Towards Classification of Covariance Matrices via Bures-Wasserstein-Based Machine Learning

Michael Zirpoli*
Department of Mathematics and Statistics
Auburn University
Auburn, AL, USA
maz0020@auburn.edu

Yuyan Yi*
Department of Mathematics and Statistics
Auburn University
Auburn, AL, USA
yzy0080@auburn.edu

Shu-Chin Lin
Institute of Statistics and Data Science
National Taiwan University
Taipei, Taiwan
scllin@ntu.edu.tw

Linqiang Ge
TSYS School of Computer Science
Columbus State University
Columbus, GA, USA
ge_linqiang@columbusstate.edu

Jingyi Zheng†
Department of Mathematics and Statistics
Auburn University
Auburn, AL, USA
jzz0121@auburn.edu

Abstract

Spatial-temporal data is a prevalent data type in biomedical domains, encompassing instances like multi-channel EEG and fMRI. In the analysis of such data, the connectivity matrix (e.g., functional connectivity derived from fMRI, covariance matrix derived from EEG) is widely extracted and analyzed. Rather than analyzing these matrices within the Euclidean space, this paper considers each matrix as a point situated on the manifold of positive semi-definite (PSD) matrices coupled with Bures-Wasserstein (BW) metric. Within this framework, two machine learning models based on the BW metric are proposed for the classification of PSD matrices on the manifold. Specifically, projection map techniques, based on the BW metric, have been introduced and integrated into machine learning models such as support vector machines and random forest. In comparison with Euclidean methods, our approach considers the geometry of the Riemannian manifold where PSD matrices reside. Moreover, compared with prevalent Affine-Invariant (AI) metrics, our framework does not require matrix regularization and is computationally efficient. To comprehensively evaluate the proposed methods, four fMRI datasets and three brain-computer interface datasets with varying dimensions and quantities have been utilized. The results demonstrate comparable and even superior performance of the proposed methods compared with Euclidean and AI-based approaches.

CCS Concepts

• **Computing methodologies**; • **Machine Learning**; • **Learning paradigms**; • **Supervised learning**; • **Supervised learning by classification**;

Keywords

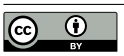
Riemannian manifold, Bures-Wasserstein distance, Fréchet Mean, Support Vector Machine, Random Forest

ACM Reference Format:
Michael Zirpoli, Yuyan Yi, Shu-Chin Lin, Linqiang Ge, and Jingyi Zheng. 2024. Towards Classification of Covariance Matrices via Bures-Wasserstein-Based Machine Learning. In *2024 the 7th Artificial Intelligence and Cloud Computing Conference (AICCC) (AICCC 2024), December 14–16, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3719384.3719386

1 Introduction

Positive semi-definite (PSD) matrices play a pivotal role in addressing various challenges in machine learning, particularly when dealing with high-dimensional data, including metric and kernel learning [1–5], medical diagnosis [6, 7], natural language processing [8–10], computer vision [11, 12], object detection [13, 14], etc. In this study, we focus on the development of machine learning algorithms for PSD matrices classification, and applying it to neuroscience data, including EEG-based Brain-Computer Interface (BCI) [15] and functional Magnetic Resonance Imaging (fMRI) [16] data. BCIs serve as a crucial link between the human brain and external devices by recording brain signals and translating them into commands for devices to execute users' imagined actions. In the analysis of BCIs and fMRI data, vital spatial and temporal characteristics are efficiently represented through covariance or correlation descriptors. In the case of BCIs, covariance matrices are utilized in EEG analysis, while fMRI analysis incorporates functional connectivity matrices expressed through Pearson's correlation and structural connectivity matrices based on the number of fibers connecting Regions of Interest (ROIs). The shared characteristic across these applications is the inherent structure taking the form of a PSD

*Equal Contribution.
†Corresponding author.



AICCC 2024, Tokyo, Japan

ACM ISBN 979-8-4007-1792-5/2024/12
https://doi.org/10.1145/3719384.3719386

matrix, its effectiveness is also highlighted in diverse classification tasks, such as image processing [17], pattern recognition [18], and computer vision [11], etc.

The classification of covariance/correlation matrices typically and initially commences in Euclidean space. In this process, the matrices are first half-vectorized since they are symmetric. Then, they are transformed into concise one-dimensional feature vectors. As the dimensionality of these matrices escalates, feature selection techniques are applied to effectively achieve dimension reduction. Following this, a classification algorithm, such as Support Vector Machines (SVM) or k-Nearest Neighbors (k-NN), is carefully chosen for the training phase. During this training stage, the classifier is fed with the vectorized and normalized covariance matrices, along with their respective labels.

This approach enhances the classifier's capability to discern intricate patterns and relationships within symmetric matrices, thereby establishing a robust foundation for accurate predictions when applied to testing data. However, it is noteworthy that matrices classification in Euclidean space possesses certain limitations, including sensitivity to the dimensionality, the assumption of linearity, and limited flexibility. In response to these constraints, symmetric matrices classification in manifold space, such as the Riemannian manifold, is purposefully designed to mitigate these shortcomings [19–21]. The Riemannian manifold provides a more flexible framework for capturing non-linear relationships, making it particularly suitable for working with these matrices, which naturally reside on such manifolds. This manifold-based approach allows for a more accurate representation of the underlying structure of symmetric matrices, proving especially advantageous in applications like neuroscience data analysis, where relationships among variables are often intricate and non-linear. Therefore, analyzing these matrices in Riemannian space allows for a more accurate characterization of brain connectivity patterns.

Recent research has conducted the symmetric matrices classification on the manifold $\mathbb{P}_n$ of $n \times n$ symmetric positive definite (SPD) matrices using Affine-Invariant (AI) Riemannian metric by treating them as positive definite matrices. Utilizing the AI distance, various statistical and computational methodologies have been developed [22-29] for the comprehensive analysis of covariance and correlation matrices, encompassing tasks such as barycenter estimation and matrix classification. However, calculating the AI distance can be computationally intensive and unstable, especially for large matrices. Moreover, in the case of high-dimensional data, the covariance/correlation matrix often exhibits positive semi-definiteness. The probability of encountering zero eigenvalues significantly rises with the increasing dimensionality of the covariance matrix, rendering the application of AI distance impractical and leading to inaccuracy in the classification of PSD matrices. In contrast, the Bures-Wasserstein (BW) metric naturally handles these challenges without requiring matrix regularization. The BW metric accounts for data variations, ensures stability with ill-conditioned matrices, and maintains computational efficiency without additional perturbations [30]. In this paper, we introduce two machine learning algorithms tailored for the classification of PSD matrices on their manifold, utilizing the BW distance as a key metric. Specifically, we first propose a BW-based classification algorithm by projecting all PSD matrices onto the tangent space, that is computed based on BW metric. Subsequently, considering the variability of feature distributions between BCI sessions, we derive the BW-based adaptive classification algorithm by adding the scalar during projection. We assess the performance of the proposed algorithms and validate the superiority of BW metric using seven real datasets, including fMRI and EEG-based BCI data.

The subsequent sections are structured as follows: In Section 2, we provide some preliminary information about the AI and BW metric and the Fréchet mean algorithm. In Section 3, we propose two BW-based projection algorithms for PSD matrices classification. In Section 4, the cross-validation classification results on five real data are presented, and we further compare the results with those publicly listed on the BCI competition leading board. In Section 5, we summarize our contributions and conclude this paper.

2 Preliminary Information

In this section, we briefly introduce the concepts and tools of Riemannian geometry for the spaces associated with symmetric positive definite (SPD) and positive semi-definite (PSD) matrices.

Let $\mathbb{S}_n$ denote the set of $n \times n$ symmetric matrices in the space of general real matrices $\mathbb{M}_n$, i.e., $\mathbb{S}_n = \{S \in \mathbb{M}_n,\ S^T = S\}$. The set of all $n \times n$ SPD matrices can be denoted as $\mathbb{P}_n = \{P \in \mathbb{S}_n, \boldsymbol{u}^T P \boldsymbol{u} > 0,\ \forall \boldsymbol{u} \in \mathbb{R}^n\}$. Similarly, the set of all $n \times n$ PSD matrices can be denoted as $\bar{\mathbb{P}}_n = \{P \in \mathbb{S}_n, \boldsymbol{u}^T P \boldsymbol{u} \geq 0, \forall \boldsymbol{u} \in \mathbb{R}^n\}$. Additionally, the application of logarithmic and exponential functions extends the compatibility of numerous methods designed for Euclidean tangent spaces to the manifold. Moreover, projection maps in both metrics are also incorporated. The tangent space at a matrix $A \in \mathbb{S}_n$, i.e., $T_A\mathbb{S}_n$, can be identified with the space $\mathbb{H}_n$ of $n \times n$ real symmetric matrices. This tangent space is Euclidean space $\mathbb{E}_n$.

2.1 Affine-Invariant (AI) Riemannian metric

For two positive-definite (PD) matrices A and $B \in \mathbb{P}_n$, the AI distance is

$$d_{AI}(A, B) = \left(\sum_{i=1}^{n} \log^2 \lambda_i \left(A^{-1} B \right) \right)^{1/2} \tag{1}$$

where $\lambda_i(A^{-1}B), i = 1, \ldots, n$, are the eigenvalues of $A^{-1}B$. B can be projected to the tangent space spanned at A, $T_A\mathbb{P}_n$, via the logarithmic map:

$$\log_A^{AI}(B) = A^{1/2} \operatorname{logm}\left(A^{-1/2} B A^{-1/2} \right) A^{1/2} \tag{2}$$

where $\operatorname{logm}(\cdot)$ denotes the logarithm of a matrix [34]. Moreover, the exponential map projects any elements, i.e., Hermitian matrix $X \in \mathbb{H}_n$, from the tangent space $T_A\mathbb{P}_n$ back to the manifold by:

$$\exp_A^{AI}(B) = A^{1/2} \operatorname{expm}\left(A^{-1/2} B A^{-1/2} \right) A^{1/2} \tag{3}$$

Where $\operatorname{expm}(\cdot)$ denotes the exponential of a matrix [31].

2.2 Bures-Wasserstein (BW) metric

The Bures-Wasserstein (BW) distance between $A, B \in \mathbb{P}_n$ is defined as [32]:

$$d_{BW} = \left\{ \operatorname{tr}(A + B) - 2\operatorname{tr}\left(A^{1/2} B A^{1/2} \right)^{1/2} \right\}^{1/2}$$

$$= \left\{ \text{tr}\,(A+B) - 2\text{tr}(AB)^{1/2} \right\}^{1/2} \quad (4)$$

The corresponding BW geodesic on $\mathbb{P}_n$ is defined as:

$$A \diamond_t B = (1-t)^2 A + t^2 B + t\,(1-t)\left[(AB)^{1/2} + (BA)^{1/2}\right] \quad (5)$$

The Wasserstein (BW) mean is the mid-point of the geodesic:

$$A \diamond B = A \diamond_{1/2} B = \frac{1}{4}\left[A + B + (AB)^{1/2} + (BA)^{1/2}\right] \quad (6)$$

The logarithm function on $\mathbb{P}_n$ under the BW metric has been described in [32], [33], and [34]. For any $A, B \in \mathbb{P}_n$, $X \in \mathbb{H}_n$, we have

$$\log_A^{BW}(X) = (AB)^{1/2} + (BA)^{1/2} - 2A, \quad (7)$$

Moreover, the concise form of the exponential map has been proposed by [38]. Let $A \diamond B$ deonte the Hadamard product of matrices A and B of the same size. Suppose $A \in \mathbb{P}_n$ has the spectral decomposition $A = U \Lambda U^*$, where U is an orthogonal matrix and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_2)$. Then for every Hermitian matrix $X \in \mathbb{H}_n$, such that $I_n + W \circ X_U$ is PSD, $X_U \coloneqq U^* X U$, we have

$$\exp_A^{BW}(X) = A + X + U\left[(W \circ X_U)\,\Lambda\,(W \circ X_U)\right] U^*, \quad (8)$$

where $W = \left(\frac{1}{\lambda_i + \lambda_j}\right)_{n \times n}$.

Unlike the AI metric, the above formulas (2.4)-(2.8) for BW distance, BW geodesic, BW mean and BW projection maps can be extended to the closure of $\mathbb{P}_n$, i.e. the set $\bar{\mathbb{P}}_n$ of $n \times n$ positive semi-definite (PSD) matrices.

2.3 Fréchet Mean Algorithm

Given a collection of matrices $A_1, \ldots, A_m \in \mathbb{S}_n$, their centroid can be estimated using the Fréchet mean with manifold based distance, e.g., d_{BW} and d_{AI}:

$$\bar{A}(A_1, \ldots, A_m) = \arg\min_{M \in \bar{\mathbb{P}}_n} \sum_{i=1}^{m} d_\circ^2(A_i, M) \quad (9)$$

The Fréchet mean is also known as the Karcher mean and the barycenter in the literature (e.g. [32, 36]). [35] proposed three methods, i.e., the inductive algorithm, the projection algorithm, and the cheap mean algorithm to estimate the Fréchet mean of PSD matrices. They suggested using BW distance along with the projection mean algorithm for the analysis of high-dimensional data, such as BCI data. Hence, in this study, we utilize the projection mean algorithm to estimate the Fréchet mean and propose SVM algorithm on the manifold equipped with BW metric.

The projection mean algorithm is constructed based on the idea that the arithmetic center of the projections of $A_1, \ldots, A_m$ onto the tangent space at the barycenter C is exactly the projection of C. Suppose the error tolerance in the stopping criteria is denoted as ϵ, the process of the algorithm on the manifold equipped with BW metric is described below.

In practice, the iteration process stops when $d_{BW}(S^{(l)}, S^{(l-1)}) \leq \epsilon$, and $S^{(l)}$ is the estimated Fréchet mean. The projection mean algorithm process on the AI manifold is similar, with the logarithm and exponential functions replaced by (2.2) and (2.3) respectively.

Algorithm 1 Fréchet Mean Algorithm

1) Set $S^{(0)} \coloneqq \frac{1}{m}\sum_{j=1}^{m} A_j$

2) Given $S^{(l)}$ for some $l \in \mathbb{N}$, update $S^{(l+1)}$ as follows:

(a) Project $\{A_1, \ldots, A_m\}$ onto the tangent space at $S^{(l)}$ using (2.7): $X_j^{(l)} \coloneqq (S^{(l)} A_j)^{1/2} + (A_j S^{(l)})^{1/2} - 2S^{(l)}$.

(b) Compute the arithmetic mean of the projection vectors: $X^{(l)} \coloneqq \frac{1}{m}\sum_{j=1}^{m} X_j^{(l)} = \frac{1}{m}\sum_{j=1}^{m}[(S^{(l)} A_j)^{1/2} + (A_j S^{(l)})^{1/2}] - 2S^{(l)}$.

(c) Update $S^{(l+1)}$ as the exponential of $X^{(l)}$ at $S^{(l)}$ using (2.11)

$S^{(l+1)} \coloneqq S^{(l)} + X^{(l)} + U^{(l)}[(W^{(l)} \circ X_U^{(l)})\Lambda^{(l)}(W^{(l)} \circ X_U^{(l)})]U^{*(l)}$,

where $S^{(l)} = U^{(l)} \Lambda^{(l)} U^{*(l)}$, $U^{(l)} \in \mathrm{U_n}$, $\Lambda^{(l)} = \text{diag}(\lambda_1^{(l)}, \lambda_2^{(2)}, \ldots, \lambda_n^{(l)})$, $W^{(l)} \coloneqq \left(\frac{1}{\lambda_i^{(l)} + \lambda_j^{(l)}}\right)_{n \times n}$,

$X_U^{(l)} = U^{*(l)} X^{(l)} U^{(l)}$.

3) The limit of $\{S^{(l)}\}_{l \in \mathbb{N}}$ serves as the approximation of the Fréchet mean with d_{BW}:

$\lim_{l \to \infty} S^{(l)} = \bar{A}(A_1, \ldots, A_m)$.(2.10)

3 Proposed Algorithms

The Support Vector Machine (SVM) [37] is a popular supervised machine learning algorithm used in BCI applications for classification, especially for its robust performance in high-dimensional feature spaces. SVM aims to find a hyperplane that best separates data points into different classes, maximizing the margin between classes. Support vectors are data points that are closest to the hyperplane. SVM can handle linear relationships which makes it particularly effective in scenarios where the data can be linearly separable.

Random Forest (RF) is a versatile machine learning algorithm widely used for classification. It builds multiple decision trees on random subsets of the data, reducing overfitting and enhancing accuracy. RF is effective in capturing nonlinear relationships, making it well-suited for complex patterns in various datasets.

In this study, we utilize both SVM and RF and propose two algorithms for PSD matrices classification using BW distance and BW projection maps.

3.1 BW-based Projection Classification

Euclidean-based classifiers cannot be performed directly on the Riemannian manifold. Instead, implementation can be achieved by utilizing the tangent space located at the Fréchet mean of the entire set of matrices $A_1, \ldots, A_m \in \bar{\mathbb{P}}_n$. Each matrix A_i is projected onto the tangent space that is obtained based on the barycenter $\bar{A}$:

$$X_i = \log_{\bar{A}}^{BW}(A_i) = \left(\bar{A} A_i\right)^{1/2} + \left(A_i \bar{A}\right)^{1/2} - 2\bar{A}. \quad (3.1)$$

After projecting all matrices onto the tangent space, we half-vectorize the projected symmetric matrices and apply a classification algorithm such as SVM or RF on the Euclidean tangent space. The full algorithm for BW-based Projection Classification is summarized in Algorithm 2.

Algorithm 2 BW-based Projection Classification

Input: A set of PSD matrices $A_1, \ldots, A_m \in \bar{\mathbb{P}}_n$, Stopping criteria $\epsilon > 0$

Output: Predicted class

1) Calculate the barycenter $\bar{A}$ such that
$\bar{A} = \arg\min_{M \in \bar{\mathbb{P}}_n} \sum_{i=1}^{m} d_{BW}^2(A_i, M)$
2) Project $A_1, \ldots, A_m$ onto tangent space $T_{\bar{A}}\bar{\mathbb{P}}_n$
$X_i = \log_{\bar{A}}^{BW}(A_i),\ i = 1, \ldots, m$
3) Half-vectorize $X_1, \ldots, X_m$ and compile into the matrix
$T^X_{m \times \frac{n \times (n+1)}{2}} \in \mathbb{E}_n$
4) Perform classification algorithm on T^X
5) Return predicted class

Furthermore, Figure 1 illustrates the pipeline of the proposed BW-based Projection Classification using SVM. This visual representation captures the sequential stages and interactions within the algorithm.

3.2 BW-based Adaptive Projection Classification

In the context of BCI, it is well known that feature distributions vary greatly between sessions [38]. This adaptation should be accomplished in an unsupervised way and is usually achieved by spatially whitening [23]. Specifically, each matrix $\bar{\mathbb{P}}_n$ will be scaled first by the estimated barycenter and then projected onto the tangent space. The full algorithm of the BW-based Adaptive Projection Classification can be found in Algorithm 3.

4 Real Data Results

To evaluate the performance of the proposed BW-based algorithms, we apply them to multiple real datasets and compare results with those obtained by some existing algorithms. We utilize three publicly accessible BCI datasets [39–41] and four fMRI connectivity datasets [42–45] to evaluate the improvement in classification performance achieved through the implementation of BW metric. Moreover, we also compare the results obtained by the proposed methods with those obtained in the BCI competition [46, 47] and literature.

Algorithm 3 BW-based Adaptive Projection Classification

Input: A set of PSD matrices $A_1, \ldots, A_m \in \bar{\mathbb{P}}_n$, Stopping criteria $\epsilon > 0$

Output: Predicted class

1) Calculate the barycenter $\bar{A}$ such that
$\bar{A} = \arg\min_{M \in \bar{\mathbb{P}}_n} \sum_{i=1}^{m} d_{BW}^2(A_i, M)$
2) Scale $A_1, \ldots, A_m$ based on $\bar{A}$
$S_i = \bar{A}^{-1/2} A_i \bar{A}^{-1/2},\ i = 1, \ldots, m$
3) Project $S_1, \ldots, S_m$ onto tangent space $T_{\bar{A}}\bar{\mathbb{P}}_n$
$X_i = \log_{\bar{A}}^{BW}(S_i), i = 1, \ldots, m$
4) Half-vectorize $X_1, \ldots, X_m$ and compile into the matrix
$A^X_{m \times \frac{n \times (n+1)}{2}} \in \mathbb{E}_n$
5) Perform classification algorithm on A^X
6) Return predicted class

4.1 Data Description

Three publicly available fMRI datasets were used: Autism Brain Imaging Data Exchange (ABIDE) dataset, Attention Deficit Hyperactivity Disorder-200 (ADHD-200) dataset, and Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, and one private fMRI dataset: Post-Traumatic Stress Disorder (PTSD) dataset. The ABIDE dataset used consists of 556 healthy control subjects, 339 subjects with Autism, and 93 with Asperger's. The ADHD-200 dataset used composes of 573 controls and three different ADHD subtypes. In particular, 208 subjects with ADHD-C, 13 with ADHD-H, and 136 with ADHD-I. The ADNI dataset used comprises 35 controls, 29 with Alzheimer's disease (AD), 34 with early mild cognitive impairment (EMCI), and 34 with late mild cognitive impairment (LMCI). Finally, the PTSD dataset contains 56 controls, 34 subjects with PTSD, and 84 subjects with both PTSD and post-concussion syndrome (PCS). More information about the datasets used can be found in [42–45] respectively. The four fMRI datasets contain multiple classes of functional connectivity (FC) built by computing Pearson's correlation connectivity matrices. Additional details regarding the processing of the Rs-fMRI data are explained in [45]. The details of the four fMRI datasets are summarized in 1 below.

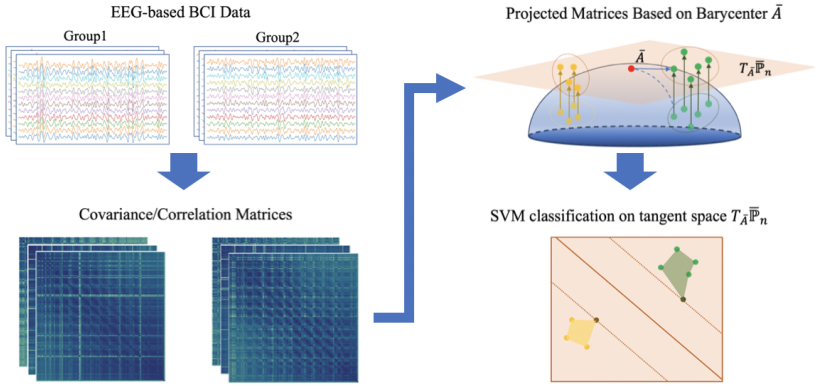


Figure 1: Graph pipeline of the BW-based projection SVM (EEG-based BCI dataset as example)

Table 1: fMRI Data Description

	ABIDE		ADHD-200		ADNI		PTSD	
Number of subjects	Control	556	Control	573	Control	55	Control	56
	Autism	339	ADHD-C	208	AD	29	PTSD	34
	Asperger	93	ADHD-H	13	EMCI	34	PTSD/PCS	84
			ADHD-I	136	LMCI	34		
	Total	988	Total	930	Total	132	Total	174
Connectivity matrix size	(200, 200)		(190, 190)		(200, 200)		(125, 125)	

Table 2: Configurations of BCI Competition

	BCI III		BCI IV
	Dataset IIIa	Dataset IVa	Dataset IIa
Number of subjects	3	5	9
Number of channels	60	118	22
Number of classes	4	2	4
Trials per class	30/45	28-224	72
Sampling rate	250Hz	1000Hz	250Hz
Filter Band	8-60Hz	8-30Hz	8-30Hz

Table 3: BCI III Dataset IVa Train and Test Trials

Subject	aa	al	av	aw	ay
Train	168	224	84	56	28
Test	112	56	196	224	252

Moreover, to encompass a broad range of scenarios commonly encountered in EEG-based BCI applications, we additionally select two datasets from BCI competition III [46] and one from BCI competition IV [47]. In BCI III, dataset IIIa [39] comprises 60-channel EEG recordings from three subjects who performed motor imagery tasks involving four classes (left hand, right hand, foot, and tongue). Dataset IVa [40] consists of 118-channel EEG recordings from five subjects engaged in two motor imagery tasks (right hand and foot). In BCI IV, dataset IIa [41] includes 22-channel EEG recordings from nine subjects performing four motor imagery tasks (left hand, right hand, feet, and tongue).

Details regarding the BCI datasets are summarized in Table 2. In general, the matrix dimensions within these three datasets span from 22 to 264, and the number of matrices varies from 10 to 200. This diverse range ensures a comprehensive and inclusive representation of scenarios commonly encountered in the analysis of neuroscience data.

4.2 Cross-Validation

Firstly, we employ a 5-fold cross-validation with 10 repetitions to compare algorithms. The comparison results for the BCI III dataset IVa are presented below as an illustrative example, showcasing cases with varying sizes of training and testing datasets (see Table 3).

Notably, this dataset presents the challenge of effectively working with a limited amount of training data. For this dataset, Support Vector Machine (SVM) classifier was employed to facilitate prediction in the projection algorithms. Subsequently, parameter tuning is carried out for each SVM algorithm through cross-validation of all relevant parameters, including the regularization parameter C, aiming to achieve the highest classification accuracy.

Table 4 provides a comprehensive summary of results obtained from five SVM algorithms across the BCI III dataset IVa. Both projection-based methods exhibit superior classification performance compared to the Euclidean-based method. This underscores the advantages of incorporating manifold analysis for the analysis of PSD matrices. Next, we compare BW-based projection SVM algorithms with AI-based projection SVM [23]. Almost all of the results obtained by the BW-based projection SVM algorithm are better than those obtained by the AI-based projection SVM algorithm, which could be attributed to the higher accuracy of barycenter estimation by the BW projection method [35]. Moreover, we compare BW-based adaptive projection SVM algorithms with AI-based adaptive projection SVM [23]. Similarly, our findings indicate superior performance from the BW-based adaptive projection SVM algorithm. Notably, we observe an enhanced accuracy achieved by the BW-based adaptive projection SVM in comparison to the BW-based projection SVM.

Additionally, we use the four fMRI datasets: PTSD, ADNI, ABIDE, and ADHD-200 to compare the results utilizing AI-based projection and BW-based projection employing 6-fold cross-validation with

Table 4: Summary of Cross-Validation Accuracy Results for BCI III IVa

Method	Euclidean	Projection		Adaptive Projection	
		AI	BW	AI	BW
Mean Accuracy	0.8432	0.8641	0.8923	0.9032	0.9184
aa	0.7440	0.8989	0.9405	0.9882	0.9941
al	0.9153	0.9065	0.9197	0.9643	0.9599
av	0.7507	0.9404	0.9882	0.7743	0.7806
aw	0.8394	0.7879	0.7864	0.8561	0.8909
ay	0.9667	0.7867	0.8267	0.9333	0.9667

Table 5: Classification Accuracy Results of fMRI datasets

	Binary				Multi-class			
Method	SVM		Random Forest		SVM		Random Forest	
	AI	BW	AI	BW	AI	BW	AI	BW
ABIDE	0.6314	0.8964	0.6688	0.7128	0.5628	0.7748	0.6254	0.6631
ADHD-200	0.6944	0.8737	0.7010	0.7039	0.6249	0.6249	0.6626	0.6548
ADNI	0.9511	1.0000	0.9886	0.8654	0.5528	0.5641	0.7582	0.7068
PTSD	1.0000	1.0000	0.9900	0.9754	0.9758	0.9754	0.8724	0.9211

10 repetitions. We conducted both binary and multi-class classification using Random Forest and SVM classifiers ensuring comprehensive evaluation. Parameter tuning was carried out through cross-validation for optimization. For SVM, the regularization parameter C was adjusted, and for random forest we optimized the number of features to consider at each split, the minimum size of leaf nodes, and the number of trees in the forest.

Feature selection was performed to enhance performance and boost accuracy. For SVM, feature selection was based on the coefficients of the trained model, while for random forest importance scores were used to select the most relevant features.

For the ABIDE dataset, the binary classification classes were Control and a combined Autism/Asperger class, while the multi-class classification included Control, Autism, and Asperger. For the ADHD-200 dataset, the binary classification involved Control and a combined ADHD class consisting of all three subtypes, with the multi-class classification featuring Control, ADHD-I, ADHD-C (excluding ADHD-H due to a low number of subjects). In the ADNI dataset, the binary classes were Control and AD, with the multi-class classification comprising Control, AD, EMCI, and LMCI. The PTSD dataset's binary classes were Control and PTSD, while the multi-class classification involved Control, PTSD, and PTSD/PCS.

Table 5 summarizes results from four projection-based algorithms for binary and multi-class classification across fMRI datasets. It compares BW-based and AI-based projection algorithms. For the binary case, the BW-based projection SVM outperforms the AI-based projection SVM. The BW-based projection random forest shows improvements over the AI-based version for the ABIDE and ADHD-200 datasets. For multi-class classification, BW-based projection results generally do better than the AI-based results. Specifically, SVM excels with the ABIDE and PTSD datasets, while random forest performs better with the ADHD-200 and ADNI datasets.

4.3 BCI Competition

We extend our comparison by evaluating the results obtained from the proposed methods against the BCI competition winner [46, 47]. To ensure a meaningful performance evaluation aligned with the competition winners, we strictly follow the identical signal preprocessing steps as specified in the guidelines for each competition dataset.

Table 6 displays the results in terms of kappa value, as it was done for the BCI competition III dataset IIIa. In this competition, the proposed BW-based algorithm achieves the mean performance of 0.7506 which ranks this method to the second place of the competition.

As shown in Table 7, the results are presented in terms of accuracy, same as the format used for the BCI competition III dataset IVa. In this competition, the mean performance of the proposed BW-based algorithm is 0.8734, securing the method a third-place ranking, just 0.0006 lower than the second one.

As depicted in Table 8, the results are presented in terms of kappa values, akin to the format used for the BCI competition IV dataset IIa. In this competition, the proposed BW-based algorithm attains a mean performance of 0.4789, securing a third-place ranking. Notably, for four subjects, the accuracy achieved by the proposed BW-based SVM surpasses the score of the competition winner.

Additionally, it is essential to note that our reported results are obtained without leveraging frequency information in contrast to the approach employed by the winner in the competition. There is potential for achieving higher performances by optimizing frequency filters.

5 Conclusion

In this paper, we introduce two innovative machine learning algorithms tailored for the classification of PSD matrices. Initially, we

Table 6: Results of the BCI Competition in BCI III Dataset IIIa

	Mean Kappa	K3b	K6b	L1b
1st	0.7926	0.8222	0.7556	0.8000
BW-based SVM	0.7506	0.9185	0.5889	0.7444
2nd	0.6872	0.9037	0.4333	0.7111
3rd	0.6272	0.9481	0.4111	0.5222

Table 7: Results of the BCI Competition in BCI III Dataset IVa

	Mean Accuracy	aa	al	av	aw	ay
1st	0.9474	0.9550	1.0000	0.8060	1.0000	0.9760
2nd	0.8740	0.8930	0.9820	0.7650	0.9240	0.8060
BW-based SVM	0.8734	0.8750	0.9821	0.8214	0.9107	0.7778
3rd	0.8454	0.8210	0.9460	0.7040	0.8750	0.8810
Literature						
Selim et al. [48]	0.85	0.866	1	0.668	0.906	0.81
Park and Chung [49]	0.845	0.741	1	0.678	0.901	0.893
Dai et al. [50]	0.792	0.681	0.939	0.685	0.884	0.749
Selim et al. [51]	0.788	0.696	0.893	0.592	0.888	0.869
Lotte and Guan [52]	0.786	0.723	0.964	0.602	0.777	0.865
Arvaneh et al. [53]	0.735	0.723	0.964	0.541	0.705	0.734
Belwafi et al. [54]	0.673	0.668	0.961	0.521	0.714	0.50

Table 8: Results of the BCI Competition in BCI IV Dataset IIa

	Mean Kappa	A01	A02	A03	A04	A05	A06	A07	A08	A09
1st	0.57	0.68	0.42	0.75	0.48	0.40	0.27	0.77	0.75	0.61
2nd	0.52	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	0.69
BW-based SVM	0.479	0.722	0.463	0.829	0.347	0.315	0.282	0.319	0.528	0.505
3rd	0.31	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44

propose the BW-based projection algorithm, where all PSD matrices undergo projection onto the tangent space computed using the BW metric. To address feature distribution variability between BCI sessions, we introduce the BW-based adaptive algorithm, incorporating a scalar during projection, termed BW adaptive projection. The performance of these proposed algorithms is evaluated, validating the superiority of the BW metric across multiple real datasets with varying dimensions and numbers of matrices.

Specifically, comparative analysis against Euclidean-based algorithms highlights the superior outcomes achieved by the proposed BW-based algorithms, underscoring the significance of Riemannian manifold analysis. Moreover, contrasting AI-based algorithms with the proposed BW-based algorithms emphasizes the enhanced accuracy provided by the BW metric. Furthermore, the consistent top-tier ranking of our proposed methods compared to winners from BCI competitions signifies their potential as state-of-the-art solutions in the field. This demonstrates not only the effectiveness of the BW-based analysis but also the broader impact and relevance of our contributions to advancing the capabilities of BCI systems. It is noteworthy that our methods' competitive performance, even without incorporating spatial and frequency filtering considerations, underscores the robustness and generalizability of the proposed algorithms across diverse BCI datasets. This resilience to variations in input data conditions further supports the adaptability and effectiveness of the BW-based approaches for real-world applications where data preprocessing challenges may arise.

Moreover, investigating spatial filtering with Common Spatial Patterns (CSP) in conjunction with our BW-based methods holds promise for synergistic performance enhancement. Subject-specific frequency filtering is another avenue to explore for personalized and effective classification outcomes. Additionally, exploring the application of alternative supervised or unsupervised machine learning algorithms on the manifold coupled with BW metric, such as K-Means, may yield valuable insights. In summary, delving into spatial and frequency filtering enhancements and diversifying machine learning algorithms applied to the manifold equipped with BW metric presents exciting possibilities for advancing BCI applications and refining the current framework.

Acknowledgments

The authors would like to thank anonymous referees for their constructive comments that improved the quality of this paper. This

paper is based upon work supported by the National Science Foundation under Grant No. 2153492.

References

[1] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine learning research*, vol. 5, no. Jan, pp. 27–72, 2004.

[2] S.Sra,"A new metric on the manifold of kernel matrices with application to matrix geometric means," *Advances in neural information processing systems*, vol. 25, 2012.

[3] M. S. Baghshah and S. B. Shouraki, "Kernel-based metric learning for semi-supervised clustering," *Neurocomputing*, vol. 73, no. 7-9, pp. 1352– 1361, 2010.

[4] J. Wang, A. Woznica, A. Kalousis *et al.*, "Metric learning with multiple kernels," *Advances in neural information processing systems*, vol. 24, 2011.

[5] F. Yger and M. Sugiyama, "Supervised logeuclidean metric learning for symmetric positive definite matrices," *arXiv preprint arXiv:1502.03505*, 2015.

[6] H. Zhu, Y. Chen, J. G. Ibrahim, Y. Li, C. Hall, and W. Lin, "Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging," *Journal of the American Statistical Association*, vol. 104, no. 487, pp. 1203–1212, 2009.

[7] C. B. Peterson, N. Osborne, F. C. Stingo, P. Bourgeat, J. D. Doecke, and M. Vannucci, "Bayesian modeling of multiple structural connectivity networks during the progression of alzheimer's disease," *Biometrics*, vol. 76, no. 4, pp. 1120–1132, 2020.

[8] J. Jagarlamudi, R. Udupa, H. Daume´ III, and A. Bhole, "Improving bilingual projections via sparse covariance matrices," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 930–940.

[9] T. Pahikkala, S. Pyysalo, J. Boberg, J. Ja¨rvinen, and T. Salakoski, "Matrix representations, linear transformations, and kernels for disambiguation in natural language," *Machine Learning*, vol. 74, pp. 133–158, 2009.

[10] W. Zhang and P. Fung, "Discriminatively trained sparse inverse covariance matrices for speech recognition," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 22, no. 5, pp. 873–882, 2014.

[11] A. Cherian and S. Sra, "Positive definite matrices: data representation and applications to computer vision," Algorithmic Advances in Riemannian Geometry and Applications: For Machine Learning, Computer Vision, Statistics, and Optimization, pp. 93–114, 2016.

[12] G. Dong and G. Kuang, "Target recognition in sar images via classification on riemannian manifolds," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 199–203, 2014.

[13] A. Y. Alaoui, Y. Tabii, R. O. H. Thami, M. Daoudi, S. Berretti, and P. Pala, "Fall detection of elderly people using the manifold of positive semidefinite matrices," *Journal of Imaging*, vol. 7, no. 7, 2021.

[14] A. Toshev, B. Taskar, and K. Daniilidis, "Shape-based object detection via boundary structure segmentation," *International journal of computer vision*, vol. 99, pp. 123–146, 2012.

[15] D. J. McFarland and J. R. Wolpaw, "Eeg-based brain–computer interfaces," *current opinion in Biomedical Engineering*, vol. 4, pp. 194–200, 2017.

[16] G. H. Glover, "Overview of functional magnetic resonance imaging," *Neurosurgery Clinics*, vol. 22, no. 2, pp. 133–139, 2011.

[17] M. Faraki, M. T. Harandi, and F. Porikli, "Image set classification by symmetric positive semi-definite matrices," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1-8.

[18] T. Graepel and R. Herbrich, "Invariant pattern recognition by semi- definite programming machines," *Advances in neural information processing systems*, vol. 16, 2003.

[19] J. Ren and X. J. Wu, "Probability distribution-based dimensionality reduction on Riemannian manifold of SPD matrices," *IEEE Access*, vol. 8, pp. 153881-153890, 2020.

[20] R. Wang, X. J. Wu, Z. Chen, C. Hu,, and J. Kittler, "SPD manifold deep metric learning for image set classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 7, pp. 8924-8938, July 2024.

[21] W. Gao, Z. Ma, W. Gan, and S. Liu, "Dimensionality reduction of SPD Data Based on Riemannian Manifold Tangent Spaces and Isometry," *Entropy*, vol. 23, no.9, p.1117, 2021.

[22] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to bci classification," in *Latent Variable Analysis and Signal Separation*, V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 629–636.

[23] ——, "Classification of covariance matrices using a riemannian-based kernel for bci applications," *Neurocomput.*, vol. 112, p. 172–178, jul 2013. [Online]. Available: https://doi.org/10.1016/j.neucom.2012.12.039

[24] A. S. M. Miah, M. R. Islam, and M. K. I. Molla, "Eeg classification for mi-bci using csp with averaging covariance matrices: an experimental study," in *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*. IEEE, 2019, pp. 1–5.

[25] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM journal on matrix analysis and applications*, vol. 29, no. 1, pp. 328–347, 2007.

[26] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, "Kernel methods on the riemannian manifold of symmetric positive definite matrices," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 73–80.

[27] Z. Huang, R. Wang, S. Shan, X. Li, and X. Chen, "Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification," in *International conference on machine learning*. PMLR, 2015, pp. 720–729.

[28] Z. Lin, "Riemannian geometry of symmetric positive definite matrices via cholesky decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 40, no. 4, pp. 1353–1370, 2019.

[29] X. Pennec, "Manifold-valued image processing with spd matrices," in *Riemannian geometric statistics in medical image analysis*. Elsevier, 2020, pp. 75–134.

[30] A. Han, B. Mishra, P.K. Jawanpuria, and J. Gao, "On riemannian optimization over positive definite matrices with the bures-wasserstein geometry," *Advances in Neural Information Processing Systems*, vol.34, pp. 8940-8953, 2021.

[31] M. Berger, "A panoramic view of riemannian geometry," 2003.

[32] R. Bhatia, T. Jain, and Y. Lim, "On the bures–wasserstein distance between positive definite matrices," *Expositiones Mathematicae*, vol. 37, no. 2, pp. 165–191, 2019.

[33] E. Massart and P.-A. Absil, "Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 41, no. 1, pp. 171–198, 2020.

[34] Y. Thanwerdas, "Riemannian and stratified geometries of covariance and correlation matrices," Ph.D. dissertation, Université Côte d'Azur, 2022.

[35] J. Zheng, H. Huang, Y. Yi, Y. Li, and S.-C. Lin, "Barycenter estimation of positive semi-definite matrices with bures-wasserstein distance," 2023.

[36] 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.

[37] S. Yue, P. Li, and P. Hao, "Svm classification: Its contents and challenges," *Applied Mathematics-A Journal of Chinese Universities*, vol. 18, pp. 332–342, 2003.

[38] P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, and K.-R. Müller, "Towards adaptive classification for bci," *Journal of neural engineering*, vol. 3, no. 1, p. R13, 2006.

[39] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery eeg data for the bci-competition 2005," *Journal of Neural Engineering*, vol. 2, no. 4, p. L14, aug 2005. [Online]. Available: https://dx.doi.org/10.1088/1741-2560/2/4/L02

[40] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Muller, "Boosting bit rates in noninvasive eeg single-trial classifications by feature combi- nation and multi-class paradigms," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 993–1002, 2004.

[41] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–graz data set a," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[42] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. Castellanos, *et al.*, "The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism," *Molecular Psychiatry*, vol. 19, pp. 659–667, 2014.

[43] ADHD Consortium, "The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience," *Frontiers in Systems Neuroscience*, vol. 6, p. 62, 2012.

[44] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack Jr., W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner, "Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization," *Neurology*, vol. 74, pp. 201–209, 2010.

[45] P. Lanka, D. Rangaprakash, M. N. Dretsch, J. S. Katz, T. S. Denney, and G. Deshpande, "Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets," *Brain Imaging and Behavior*, vol. 14, no. 6, pp. 2378–2416, 2020.

[46] B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Scholgl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, "The bci competition iii: Validating alternative approaches to actual bci problems," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 14, no. 2, pp. 153-159, 2006.

[47] M. Tangermann, K. R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K.J. Miller, G. Mueller-Putz, *et al.* "Review of the bci competition iv." *Frontiers in neuroscience*, p. 55, 2012.

[48] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A csp \ am-ba-svm approach for motor imagery bci system," *Ieee Access*, vol. 6, pp. 49 192-49 208, 2018.

[49] Y. Park and W. Chung, "Bci classification using locally generated csp features," in *2018 6th International Conference on Brain-Computer Interface (BCI)*. IEEE, 2018, pp. 1-4.

[50] M. Dai, D. Zheng, S. Liu, and P. Zhang, "Transfer kernel common spatial patterns for motor imagery brain-computer interface classification," *Computational and mathematical methods in medicine*, vol. 2018, 2018.

[51] S. Selim, M. Tantawi, H. Shedeed, and A. Badr, "Reducing execution time for real-time motor imagery based bci systems," in *Proceedings of the International*

Conference on Advanced Intelligent Systems and Informatics 2016 2. Springer, 2017, pp. 555-565.

[52] F. Lotte and C.Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *IEEE Transactions on biomedical Engineering*, vol. 58, no. 2, pp. 355-362, 2010.

[53] M. Arvaneh, C. Guan K. K. Ang, and H. C. Quek, "Spatially sparsed common spatial pattern to improve bci performance," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 2412-2415.

[54] K. Belwafi, O. Romain, S. Gannoini, F. Ghaffari, R. Djemal, and B. Ouni, "An embedded implementation based on adaptive filter bank for brain-computer interface systems," *Journal of neuroscience methods*, vol. 305, pp. 1-16, 2018.