2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Is Riemannian Geometry Better than Euclidean in Averaging Covariance Matrices for CSP-based Subject-Independent Classification of Motor Imagery?

Yassawe Kainolda[1], Berdakh Abibullaev[2], Reza Sameni[3], Amin Zollanvari[1]

Abstract— **Common Spatial Pattern (CSP) is a popular feature extraction algorithm used for electroencephalogram (EEG) data classification in brain-computer interfaces. One of the critical operations used in CSP is taking the average of trial covariance matrices for each class. In this regard, the arithmetic mean, which minimizes the sum of squared Euclidean distances to the data points, is conventionally used; however, this operation ignores the Riemannian geometry in the manifold of covariance matrices. To alleviate this problem, Fréchet mean determined using different Riemannian distances have been used. In this paper, we are primarily concerned with the following question: Does using the Fréchet mean with Riemannian distances instead of arithmetic mean in averaging CSP covariance matrices improve the subject-independent classification of motor imagery (MI)? To answer this question we conduct a comparative study using the largest MI dataset to date, with 54 subjects and a total of 21,600 trials of left- and right-hand MI. The results indicate a general trend of having a statistically significant better performance when the Riemannian geometry is used.**

I. INTRODUCTION

Brain-Computer Interface (BCI) systems aim to infer different brain activity patterns of a user as accurately as possible and translate them into appropriate commands [1].

Most studies use electroencephalography (EEG) signals generated by the collective action of millions of cortical cells in the development of BCI systems. The key component of such BCI systems are the signal processing and classification methods that allow the extraction of discriminative features of EEG in which human thoughts are best encoded. Although there has been significant progress in the field, most studies still focus on subject-dependent classification of user mental states from EEG data [2], [3]. A subject-independent classification of users' mental states remains one of the major challenges in designing practical BCIs. This is due to enormous variability in data; in other words, EEG distribution is highly variable between different users, experimental sessions, and even between different trials [4]. As a result, a classifier model trained on one subject's data will not generalize to the data of another subject. Such variability dramatically affects most BCI performances and requires tedious subject-specific calibration and classifier modeling. This study focuses on the problem of subject-independent classification of EEG recordings acquired from motor imagery (MI) experiments.

Common Spatial Pattern (CSP) is one of the most popular feature extraction methods for EEG data, which creates a set of spatial filters that transform the data to be more discriminative in terms of variances [5], [6]. It is instrumental in the MI paradigm, much of useful EEG features in the spatial domain, since brain areas responsible for different limb movements are spatially separated.

CSP works by simultaneously diagonalizing the average covariance matrices of trials of each class. Most commonly, average covariance matrices are estimated by arithmetic averaging. Arithmetic averaging, however, suffers from a series of problems, the most significant of which is the swelling effect [7], resulting in the determinant of the average covariance matrix (which is a measure of the overall scatter of the samples in the feature space), being much larger than determinants of individual matrices that comprised the average. Such a distortion is an undesirable artifact of the computation. Since covariance matrices are by construction symmetric positive definite (SPD) and therefore lie on an SPD manifold, their underlying geometry is better captured by Riemannian methods [8], [9]. This work applies that reasoning and compares the results produced by CSP with arithmetic averaging and CSP with Riemannian means in the subject-independent classification framework.

The rest of the paper is organized as follows: in Section II the relevant concepts of the Common Spatial Pattern algorithm and Riemannian geometry will be reviewed, in Section III the methodology of the experiment will be explained, and Section IV presents the results of the experiment.

II. BACKGROUND

A. Common Spatial Pattern

Let $\mathbf{X}_l$ denote the l^{th} trial data matrix of size $C \times N$ where C is the number of EEG channels and N is the number of observations. The spatial covariance matrix for each trial is then

$$\mathbf{C}_l = \frac{\mathbf{X}_l \mathbf{X}_l^T}{\text{tr}\left(\mathbf{X}_l \mathbf{X}_l^T\right)}, \tag{1}$$

*The work was partially supported by the Nazarbayev University Faculty Development Competitive Research Grants Program under grant number 021220FD1151.

[1]Yassawe Kainolda and Amin Zollanvari are with the Electrical and Computer Engineering Department, School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan {yassawe.kainolda, amin.zollanvari}@nu.edu.kz

[2]Berdakh Abibullaev is with the Robotics and Mechatronics Department, School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan berdakh.abibullaev@nu.edu.kz

[3]Reza Sameni is with the Department of Biomedical Informatics, School of Medicine, Emory University, GA, USA rsameni@dbmi.emory.edu

where T and tr(.) denote the transpose and trace operators, respectively. In the case of a two-class paradigm, CSP performs a linear transformation that maps the data from the original sensor space to a new surrogate feature space, in which the variance of one class is maximized while the variance of the other is minimized [10]. To achieve this transformation, CSP finds a set of M spatial filters given by an $C \times M$ matrix $\mathbf{W}$ to linearly transform an input signal. Finding $\mathbf{W}$ involves solving the generalized eigenvalue problem, which involves the class-specific "average" covariance matrices given by the usual arithmetic mean [10]:

$$\mathbf{\Sigma}_{i,a} = \frac{1}{T_i} \sum_{l=1}^{T_i} \mathbf{C}_l, \tag{2}$$

where $i = 1, 2$ denote class indexes, T_i is the number of trials in class i and the subscript "a" is used to denote the use of arithmetic mean. An input data matrix $\mathbf{X}$ is then transformed as

$$\mathbf{S} = \mathbf{W}_{\text{cropped}}^T \mathbf{X}, \tag{3}$$

where $\mathbf{W}_{\text{cropped}}$ is a matrix of the shape $C \times (2n)$, which in order to achieve the most separability it consists of only the first and the last n columns of $\mathbf{W}$ [10]—it is common to choose a small value of n (e.g., $n = 3$). The feature vector $f = [f_1, f_2, , ..., f_{2n}]$ is obtained as :

$$f_r = \log \frac{\text{var}(s_r)}{\sum_{k=1}^{2n} \text{var}(s_k)}, \tag{4}$$

where var$(\cdot)$ denotes the variance and s_r is r^{th} row of $\mathbf{S}$, $r = 1, ..., 2n$. This feature vector is then used as the input to a classifier of choice.

B. Riemannian Mean

The use of Euclidean geometry in CSP is manifested through the arithmetic mean (2) [11]. However, employing Euclidean geometry in manipulating SPD covariance matrices has multiple drawbacks: 1) it forms a non-complete space; that is, Euclidean computations in general may lead to indefinite matrices [12]; 2) arithmetic mean is not invariant to matrix inversion, which is an undesirable property in some applications [13]; and 3) it leads to the well-known swelling effect, which implies the determinant of the arithmetic mean matrix may become larger than the original determinants [7] – in data analysis this may add harmful spurious variations to the data [12].

The aforementioned issues are due to the application of Euclidean geometry in the space of SPD matrices, which is a differentiable Riemannian manifold [14]. To overcome difficulties associated with the arithmetic mean of SPDs in (2), one may use instead the Fréchet mean (also known as Riemannian mean) defined as [7]

$$\mathbf{\Sigma}_{i,r} = \underset{\mathbf{C} \in P(C)}{\text{argmin}} \sum_{l=1}^{T_i} \delta_r^2(\mathbf{C}, \mathbf{C}_l)\,, \tag{5}$$

where $P(C)$ is the set of all $C \times C$ SPD matrices, $\mathbf{C}_1, ..., \mathbf{C}_{T_i}$ is the set of given SPD covariance matrices for T_i trials, $\delta_r(\mathbf{A}, \mathbf{B})$ is a Riemannian distance between two SPD matrices $\mathbf{A}$ and $\mathbf{B}$, and the subscript "r" is used to denote the use of Riemannian mean. Depending on the choice of scalar multiplication to equip the tangent spaces, two Riemannian distances are commonly used: 1) the Log-Euclidian metric (LEM) [7]; and 2) the Affine-Invariant Riemannian metric (AIRM) [15].

Formally, for two SPD matrices $\mathbf{A}$ and $\mathbf{B}$, LEM distance $\delta_{\text{LEM}}(\mathbf{A}, \mathbf{B})$ is given by [7]:

$$\delta_{\text{LEM}}(\mathbf{A}, \mathbf{B}) = \| \log(\mathbf{A}) - \log(\mathbf{B}) \|_{\mathcal{F}}, \tag{6}$$

where $\log(\mathbf{A})$ is the matrix logarithm, $\|\mathbf{A}\|_{\mathcal{F}}$ is the Frobenius norm of $\mathbf{A}$ defined as $\sqrt{\text{tr}(\mathbf{A}\mathbf{A}^T)}$. Using LEM distance leads to a closed-form solution for calculating the mean as follows [7]:

$$\mathbf{\Sigma}_{i,\text{LEM}} = \exp(\frac{1}{T_i} \sum_{l} \log(\mathbf{C}_l)), \tag{7}$$

where $\exp(\cdot)$ denotes the matrix exponential. AIRM distance, on the other hand, is defined as [16]

$$\delta_{\text{AIRM}}(\mathbf{A}, \mathbf{B}) = \| \log(\mathbf{A}^{-1/2} \mathbf{B} \mathbf{A}^{-1/2}) \|_F. \tag{8}$$

In contrast with the LEM distance, the Fréchet mean determined using AIRM distance has no closed-form solution. In this regard, optimization methods such as incremental Fréchet mean estimation can be used to find the mean [17].

III. Material and Methods

A. Dataset

The dataset consists of EEG recordings of 54 subjects recorded using 62 Ag/AgCl electrodes with a sampling rate of 1000 Hz. Electrodes were placed according to the international 10-20 system [18]. The dataset contains two motor imagery classes: 1) imagined movement of the left hand; and 2) imagined movement of the right hand. Furthermore, the dataset is organized in two sessions, where each session contains 200 trials for each class and subject. For more information on the experimental protocol, the readers are referred to [18].

There are in total 21600 trials collected across 54 subjects in this dataset, which to the best of our knowledge makes it the largest publicly available EEG MI dataset to date. For the sake of comparison, there are a total of 5184 trials across 9 subjects in BCI Competition IV 2a [19] and 10400 trials across 52 subjects in [20] datasets, which have been frequently used in EEG-based MI studies before [21], [22], [23].

In our simulations, we used the 20 channels from the motor cortex region that were used in [18] (FC-5/3/1/2/4/6, C-5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6). The placement of these channels is shown in Fig. 1. The data was further bandpass filtered within 8-30 Hz using a fifth order Butterworth filter.

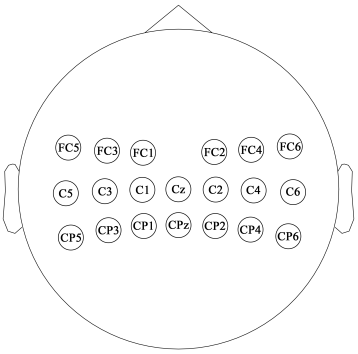

Fig. 1. Electrode placement of the 20 used channels.

B. Methods

Depending on the approach used to compute the class-specific average covariance matrices, we use and compare the following three types of CSP feature extraction: I) CSP with $\mathbf{\Sigma}_{i,a}$ computed from (2); II) CSP with $\mathbf{\Sigma}_{i,\text{LEM}}$ computed from (7); and III) CSP with $\mathbf{\Sigma}_{i,\text{AIRM}}$, which is obtained by using $\delta_{\text{AIRM}}(\mathbf{A},\mathbf{B})$ in (5). Hereafter, we refer to these methods as conventional CSP, LEM-CSP, and AIRM-CSP, respectively, whereas by Riemannian-CSP we refer to both LEM-CSP and AIRM-CSP.

In order to simulate the subject-independent classification, we use a Leave-One-Subject-Out Cross-Validation (LOSO-CV) procedure where we successfully hold out the observations (trials) collected from each subject, construct a classifier using the pooled set of observations from the remaining subjects, and test the performance of the constructed classifier on data from the held-out subject. To construct our classifiers, each training and test observation was transformed into a feature vector using (4) for $n = 2, ..., 5$. These feature vectors were used as the input to linear discriminant analysis (LDA) classifier [24].

IV. Results

Here, we compare the subject-independent classification performance obtained using the Riemannian-CSP with respect to the conventional CSP. In this regard, we first estimate the distributions of the difference between the LOSO-CV classification accuracy obtained using Riemannian-CSP and conventional CSP (i.e., empirical deviation distribution). The deviation distributions provide a comprehensive picture of the conducted comparisons in each setting.

Fig. 2 shows the plots of the empirical deviation distribution across session 1 and 2 for different n. The plots were obtained by fitting a beta density to the raw histograms of classification accuracies estimated using all test observations for each of the 54 subjects used as part of the LOSO-CV procedure. Therefore, the density of a point larger (less) than zero shows the likelihood of an improvement in classification accuracy equivalent to that point obtained using the Riemannian-CSP (conventional CSP) with respect to the conventional CSP (Riemannian-CSP). As a result, the area of the shaded region (the grey region) in each density plot is an estimate of the *probability* of achieving a higher classification accuracy using the Riemannian-CSP with respect to the conventional CSP (we simply refer to this probability as the probability of improvement denoted as P_I). As we can see in Fig. 2, out of 8 scenarios (2 sessions $\times$ 4 values of n) to compare the LEM-CSP and the conventional CSP, in 7 cases $P_I > 54\%$ with 4 cases $P_I > 60\%$. This shows a clear advantage of the LEM-CSP with respect to the conventional CSP. We observe a similar situation when AIRM-CSP is used. In this case, out of 8 similar scenarios, in 7 cases $P_I > 53\%$ with 4 cases $P_I > 60\%$.

For a more conventional comparison, for each scenario we also recorded the the average classification accuracy obtained across all 54 subjects and for session 1 and 2 ($\pm$ standard deviations) in Tables I and II, respectively. As we can see in these tables, both the AIRM-CSP and the LEM-CSP show a larger average accuracy than the conventional CSP in 7 out of 8 scenarios. The results of a paired Wilcoxon signed rank test show that in 7 and 6 out of 8 scenarios, the improvement observed by LEM-CSP and AIRM-CSP, respectively, with respect to the conventional CSP is statistically significant at a significance level of 0.1 (see Table III).

TABLE I
AVERAGE SUBJECT-INDEPENDENT ACCURACIES FOR SESSION 1 ($\pm$ STANDARD DEVIATIONS)

n	Conventional CSP	AIRM CSP	LEM CSP
2	62.04 $\pm$ 12.24 %	65.35 $\pm$ 13.80 %	**65.63** $\pm$ **13.87** %
3	62.51 $\pm$ 13.06 %	**65.56** $\pm$ **13.71** %	65.00 $\pm$ 13.60 %
4	62.60 $\pm$ 13.05 %	**65.31** $\pm$ **13.96** %	65.11 $\pm$ 13.72 %
5	62.83 $\pm$ 12.99 %	**65.26** $\pm$ **14.02** %	64.88 $\pm$ 13.58 %

TABLE II
AVERAGE SUBJECT-INDEPENDENT ACCURACIES FOR SESSION 2 ($\pm$ STANDARD DEVIATIONS)

n	Conventional CSP	AIRM CSP	LEM CSP
2	66.25 $\pm$ 15.62 %	67.05 $\pm$ 15.42 %	**67.19** $\pm$ **15.18** %
3	66.63 $\pm$ 15.72 %	**67.26** $\pm$ **15.34** %	67.34 $\pm$ 15.08 %
4	66.05 $\pm$ 15.64 %	**67.16** $\pm$ **14.95** %	67.01 $\pm$ 14.73 %
5	**67.76** $\pm$ **15.24** %	67.07 $\pm$ 14.78 %	66.25 $\pm$ 14.92 %

TABLE III
OUTPERFORMANCE P-VALUES OF RIEMANNIAN VS CONVENTIONAL CSPS, FOR SESSIONS 1 AND 2

n	Session 1		Session 2	
	AIRM CSP	LEM CSP	AIRM CSP	LEM CSP
2	< 0.001	< 0.001	0.041	0.013
3	< 0.001	0.002	0.136	0.089
4	0.006	0.005	0.078	0.097
5	0.004	0.018	0.595	0.955

V. Conclusion

In this work, we compared the effect of using the Fréchet mean with Riemannian distances with respect to the arithmetic mean in averaging spatial covariance matrices on the performance of CSP-based subject-independent classifiers of

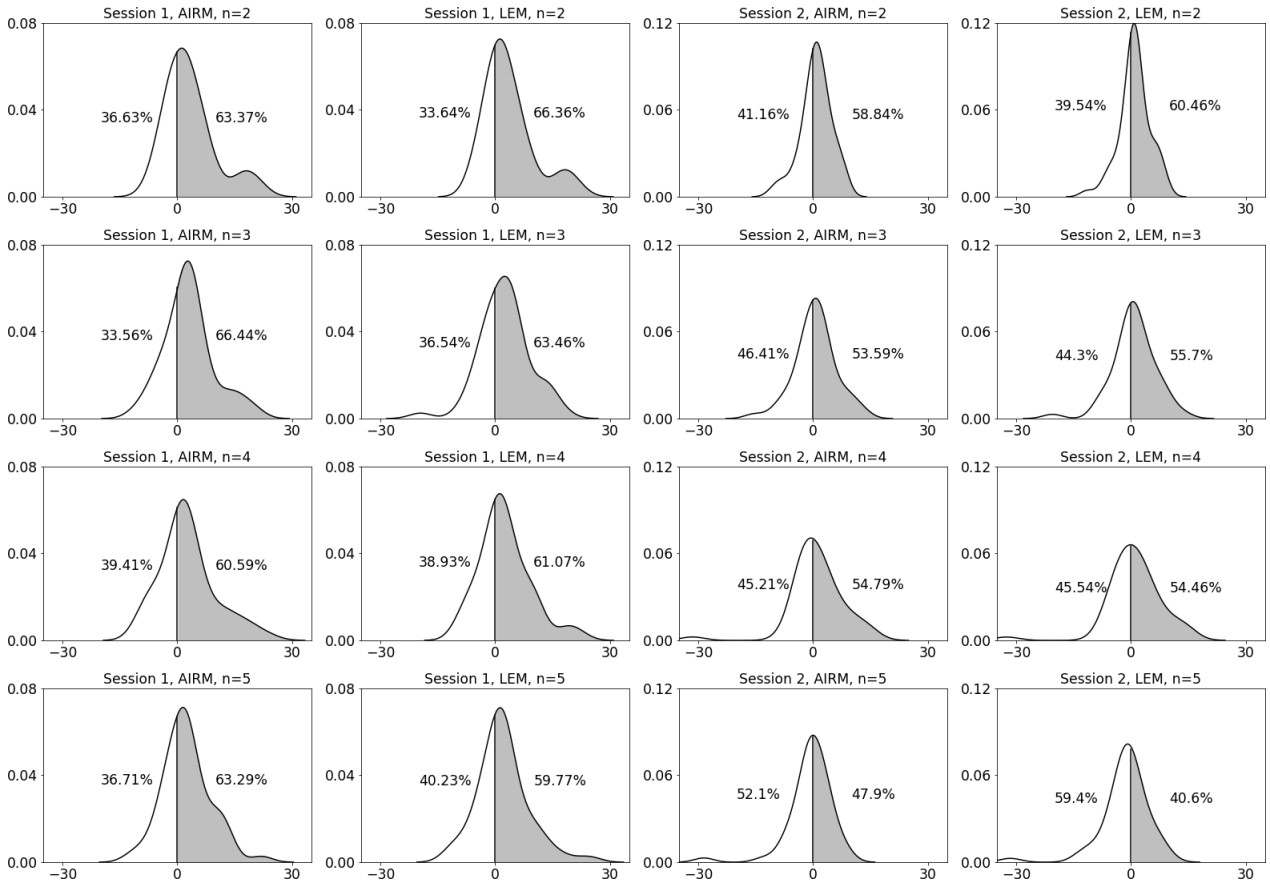

Fig. 2. Empirical distribution of the difference between classification accuracies achieved by Riemannian CSP and conventional CSP for Session 1 (from left to right, columns 1 and 2) and Session 2 (columns 3 and 4). The areas of the shaded and unshaded regions are shown next to the regions. The shaded region in each plot shows the probability of achieving a higher classification accuracy obtained using the Riemannian-CSP with respect to the conventional CSP.

motor imagery. Using the largest MI dataset collected to date, we showed that for a common number of CSP filters, using the Riemannian mean generally leads to a higher subject-independent classification accuracy with respect to the arithmetic mean. However, in this study we did not examine the effect of dimensionality (number of channels) on the joint behaviour of Riemannian mean and the classification performance. It has been previously observed that in subject-dependent classification, having a large number of channels is not in favour of Riemannian CSP [25]. Other aspects such as data completion and robustness to corrupted/missing data segments also remain as an open problems [26]. We leave the investigation of these effects in the subject-independent context for future studies.

REFERENCES

[1] M. Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, and *et al.*, "The brain-computer interface cycle," *J. Neural. Eng.*, vol. 6, no. 4, pp. 1–9, 2009.

[2] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, Jun. 2007.

[3] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *J. Neural Eng.*, vol. 15, no. 3, p. 031005, 2018.

[4] S. Saha and M. Baumert, "Intra-and inter-subject variability in EEG-based sensorimotor brain computer interface: a review," *Frontiers in computational neuroscience*, vol. 13, p. 87, 2020.

[5] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2007.

[6] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 2008, pp. 2390–2397.

[7] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Geometric means in a novel vector space structure on sysmetric positive-definite matrices," *SIAM J. Matrix Analysis Applications*, vol. 29, pp. 328–347, 01 2006.

[8] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to BCI classification," in *Latent Variable Analysis and Signal Separation*. Springer Berlin Heidelberg, 2010, pp. 629–636. [Online]. Available: https://doi.org/10.1007/978-3-642-15995-4_78

[9] M. Congedo, P. L. C. Rodrigues, and C. Jutten, "The Riemannian minimum distance to means field classifier," in *8th Graz Brain-Computer Interface Conference 2019*, 2019.

[10] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, pp. 441–6, 01 2001.

[11] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Common spatial pattern revisited by Riemannian geometry," in *2010 IEEE International Workshop on Multimedia Signal Processing*, 2010, pp. 472–476.

[12] F. Yger, M. Berar, and F. Lotte, "Riemannian Approaches in Brain-

Computer Interfaces: A Review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1753–1762, 2017.

[13] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Fast and simple calculus on tensors in the log-euclidean framework," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, J. S. Duncan and G. Gerig, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 115–122.

[14] M. Moakher and M. Zerai, "The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 171–187, 06 2011.

[15] P. T. Fletcher and S. Joshi, "Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2004, pp. 87–98. [Online]. Available: https://doi.org/10.1007/978-3-540-27816-0_8

[16] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Analysis Applications*, vol. 26, pp. 735–747, 01 2005.

[17] G. Cheng, J. Ho, H. Salehian, and B. C. Vemuri, "Recursive Computation of the Fréchet Mean on Non-positively Curved Riemannian Manifolds with Applications," in *Riemannian Computing in Computer Vision*. Springer International Publishing, 2016, pp. 21–43. [Online]. Available: https://doi.org/10.1007/978-3-319-22957-7_2

[18] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy," *GigaScience*, vol. 8, 01 2019.

[19] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–graz dataset A," *Inst. Knowl. Discov. Graz Univ. Technol.*, vol. 16, pp. 1–6, 2008.

[20] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. Jun, "EEG datasets for motor imagery brain computer interface," *GigaScience*, vol. 6, 05 2017.

[21] X. Zhao, J. Zhao, C. Liu, and W. Cai, "Deep neural network with joint distribution matching for cross-subject motor imagery brain-computer interfaces," *BioMed Research International*, vol. 2020, pp. 1–15, 02 2020.

[22] D. Li, J. Wang, J. Xu, and X. Fang, "Densely Feature Fusion Based on Convolutional Neural Networks for Motor Imagery EEG Classification," *IEEE Access*, vol. 7, pp. 132 720–132 730, 2019.

[23] S. Kumar, A. Sharma, and T. Tsunoda, "Brain wave classification using long short-term memory network based optical predictor," *Scientific Reports*, vol. 9, 06 2019.

[24] T. Anderson, "Classification by multivariate analysis," *Psychometrika*, vol. 16, no. 1, pp. 31–50, Mar 1951. [Online]. Available: https://doi.org/10.1007/BF02313425

[25] F. Yger, F. Lotte, and M. Sugiyama, "Averaging covariance matrices for EEG signal classification based on the CSP: An empirical study," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, 2015, pp. 2721–2725.

[26] P. L. C. Rodrigues, M. Congedo, and C. Jutten, "A data imputation method for matrices in the symmetric positive definite manifold," in *XXVIIème colloque GRETSI (GRETSI 2019)*, 2019.