Riemannian Transfer Learning in Motor Imagery Decoding: Reproducibility and Standardized Benchmarks

Vishwa Ramkumar
Department of Biotechnology
Indian Institute of Technology, Madras
Chennai, Tamil Nadu, India
be20b039@smail.iitm.ac.in

Chandrashekar Lakshminarayanan
Department of Computer Science
Indian Institute of Technology, Madras
Chennai, Tamil Nadu, India
chandrashekar@cse.iitm.ac.in

Abstract

Motor imagery (MI)-based brain-computer interfaces (BCIs) hold significant potential for rehabilitation and assistive technologies. However, their widespread adoption is hindered by high inter-subject variability in electroencephalogram (EEG) signals, necessitating extensive calibration for new users. Transfer learning (TL) methods overcome this by leveraging data from existing subjects to reduce the calibration time. However, the lack of standard evaluation protocols in EEG-MI TL research makes it challenging to compare different approaches fairly. Moreover, the lack of availability of codebases adds to the issue of reproducibility. In this paper, we propose a standardized evaluation protocol to compare key transfer learning techniques across cross-session and cross-subject scenarios. We further conduct ablation studies focusing on signal length and preprocessing parameters to quantify the sensitivity of the algorithms to signal and noise variability. Finally, we present Python implementations of the methods for reproducibility.

1 Introduction

Brain-Computer Interfaces (BCIs) enable direct communication between brain and devices. Motor Imagery (MI) BCIs, which decode imagined movements from EEG, hold promise for rehabilitation and assistive use. Classical methods such as Minimum Distance to Riemannian Mean (MDRM) [1] and Common Spatial Patterns (CSP) [2] with Linear Discriminant Analysis (LDA) [3, 4] perform well within sessions but degrade across sessions and subjects, requiring costly calibration. Transfer learning (TL) mitigates this. Riemannian Alignment (RA) whitens trials via resting-state means before MDRM. Euclidean Alignment (EA) [5] uses the global mean in Euclidean space, while Log-Euclidean Alignment MDRM (LEA-MDRM) [6] relies on log-Euclidean means of active trials. Tangent Space Alignment (TSA) [7] applies tangent mapping with class-wise Procrustes rotations, and Riemannian Procrustes Analysis (RPA) [8] aligns by recentering, scaling, and rotation. Manifold Embedded Knowledge Transfer (MEKT) [9] aligns covariance matrices to a shared reference, projects to tangent space, and reduces distributional shifts while preserving geometry. Yet, the field lacks standardized protocols and open codebases, hindering reproducibility and fair comparison. Our study addresses these gaps through the following contributions:

- **Standardized Evaluation:** We benchmark key transfer learning (TL) methods using a unified protocol for cross-session and cross-subject scenarios.
- **Ablation Studies:** We analyze signal length and preprocessing choices to assess algorithm sensitivity to signal and noise variability.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Learning from Time Series for Health.

- **Open-Source Code:** We release Python implementations and the benchmarking codebase to ensure reproducibility and support future work.

2 Related Work

To mitigate domain shifts from inter-subject and inter-session variability in EEG signals, classical and Riemannian transfer learning (TL) techniques are widely used in motor imagery (MI) decoding. However, inconsistent evaluation protocols hinder fair comparisons. Benchmarking efforts such as MOABB [10] enable reproducible BCI evaluations, with a recent large-scale study [11] testing 30+ pipelines across 36 datasets but restricted to within-session settings. EEG-FM-Bench [12] extends this to foundation models. Yet, systematic studies of Riemannian TL methods in cross-subject and cross-session MI decoding, including ablation on preprocessing, remain lacking. Our work addresses this gap with standardized evaluations and open-source implementations to ensure reproducibility.

3 Motor Imagery datasets

We used the BCI Competition IV datasets 2a [13] and 2b [14], standard MI benchmarks with multi-session data enabling cross-session and cross-subject evaluation (see Appendix for details).

4 Methods: Standardizing benchmarks

The selected methods span key paradigms in EEG-based motor imagery transfer learning, covering Euclidean, Riemannian, and tangent space approaches. We included CSP+LDA, MDRM, RA-MDRM, EA, LEA, TSA, RPA, and MEKT to enable comparison between simple and advanced techniques. To ensure rigor, we first reproduced algorithm results 5–12, confirming close agreement. Benchmarks were then conducted in two scenarios: cross-session (training on Session 1, testing on Session 2) and cross-subject (leave-one-subject-out). Both 4-class (left hand, right hand, foot, tongue) and 2-class (left vs. right hand) tasks were evaluated. Trials spanned 0.5–3.5s post-cue, covering the motor imagery window. Signals were band-pass filtered to 8–30 Hz with a 50th-order filter, targeting mu and beta rhythms. Preprocessing followed prior consensus, with ablation confirming utility. For methods with tunable hyperparameters, values were adopted directly from source papers.

4.1 Benchmark 1: 2 Class vs 4 Class: Cross-Session

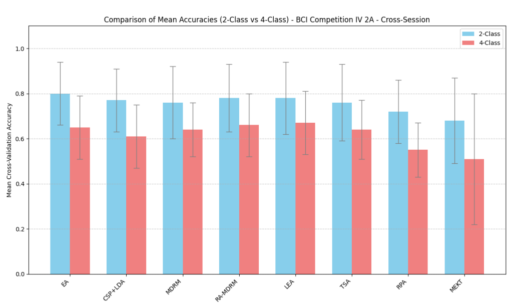

Figure 1: Comparison of mean accuracies in the cross-session scenario, on the BCI Competition IV 2A dataset. EA and MEKT achieve the highest and lowest mean accuracies respectively.

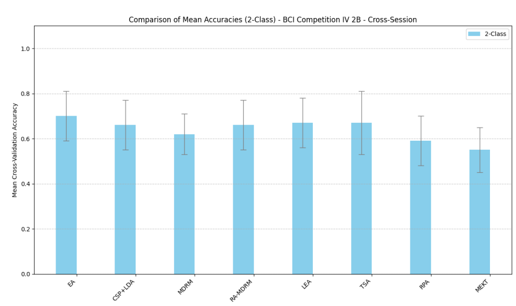

Figure 2: Comparison of mean accuracies in the cross-session scenario, on the BCI Competition IV 2B dataset. EA and MEKT are again the best and worst performers respectively, with LEA slightly behind EA.

In dataset 2a cross-session 13 14, most TL methods perform well 1, especially in 2-class, showing adaptability to session shifts. EA achieves the highest mean 2-class accuracy, slightly surpassing RA-MDRM and LEA, confirming the benefit of Euclidean alignment before feature extraction. Accuracy drops notably in 4-class, reflecting higher task difficulty. MEKT performs worst in both tasks, suggesting sensitivity to alignment dataset size. CSP+LDA (0.77 / 0.61) is competitive but generally trails alignment methods. Large subject-wise SDs indicate strong inter-subject variability. For dataset 2b 17, 2-class cross-session accuracies 2 are lower than 2a, likely due to fewer EEG channels (3 vs. 22). EA (0.70) again leads, slightly ahead of LEA, TSA, and RA-MDRM. CSP+LDA performs comparably to several alignment methods, suggesting weaker session shifts or limited alignment gains with few channels. MDRM (0.62) and MEKT (0.55) are lowest, with MEKT most affected by reduced spatial data. Variability remains high (SD $\sim$0.10–0.14) but slightly lower than in 2a, confirming persistent subject differences.

4.2 Benchmark 2: 2 Class vs 4 Class: Cross-Subject

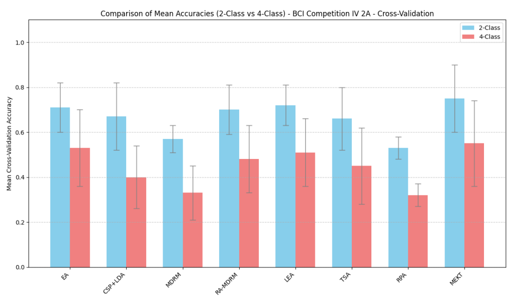

Figure 3: Comparison of mean accuracies in the cross-subject scenario, on the BCI Competition IV 2A dataset. MEKT performs the best in both the 2-class and 4-class scenarios with more data available.

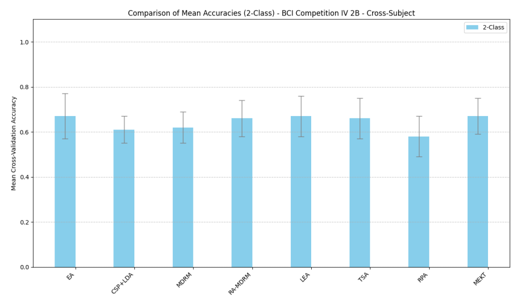

Figure 4: Comparison of mean accuracies in the cross-subject scenario, on the BCI Competition IV 2B dataset. Performance differences between models are less pronounced due to the dataset's low channel complexity.

In dataset 2a cross-subject 15 16, accuracies drop compared to cross-session 3, reflecting higher inter-subject variability. MEKT performs best (0.75 / 0.55), showing its manifold embedding and joint adaptation are well-suited to bridging subject differences. LEA (0.72 / 0.51) and EA (0.71 / 0.53) also rank high, surpassing RA-MDRM (0.70 / 0.48), MDRM (0.57 / 0.33), and CSP+LDA (0.67 / 0.40), confirming the importance of alignment. RPA (0.53 / 0.32) lags, especially in 4-class, suggesting Procrustes limitations under high variability. The sharp 4-class drop (e.g., MEKT 0.75 $\rightarrow$ 0.55) underscores the difficulty of cross-subject multi-class MI BCI. For dataset 2b 18, 2-class cross-subject accuracy also degrades but less than in 2a, likely due to task simplicity. EA, LEA, and MEKT each reach 0.67 4, while RA-MDRM (0.66) and TSA (0.66) perform similarly, ahead of CSP+LDA (0.61), MDRM (0.62), and RPA (0.58). The smaller gap between best methods and baseline than in 2a may reflect channel limitations reducing alignment gains. MEKT's strength here, unlike cross-session, highlights its ability to handle distribution shifts with more data.

5 Ablation study

An ablation study is conducted to identify the optimal EEG time window for motor imagery (MI) analysis, as neural responses evolve dynamically after cue onset. The choice of start–end points determines which phases of the motor imagery response is captured. Short windows may miss information, while longer ones risk adding noise, and algorithms vary in sensitivity to these dynamics. We test multiple windows ((0.5–1.5), (0.5–2.5), (0.5–3.5), (1.5–2.5), (1.5–3.5), (2.5–3.5) s) to evaluate each method's reliance on temporal characteristics. We also examine filtering, a standard preprocessing step to improve SNR by isolating MI-related frequency bands. Since MI is linked to mu (8–13 Hz) and beta (14–30 Hz) rhythms via ERD/ERS, we compare unfiltered signals with the standard MI band (8–30 Hz) and narrower alpha (8–12 Hz) and beta (13–30 Hz) filters. This assesses the importance of spectral components for classification and how algorithm performance depends on them. Together, these analyses reveal algorithm robustness to temporal and spectral variations and their reliance on discriminative features.

Table 1: Mean Accuracy ± Standard Deviation across varying time windows: Dataset 2a. Using the full trial length significantly improves performance.

Time Window Method	(0.5,1.5)	(0.5,2.5)	(0.5,3.5)	(1.5,2.5)	(1.5,3.5)	(2.5,3.5)
EA	0.60 ± 0.16	0.64 ± 0.16	**0.65 ± 0.14**	0.57 ± 0.15	0.60 ± 0.14	0.49 ± 0.11
CSP+LDA	0.57 ± 0.15	0.60 ± 0.13	**0.61 ± 0.14**	0.55 ± 0.13	0.57 ± 0.12	0.47 ± 0.09
MDRM	0.58 ± 0.12	0.62 ± 0.11	**0.64 ± 0.12**	0.54 ± 0.11	0.58 ± 0.12	0.49 ± 0.12
RA-MDRM	0.60 ± 0.13	**0.67 ± 0.13**	0.66 ± 0.14	0.58 ± 0.13	0.61 ± 0.12	0.50 ± 0.11
LEA	0.62 ± 0.14	**0.67 ± 0.13**	**0.67 ± 0.14**	0.58 ± 0.14	0.62 ± 0.13	0.52 ± 0.12
TSA	0.56 ± 0.16	0.61 ± 0.16	**0.62 ± 0.14**	0.52 ± 0.14	0.56 ± 0.15	0.46 ± 0.13
RPA	0.51 ± 0.11	0.53 ± 0.13	**0.56 ± 0.14**	0.46 ± 0.11	0.49 ± 0.11	0.45 ± 0.10
MEKT	0.47 ± 0.23	**0.51 ± 0.27**	**0.51 ± 0.29**	0.46 ± 0.23	0.49 ± 0.26	0.45 ± 0.19

Table 2: Mean Accuracy ± Standard Deviation across varying time windows: Dataset 2b. Performance difference isn't significant between the windows due to low complexity of the dataset.

Time Window Method	(0.5,1.5)	(0.5,2.5)	(0.5,3.5)	(1.5,2.5)	(1.5,3.5)	(2.5,3.5)
MDRM	0.62 ± 0.11	**0.63 ± 0.11**	0.62 ± 0.09	0.58 ± 0.10	0.59 ± 0.07	0.56 ± 0.07
RA-MDRM	0.65 ± 0.10	**0.66 ± 0.12**	**0.66 ± 0.11**	0.61 ± 0.09	0.61 ± 0.09	0.57 ± 0.07
EA	0.67 ± 0.11	0.68 ± 0.11	**0.70 ± 0.11**	0.62 ± 0.10	0.64 ± 0.09	0.60 ± 0.07
LEA	0.65 ± 0.11	**0.67 ± 0.11**	**0.67 ± 0.11**	0.61 ± 0.10	0.62 ± 0.08	0.58 ± 0.06
CSP+LDA	0.63 ± 0.11	0.65 ± 0.13	**0.66 ± 0.11**	0.60 ± 0.10	0.62 ± 0.09	0.59 ± 0.06
MEKT	0.54 ± 0.07	0.55 ± 0.08	0.55 ± 0.10	0.54 ± 0.07	**0.57 ± 0.09**	0.56 ± 0.09
TSA	0.63 ± 0.12	**0.68 ± 0.11**	0.65 ± 0.13	0.59 ± 0.12	0.58 ± 0.12	0.56 ± 0.07
RPA	0.60 ± 0.12	0.60 ± 0.13	**0.64 ± 0.09**	0.53 ± 0.13	0.55 ± 0.07	0.55 ± 0.09

5.1 Ablation study 1: 4-Class Cross-Session: Variable Input signal length

In dataset 2a 1, longer windows covering the full MI period improve accuracy, with (0.5–3.5s) giving the best or near-best results (EA 0.65, RA-MDRM 0.66, LEA 0.67, MDRM 0.64). The shortest late window (2.5–3.5s) performs worst, confirming reduced discriminative power later. RA-MDRM and LEA slightly exceed EA and CSP+LDA, while MEKT and RPA remain weakest. In dataset 2b 2, trends are similar but weaker: (0.5–3.5s) is best for EA (0.70) and CSP+LDA (0.66), (0.5–2.5s) for TSA (0.68), and late-only windows lowest (EA 0.60, LEA 0.58). MEKT stays lowest overall (0.55–0.57) with little sensitivity. Flatter patterns than 2a suggest fewer channels and two classes reduce window choice impact, aside from avoiding short/late segments.

5.2 Ablation study 2: Cross-Session: Pre-processing filters

Table 3: Mean Accuracy ± SD for varying preprocessing filters: Dataset 2a. Most discriminative information lies in the beta band, though some is also in the alpha band.

Preprocessing Method	None	Standard	Alpha	Beta
EA	0.53 ± 0.14	**0.65** ± 0.14	0.56 ± 0.15	**0.65** ± 0.16
CSP+LDA	0.50 ± 0.15	**0.61** ± 0.14	0.52 ± 0.14	0.60 ± 0.14
MDRM	0.55 ± 0.15	**0.64** ± 0.12	0.58 ± 0.16	0.63 ± 0.11
RA-MDRM	0.59 ± 0.12	0.66 ± 0.14	0.61 ± 0.16	**0.67** ± 0.13
LEA	0.60 ± 0.12	**0.67** ± 0.14	0.61 ± 0.16	**0.67** ± 0.13
TSA	0.53 ± 0.15	**0.63** ± 0.14	0.57 ± 0.17	**0.63** ± 0.13
RPA	0.49 ± 0.10	**0.55** ± 0.13	0.52 ± 0.13	0.52 ± 0.11
MEKT	0.46 ± 0.23	**0.51** ± 0.29	0.50 ± 0.26	**0.52** ± 0.27

Table 4: Mean Accuracy ± SD for varying preprocessing filters: Dataset 2b. Performance differences are smaller due to fewer available channels.

Preprocessing Method	None	Standard	Alpha	Beta
EA	0.58 ± 0.06	**0.70** ± 0.11	0.68 ± 0.10	0.65 ± 0.12
CSP+LDA	0.60 ± 0.08	**0.66** ± 0.11	0.65 ± 0.11	0.61 ± 0.09
MDRM	0.56 ± 0.07	**0.62** ± 0.09	0.61 ± 0.07	0.61 ± 0.07
RA-MDRM	0.59 ± 0.08	0.66 ± 0.11	**0.67** ± 0.08	0.64 ± 0.10
LEA	0.61 ± 0.09	**0.67** ± 0.11	**0.67** ± 0.09	0.65 ± 0.09
TSA	0.61 ± 0.09	**0.67** ± 0.11	**0.67** ± 0.10	0.64 ± 0.09
RPA	0.56 ± 0.07	**0.60** ± 0.10	0.55 ± 0.07	**0.60** ± 0.08
MEKT	0.54 ± 0.06	0.55 ± 0.10	**0.57** ± 0.09	0.54 ± 0.08

Filtering improves accuracy in both datasets. In 2a 3, the 'Standard' 8–30 Hz band boosts performance over 'None' (EA 0.53→0.65, LEA 0.60→0.67). EA, RA-MDRM, and LEA also benefit from the 'Beta' band, while 'Alpha' is consistently weaker, confirming beta activity as most discriminative. CSP+LDA peaks with 'Standard,' while MEKT and RPA remain lowest overall. In 2b 4, gains are smaller due to fewer channels and simpler 2-class discrimination. EA (0.70) again leads, followed by LEA, TSA, and RA-MDRM. CSP+LDA shows robustness, with 'None' (0.60) ≈ 'Beta' (0.61), still surpassing MEKT and RPA.

6 Discussion

This work contributes to EEG-based motor imagery BCIs by establishing a standardized evaluation protocol and open codebase for comparing key transfer learning techniques. Comparative analysis on BCI Competition IV 2a and 2b highlights distinct performance profiles: EA excelled in cross-session tasks, showing that Euclidean alignment can be both competitive and efficient, while MEKT performed best in cross-subject transfer, reflecting its sophisticated distribution alignment. These results emphasize that different strategies suit different domain shifts. Ablation studies showed performance peaks when using longer MI windows (0.5–3.5s post-cue) and standard mu/beta filtering (8–30Hz), which improves SNR and accuracy across methods. By releasing implementations, we directly address reproducibility and enable fair comparisons. Few limitations remain. Findings are based on two standard yet widely used datasets (2a, 2b); broader validation on datasets with different channels, subjects, equipment, or paradigms (e.g., ERP, SSVEP) is needed. Algorithm selection covered key Euclidean, Riemannian, and tangent space methods, but other approaches may yield different trade-offs. Evaluation was offline only; real-time BCIs pose additional challenges such as computational constraints and online adaptation, not considered here. Finally, we focused solely on cross-session and cross-subject transfer as the most practically relevant scenarios, leaving metrics like computational cost outside the scope of this paper.

References

[1] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012. doi: 10.1109/TBME.2011.2172210.

[2] Herbert Ramoser, Johannes Muller-Gerking, and Gert Pfurtscheller. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE transactions on rehabilitation engineering*, 8(4):441–446, 2000.

[3] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[4] Shang-Lin Wu, Chun-Wei Wu, Nikhil R. Pal, Chih-Yu Chen, Shi-An Chen, and Chin-Teng Lin. Common spatial pattern and linear discriminant analysis for motor imagery classification. In *2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, pages 146–151, 2013. doi: 10.1109/CCMB.2013.6609178.

[5] He He and Dongrui Wu. Transfer learning for brain-computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2020.

[6] Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, and Yannick Berthoumieu. Transfer learning: a riemannian geometry framework with applications to brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, 2018.

[7] Alexandre Bleuze, J'er'emie Mattout, and Marco Congedo. Tangent space alignment: Transfer learning for brain-computer interface. *Frontiers in Human Neuroscience*, 16:1049985, 2022. doi: 10.3389/fnhum.2022.1049985.

[8] Pedro L. C. Rodrigues, Christian Jutten, and Marco Congedo. Riemannian procrustes analysis: Transfer learning for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 66(8):2390–2401, 2018.

[9] Dongrui Wu, Yifan Xu, and Bao-Liang Lu. Manifold embedded knowledge transfer for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28 (5):1117–1127, 2020.

[10] Vinay Jayaram and Alexandre Barachant. Moabb: trustworthy algorithm benchmarking for bcis. *Journal of Neural Engineering*, 15(6):066011, sep 2018. doi: 10.1088/1741-2552/aadea0. URL `https://dx.doi.org/10.1088/1741-2552/aadea0`.

[11] Sylvain Chevallier, Igor Carrara, Bruno Aristimunha, Pierre Guetschel, Sara Sedlar, Bruna Lopes, Sebastien Velut, Salim Khazem, and Thomas Moreau. The largest eeg-based bci reproducibility study for open science: the moabb benchmark, 2024. URL `https://arxiv.org/abs/2404.15319`.

[12] Wei Xiong, Jiangtong Li, Jie Li, and Kun Zhu. Eeg-fm-bench: A comprehensive benchmark for the systematic evaluation of eeg foundation models, 2025. URL `https://arxiv.org/abs/2508.17742`.

[13] Clemens Brunner, Robert Leeb, and Gernot Müller-Putz. Bci competition 2008–graz data set a, 2024. URL `https://dx.doi.org/10.21227/katb-zv89`.

[14] Robert Leeb, Clemens Brunner, G Müller-Putz, A Schlögl, and GJGUOT Pfurtscheller. Bci competition 2008–graz data set b. *Graz University of Technology, Austria*, 16:1–6, 2008.

Appendix A.

BCI Competition IV 2a dataset: This dataset comprises EEG recordings from 9 subjects performing four distinct motor imagery tasks (left hand, right hand, both feet, and tongue movements) across two sessions held. Each session contained 6 runs of 48 trials (12 per class), totaling 288 trials per session. During the experimental paradigm, subjects seated in an armchair viewed a fixation cross $(t = 0s)$, heard a warning tone, then saw a directional arrow cue $(t = 2s)$ indicating which motor imagery task to perform until the cross disappeared $(t = 6s)$, with no feedback provided. The data was sampled at 250 Hz, bandpass-filtered between 0.5-100 Hz, and an additional 50 Hz notch filter to eliminate line noise.

BCI Competition IV 2b dataset: This dataset (BCIC IV 2b) comprises EEG recordings from 9 right-handed subjects performing two distinct motor imagery tasks (left hand vs. right hand movements) across five sessions(we used the first two sessions). Each screening session contained six runs of ten trials each (five per class), totaling 60 trials per session. During the experimental paradigm, subjects seated in an armchair viewed a fixation cross, heard a warning tone $(t = 2s)$, then saw a directional arrow cue (left or right) presented for 1.25 seconds (starting $t = 3s$), followed by a 4-second motor imagery period with no feedback provided. The data from 3 bipolar EEG channels (C3, Cz, C4) was sampled at 250 Hz, bandpass-filtered between 0.5-100 Hz, and included an additional 50 Hz notch filter to eliminate line noise.

Code: Anonymous Github

Table 5: Comparison of CSP+LDA and recreated accuracy per subject

Subject	**CSP+LDA**	**Ours**
S1	0.78	0.80
S2	0.45	0.53
S3	0.82	0.85
S4	0.59	0.57
S5	0.40	0.42
S6	0.50	0.54
S7	0.81	0.79
S8	0.69	0.81
S9	0.77	0.82
Mean	0.65 ± 0.17	0.68 ± 0.18

Table 6: Comparison of MDRM baseline with recreated accuracy scores.

Subject	**MDRM**	**Recreated Accuracy**
Subject 1	0.78	0.78
Subject 2	0.44	0.47
Subject 3	0.77	0.78
Subject 4	0.55	0.58
Subject 5	0.44	0.41
Subject 6	0.47	0.48
Subject 7	0.72	0.76
Subject 8	0.75	0.75
Subject 9	0.77	0.76
Mean	**0.63 ± 0.15**	**0.64**

Table 7: Comparison of expected accuracy (RA-MDRM) and recreated method's accuracy across subjects.

Subject	RA-MDRM	RA-MDRM (ours)
1	0.79	0.77
2	0.54	0.52
3	0.77	0.80
4	0.54	0.62
5	0.46	0.48
6	0.45	0.51
7	0.76	0.78
8	0.79	0.79
9	0.75	0.73
Mean	**0.65**	**0.66**

Table 8: LEA - MDRM - Affine Transformed Accuracy Across Subjects

Subject	Accuracy
Subject 1	0.7951
Subject 2	0.5243
Subject 3	0.8056
Subject 4	0.6181
Subject 5	0.4688
Subject 6	0.5174
Subject 7	0.7847
Subject 8	0.8021
Subject 9	0.7292
Mean	**0.6717**
Expected Mean	**0.66**

Table 9: Expected accuracy vs. recreated test accuracy - EA method

Subject	Expected Accuracy	EA (ours)
Subject 1	0.88	0.84
Subject 2	0.56	0.53
Subject 3	0.99	0.94
Subject 4	0.74	0.74
Subject 5	0.50	0.52
Subject 6	0.65	0.72
Subject 7	0.69	0.74
Subject 8	0.90	0.88
Subject 9	0.73	0.78
Mean	**0.74**	**0.75**

Table 10: Per-subject Cross-Session accuracy and mean accuracy across 9 subjects - RPA method.

Subject	Accuracy
Subject 1	0.7500
Subject 2	0.5172
Subject 3	0.8276
Subject 4	0.5431
Subject 5	0.5776
Subject 6	0.5948
Subject 7	0.7241
Subject 8	0.7328
Subject 9	0.8621
Mean Accuracy	**0.6810**
Trimmed Mean Accuracy (Excl. 4 Lowest)	**0.7793**
Expected Trimmed Mean Accuracy	**0.7848**

Table 11: Per-subject cross-session accuracy and mean accuracy across 9 subjects - TSA Method.

Subject	Accuracy
Subject 1	0.7069
Subject 2	0.5000
Subject 3	0.9310
Subject 4	0.6638
Subject 5	0.4914
Subject 6	0.5000
Subject 7	0.6121
Subject 8	0.9655
Subject 9	0.9397
Mean Accuracy	**0.7011**
Trimmed Mean Accuracy (Excl. 2 Lowest)	**0.7814**
Expected Trimmed Mean Accuracy	**0.7856**

Table 12: Per-subject accuracy and mean accuracy across 9 subjects - MEKT Method

Subject	Accuracy
Subject 1	0.8681
Subject 2	0.5000
Subject 3	0.9514
Subject 4	0.7431
Subject 5	0.5833
Subject 6	0.6944
Subject 7	0.6389
Subject 8	0.9444
Subject 9	0.8056
Mean Accuracy	**0.7477**
Expected Accuracy	**0.7631**

Table 13: Comparison of 2-Class Test Accuracies - Cross-Session: Dataset 2a

Subject	EA	CSP+LDA	MDRM	RA-MDRM	LEA	TSA	RPA	MEKT
Subject 1	0.90	0.90	**0.91**	0.86	0.87	0.87	0.81	0.51
Subject 2	0.59	0.53	0.55	0.59	0.56	0.58	0.57	**0.65**
Subject 3	**0.99**	0.92	0.87	0.98	0.98	0.97	0.91	0.97
Subject 4	0.72	0.72	**0.78**	0.77	**0.78**	0.75	0.58	0.55
Subject 5	0.59	0.58	0.57	0.51	0.51	0.48	0.51	**0.65**
Subject 6	0.72	**0.74**	0.68	0.70	0.69	0.66	0.63	0.44
Subject 7	**0.84**	0.76	0.55	0.72	0.76	0.69	0.74	0.81
Subject 8	0.97	0.97	**0.99**	0.96	0.97	0.97	0.82	0.97
Subject 9	0.87	0.84	**0.92**	0.90	**0.92**	0.91	0.88	0.54
Mean ± SD	**0.80±0.14**	0.77±0.14	0.76±0.16	0.78±0.15	0.78±0.16	0.76±0.17	0.72±0.14	0.68±0.19

Table 14: Comparison of 4-Class Test Accuracies - Cross-Session: Dataset 2a

Subject	EA	CSP+LDA	MDRM	RA-MDRM	LEA	TSA	RPA	MEKT
Subject 1	0.76	0.78	**0.81**	0.77	0.80	0.78	0.63	0.21
Subject 2	0.52	0.48	0.51	0.55	0.56	0.50	0.38	**0.67**
Subject 3	0.77	0.68	0.70	0.79	0.81	0.80	0.63	**0.89**
Subject 4	0.66	0.62	**0.67**	0.66	0.64	0.60	0.56	0.23
Subject 5	0.38	0.33	0.44	0.42	0.42	**0.47**	0.38	0.43
Subject 6	**0.51**	**0.51**	0.49	0.48	0.48	0.43	0.42	0.23
Subject 7	0.80	0.62	0.60	0.76	0.77	0.69	0.61	**0.84**
Subject 8	0.81	0.82	0.74	0.77	0.80	0.73	0.68	**0.91**
Subject 9	0.67	0.64	**0.76**	**0.76**	0.73	0.73	0.66	0.22
Mean ± SD	0.65±0.14	0.61±0.14	0.64±0.12	0.66±0.14	**0.67±0.14**	0.64±0.13	0.55±0.12	0.51±0.29

Table 15: Comparison of 2-Class Test Accuracies - Cross subject: Dataset 2a

Subject	EA	CSP+LDA	MDRM	RA-MDRM	LEA	TSA	RPA	MEKT
Subject 1	0.65	0.72	0.62	0.72	0.73	0.64	0.49	**0.87**
Subject 2	**0.59**	0.54	0.49	0.54	0.56	0.43	0.46	0.50
Subject 3	0.92	0.92	0.63	0.81	0.83	0.79	0.62	**0.95**
Subject 4	0.72	0.66	0.67	0.62	0.69	0.65	0.55	**0.74**
Subject 5	0.53	0.44	0.50	0.62	**0.65**	0.44	0.53	0.58
Subject 6	0.68	0.55	0.50	0.69	**0.70**	0.69	0.47	0.69
Subject 7	0.65	0.60	0.53	0.59	0.66	**0.71**	0.51	0.64
Subject 8	0.83	0.91	0.60	0.87	0.86	0.84	0.58	**0.94**
Subject 9	0.80	0.69	0.57	0.79	0.78	0.76	0.60	**0.81**
Mean ± SD	0.71±0.11	0.67±0.15	0.57±0.06	0.70±0.11	0.72±0.09	0.66±0.14	0.53±0.05	**0.75±0.15**

Table 16: Comparison of 4-Class Test Accuracies - Cross subject: Dataset 2a

Subject	EA	CSP+LDA	MDRM	RA-MDRM	LEA	TSA	RPA	MEKT
Subject 1	0.69	0.55	0.53	0.64	0.64	0.64	0.38	**0.72**
Subject 2	**0.27**	0.24	0.24	0.23	0.26	0.20	0.25	0.23
Subject 3	0.74	0.49	0.55	0.62	0.67	0.68	0.34	**0.80**
Subject 4	0.42	0.33	0.34	0.41	0.41	0.29	0.31	**0.45**
Subject 5	0.30	0.24	0.25	**0.37**	0.35	0.32	0.24	0.33
Subject 6	**0.42**	0.29	0.25	0.32	0.37	0.35	0.29	0.40
Subject 7	0.60	0.34	0.33	0.38	**0.61**	0.41	0.34	0.56
Subject 8	0.69	0.59	0.26	0.69	0.70	0.62	0.33	**0.78**
Subject 9	0.62	0.56	0.25	0.61	0.61	0.57	0.40	**0.65**
Mean ± SD	0.53±0.17	0.40±0.14	0.33±0.12	0.48±0.15	0.51±0.15	0.45±0.17	0.32±0.05	**0.55±0.19**

Table 17: Comparison of 2-Class Test Accuracies - Cross-Session: Dataset 2b

Subject	EA	CSP+LDA	MDRM	RA-MDRM	LEA	TSA	RPA	MEKT
Subject 1	0.69	0.57	0.52	0.69	0.68	**0.75**	0.60	0.55
Subject 2	0.56	0.54	0.53	0.52	**0.58**	0.39	0.57	0.47
Subject 3	0.58	**0.60**	0.57	0.58	0.57	**0.60**	0.44	0.46
Subject 4	**0.90**	0.89	0.82	0.84	0.84	0.85	0.78	0.58
Subject 5	0.83	0.63	0.65	**0.84**	**0.84**	0.81	0.63	0.49
Subject 6	**0.79**	**0.79**	0.70	0.69	0.75	0.75	0.74	0.77
Subject 7	0.58	0.59	**0.62**	0.61	0.58	0.55	0.51	0.52
Subject 8	**0.67**	0.64	0.61	0.53	0.61	0.66	0.52	0.65
Subject 9	**0.68**	**0.68**	0.60	0.60	0.59	0.62	0.51	0.47
Mean ± SD	**0.70±0.11**	0.66±0.11	0.62±0.09	0.66±0.11	0.67±0.11	0.67±0.14	0.59±0.11	0.55±0.10

Table 18: Comparison of 2-Class Test Accuracies - Cross-Subject: Dataset 2b

Subject	EA	CSP+LDA	MDRM	RA-MDRM	LEA	TSA	RPA	MEKT
Subject 1	0.77	0.75	**0.78**	0.72	**0.78**	0.76	0.56	**0.78**
Subject 2	0.64	0.60	0.62	0.60	0.65	**0.66**	0.57	**0.66**
Subject 3	0.59	0.54	0.56	0.59	**0.61**	0.60	0.50	0.58
Subject 4	**0.87**	0.60	0.64	0.86	**0.87**	0.85	0.73	0.81
Subject 5	0.66	0.57	0.57	0.64	0.65	**0.67**	0.52	0.63
Subject 6	0.74	0.62	0.68	0.68	0.71	0.66	0.70	**0.75**
Subject 7	0.58	0.58	0.57	**0.62**	0.61	0.59	0.54	**0.62**
Subject 8	0.55	0.56	0.57	0.59	0.54	0.56	**0.63**	0.57
Subject 9	0.64	0.63	0.59	0.59	0.62	0.60	0.44	**0.65**
Mean ± SD	**0.67±0.10**	0.61±0.06	0.62±0.07	0.66±0.08	**0.67±0.09**	0.66±0.09	0.58±0.09	**0.67±0.08**