ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Non-Linear Activation Functions for Deep Riemannian Neural Networks

Lucas H. dos Santos[1], João G. P. Barbon[2], Sylvain Chevallier [3], Denis G. Fantinato[2] *

1- CMCC - Federal University of ABC
Av. dos Estados, 5001 Santo André - Brazil

2- FEEC - Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 400 Campinas - Brazil
3 - TAU, LISN, University Paris-Saclay
rue René Thom, 01190 Gif-sur-Yvette - France

Abstract. In the context of EEG-based Brain-Computer Interfaces (BCIs), Deep Riemannian Neural Networks (DRNNs) have emerged as a state-of-the-art framework, particularly in classifying motor imagery. A crucial component of these networks is the activation function, which must preserve the manifold's geometry. The ReEig function is the prevailing choice, providing a foundational but potentially limited nonlinear transformation. This work investigates whether alternative activation functions can improve the performance of DRNNs. We conduct a comparative analysis of the standard ReEig function against four alternatives – cosh, sinh, ReLU, and SiLU – within the SPDNet and EE(G)-SPDNet architectures. The experiments are performed on three public motor imagery datasets: BCI Competition IV2a, PhysioNetMI, and Cho. The results consistently indicate that alternative nonlinear functions perform better than the conventionally used ReEig, achieving superior classification accuracy.

1 Introduction

Brain-Computer Interfaces (BCIs) based on electroencephalography (EEG) have emerged as a promising technology for a direct communication between the human brain and external devices [1]. In the context of EEG-based BCIs for motor imagery (MI) tasks, a crucial stage is the classification of neural patterns, in which traditional machine learning approaches often rely on handcrafted features extracted from the EEG signals. A possible approach is to extract EEG covariance matrices, which are inherently symmetric positive definite (SPD) and thus reside on a smooth Riemannian manifold. Within this geometric structure, techniques grounded in Riemannian Geometry (RG) have demonstrated remarkable performance by operating directly on the manifold of SPD matrices [2], preserving the intrinsic data structure and leading to more robust classifiers. Recently, this paradigm has been improved by the development of Deep Riemannian Neural Networks (DRNNs), which combine the geometric fidelity of Riemannian approaches with the powerful hierarchical representation learning of deep learning [3]. These networks, such as SPDNet [4], map SPD matrices through a series

*This study was financed by the São Paulo Research Foundation (FAPESP), Brazil. Process Number #2020/09838-0, #2023/00640-1 and #2023/00049-1.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

of Riemannian layers, enabling end-to-end learning on the manifold and pushing the state-of-the-art in EEG decoding accuracy [3].

While DRNNs represent a significant leap forward, their architectural components, particularly the activation functions, are not yet fully explored. A critical element of these networks is the need for activation functions that maintain the manifold structure of the data, as standard nonlinearities like ReLU would violate the SPD property. Currently, the ReEig (rectifying eigenvalues) function is widely adopted as the main nonlinearity in DRNNs, introducing a nonlinear capacity by thresholding the eigenvalues of the SPD matrix. However, the ReEig function provides a relatively limited form of nonlinear transformation, potentially constraining the network's ability to learn complex, discriminative features on the manifold. In that sense, in this work, we intend to compare ReEig with a set of four alternative activation functions for SPDNet and EE(G)-SPDNet: cosh, sinh, ReLU and SiLU. The accuracy for each case is evaluated on three public MI datasets: BCI Competition IV 2a [5], PhysioNetMI [6] and Cho [7].

2 Deep Riemannian Neural Networks (DRNNs)

EEG signals are usually represented as multivariate time series $\mathbf{X} \in \mathbb{R}^{c \times T}$, composed of c electrodes (or channels) and T time samples [8]. The sample covariance estimator (SCM) estimates its covariance matrix $\mathbf{C}_{\mathbf{X}}$, which is an element of the manifold of the set of Symmetric Positive-Definite (SPD) matrices: $\mathcal{S}_c^{++} = \{\mathbf{C} \in \mathbb{R}^{c \times c} : \mathbf{C} = \mathbf{C}^\top, \mathbf{C} \succ 0\}$ that forms a well-defined Riemannian geometry (RG) [9]. The analysis of EEG signals with RG showed to be very promising [2] and inspired the construction of Deep Riemannian Neural Networks (DRNNs), a family of deep learning architectures based on data residing in the SPD manifold. The seminal SPDNet [4] introduced layers that preserve such manifold structure: BiMap, ReEig and LogEig.

Analogously to a fully connected layer, the BiMap uses a bilinear transformation $\mathbf{X}_{out} = \mathbf{W}\mathbf{X}_{in}\mathbf{W}^\top$ with $\mathbf{W}$ constrained to be semi-orthogonal, preserving the SPD structure. The ReEig layer is applied as a non-linear activation by clamping eigenvalues below a small threshold ϵ, preventing numerical loss of positive definiteness. Lastly, the LogEig layer computes the matrix logarithm, $\mathbf{X}_{out} = \mathbf{Q}\log(\mathbf{\Lambda})\mathbf{Q}^\top$, mapping the SPD matrix to the tangent space, allowing the learned geometric features to be processed by standard Euclidean-based components.

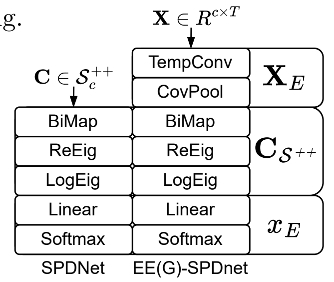

Fig. 1: Architecture of the analyzed DRNNs.

Models such as EE(G)-SPDNet [3], described in Figure 1, extended the SPDNet architecture by adding temporal and spatial modules preceding the SPD blocks. In both those the authors also repeat the layers BiMap and ReEig, with EE(G)-SPDNet building a sequence of three blocks of this repetition. However, by relying on the same ReEig "activation" many DRNNs architectures might have a limited nonlinear capacity. This motivate us to explore more impactful

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

SPD-preserving activation functions.

3 Activation Functions

In Euclidean deep learning, activation functions are fundamental for a network's ability to learn complex, non-linear patterns [10, 11]. For the SPD manifold, the commonly used ReEig layer (eigenvalue clamping) has often been described as an activation layer. However, in architectures based on the BiMap, for instance, there is an enforced positive definiteness by construction, reducing ReEig to serve mainly a numerical role, preventing loss of positive definiteness due to floating-point errors, rather than introducing the nonlinearity expected from this class of layers. Furthermore, there is a wide range of available activations in the Euclidean domain whose applications are effective and widely employed, but have seen few to none use in DRNNs.

The core of spectral operations as ReEig and LogEig is due to eigendecomposition of the symmetric matrix $\mathbf{C}$ from which the eigenvalues are modified: $\mathbf{C} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$, where $\mathbf{Q}$ is an orthogonal matrix of eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues. Since only the eigenvalues are modified while the eigenvectors remain unchanged, the geometric structure of the data is preserved. However, acting only on eigenvalues can limit the power of the DRNNS, since the goal is to apply nonlinearities to all the entries, analogous to Euclidean activations.

An alternative is an elementwise (Hadamard) activation function that provably preserves positive semi-definiteness (PSD). From Schoenberg's theorem, an elementwise function $f : \mathbb{R} \to \mathbb{R}$ preserves PSD if and only if its Maclaurin series ($f(t) = \sum_{k=0}^{\infty} a_k t^k$) has nonnegative coefficients ($a_k \geq 0$). Consequently, for any PSD matrix $\mathbf{X}$ the elementwise application is $f^\circ(\mathbf{X}) = [f(X_{ij})] = \sum_{k=0}^{\infty} a_k \mathbf{X}^{\circ k} \succeq 0$ where $\mathbf{X}^{\circ k}$ is the k-th Hadamard power. For SPD matrices ($\mathbf{X} = \mathbf{X}^\top$ and $\mathbf{X} \succ 0$), a suitable choice of f ensures the eigenvalues remain strictly positive, preserving the symmetric positive definiteness. Since the Schur product theorem is also valid for SPD matrices, we see that each term $\mathbf{X}^{\circ k}$ is SPD and, consequently, the above series $f(\mathbf{X})$ remains SPD. These theorems provide a practical family of SPD-preserving elementwise activations for DRNNs, such as exp, sinh and cosh [12]. This also demonstrates that functions such as cos or tanh, which have negative coefficients, are not suitable for DRNNs as they cannot preserve PSD.

Another approach is to rely onto the logarithm and exponential map from RG. These enable pushing covariance matrices into the tangent space (with the LogEig layer, for instance), applying an activation function, and mapping back to the Riemannian manifold (with the exponential map). By definition, this is guaranteed to maintain the positive definiteness, while also allowing the use of many known activation functions.

Motivated by these results, we analyze the replacement of the ReEig layer to four alternative functions: sinh, cosh, ReLU [13], and SiLU (Swish) [14]. Although sinh and cosh are theoretically SPD-preserving according to Schoenberg's theorem, their actual impact on DRNNs remains unexplored. ReLU is included as the main reference in the literature related to activations in Euclidean deep learning models, and SiLU is one of its variations, which has shown superior performance in certain scenarios [14]. We also explore the use of diagonal loading ($\mathbf{C}_{DiagLoad} = \mathbf{C} + \epsilon\mathbf{I}$, where $\epsilon = 1e^{-4}$), equivalent to no activation function and only serving to maintain positive definiteness.

By modifying SPDNet and EE(G)-SPDNet, we aim to evaluate how truly nonlinear and SPD-preserving activations affect the performance of DRNNs across multiple Motor Imagery datasets. Due to numerical instabilities (mainly bad convergence), we also modified both DRNNs to have only one BiMap-ReEig block.

4 Datasets

We evaluated the proposed modifications on three public motor imagery (MI) EEG datasets: BCI Competition IV2a (BNCI2014-001), PhysioNetMI and Cho. These datasets contain recordings from 9, 109, and 52 subjects, respectively. Due to anomalies or missing trial in the recordings, 6 PhysioNetMI subjects were excluded [15]) and 3 Cho subjects were also removed [15]). Each subject performed multiple MI tasks (e.g., left/right hand movements), resulting in 4 classes for BNCI2014-001, 5 for Physionet (including resting state) and 2 for Cho dataset.

Following benchmarking practices [16], the full trial duration was used. Data were band-pass filtered between 8-32 Hz and channel-wise normalized. A cross-session evaluation was used on BNCI2014-001, while within-session was used on PhysioNetMI and Cho. All experiments were repeated with three random seeds, and results are reported as averages. Training was conducted using the Riemannian ADAM optimizer, with a batch size of 64 and 300 epochs, keeping hyperparameters consistent across subjects and datasets.

5 Results

Table 1 shows the results obtained in the experiments. As expected, diagonal loading and ReEig achieved similar accuracies, both mainly ensuring numerical stability rather than adding non-linearity. This confirms that within BiMap based architectures, the ReEig layer acts mainly as a stabilization mechanism.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Model	Activation	BNCI2014-001	Cho2017	PhysionetMI	Avg.
SPDNet	ReEig	52.89	61.06	52.19	55.38
	DiagLoad	53.67	61.12	52.41	55.73
	Cosh	54.54	**64.26**	**54.09**	**57.63**
	Sinh	**54.84**	63.55	53.51	57.30
	ReLU	50.37	58.11	48.73	52.40
	SiLU	52.64	60.79	51.41	54.95
EE(G)-SPDNet	ReEig	70.52	60.71	47.24	59.49
	DiagLoad	70.87	61.04	47.19	59.70
	Cosh	64.04	**68.36**	**53.57**	**61.99**
	Sinh	**71.78**	64.39	49.33	61.83
	ReLU	69.27	49.67	36.22	51.72
	SiLU	70.41	56.17	44.93	57.17

Table 1: Accuracy (%) comparison for SPDNet and EE(G)-SPDNet.

The best overall results were obtained with cosh for both SPDNet and EEG(G)-SPDNet models, with a single exception where sinh outperformed it for the latter model on the BCI Competition IV 2a dataset. Otherwise, the results remained consistently competitive across all datasets, leading to an overall average performance close to that of cosh. This consistency is further illustrated in Figure 2, particularly for the Cho2017 and PhysionetMI datasets, where cosh and sinh demonstrated higher mean accuracies compared to the other activation functions. This behavior suggests that the SPD-preserving and nonlinearities introduced by these activations not only improved performance but also contributed to a more stable training.

In contrast, ReLU and SiLU underperformed across all datasets, specially for ReLU on PhysionetMI, which achieved the lowest overall accuracy, 36.22%. In some other cases, their performance collapsed into a narrow range (Figure 2), possibly indicating that the chosen hyperparameters were suboptimal for achieving a good convergence on those cases.

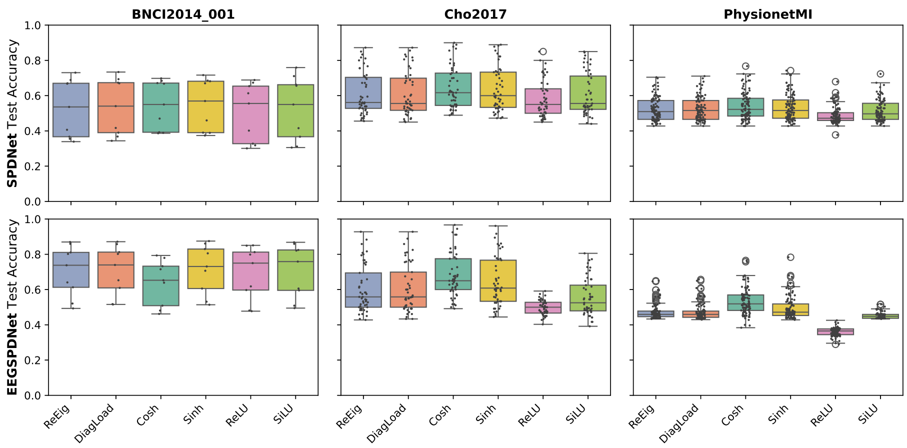

Fig. 2: Accuracy (%) distribution by activation function.

6 Conclusion

This work investigated through a comparative analysis whether alternative activation functions can improve the performance of DRNNs. Four alternatives for ReEig function were considered for the SPDNet and EE(G)-SPDNet architectures: cosh, sinh, ReLU, and SiLU. The results obtained on three public motor imagery datasets consistently indicate that alternative nonlinear functions perform better than the conventionally used ReEig, achieving superior classification accuracy. The best overall results were obtained with cosh, followed by sinh. Such behavior indicates that alternative SPD-preserving activations improve performance while also maintaining stability during training.

References

[1] Xiaotong Gu, Zehong Cao, Alireza Jolfaei, Peng Xu, Dongrui Wu, Tzyy-Ping Jung, and Chin-Teng Lin. Eeg-based brain-computer interfaces (bcis): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications, 2020.

[2] Imad Eddine Tibermacine, Samuele Russo, Ahmed Tibermacine, Abdelaziz Rabehi, Bachir Nail, Kamel Kadri, and Christian Napoli. Riemannian geometry-based eeg approaches: A literature review. *arXiv preprint arXiv:2407.20250*, 2024.

[3] Daniel Wilson, Robin T Schirrmeister, Lukas AW Gemein, and Tonio Ball. Deep riemannian networks for end-to-end eeg decoding. *Imaging Neuroscience*, 3, 2025.

[4] Zhiwu Huang and Luc Van Gool. A riemannian network for spd matrix learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[5] Clemens Brunner, Robert Leeb, Gernot Müller-Putz, Alois Schlögl, and Gert Pfurtscheller. BCI Competition 2008–Graz data set A. *Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology*, 16, 2008.

[6] PhysioToolkit PhysioBank. Physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101, 2000.

[7] Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon, and Sung Chan Jun. Eeg datasets for motor imagery brain–computer interface. *GigaScience*, 6, 2017.

[8] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15, 2018.

[9] Rajendra Bhatia. Positive definite matrices. *Princeton university press*, 2000.

[10] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. Activation functions in deep learning: A comprehensive survey and benchmark, 2022.

[11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep feedforward networks. In *Deep Learning*. MIT Press, Cambridge, MA, 2016.

[12] Tong Zhang, Wenming Zheng, Zhen Cui, Yuan Zong, Chaolong Li, Xiaoyan Zhou, and Jian Yang. Deep manifold-to-manifold transforming network for skeleton-based action recognition. *IEEE transactions on multimedia*, 22, 2020.

[13] Vinod Nair and Geoffrey Hinton. Rectified linear units improve restricted boltzmann machines vinod nair. volume 27, 2010.

[14] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions, 2017.

[15] Yassine El Ouahidi, Vincent Gripon, Bastien Pasdeloup, Ghaith Bouallegue, Nicolas Farrugia, and Giulia Lioi. A strong and simple deep learning baseline for bci mi decoding, 2024.

[16] Sylvain Chevallier, Igor Carrara, Bruno Aristimunha, Pierre Guetschel, Sara Sedlar, Bruna Lopes, Sebastien Velut, Salim Khazem, and Thomas Moreau. The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark. *arXiv preprint arXiv:2404.15319*, 2024.