

A Survey on Transfer Learning for Brain Computer Interfaces using Data Mining Techniques

Manisha Thombare¹, Radhika Chandwadkar²

Assistant Professor, Department of Computer Engineering, MVPS's KBTCOE Nasik, Maharashtra, India¹

Assistant Professor, Department of Computer Engineering, MVPS's KBTCOE Nasik, Maharashtra, India²

ABSTRACT: The major challenge during this paper targets in developing sensible EEG-based brain-computer interfaces (BCIs): the way to address individual variations so higher learning performance may be obtained for a brand new subject, with minimum or perhaps no subject-specific data? Methods: we have a tendency to propose a unique approach to align electroencephalogram trials from completely different subjects within the Euclidean space to create them additional similar and thence improve the training performance for a brand new subject. An approach has three properties: 1) It aligns the engineering G trials directly among the Euclidean space, and any signal method, feature extraction and machine learning algorithms can then be applied to the aligned trials; 2) its procedure worth is unbelievably low; and, 3) it's unsupervised and does not would love any label information from the new subject. It results each offline and simulated on-line experiments on motor imagination classification and event-related potential classification verified that our planned approach outperformed a progressive Riemannian space knowledge alignment approach, and several other approaches while not knowledge alignment. Conclusion: The planned Euclidean space electroencephalogram knowledge alignment approach will greatly facilitate transfer learning in BCIs. Significance: Our planned approach is effective, efficient, and simple to implement. It may be a vital pre-processing step for EEG-based BCIs.

KEYWORDS: Text detection, Morphological operations, Connected component labelling.

I. INTRODUCTION

A brain-computer interface (BCI) [11],[12] could be a communication pathway for a user to act with his/her surroundings by exploitation brain signals, that contain info regarding the user's state or intentions. Graphical record (EEG) is that the most well liked input in BCI systems. Motor imaging (MI) and event-related potentials (ERPs) square measure 2 common approaches of EEG-based BCIs, and additionally the main target of this paper. For MI-based BCIs, the user has to imagine the movements of his/her body elements (e.g., hands, feet, and tongue), that causes modulations of brain rhythms within the concerned animal tissue areas. So, the imagination of various movements is distinguished from the spatial localization of various activity rhythm modulations, and so went to management external devices. For ERP-based BCIs, the user is stimulating d by a majority of common stimuli (non-target) and a tiny low.

Recently, the appliance scope of BCIs has been extended to able folks and EEG becomes the foremost in style signalling as a result of it's simple and safe to amass, and has high temporal resolution. However, EEG measures the terribly weak brain electrical signals from the scalp, which ends up in poor spatial resolution and low ratio [4]. Consequently, refined signal process and machine learning algorithms square measure required in EEG-based BCI systems to decrypt the EEG signal, particularly for single-trial classification of EEG signals in real-world applications. Sometimes the EEG signals square measure initial band-pass filtered and spatially filtered to extend the ratio, and so discriminative options square measure extracted, that square measure next fed into machine learning algorithms like Linear Discriminator Analysis (LDA) and Support Vector Machine (SVM) [3] for classification. Since the variance matrices square measure interchangeable positive definite (SPD) and lie on a Riemannian manifold, a well-liked approach is to look at every variance matrix as some extent within the Riemannian area, and use its geodesic to the Riemannian mean as a feature in classification. This approach is named the Minimum Distance (MDRM) classifier [1].

For each MI and ERP, the classification is performed within the Riemannian area, whose geodesic computation is far a lot of sophisticated, long, and unstable than the space calculation within the metric space. During this we have a tendency to propose a replacement EEG knowledge alignment approach within the metric space that has the subsequent fascinating characteristics:



1) It transforms and aligns the EEG trials within the metric space, and any signal process, feature extraction and machine learning algorithms will then be applied to the aligned trials. On the contrary, RA aligns the variance matrices (instead of the EEG trials themselves) within the Riemannian area, and therefore an ensuing classifier should be ready to treat the variance matrices directly, whereas there square measure only a few such classifiers.

2) It is computed many times quicker than RA.

3) It solely needs untagged EEG trials however doesn't want any label info from the new subject; therefore, it is employed in utterly unsupervised learning. The effectiveness of our projected approach is then incontestable in 2 BCI classification scenarios:

1) Offline unsupervised classification, during which untagged EEG trials from a replacement subject square measure offered, and that we have to be compelled to label them by creating use of auxiliary labelled knowledge from alternative subjects.

2) Simulated on-line supervised classification, during which a tiny low range of labelled EEG epochs from a replacement subject square measure obtained consecutive on-the-fly and a classifier is trained from them and auxiliary labelled knowledge from alternative subjects to label future incoming epochs from the new subject.

The remainder of this paper is organized as follows: Section II introduces the RA-MDRM approach within the Riemannian area. Section III proposes our metric space knowledge alignment approach. Section IV introduces the 3 datasets employed in our experiments, together with 2 MI datasets and one ERP dataset. Sections V and VI compare the performance of our approach with RA-MDRM in offline and simulated on-line learning, severally. Finally, Section VII attracts conclusion and points out some future analysis directions.

II. RELATED WORK

The variance matrices of graphical record trials area unit SPD, and belong a Riemannian house rather than a metric space. Since the variance matrices directly encipher the special info of the graphical record trials, and by fitly augmenting the graphical record trials (such as in ERP classification) they'll conjointly encipher the temporal info, we are able to perform graphical record classification directly supported the variance matrices. This section introduces the MDRM classifier, that assigns an effort to the category whose Riemannian mean is that the nearest to its variance matrix, and conjointly a Riemannian house variance matrix alignment approach (RA).

A. Riemannian Distance

The Riemannian distance between 2 SPD matrices P_1 and P_2 is termed the geodesic, that is that the minimum length of a curve connecting them on the Riemannian manifold.

$$\delta(P_1, P_2) = \|\log(P_1^{-1} P_2)\|_F = \left[\sum_{r=1}^R \log^2 \lambda_r \right]^{\frac{1}{2}}, \quad (1)$$

Where the subscript F denotes the Frobenius norm, and $\lambda_r (r = 1, 2, \dots, R)$ are the real eigenvalues of $P_1^{-1} P_2$.

The Riemannian distance between two SPD matrices P_1 and P_2 remains underneath linear invertible transformation:

$$\delta(C^r P_1 C, C^r P_2 C) = \delta(P_1, P_2), \quad (2)$$

Where C is Associate invertible matrix. This property of the Riemannian distance is named harmony invariability.

A. Riemannian Mean

The mean of a collection of SPD matrices will be computed within the Euclidean space as their first moment, and additionally within the Riemannian area because the Riemannian mean (geometric mean), outlined because the matrix minimizing the ad of the square Riemannian distances:

$$\varrho(P_1, \dots, P_N) = \arg \min_P \sum_{n=1}^N \delta^2(P, P_n). \quad (3)$$

There is no closed-form resolution to (3), Associate in its sometimes computed by an unvarying gradient descent rule [8].

B. MDRM

The MDRM classifier [1], [7] initial computes the mathematician mean of every category from the variance matrices of the labeled coaching trials, then assigns every check trial to the category whose mathematician mean is that the highest to its variance matrix, i.e.,

$$g(\Sigma) = \arg \min_c \delta(\Sigma, \bar{\Sigma}^c), \quad (4)$$

where Σ is that the variance matrix of the check trial, $\bar{\Sigma}^c$ is that the mathematician mean of sophistication c , and $g(\Sigma)$ is that the foretold category label of Σ .

C. RA-MDRM

Zanini et al. planned a unique TI approach within the mathematician house, observed during this paper as MDRM, to improve the performance of the MDRM classifier by utilizing auxiliary information from alternative sessions and/or subjects when there area unit solely a few labeled trials from a new subject. Since the variance matrices of the trials area unit the input to MDRM, RA-MDRM aims to align the variance matrices from completely different sessions/subjects to convey them a typical reference assumes that “different offer configurations and conductor positions induce shifts of variance matrices with reference to a reference (resting) state, but that when the brain is engaged in an exceedingly very specific task, variance matrices yield the SPD manifold within the same direction.” Then RA-MDRM centers “the variance matrices of each session/subject with regard to a reference variance matrix in order that what we tend to observe is merely the displacement with regard to the reference state thanks to the task.” More specifically, RA-MDRM 1st computes the variance matrices of some resting trials, $k_i = 1$, within which the topic isn't performing arts any task, so computes the mathematician mean \bar{R} of those matrices. \bar{R} is then used because the reference matrix in RA-MDRM to cut back the inter-session/subject variability by the subsequent transformation:

$$\tilde{\Sigma}_i = \bar{R}^{-1/2} \Sigma_i \bar{R}^{-1/2}, \quad (5)$$

Where Σ_i is that the variance matrix of the i^{th} trial and $\tilde{\Sigma}_i$ is that the corresponding aligned variance matrix. Equation (5) makes the reference state of various sessions/ subjects targeted at the scalar matrix. This transformation wouldn't modification the gap between the variance matrices happiness to constant session/subject owing to the congruousness Invariance property in (2) however makes the variance matrices of various sessions/subjects yield the mathematician manifold in several directions with regard to the corresponding reference matrices, and thus reduces the cross-session/subject variations. Used the non-target stimuli because the resting state in ERP, which needs that some labeled trials from the new subject, should be notable. That is, in ERP,

$$\bar{R} = \arg \min_R \sum_{i \in I} \delta^2(R, \Sigma_i), \quad (6)$$

Where I is that the index set of the non-target trials. RA-MDRM may be applied to each MI and ERP data; but, there's a crucial distinction in building variance matrices in these 2 paradigms.

III. DATASETS

This section introduces 2 MI datasets and one ERP dataset utilized in our experiments.

A. MI Datasets

Two MI datasets from BCI Competition IV1 were used. Their experimental paradigms were similar: In every session a theme weekday in a very comfy chair ahead of a laptop. At the start of an effort, a fixation cross appeared on the black screen to prompt the topic to be ready. An instant later, associate arrow inform to a particular direction was

bestowed as a visible cue for a couple of seconds. Then the visual cue disappeared from the screen and a brief break followed till future trial began. The primary dataset2 (Dataset one [5]) was recorded from seven healthy subjects. Continuous 59-channel EEG signals were uninherited for 3 phases: activity, evaluation, and special feature.

Each subject had one hundred trials from every category within the activity part. The second MI informationset3 (Dataset 2a) consisted of EEG data from 9 healthy subjects. Every subject was taught to perform four completely different MI tasks, particularly the imagination of the movement of the hand, right hand, both feet, and tongue. 22-channel EEG signals and 3-channel EOG signals were recorded at 250Hz. A coaching part associated an analysis part were recorded on completely different days for every subject. Here we tend to solely use the EEG information from the coaching part, including complete marker info. A causative band-pass filter (50-order linear part overacting window FIR filter designed by Matlab perform fir1, 6dB cut-off frequencies at [8, 30] Hz) was applied to get rid of muscle artefacts, line-noise contamination and DC drift. Next, we tend to extracted EEG signals between [0.5, 3.5] seconds when the cue look as our trials for each datasets. EEG signals between [4.25, 5.25] seconds when the cue look were extracted as resting states.

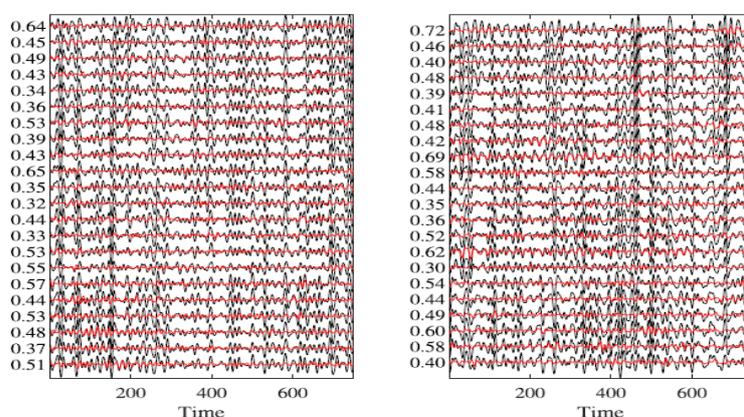


Fig. 1. EEG trials before (black curves) and after (red curves) EA. Each Row is a different channel.

B. ERP Dataset

We used associate RSVP dataset from PhysioNet4 [9] for ERP classification. It contained EEG information from eleven healthy subjects upon fast presentation of pictures at five, 6, and ten rates. Every subject was sitting ahead of a laptop showing a series of pictures apace. Footage the photographs} were aerial pictures of London falling into 2 classes, particularly target pictures and non-target pictures. Target pictures contained an at random turned and positioned heavier-than-air craft that had been picture realistically superimposed, and non-target pictures didn't contain airplanes.

For each presentation rate and subject there have been 2 sessions depicted by “a” and “b”, that indicated whether or not the primary image was “target” or “non-target”, severally. Here we tend to use the five rate version (five pictures per second) in Session a. the amount of samples for various subjects varied between 368 and 565, and therefore the target to non-target quantitative relation was around 1:9.

The continuous EEG information had been band pass filtered between [0.15, 28] Hz. we tend to down sampled the EEG signal from 2048Hz to 64Hz, and epoched every trial to the [0, 0.7] second interval time-locked to the information onset.

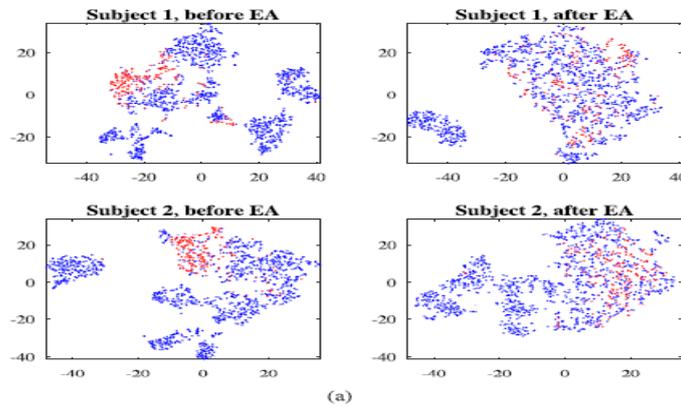


Fig. 2(a): *t*-SNE visualization of the first two subjects before and after EA. MI Dataset 1

C. DATA VISUALIZATION

It's fascinating to check however the graph trials are changed by EA. Fig. 1 shows two examples (one for hand imaging, and therefore the alternative for right) from Subject one in dataset 2a. The black and red curves EEG signals before and once EA, severally, and therefore the vertical axis numbers show their correlations. The magnitudes of the graph signals are smaller and additional uniform once Ea, and therefore the graph signals before and once EA usually have low correlation. To check however EA reduces individual variations, every time it tends to picked trials from one subject because the check set, and combined trials from all remaining subjects because the coaching set. Fig. 2(a) shows the *t*-SNE visual image of the primary 2 subjects in MI Dataset 1, every row akin to a unique check subject.

The red dots are trials from the check subject, and therefore the blue dots from the coaching subjects. In every row, the left plot shows the trials before Ea, and therefore the right once Ea. Corresponding visual image results for the primary 2 subjects in MI dataset 2a and ERP are shown in Figs. 2(b) and 2(c), severally.

IV. PERFORMANCE EVALUATION: OFFLINE UNSUPERVISED CLASSIFICATION

This section presents the performance comparison of Ea with alternative approaches on each MI and ERP datasets in offline unattended classification.

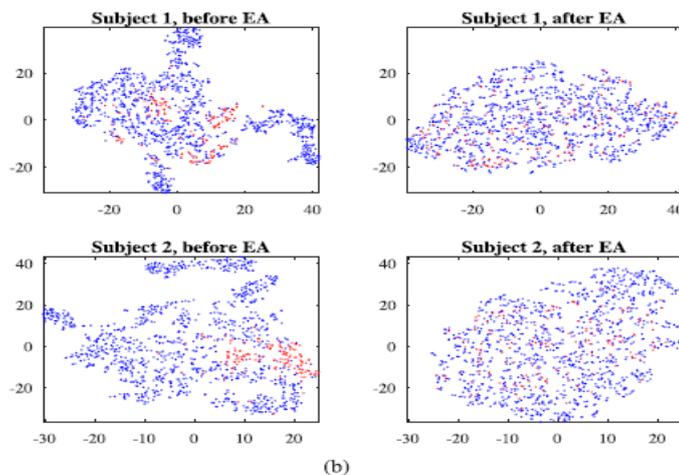


Fig. 2(b): *t*-SNE visualization of the first two subjects before and after EAMI Dataset 2a;

A. Offline Unsupervised Classification

In every dataset, there were multiple subjects, and every subject was 1st aligned severally, either within the Riemannian area victimization (5), or within the metric space victimization (11). Since we tend to had access to any or all graph

recordings in offline classification, all trials or resting epochs between all trials were used to estimate the reference matrices. We then used leave-one-subject-out cross validation to judge the classification performance: every time we tend to pick one subject because the new subject (test set), combined graph trails from all remaining subjects because the training set to make the classifier, then tested the classifier on the new subject.

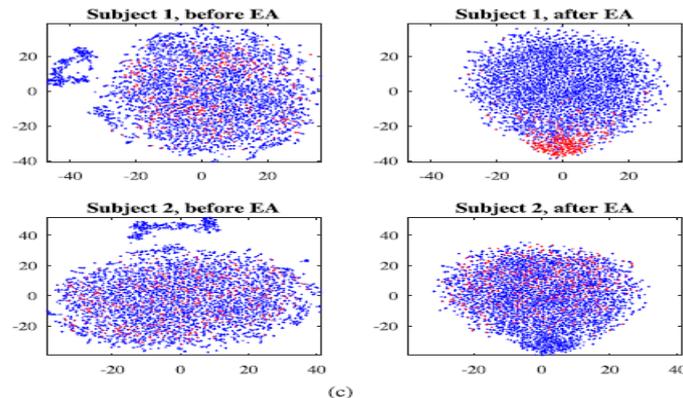


Fig. 2(c): *t*-SNE visualization of the first two subjects before and after EA. ERP. Red dots: trials from the test subject; blue dots: trials from the training subjects.

B. MI Datasets Offline Classification

We first tested Ea on the 2 MI datasets, and compared its performance with RA-MDRM. Within the metric space, after EA, we tend to use CSP [11] for abstraction filtering and LDA for classification. Additionally, the subsequent four approaches were compared:

- 1) MDRM: the essential MDRM classifier, as introduced in Section II-C. It doesn't embrace any information alignment.
- 2) RA-MDRM: It is that the approach introduced in Section II-D, that first aligns the variance matrices within the Riemannian area, and then performs MDRM.
- 3) CSP-LDA: it's a typical metric space classification approach for MI that spatially filters the graph trials by CSP then classifies them by LDA. It doesn't embrace any information alignment.
- 4) Ea-CSP-LDA: It first aligns the graph trials within the metric space by EA (Section III), then performs CSP filtering and LDA classification.

The classification accuracies of the four approaches are conferred in Fig. three that show that:

- 1) RA-MDRM outperformed MDRM on fifteen out of the sixteen subjects, suggesting that RA was effective.
- 2) EA-CSP-LDA additionally outperformed CSP-LDA on fourteen out of the sixteen subjects, suggesting that the projected Ea was additionally effective.
- 3) EA-CSP-LDA outperformed RA-MDRM on eleven out of the sixteen subjects, suggesting that the projected Ea, that permits the employment of a good variety of metric space signal process and machine learning approaches.

Finally, it's worth noting that for a tiny low variety of subjects (e.g., Subjects four and nine in Dataset 2a), Ea really degraded the classification accuracy. Some potential reasons are explained at the top of the paper, and can be investigated in our future analysis.

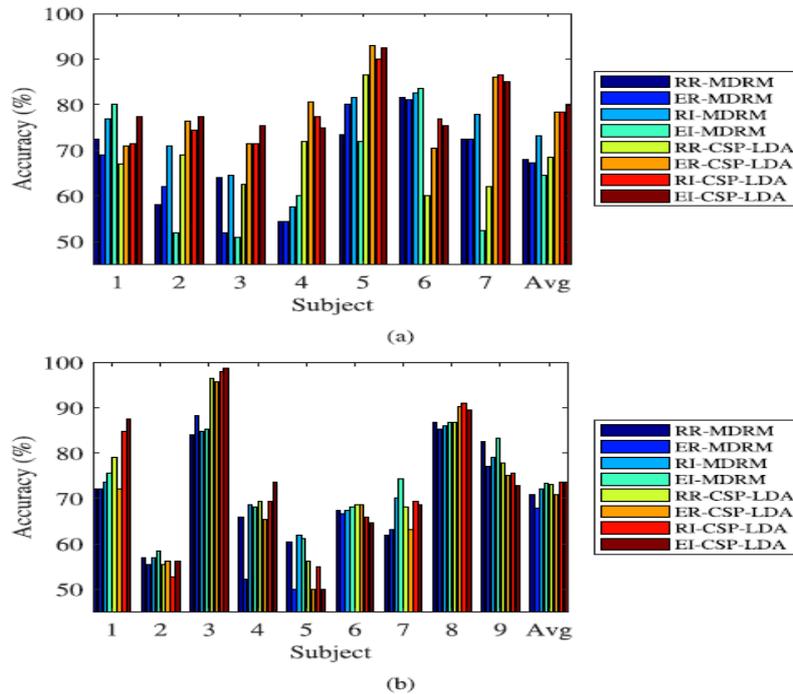


Fig. 3: Comparison of different reference matrices on the MI datasets. (a) Dataset 1 (b) Dataset 2a. RR: Riemannian mean of the resting trials; ER: Euclidean mean of the resting trials; RI: Riemannian mean of all Imagery trials; EI: Euclidean mean of all imagery trials.

D. Discussion: Different Choices of the Reference Matrix

Reference matrix estimation features a direct impact on the performance of the alignment algorithms. RA uses the Riemannian mean of the resting variance matrices for MI classification, and therefore the Riemannian mean of the non-target variance matrices for ERP classification [see (6)]. Ea estimates the reference matrix from all trials by (10), whose procedure is that the same for each MI and ERP classification.

In summary, the reference matrix is often calculable from 2 sorts of trials for MI classification: 1) the resting trials that the topic isn't activity any task; and, 2) the {imagination| representational process} trials that the topic is activity a motor imagery task. What is more, the reference matrix is often computed because the Riemannian mean or the geometer mean. Therefore we've got four potential combinations:

Riemannian mean of the resting trials (RR), geometer mean of the resting trials (ER), Riemannian mean of all imaging trials (RI), and geometer mean of all imaging trials (EI).

This subdivision compares the performances of the higher than four reference matrices. The results are shown in Figs. 5(a) and 5(b) for MI Datasets one and 2a, severally. They show that:

1) on the average RI-MDRM outperformed RR-MDRM, and EI-CSP-LDA outperformed ER-CSP-LDA, on each datasets, suggesting that estimating the reference matrix from all imaging trials would be higher than victimization all resting trials.

2) On the average across all sixteen subjects, EI achieved the most effective performance for CSP-LDA, and Little Rhody achieved the most effective performance for MDRM. This is in step with our expectation: MDRM operates within the Riemannian area, thence the Riemannian mean may provides a additional correct estimation of mean variance matrices than the geometer mean; on the opposite hand, CSP-LDA operates within the metric space, that the geometer mean sounds additional cheap.

3) On the average across all sixteen subjects, EI-CSP-LDA outperformed RI-MDRM, suggesting that Ea was advantageous to RA even after they each used the most effective reference matrix.

V. CONCLUSION

The automatic text detection technique is enforced from a picture for inpainting. The algorithmic program with success detects the text region from the image that consists of mixed text-picture-graphic regions. The algorithmic program are applied on several pictures and located that it with success notices the text region.

REFERENCES

- [1] A. Barachant *et al.*, "Multiclass brain-computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012.
- [2] A. Barachant and M. Congedo, "A plug & play P300 BCI using information geometry," 2014, arXiv:1409.0107.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
- [4] B. Blankertz *et al.*, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.
- [5] B. Blankertz *et al.*, "The non-invasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
- [6] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.
- [7] M. Congedo *et al.*, "A new generation of brain-computer interface based on Riemannian geometry," 2013, arXiv:1310.8115.
- [8] P. T. Fletcher and S. Joshi, "Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors," in *Lecture Notes in Computer Science*, vol. 3117, pp. 87–98, 2004.
- [9] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [10] A. Gretton *et al.*, "A kernel method for the two-sample-problem," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2007, pp. 513–520.
- [11] B. J. Lance *et al.*, "Brain-computer interface technologies in the coming decades," *Proc. IEEE*, vol. 100, no. 3, pp. 1585–1599, May 2012.
- [12] J. R. Wolpaw *et al.*, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.