ samples from S_j. By contrast, RTCSP-SSF and RTCSP-Combine are trained using both the $\lfloor pN \rfloor$ samples from S_j and the additional $N(k-1)$ training samples provided by the source subjects, which are used in our custom multiple subject transfer procedure. After the spatial filters are obtained, performance is evaluated on a separate test set from S_j, which was recorded in a different session to ensure cross-session generalization.

This procedure is repeated for varying values of p, thereby producing a performance curve for each subject in the dataset. Finally, the curves are averaged across all subjects to produce the aggregate results shown in Figure 1. For readers interested in the subject-level results, the corresponding individual curves are provided in Appendix B. In all experiments that involved CSP and its variants, three pairs of spatial filters were used.

A.2 RTCSP-Combine Algorithm

Originally, a more surface-level explanation of RTCSP-Combine was provided. The key difference from RTCSP-SSF is that RTCSP-Combine generates multiple CSP-based spatial filters $W_{S_1}, \ldots, W_{S_K}$, each aligned to the target subject's filter W_T. These filters are **used to generate additional training examples** and not to add dimensions for an LDA classifier.

Specifically, for a target subject with M training trials, each aligned filter W_{S_i} is applied to X_T to produce M new feature vectors. Repeating this for all K source filters and including the original M trials produces a total of $M(K+1)$ training examples for the LDA classifier. If each trial has C channels and n pairs of eigenvectors are selected per filter, then each $W_{S_i} \in \mathbb{R}^{C \times 2n}$, and applying W_{S_i} to X_T yields feature vectors in $\mathbb{R}^{2n}$. These features can be combined because all filters are aligned to the same target subject, ensuring consistency in the resulting feature distribution.

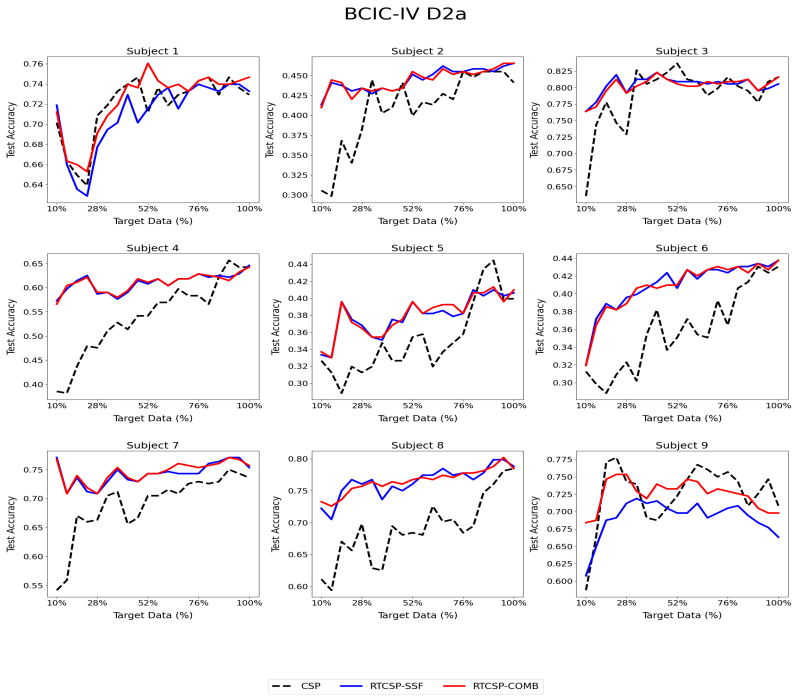
At test time predictions are made using the spatial filter learned with RTCSP-SSF applied to the unseen test trials. In the below algorithm 'Align' refers to our alignment method described in a single subject transfer that serves as the basis for both RTCSP-SSF and RTCSP-Combine.

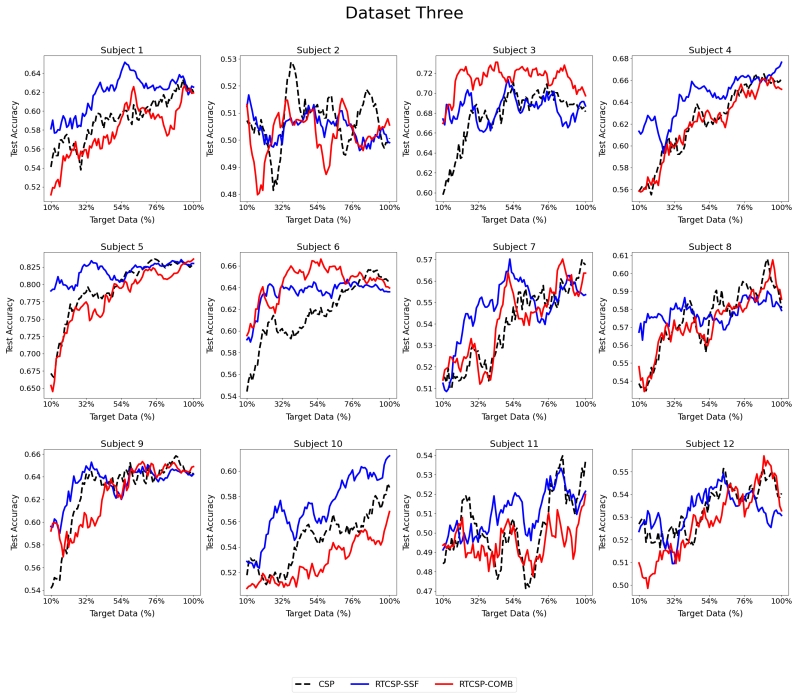
Algorithm 1 RTCSP-Combine

Input: Source subjects $S_1, \ldots, S_K$
Target subject S_T

```
Initialize TrainFeatures = []
Initialize TrainLabels = []
for i = 1 to K do
    Σ_{S_i^al} = Align(Σ_{S_i}, Σ_T)
    W_{S_i} = CSP(Σ_{S_i^al}, y_{S_i})
    f_{S_i} = getLogVarFeatures(X_T, W_{S_i})
    TrainFeatures.append(f_{S_i})
    TrainLabels.append(y_T)
end for
W_T = CSP(Σ_T, y_T)
f_T = getLogVarFeatures(X_T, W_T)
TrainFeatures.append(f_T)
TrainLabels.append(y_T)
clf = LDA(TrainFeatures, TrainLabels)
```

B Subject-Specific Performance Curves





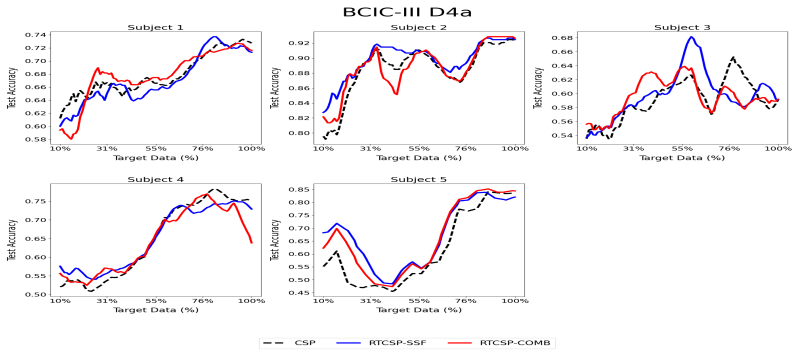


Figure 3: Subject-specific performance curves for BCIC-IV 2a, Dataset Three, and BCIC-III D4a. Each curve shows the classification accuracy for a given subject as the fraction of available target training data increases. CSP, RTCSP-SSF, and RTCSP-Combine are compared. A moving average (window size 9) was applied to Dataset Three and BCIC-III D4a for clarity.

C Additional Experiments

The objective of CSP is to learn a spatial filter that projects a task relevant EEG signal to a space where variance is maximized for one class, while it is minimized for the other. We have explored the end result of our methods, that is, test accuracy for some subject when varying amounts of training data are available, but we have yet to analyze how our methods affect the resulting spatial filters found when data availability varies, and whether or not the resulting spatial filters are able to create more discriminative features. We dedicate this section to explain the experiment we have conducted to explore this question. An explanation will be provided for one run of the experiment, with one condition, as in, what percentage of the target subjects training data was used, with the understanding that this procedure is simply repeated to obtain a distribution of the statistic we are interested in: how well RTCSP-SSF generalizes to unseen data compared to standard CSP when different amounts of data are available.

Suppose we have subjects $\{S_1, S_2, \dots, S_n\}$. For simplicity sake assume that each subject S_i has an associated dataset with shape $N \times C \times T$, that is, N blocks of EEG that represent a motor imagery execution, with C channels, and T samples per channel. Additionally, suppose that we have some target subject S_k, and we are interested in the condition where we use only p% of the available training data from subject S_k, thus, a total of $\left\lceil \frac{p}{100} N \right\rceil$ motor imagery trials to obtain a spatial filter with. If S_k is the target subject then the source subjects are $\{S_1, S_2, \dots, S_n\} \setminus \{S_k\}$. We do not limit the training data on the remaining $n-1$ source subjects.

To compare standard CSP in low training data scenarios with RTCSP-SSF, we obtain two spatial filters. One using standard CSP with the limited $\left\lceil \frac{p}{100} N \right\rceil$ training motor imagery trials, and we call this spatial filter W_{base}. The other spatial filter results from the RTCSP-SSF method and will use $\left\lceil \frac{p}{100} N \right\rceil + N(n-1)$ total motor imagery trials to obtain a spatial filter with. This spatial filter will be referred to as W_{RT}. Suppose now that the target subject S_k has N_{te} test trials recorded from another session that were not used at all when obtaining W_{base} and W_{RT}. We are interested in how well W_{base} and W_{RT} are able to generalize to these unseen trials and quantifying this for hypothesis testing purposes.

We label these unseen test trials as $X^{(1)}, \dots, X^{(N_{te})}$ and map each one to the CSP space using W_{base} and W_{RT}. For simplicity purposes consider a generic CSP-based spatial filter W. We can map $X^{(i)}$ to the CSP space with the transform $X_{CSP}^{(i)} = W' X^{(i)}$. In this experiment we used one pair of spatial filters, thus $X_{CSP}^{(i)}$ has two channels. Suppose that the variance of these two channels are $v_1^{(i)}$ and $v_2^{(i)}$, then the statistic we are interested in is,

$$\frac{\max(v_1^{(i)}, v_2^{(i)})}{\min(v_1^{(i)}, v_2^{(i)})}$$

which is directly related to the CSP objective function.

Using this statistic, we then compute the "mean-variance-ratio" (MVR) across all the test trials $X^{(1)}, \dots, X^{(N_{te})}$

$$MVR(W, S_k) = \frac{1}{N_{te}} \sum_{i=1}^{N_{te}} \frac{\max(v_j^{(i)})}{\min(v_j^{(i)})}, j = 1, 2$$

The S_k in the argument of MVR referring to the fact that the above quantity is being calculated for test trials of subject S_k , and the W referring to what spatial filter is being used to project test trials into CSP space. This quantity should be obtained for all subjects, alternating between which subject is the target, which will also change which subjects are used as sources. Once we have this quantity for all subjects in a dataset, we simply average it to obtain our test statistic for one run of the experiment using the spatial filter W. For subjects $S_1, \ldots, S_n$ our statistic for a single spatial filter W can be formulated as $MVR(W) = \frac{1}{n} \sum_{i=1}^{n} MVR(W, S_i)$.

The quantity should be computed for both $W = W_{base}$ (standard CSP) and $W = W_{RT}$ (RTCSP-SSF). Due to the variability of this quantity, especially when less training data is used, this experiment was ran 50 times for each value of p, and during hypothesis testing we compare the distribution of $MVR(W_{base})$ and $MVR(W_{RT})$, with the alternative hypothesis that $MVR(W_{RT}) > MVR(W_{base})$, meaning that RTCSP-SSF is able to map to a feature space where the resulting variance based features will be more discriminative on unseen data, which translates to more effective classification. We tested this hypothesis across nine different percentages of training data to use from a target subject, thus to conservatively account for multiple comparisons we set our alpha level to $0.05/9$.

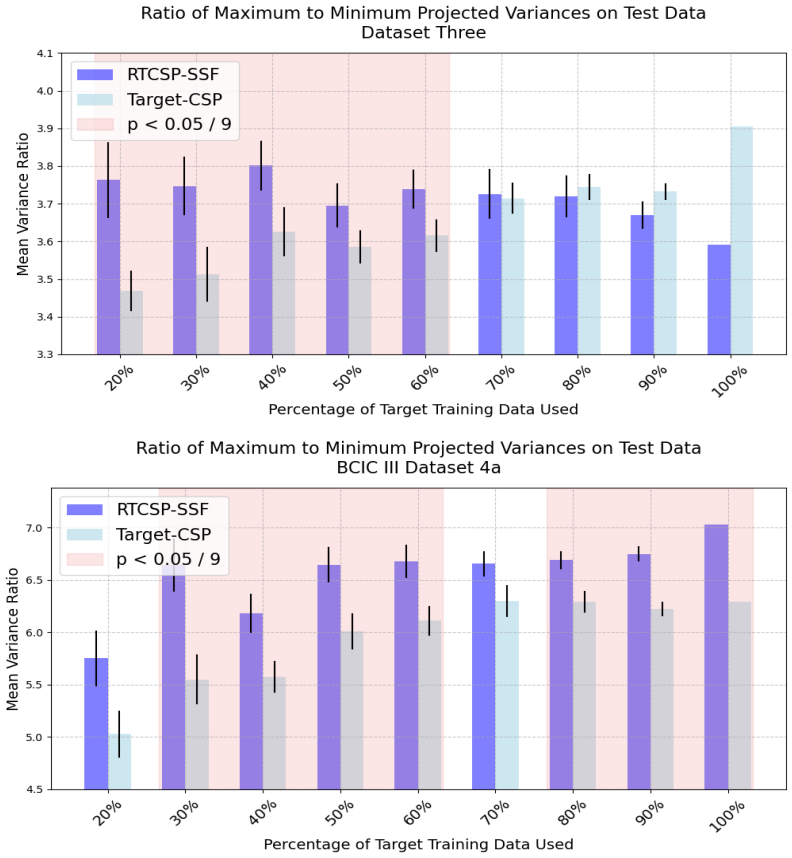


Figure 4: The ratio of maximum to minimum projected variances on datasets three and BCIC-III 4a, averaged over trials and subjects. This quantity reflects a spatial filter's ability to generalize and create discriminative features on unseen data, enabling more accurate classification. Standard CSP requires a larger amount of training data to generalize effectively, whereas RTCSP-SSF achieves similar performance with substantially less data. Error bars represent the standard error across 50 runs. Refer to Fig. 5 for the distribution corresponding to the left-most bars (20% of training data used).

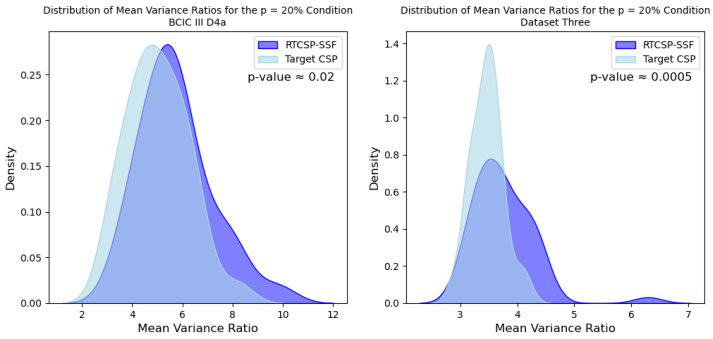


Figure 5: Example of 50 runs of the described experiment when 20% of target training data is used. These distributions correspond with the left most bars on figure 4.