Binary Models for Motor-Imagery Brain–Computer Interfaces: Sparse Random Projection and Binarized SVM

Michael Hersche, Luca Benini, Abbas Rahimi
Integrated Systems Laboratory, ETH Zurich, Switzerland
Emails: {hersche, benini, abbas}@iis.ee.ethz.ch

***Abstract*—Successful motor imagery brain–computer (MI-BCI) algorithms typically rely on a large number of features used in a classifier with real-valued weights that render them unsuitable for real-time execution on a resource-limited device. We propose a new method that randomly projects a large number of real-valued Riemannian covariance features to a binary space, where a linear SVM classifier can be learned with binary weights too. Flexibly increasing the dimension of binary embedding achieves almost the same accuracy (≤1.27% lower) compared to all models with float16 in 4-class and 3-class MI, yet delivering a more compact model with simpler operations to execute.**

***Index Terms*—EEG, motor imagery, embedding, sparse random projection, binarized SVM, Hamming distance.**

I. Introduction

Brain–computer interfaces (BCIs) enable a communication channel between a user and an external device through intentional modulation of brain signals, e.g., motor imagery (MI) movement of a part of body [1]. A BCI aims to recognize human intentions from the analysis of spatiotemporal neural activity, typically recorded non-invasively by a number of electroencephalogram (EEG) electrodes. Such information can enable control games [2], [3], drive a wheelchair [4], and even motor rehabilitation after stroke [5].

Accurate EEG decoding of MI is a challenging task due to inter- and intra-subject variabilities [6], [7]. To deal with the high variability of EEG signals between subjects, most approaches train a personalized model per subject [8]–[11]. Some approaches use well-known common spatial patterns [10], or Riemannian covariance features [11] followed by a SVM or LDA classifier, while others such as Shallow ConvNet [8] strive to learn end-to-end representation using deep learning algorithms. Among them unsupervised Riemannian features with linear SVM classifier [9] achieve the highest average classification accuracy (75.47%) among nine subjects on the 4-class BCI competition IV-2a dataset [12]. This multiscale approach however extracts a large number of real-valued Riemannian covariance features increasing the number of weights and complexity of a classifier.

One viable option is to transform those features to binary space with distance-preserving methods such as sparse random projection [13]. Interestingly, the weights in the matrix of random projection do not need to be stored (i.e., can be *rematerialzed* by a random function on the fly), or can be realized by emerging memristor [14] and optical [15] devices. A readout function layer can then effectively analyze the projected features for various classification tasks, e.g., in EEG [16], electrocardiography (ECG) signals [17], [18], and electrocorticography (ECoG) [19].

In this paper, we propose to embed real-valued Riemannian covariance features effectively to d-dimensional binary Hamming space using bipolar sparse random projections. In the binary space, a linear SVM is trained and, for the first time, binarized such that classification is solely based on computationally efficient Hamming distance calculations. Our method is evaluated on multi-spectral real-valued Riemannian features from the 4-class MI-dataset of the BCI competition IV-2a [12] as well as a new 3-class MI-dataset based on experiments in [2]. Maintaining the dimension of original real-valued features, our method binarizes them alongside SVM weights, and achieves 7.92% and 2.00% higher accuracy in the datasets compared to plain binarized SVM without random projection. Our method allows to flexibly increase the dimension of binary space to compensate for accuracy loss yet delivering a more compact and simpler model to execute: increasing the dimension by 5.7× surpasses the accuracy of SVM float16 and results in 1.92× smaller model for 3-class; similarly, 9.2× increased dimension achieves 1.31× smaller model than the Shallow ConvNet at the same accuracy, and 1.51× smaller model than SVM float16 at 1.27% lower accuracy for 4-class. We also provide the source code[1].

II. Methods

This section presents the main contribution of the paper, which is to classify MI-EEG signals using a multi-spectral Riemannian feature extractor, random projection, and a binarized SVM, shown in Fig. 1.

A. Riemannian Covariance Features

Traditional BCIs extract features from the EEG signal using spectral-power features in connection with a spatial filter, which is better known as filter bank common spatial pattern (FBCSP) [10]. Here we use a more recent approach which is to directly manipulate spatial EEG covariance matrices using the dedicated Riemannian geometry [20]. First, we estimate the covariance matrix $\mathbf{C}$ from the multi-channel EEG signal

This project was supported in part by ETH Research Grant 09 18-2, by the Hasler Foundation under project no. 18082, and by EU's H2020 under grant no. 780215.

[1]Code available under https://github.com/MHersche/HDembedding-BCI

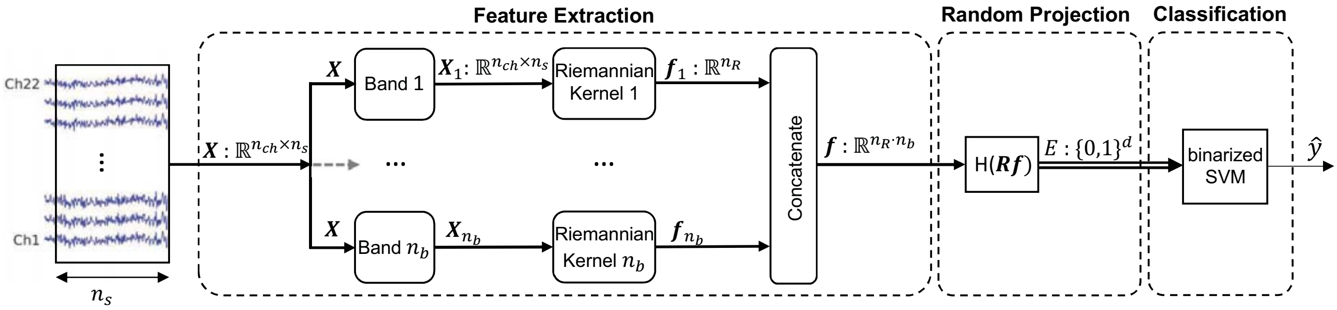


Fig. 1: Overall architecture for binarized learning and classification of EEG signals with Feature Extraction, Random Projection, and Classification. The EEG signal $\boldsymbol{X}$ of one temporal window with n_s samples and n_{ch} channels is processed at the time. Every EEG channel is divided into n_b frequency bands ($b_1 - b_{n_b}$) using second order Butterworth band pass filters. A Riemannian covariance kernel computes spatial energy features which are concatenated and binarized using sparse random projection. Binary features E are classified with a binarized SVM.

$\mathbf{X}$ with n_{ch} channels and n_s samples:

$$\boldsymbol{C} = \frac{1}{n_s - 1}(\boldsymbol{X}\boldsymbol{X}^T + \alpha \boldsymbol{I}_{n_{ch}}), \tag{1}$$

where $\boldsymbol{I}_{n_{ch}}$ is the $n_{ch} \times n_{ch}$ identity matrix and α a regularization constant ensuring positive definiteness of the estimated covariance matrices set to 0.1. The Riemannian kernel K calculates $n_R = n_{ch}(n_{ch}+1)/2$ output features based on the input covariance matrix $\mathbf{C}$:

$$K : \mathbb{R}^{n_{ch}\times n_{ch}} \to \mathbb{R}^{n_R}, \tag{2}$$

and is defined as

$$\boldsymbol{f} = \text{vect}\Big(\text{logm}\Big(\boldsymbol{C}_{ref}^{-1/2}\boldsymbol{C}\boldsymbol{C}_{ref}^{-1/2}\Big)\Big), \tag{3}$$

where logm(.) is the matrix logarithm and vect(.) the ℓ_2-norm preserving half vectorization of a matrix. The reference covariance matrix $\mathbf{C}_{ref}$ is the average over all covariance matrices in the training set. The Riemannian kernel does not need labeled data and is therefore unsupervised.

In analogy to frequency band common spatial pattern (FBCSP), the set of Riemannian features is extended to multi-spectral features by using multiple Riemannian kernels on different frequency bands of the multi-channel EEG signal. The signal is divided into multiple frequency bands using a filter bank. A separate Riemannian kernel is used with $\mathbf{C}_{ref}$ computed solely on the corresponding frequency band. A recent work [9] with high classification accuracy suggests to use 43 overlapping frequency bands within the 4–40 Hz band with bandwidths varying between 2–32 Hz.

B. Sparse Bipolar Random Projection

An embedding is a representation for which the computation of distances directly gives an estimate of the distances in their initial representation [13]. The building of such representations is provided by binary locality-sensitive hashing (LSH) functions, which ensure that similar elements are statistically likely to be embedded into the same value [17]. Once mapped to the binary Hamming space, the similarity is computed with the Hamming distance.

Here, we use random projections to embed real valued feature vectors to the binary Hamming space. Random projections are usually used for dimensionality reduction in the Euclidean space [21]. The Johnson-Lindenstrauss lemma [22] ensures distances between two points in the projected space to be preserved if the output dimension is suitably high. Such projections deal with embeddings between Euclidean spaces. However, here the data is projected to a high-dimensional Hamming space. Recently, it has been shown [13] that random projections can indeed project data to a high-dimensional Hamming space while preserving the distance between points with success in monitoring arterial blood pressure via ECG signals [17], [18].

Random projection to binary space is defined as

$$E = \text{H}(\boldsymbol{R}\boldsymbol{f}), \tag{4}$$

where H(.) is the component-wise Heavyside step function and $\boldsymbol{R} \in \mathbb{R}^{d\times n_R}$ the projection matrix [13]. Usually, the components $r_{i,j}$ of $\boldsymbol{R}$ are drawn from an i.i.d. Gaussian normal distribution ($r_{i,j} \sim \mathcal{N}(0,1)$). However, the Gaussian projection matrix can be replaced by a much simpler one such as the sparse bipolar random matrix [23]:

$$r_{i,j} = \begin{cases} +1 & \text{with probability } s/2 \\ 0 & \text{with probability } 1-s \\ -1 & \text{with probability } s/2, \end{cases} \tag{5}$$

where $s \in [0,1]$ is the relative sparsity. Achlioptas [23] has shown that by using a sparsity of $s = 2/3$ this projection comes without any sacrifice in the quality of embedding compared to the plain Gaussian projection. Moreover, this distribution reduces the computational cost, since the projection requires only additions and subtractions. The sparsity s allows to further reduce the number of operations as many entries of the projection matrix are zero. Furthermore, the projection matrix does not need to be restored from a memory as it can be cheaply recomputed. The same projection matrix is also shared among all the frequency bands to reduce the memory footprint during a full parallel projection. In our case we could reduce the sparsity to $s = 1/10$ without loosing performance in classification accuracy.

Random entries of the projection matrix do not need to be stored permanently, but can be efficiently *regenerated during operation* with a random number generator. This process is repeatable and requires an arbitrary seed, which requires negligible 32-bit storage. Thus, the use of random projections is not increasing the memory footprint for storing a model on an embedded device [24].

C. Binarized SVM

This section describes how linear SVM is binarized to do binary inference solely based on Hamming distance computations. The decision function of the original linear SVM without bias is defined as

$$\hat{y} = \underset{i=1,\dots,n_{cl}}{\text{argmax}} < \mathbf{w}_i, \mathbf{f} >, \tag{6}$$

where $\mathbf{w}_i \in \mathbb{R}^d$ is the learned support vector of class i. For training the linear SVM—still in full float32 precision—on binary features, we map all elements in $E \in \{0,1\}^d$ to bipolar values $\{-1,1\}^d$. The learned support vectors are then binarized using the component-wise Heavyside step function:

$$W_i = \text{H}(\mathbf{w}_i) \quad i = 1, ..., n_{cl} \tag{7}$$

During inference, the binarized SVM classifies a binary vector E by searching for the binary support vector with smallest Hamming distance to E:

$$\hat{y} = \underset{i=1,\dots,n_{cl}}{\text{argmin}} \ \text{d}_H(W_i, E), \tag{8}$$

where $\text{d}_H(.)$ is the Hamming distance.

III. Experimental Results

In this section, we assess the proposed methods on the 4-class MI dataset from BCI competition IV2a [12] and on an additional 3-class MI dataset [2]. An ℓ_2-regularized linear SVM performed best on the 4-class dataset with multi-spectral Riemannian features [9] and serves as baseline classifier. An LDA with automatic shrinkage, commonly used in EEG classification [25], is used as second baseline. We measure the classification accuracy as the ratio between correct classified trials over the total number of trials.

A. Dataset description

a) 4-class MI dataset. The BCI Competition IV-2a dataset [12] consists of EEG data from 9 different subjects with four different MI tasks, namely the imagination of the movement of the left hand, right hand, both feet, and tongue. Two sessions were recorded on two different days. For each subject a session consists of 72 trials per class yielding 288 trials in total. One session is used for training and the other for testing exclusively. The signal was recorded with 22 EEG electrodes, bandpass filtered between 0.5 Hz and 100 Hz, and sampled with 250 Hz. The Riemannian feature extraction uses $n_b = 43$ overlapping frequency bands in the range between 4–40 Hz with bandwidths varying within 2–32 Hz described in [9]. When using 22 EEG channels the number of Riemannian features per frequency band is $n_R = 22(22+1)/2 = 253$, which gives a total of $n_b \cdot n_R = 43 \cdot 253 = 10,879$ features.

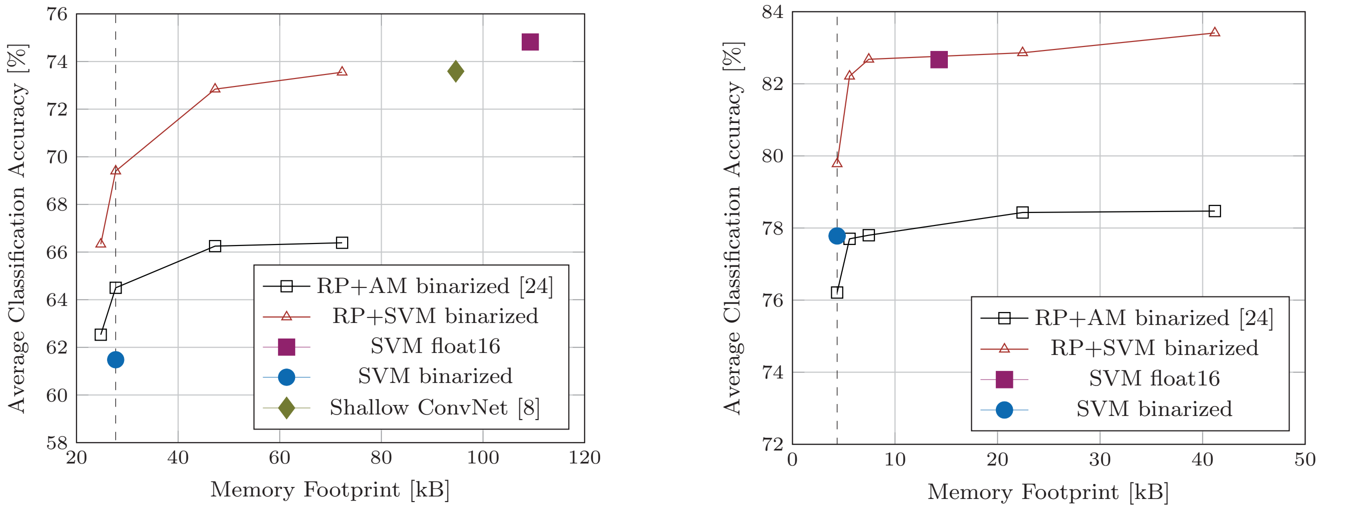
b) 3-class MI dataset. The second dataset is based on the study in [2], and consists of 3-class motor imagery (left hand, right hand, and both feet) from 5 different subjects. Every subject participated in four sessions recorded on the same day with 45 trials per session. The signal was recorded with 16 EEG electrodes with a sampling frequency of 512 Hz. For testing, 4-fold cross validation is applied, using three sessions for training and one for testing. In this dataset, we find that using only $n_b = 13$ frequency bands in the range of 4–30 Hz with bandwidth 2 Hz and a temporal window of 4 s length are sufficient. The use of more frequency bands does not hurt the classification performance, however, it increases the computation time. The resulting number of Riemannian features per frequency band is $n_R = 16(16+1)/2 = 136$ and the total number of features $1,768$. The highest classification accuracy reported on this dataset is 76% using power spectral density features with a Gaussian classifier [2].

TABLE I: Summary of best classification accuracies (%) on the 4-class MI BCI competition IV2a test set and the 4-fold cross-validation accuracy on a 3-class MI data set. Baseline classifier with float16 precision are a ℓ_2-regularized linear SVM (c = 0.1) and LDA with automatic shrinkage. All classifier are fed with the same multi-spectral Riemannian features. Classifier and features are either directly binarized in original space (SVM binarized and LDA binarized) or features are projected to d-dimensional binary Hamming space and classified with binarized SVM (RP+SVM binarized). p-value reports the significance of Wilcoxon signed-rank test with respect to linear SVM with float16 precision.

		SVM float16	LDA float16	SVM binarized	LDA binarized	RP+SVM binarized
	d	10,879	10,879	10,879	10,879	100,000
4-class MI[(a)]	A1	91.81	88.26	78.65	78.29	90.46
	A2	51.59	58.66	45.58	44.88	53.96
	A3	83.52	82.78	68.13	71.79	79.16
	A4	73.25	53.51	57.89	56.58	71.49
	A5	63.41	59.42	42.03	40.94	65.18
	A6	59.07	57.21	47.91	50.23	56.98
	A7	86.64	89.53	71.12	73.29	82.42
	A8	81.55	81.92	71.59	75.65	79.63
	A9	82.58	77.65	70.45	73.86	82.65
	Avg	**74.82**	**72.10**	**61.48**	**62.83**	**73.55**
	Std	12.37	13.10	12.01	13.09	11.17
	p-value	-	0.260	0.008	0.008	0.214
	d	1,768	1,768	1,768	1,768	10,000
3-class MI[(b)]	B1	77.22	81.67	66.67	68.89	78.17
	B2	82.78	81.67	78.33	78.33	81.89
	B3	89.44	90.56	86.11	83.89	86.61
	B4	98.33	98.33	95.00	94.44	97.50
	B5	65.56	68.33	62.78	61.67	69.22
	Avg	**82.67**	**84.11**	**77.78**	**77.44**	**82.68**
	Std	10.11	9.17	10.91	10.43	8.54
	p-value	-	0.144	0.042	0.042	0.893

B. MI classification

Table I compares the classification accuracy on the 4- and 3-class datasets for float16 precision linear SVM and LDA with different binary classifiers. The linear SVM achieves 74.82% and 82.67% average classification accuracy on the 4- and 3-class dataset, respectively. Similar results are observed with LDA, where it performs particularly best on 3-class



(a) 4-class MI

(b) 3-class MI

Fig. 2: Average classification accuracy (%) vs. memory footprint for the whole model using different classifiers. Proposed random projection with binarized SVM (RP+SVM binarized) is compared with another binarized classifier with associative memory learning [24] (RP+AM binarized), a float16 SVM on original features, a binarized SVM on original but binarized features (SVM binarized), and Shallow ConvNet [8]. All non-binary parameters are in float16, including Riemannian feature extractor and Shallow ConvNet weights.

MI with highest accuracy of 84.11%, which is an improvement of 8.11% compared to state-of-the-art accuracy on this dataset [2].

In a first step, the baseline classifier and features are binarized in their original space applying the Heavyside step function directly on features and support vectors. This results in significant loss of 13.34% and 4.89% in accuracy for binarized SVM, and for binarized LDA 9.27% and 6.67%, relative to their corresponding float16 classifier. However, this performance loss due to binarized classification can be recovered when applying our proposed method using random projection to binary Hamming space and binarized SVM. When comparing to SVM with float16, our RP+SVM binarized is only 1.27% lower in 4-class MI and even the same on 3-class MI.

C. Memory footprint

Fig. 2 compares the performance of all classifier, considering not only the classification accuracy but also the memory footprint required to store learned parameters of the whole model. Here, we include another binarized classifier [24] which uses random projections as well, but encodes the projected binary vectors per frequency band using holographic superposition (RP+AM binarized). The binary vectors are classified using an associate memory (AM). Moreover, the Shallow ConvNet [8] with 47,324 float16 parameters is considered in 4-class MI task.

The output dimension of the random projection is varied between $d = 1000 - 100,000$, which has direct impact on the required memory footprint. Generally, accuracy of binarized classifier improves significantly in higher dimensions, especially when using the proposed binarized SVM readout. When fixing the dimension to the number of Riemannian features in 4-class MI (i.e., $d = 10,879$ or memory footprint of 27.71 kB), the simple SVM binarized achieves lower accuracy compared to both RP methods. This supports the necessity of RP when doing binarized classification. At memory footprint of 72.27 kB ($d = 100,000$), RP+SVM binarized achieves an accuracy of 73.55% which is 1.27% lower than float16 SVM, but it requires 1.51× lower memory footprint. Compared to Shallow ConvNet with 73.59% accuracy and 94.65 kB memory footprint, RP+SVM binarized reduces the memory footprint by 1.31× at the same accuracy.

Similar trends are observed on 3-class MI, however, here RP+SVM binarized outperforms simple SVM binarized at the same memory footprint. On this dataset, RP+SVM binarized at memory footprint of 7.44 kB ($d = 10,000$) even achieves the same accuracy as SVM float16, but memory footprint is reduced by 1.92×.

IV. Conclusion

In this paper, we propose to *binarize* multi-spectral Riemannian features from MI-EEG signals with sparse bipolar random projection and to do classification with binarized SVM readout, which boils down to computational efficient Hamming distance calculation. Experimental results show that our method binarizes real-valued features in the same dimensionality with 7.42% accuracy loss for 4-class task MI, and 5.74% for 3-class MI task, compared to SVM models in float16. Further increasing the dimensionality in binary space improves the accuracy of the binary model, which results in 1.27% lower accuracy but at 1.31× lower memory footprint in 4-class MI, and the same accuracy but 1.92× lower memory footprint in 3-class MI.

References

[1] R. A. Ramadan and A. V. Vasilakos, "Brain computer interface: control signals review," *Neurocomputing*, vol. 223, pp. 26–44, 2017.

[2] S. Saeedi, R. Chavarriaga, R. Leeb, and J. D. R. Millan, "Adaptive Assistance for Brain-Computer Interfaces by Online Prediction of Command Reliability," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 32–39, 2016.

[3] S. Perdikis, L. Tonin, S. Saeedi, C. Schneider, and J. d. R. Millán, "The Cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users," *PLOS Biology*, vol. 16, no. 5, p. e2003787, 2018.

[4] T. Carlson and J. Del R. Millan, "Brain-controlled wheelchairs: A robotic architecture," *IEEE Robotics and Automation Magazine*, vol. 20, no. 1, pp. 65–73, 2013.

[5] A. Ramos-Murguialday, D. Broetz, M. Rea, L. Läer, Yilmaz, F. L. Brasil, G. Liberati, M. R. Curado, E. Garcia-Cossio, A. Vyziotis *et al.*, "Brain-machine interface in chronic stroke rehabilitation: A controlled study," *Annals of Neurology*, vol. 74, no. 1, pp. 100–108, 2013.

[6] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz *et al.*, "Review of the BCI Competition IV." *Frontiers in neuroscience*, vol. 6, p. 55, 2012.

[7] 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," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[8] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[9] M. Hersche, T. Rellstab, P. D. Schiavone, L. Cavigelli, L. Benini, and A. Rahimi, "Fast and Accurate Multiclass Inference for MI-BCIs Using Large Multiscale Temporal and Spectral Features," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 1690–1694.

[10] Kai Keng Ang, Zhang Yang Chin, Haihong Zhang, and Cuntai 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.

[11] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, 2013.

[12] C. Brunner, R. Leeb, G. R. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008 - Graz data set A," http://bnci-horizon-2020.eu/database/data-sets.

[13] D. A. Rachkovskij, "Binary Vectors for Fast Distance and Similarity Estimation," *Cybernetics and Systems Analysis*, vol. 53, no. 1, pp. 138–156, 2017.

[14] C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu, "Reservoir computing using dynamic memristors for temporal information processing," *Nature Communications*, vol. 8, no. 1, p. 2204, 2017.

[15] A. Saade, F. Caltagirone, I. Carron, L. Daudet, A. Dremeau, S. Gigan, and F. Krzakala, "Random projections through multiple optical scattering: Approximating Kernels at the speed of light," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 6215–6219.

[16] P. Tan, W. Sa, and L. Yu, "Applying Extreme Learning Machine to classification of EEG BCI," in *2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE, 2016, pp. 228–232.

[17] Y. B. Kim and U.-M. O'Reilly, "Large-scale physiological waveform retrieval via locality-sensitive hashing," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 5829–5833.

[18] ——, "Analysis of locality-sensitive hashing for fast critical event prediction on physiological time series," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 783–787.

[19] X.-m. Zhang, Y.-x. Dai, X.-b. Xu, and T.-t. He, "Binary Classification on ECoG Signals Using Optimized Extremely Learning Machine," *DEStech Transactions on Computer Science and Engineering*, pp. 521–531, 2017.

[20] 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.

[21] E. Bingham and H. Mannila, "Random projection in dimensionality reduction," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*. New York, New York, USA: ACM Press, 2001, pp. 245–250.

[22] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Contemporary mathematics*, vol. 26, no. 1, pp. 189–206, 1984.

[23] D. Achlioptas and Dimitris, "Database-friendly random projections," in *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '01*. New York, New York, USA: ACM Press, 2001, pp. 274–281.

[24] M. Hersche, J. d. R. Millán, L. Benini, and A. Rahimi, "Exploring Embedding Methods in Binary Hyperdimensional Computing: A Case Study for Motor-Imagery based Brain-Computer Interfaces," *arXiv preprint arXiv:1812.05705*, 2018.

[25] F. Lotte and Cuntai 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, 2011.