

Article

Not peer-reviewed version

Kernel-based Regularized EEGNet using Centered Alignment and Gaussian Connectivity for Motor Imagery Discrimination

[Mateo Tobon-Henao](#) ^{*}, [Andrés Álvarez-Meza](#) ^{*}, [Germán Castellanos-Domínguez](#)

Posted Date: 6 July 2023

doi: [10.20944/preprints202307.0406.v1](https://doi.org/10.20944/preprints202307.0406.v1)

Keywords: brain computer interfaces; electroencephalography; motor imagery; regularizer; centered kernel alignment; functional connectivity; deep learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Kernel-Based Regularized EEGNet Using Centered Alignment and Gaussian Connectivity for Motor Imagery Discrimination

Mateo Tobón-Henao¹ , Andrés Marino Álvarez-Meza¹ , and Cesar German Castellanos-Dominguez¹ 

¹ Signal Processing and Recognition Group, Universidad Nacional de Colombia, 170003 Manizales, Colombia; {mtobonh;amalvarezme;cgcastellanosd}@unal.edu.co (M.T.-H.; A.A.-M.; C.G.C-D.)

* Correspondence: mtobonh@unal.edu.co

Abstract: Brain-computer interfaces (BCI) from electroencephalography (EEG) provide a practical approach to support human-technology interaction. In particular, motor imagery (MI) is a widely-used BCI paradigm that guides the mental trial of motor tasks without physical movement. Here, we present a deep learning methodology, named Kernel-based Regularized EEGNet (KREEGNet), leveled on Centered Kernel Alignment and Gaussian Functional Connectivity, explicitly designed for EEG-based MI classification. The approach proactively tackles the challenge of intra-subject variability brought on by noisy EEG records and the lack of spatial interpretability within end-to-end frameworks applied for MI classification. KREEGNet is a refinement of the widely accepted EEGNet architecture, featuring an additional kernel-based layer for regularized Gaussian functional connectivity estimation based on CKA. The superiority of KREEGNet is evidenced by our experimental results from binary and multi-class MI classification databases, outperforming the baseline EEGNet and other state-of-the-art methods. Further exploration of our model's interpretability is conducted at individual and group levels, utilizing classification performance measures and pruned functional connectivities. Our approach is a suitable alternative for interpretable end-to-end EEG-BCI based on deep learning.

Keywords: brain computer interfaces; electroencephalography; motor imagery; regularizer; centered kernel alignment; functional connectivity; deep learning

1. Introduction

Brain-Computer Interface (BCI) has emerged as a cutting-edge technology that directly connects the human brain and external devices, bridging the ultimate frontier between humans and computers [1]. This breakthrough technology has enabled people with neuromotor disorders, nervous system injuries, or limb amputations to control machines using their brains, as no peripheral nerves or muscles are involved in the process [2]. Motor Imagery (MI) is one of the essential branches of BCIs control paradigms, which allows users to control robots or external machines merely by imagining movement without the intervention of peripheral nerves [3]. Regarding this, BCI technology has significant potential in motor function rehabilitation [4], assistance [5], and other areas, sparking extensive discussions on MI-based approaches [6].

Acquiring brain activity is a critical aspect of MI-based BCIs, and multi-channel time series signals such as EEG are commonly preferred due to their high time resolution, cost-effectiveness, and user-friendliness compared to other neuroimaging methods [7]. Moreover, using multi-channel time series signals in MI tasks is essential as it captures the activation of multiple brain regions, enabling a comprehensive understanding of complex neural activity [8]. These signals facilitate exploring Functional Connectivity (FC) and coordinated patterns between brain regions during MI while reducing noise and artifacts through redundancy and robust signal processing techniques [9]. Nonetheless, the insufficient functioning of MI EEG-based BCIs can have severe consequences for individuals relying on these devices. In fact, suboptimal performance can lead to frustration, inaccuracy, and reduced functionality [10].

Hence, aiming to enhance effectiveness, it is necessary to prioritize transparency in BCIs. This can result in improved operational efficiency and smoother integration of BCI technology into daily life, ultimately enriching the quality of life for individuals with motor disabilities [11]. Still, the factors contributing to the limited usefulness of MI tasks are complex and varied. Inter-subject variability is a noteworthy aspect that contributes to poor performance. In this sense, the subject mental state, attention, and fatigue can also substantially influence [12]. Also, the quality of electrical activity patterns generated by the brain plays a crucial role in controlling external devices [13]. However, these patterns exhibit substantial variation among subjects, even under identical stimuli or conditions [14]. Various factors, including gender, age, lifestyle, neurophysiological and psychological parameters, genetic differences, and cognitive processes, contribute to this variability [15]. Such diversities in brain patterns result in performance fluctuations, impeding the development of reliable and accurate BCIs [16].

Moreover, noise in EEG signals significantly contributes to this variability, obscuring underlying neural activity [17]. Notably, noisy records can originate from diverse sources, such as electromagnetic interferences, movement artifacts, individual skull thickness, and conductivity differences [18,19]. These unwanted signals make it difficult to identify the neural activity patterns that drive BCI performance accurately. Additionally, the need for interpretability in BCIs poses a critical challenge, hindering the identification of different patterns between high-performing and low-performing subjects. The difficulty in interpreting MI EEG-based BCIs and understanding their decision-making procedures complicates devising and enriching MI functionality [20].

In recent years, several methods have been proposed to enhance the performance of BCIs during the preprocessing and feature extraction stages. The preprocessing methods aim to mitigate the impact of low Signal-to-Noise Ratio (SNR) caused by environmental and physiological artifacts such as electrical noise, eye and muscle movements, heart activity, and respiration [21]. Additionally, the preprocessing stage seeks to tackle the low spatial resolution challenge caused by the volume conduction effect [22]. Also, artifacts in EEG signals can be removed using regression-based techniques, which use linear approaches to remove the noise [23]. Band-pass and notch filters can also eliminate electrical and environmental noise and frequency bands where neurophysiological information is irrelevant [24]. Blind source separation techniques, such as Canonical Correlation Analysis (CCA), Principal Component Analysis (PCA), and Independent Component Analysis (ICA), are commonly used to decompose the contaminated EEG into statistically independent components to remove or correct the artifact [25]. Of note, ICA is recognized for its success in eliminating various types of

artifacts [26]. Furthermore, different spatial filters have been proposed to overcome the volume conduction issue, including the Common Average Reference (CAR) and the Surface Laplacian (SL). The CAR spatial filter subtracts the average electrical activity measured across all sensors from each sensor to reduce the recorded noise [27]. Nevertheless, it does not address sensor-specific noise and may introduce noise into an otherwise clean sensor [28]. In contrast, SL aspires to remove the common brain activity of neighboring sensors due to the volume conduction effect, which improves local topographical features, facilitates sensor-level connectivity analysis, and helps to enhance the SNR [29]. Despite the effectiveness of these methods, applying them to all subjects regardless of the individual noise level can be detrimental to subjects with clean EEG [30].

On the other hand, the feature extraction strategies seek to transform the raw EEG signals into relevant brain patterns independent of subject-specific differences. This approach allows for identifying common patterns across individuals, improving the generalizability of BCI systems. Feature extraction techniques can be broadly categorized into time, time-frequency, and spatial approaches. In the time domain, amplitude modulation [31] and time-domain analysis of variance [32] are widely used to extract features related to the amplitude and timing of specific EEG components, providing insights into the underlying neural processes involved in MI. These features enable the identification of significant differences between classes that can be used to classify the signals effectively. In the time-frequency domain, wavelet transform [33] is a commonly used method that analyzes the changes in the frequency content of the EEG signal over time. This method provides information about the temporal dynamics of neural processes during MI, including evoked-related algorithms and intertrial coherence to capture the temporal evolution during the MI task [34]. Common Spatial Patterns (CSP) and FCs are standard methods for feature extraction in the spatial domain. CSP projects the EEG signals into a lower dimensional space using a set of learned spatial filters that enhance the differences between MI classes [35]. FCs capture the similarity between EEG channels, providing information on which brain regions interact when a subject performs the MI task [36]. However, choosing the appropriate feature extraction method for the MI task is challenging, as it demands considerable subject-matter expertise and prior knowledge about the anticipated EEG signal [37]. Moreover, the specificity of the EEG signals' preprocessing steps for the interesting feature could exclude potentially relevant patterns from the analysis [38].

Nowadays, deep learning methods have emerged as a promising approach to overcoming the limitations of traditional methods in addressing MI inter-subject variability by automating the preprocessing and extracting relevant features from EEG signals within an end-to-end framework [39]. In particular, models such as EEGNet, ShallowConvNet, DeepConvNet, Graph Convolution Neural Networks (GCN) [40,41], and EEG-transformer [42] have great potential to tackle EEG-based MI challenges. EEGNet and ShallowConvNet utilize convolutional layers to extract spatial and temporal patterns from EEG data. However, EEGNet may need help with capturing long-range temporal dependencies [43], while ShallowConvNet may not be as effective as deeper architectures in capturing complex patterns. DeepConvNet excels at capturing spatial and temporal patterns but requires much training data to avoid overfitting [44]. GCNs capture spatial relationships between electrodes by aggregating information from neighboring nodes in the graph. Nonetheless, they are sensitive to graph construction from EEG signals [45]. Recently, transformer-based models like EEG-transformer have been adept at processing variable-length sequences by employing a self-attention mechanism to capture dependencies between different

segments. Nonetheless, these models come with higher computational costs and require large amounts of samples [46].

Overall, deep learning can be prone to overfitting when they are too complex or the training data is noisy and insufficient [47]. Diverse regularization techniques have been proposed to tackle this issue. For example, domain adaptation aims to reduce variability across different subjects by learning a mapping between source and target spaces [48]. Yet, it requires a substantial amount of labeled data from both domains [49]. Multi-task learning leverages information from related tasks to improve the performance of individual tasks [50]. Nevertheless, it assumes the availability of multiple related tasks, which may need to be more practical in specific scenarios [51]. Dropout and batch normalization are also helpful techniques that can reduce overfitting. The former randomly drops out a fraction of neurons during training to enhance the model's ability to learn robust features [52]. The latter normalizes input features across subjects to enhance network stability and convergence [53]. However, both techniques can increase computational requirements, and their performance can be sensitive to hyperparameters tuning and noisy samples [54,55]. FC-based regularizers introduce a penalty term to obtain low-rank or sparse connectivity matrices, reducing the impact of MI inter-subject variability [28]. Regardless, these regularizers assume a smooth or sparse connectivity structure of the brain, which may not always hold in practice [56].

Here, we introduce a novel deep learning approach for EEG-based MI classification: Kernel-based Regularized EEGNet (KREEGNet). Our approach addresses the challenges posed by intra-subject variability in noisy EEG records and the lack of spatial interpretability in existing end-to-end frameworks used for MI classification. KREEGNet enhances the well-established EEGNet architecture, incorporating a twofold approach: i) a kernel-based layer for Gaussian functional connectivity estimation is coupled within the EEGNet architecture, ii) a Centered Kernel Alignment (CKA) loss is associated with conventional Cross-Entropy measure for deep learning classification to deal with noisy EEG records while preserving the spatial interpretability based on kernel mappings. Through experimentation on binary and multi-class MI classification databases, we demonstrate the superiority of KREEGNet over the baseline EEGNet and other state-of-the-art methods. Moreover, we explore the interpretability of our model at both individual and group levels, employing classification performance measures and pruned functional connectivities. Our findings highlight KREEGNet as a promising and interpretable deep learning approach for EEG-based BCI systems.

The agenda is as follows: Section 2 describes the materials and methods. Sections 3 and 4 present the experiments and discuss the results. Finally, Section 5 outlines the concluding remarks.

2. Materials and Methods

2.1. Centered Kernel Alignment Fundamentals

Let $X \subset \mathcal{X}$, $Y \subset \mathcal{Y}$ be a pair of random variables holding samples $x \in X$ and $y \in Y$, respectively. The kernels $\kappa_X : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ and $\kappa_Y : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ can be defined to code nonlinear relationships among samples from positive definite functions, yielding:

$$\kappa_X(x, x') = \langle \phi_X(x), \phi_X(x') \rangle_{\mathcal{H}_X}, \quad (1)$$

$$\kappa_Y(y, y') = \langle \phi_Y(y), \phi_Y(y') \rangle_{\mathcal{H}_Y}, \quad (2)$$

where $\phi_X : \mathcal{X} \rightarrow \mathcal{H}_X$ and $\phi_Y : \mathcal{Y} \rightarrow \mathcal{H}_Y$, being \mathcal{H}_X and \mathcal{H}_Y the resulting Reproducing Kernel Hilbert Spaces (RKHS). Hence, the statistical alignment between κ_X and κ_Y , $\rho_{CKA}(X, Y) \in [0, 1]$, referred to as Centered Kernel Alignment (CKA), is calculated by taking the normalized inner product between them and averaging it across all pairs of realizations, as shown below [57,58]:

$$\rho_{CKA}(X, Y) = \frac{\mathbb{E}_{XY}\{\tilde{\kappa}_X(x, x')\tilde{\kappa}_Y(y, y')\}}{\sqrt{\mathbb{E}_X\{\tilde{\kappa}_X(x, x')\}\mathbb{E}_Y\{\tilde{\kappa}_Y(y, y')\}}}, \quad (3)$$

where $x, x' \in X$ and $y, y' \in Y$, $\mathbb{E}\{\cdot\}$ is the expectation operator, and $\tilde{\kappa}_Z$ stands for centered kernel aiming to provide translation invariance, as follows:

$$\tilde{\kappa}_Z(z, z') = \kappa_Z(z, z') - \mathbb{E}_z\{\kappa_Z(z, z')\} - \mathbb{E}_{z'}\{\kappa_Z(z, z')\} + \mathbb{E}_{zz'}\{\kappa_Z(z, z')\}, \quad (4)$$

which is defined for a given $Z \subset \mathcal{Z}$ with samples $z, z' \in Z$. In practical applications, when provided with a set of input-output pairs $\{\mathbf{x}_n \in \mathbb{R}^P, \mathbf{y}_n \in \mathbb{R}^Q\}_{n=1}^N$, we can compute the kernel matrices $\mathbf{K}_X, \mathbf{K}_Y \in \mathbb{R}^{N \times N}$ as: $K_X[n, n'] = \kappa_X(\mathbf{x}_n, \mathbf{x}_{n'})$ and $K_Y[n, n'] = \kappa_Y(\mathbf{y}_n, \mathbf{y}_{n'})$. Utilizing Eqs. (3) and (4), we can calculate the empirical estimate for the CKA alignment $\hat{\rho}_{CKA}(\mathbf{K}_X, \mathbf{K}_Y) \in [0, 1]$:

$$\hat{\rho}_{CKA}(\mathbf{K}_X, \mathbf{K}_Y) = \frac{\langle \tilde{\mathbf{K}}_X, \tilde{\mathbf{K}}_Y \rangle_F}{\sqrt{\|\tilde{\mathbf{K}}_X\|_F, \|\tilde{\mathbf{K}}_Y\|_F}}, \quad (5)$$

where $\|\cdot\|_F$ and $\langle \cdot, \cdot \rangle_F$ are the Frobenius norm and inner product, respectively. Besides, the centered kernel matrices in Eq. (5) are calculated as: $\tilde{\mathbf{K}}_X = \mathbf{H}\mathbf{K}_X\mathbf{H}$ and $\tilde{\mathbf{K}}_Y = \mathbf{H}\mathbf{K}_Y\mathbf{H}$, with $\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{1}^\top\mathbf{1}$ (\mathbf{I} and $\mathbf{1}$ are the identity matrix and the all-one vector of proper size, respectively). As a result, the alignment described in Eq. (5) serves as a data-driven estimator, enabling us to quantify the similarity between the random variables X and Y .

2.2. Gaussian Functional Connectivity from EEG Records

Let us examine a collection of multi-channel EEG recordings referred to as $\{\mathbf{X}_n \in \mathbb{R}^{C \times T}\}_{n=1}^N$, where C denotes the number of channels, T represents the samples within EEG recordings, and N the number of trials.

Next, let us consider two EEG channels of a given trial $\mathbf{x}_c, \mathbf{x}_{c'} \in X$, with $c, c' \in \{1, 2, \dots, C\}$, a pairwise correlation between the EEG channels can be computed as:

$$\hat{\rho}_L(\mathbf{x}_c, \mathbf{x}_{c'}) = \frac{1}{T} \langle \mathbf{x}_c, \mathbf{x}_{c'} \rangle_2, \quad (6)$$

where $\langle \cdot, \cdot \rangle_2$ stands for the inner product. The pairwise linear relationships in Eq. (6) allow computing functional connections between EEG channels as an undirected graph representation.

However, we can effectively capture the nonlinear interactions among various channels by operating a generalized stationary kernel that transforms the input space into an RKHS. This approach enables us to obtain a more precise depiction of the underlying neural activity. Moreover, employing a stationary kernel guarantees that the proposed technique can effectively capture the temporal dynamics of EEG signals.

Given these considerations, the Gaussian kernel is widely preferred in pattern analysis and machine learning. It can approximate any function and offers mathematically tractable properties [59]. Therefore, it is an excellent choice for computing pairwise connections as a Gaussian-based Functional Connectivity (GFC) measure from the kernel function $\kappa_G : \mathbb{R}^T \times \mathbb{R}^T \rightarrow [0, 1]$, as [60]:

$$\kappa_G(\mathbf{x}_c - \mathbf{x}_{c'}; \gamma) = \exp\left(-\frac{1}{2}\gamma\|\mathbf{x}_c - \mathbf{x}_{c'}\|_2^2\right), \quad (7)$$

where $\|\cdot\|_2$ denotes the 2-norm operator and $\gamma \in \mathbb{R}^+$ represents a scale parameter. The inclusion of a Gaussian function in Eq. (7) facilitates the accurate and efficient calculation of the nonlinear interactions between \mathbf{x}_c and $\mathbf{x}_{c'}$.

2.3. KREEGNet: Kernel-Based Regularized EEG Network

Let us consider an input-output set consisting of multi-channel EEG records and labels denoted as $\{\mathbf{X}_n \in \mathbb{R}^{C \times T}, \mathbf{y}_n \in \{0, 1\}^Q\}_{n=1}^N$. Here, \mathbf{y}_n gathers the target labels for MI tasks encoded using the one-hot encoding (with Q classes being considered). Our Kernel-based Regularized EEG Network (KREEGNet), an enhanced version of the well-known EEGNet [61], enables accurate prediction of the MI label $\hat{\mathbf{y}} \in [0, 1]^Q$ for a given EEG trial \mathbf{X} . This prediction is accomplished through two primary blocks. Initially, the class membership prediction is performed as follows:

$$\hat{\mathbf{y}} = (\varphi_Q \circ \varphi_T \circ \varphi_C \circ \varphi_F)(\mathbf{X}), \quad (8)$$

where notation $\varphi(\tilde{\mathbf{X}}) = \xi_\varphi(\mathbf{W}_\varphi \otimes \tilde{\mathbf{X}} + \mathbf{b}_\varphi)$ stands for deep learning-based layer mapping, \circ is the function composition operator, \otimes is the tensor product, e.g., convolutional or fully connected layer-based operations. Besides, $\tilde{\mathbf{X}}$ is a given network's feature map of proper size, $\mathbf{W}_\varphi, \mathbf{b}_\varphi$ gather the weight matrix and bias vector of the layer, and $\xi(\cdot)$ is a nonlinear activation function. Namely, each layer function in Eq. (8) is described as:

- $\varphi_F : \mathbb{R}^{C \times T} \rightarrow \mathbb{R}^{\check{F}, C, T}$ is a convolutional layer holding \check{F} filters, a batch normalization, and a linear activation.
- $\varphi_C : \mathbb{R}^{\check{F} \times C \times T} \rightarrow \mathbb{R}^{\alpha \check{F}, C, \frac{T}{4}}$ is a depthwise convolutional layer holding ELU activation (α gathers the number of spatial filters), followed by an average pooling and a dropout operation.
- $\varphi_T : \mathbb{R}^{\alpha \check{F}, C, \frac{T}{4}} \rightarrow \mathbb{R}^{\check{F}', \frac{T}{32}}$ is a separable convolutional layer with ELU activation (\check{F}' is the number of pointwise filters), setting a batch normalization, an average pooling, and a dropout.
- $\varphi_Q : \mathbb{R}^{\check{F}', \frac{T}{32}} \rightarrow [0, 1]^Q$ is a fully connected classification layer fixing a flatten operation and a softmax activation.

In turn, a kernel-based regularizer is applied by properly computing the data-driven GFC:

$$\mathbf{K}^\dagger = (\tilde{\kappa} \circ \varphi_{\check{F}})(\mathbf{X}), \quad (9)$$

where $\varphi_{\check{F}}$ is defined as in Eq. (8), $\tilde{\kappa} : \mathbb{R}^{\check{F}, C, T} \rightarrow [0, 1]^{\check{F}, C, C}$ extracts the GFC among EEG channels (see Eq. 7) along each of the \check{F} filters, and $\mathbf{K}^\dagger \in [0, 1]^{\check{F}, C, C}$.

Furthermore, the parameter set θ , stacking the weight matrices and bias vectors in Eq. (8), and the scale parameter γ of the GFC in Eqs. (7) and (9), is optimized using a gradient descent-based framework with back-propagation [62]:

$$\theta^* = \arg \min_{\theta} \frac{(1 - \lambda)}{N} \sum_{n=1}^N CE(\mathbf{y}_n, \hat{\mathbf{y}}_n(\theta)) - \frac{\lambda}{\check{F}} \sum_{f=1}^{\check{F}} \hat{\rho}_{CKA}(\check{\mathbf{K}}_f(\theta), \mathbf{K}_\delta), \quad (10)$$

where $\lambda \in [0, 1]$ is a trade-off hyperparameter and $CE(\cdot, \cdot)$ stands for the cross-entropy loss defined as:

$$CE(\mathbf{y}_n, \hat{\mathbf{y}}_n(\theta)) = - \sum_{q=1}^Q y_{nq} \log (\hat{y}_{nq}(\theta)), \quad (11)$$

with $y_{nq} \in \mathbf{y}_n$ and $\hat{y}_{nq}(\theta) \in \hat{\mathbf{y}}_n$. Moreover, the kernel-matrix $\check{\mathbf{K}}_f(\theta) \in [0, 1]^{N \times N}$ is computed as:

$$\check{\mathbf{K}}_f[n, n'] = \left\langle \text{triu}(\mathbf{K}_n^\dagger(\theta; f)), \text{triu}(\mathbf{K}_{n'}^\dagger(\theta; f)) \right\rangle_2 \quad (12)$$

where $\text{triu}(\mathbf{K}^\dagger(\theta; f)) \in [0, 1]^{C \frac{(C-1)}{2}}$ holds the upper triangular matrix of the GFC stored in \mathbf{K}^\dagger for filter f . Likewise, the target kernel matrix $\mathbf{K}_\delta \in \{0, 1\}^{N \times N}$ is built as:

$$K_\delta[n, n'] = \delta(|\mathbf{y}_n - \mathbf{y}_{n'}|_1), \quad (13)$$

being $\delta(\cdot)$ the delta function and $|\cdot|_1$ the 1-norm.

The optimization problem outlined in Eq. (10) enables the training of our KREEGNet for MI discrimination. Figure 1 summarizes the KREEGNet pipeline. To ensure numerical stability and simplicity, the GFC scale parameter is learned as a mapping of 10^γ . It is worth mentioning that a preprocessing stage is included to align the various database conditions, such as sample frequency, band-pass filtering, and EEG window size.

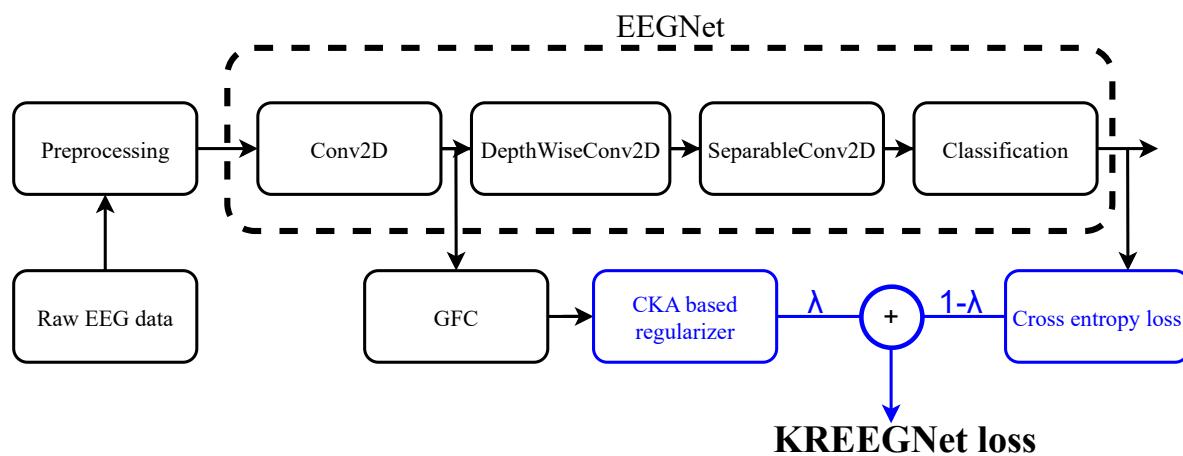


Figure 1. KREEGNet pipeline for Motor Imagery classification from EEG records.

3. Experimental Set-Up

We provide a comprehensive overview of the pipeline used to develop and evaluate the KREEGNet model for motor imagery discrimination. It includes an analysis of the datasets used, the training phase of the model, and the techniques employed to assess the proposal's effectiveness.

3.1. Datasets Description

In order to evaluate the effectiveness of our KREEGNet, we conducted tests on two well-known databases that involve motor-related tasks.

DBI: BCI Competition 2008 - Graz Dataset 2a (<http://www.bbci.de/competition/iv/index.html> - accessed on 1 April 2023). The dataset contains EEG data from nine subjects who participated in a motor imagery paradigm consisting of four tasks: imagining the movement of the left hand, right hand, both feet and the tongue (four class problem). The data were collected in two sessions on different days, comprising six runs with 48 trials per run (12 for each class). This resulted in a total of 288 trials per session. A short acoustic warning and a cross on a black screen signaled the start of each trial, which lasted seven seconds. Then, at two seconds, a visual cue appeared on the screen for 1.25 seconds, indicating which MI task to perform until the cross disappeared at six seconds. Next, a short break followed, and the screen went black. The EEG data were recorded using a 22-channel Ag/AgCl electrode montage based on the 10/20 system. In addition, three EOG electrodes were also used to record ocular artifacts. The signals were sampled at 250 Hz and bandpass-filtered between 0.5 and 100 Hz, with a 50 Hz Notch filter applied. The datasets for each subject and session were stored in the General Data Format for biomedical signals, with one file per subject and session.

DBII: GiGaScience (<http://gigadb.org/dataset/100295> - accessed on 1 April 2023). It includes EEG data from 52 healthy subjects, although only 50 are available for evaluation. The data were acquired in one session using the MI experimental paradigm with two classes (left and right hands). Each session comprised five or six runs with 100 or 120 trials per class. Moreover, each trial lasted seven seconds, starting with a black screen with a fixation cross within two seconds. A cue instruction appeared randomly on the screen within three seconds, prompting the subject to perform the indicated MI task. The trial ended with a blank screen and a 4.1 to 4.8

seconds break. The EEG data were collected using a Biosemi ActiveTwo system with 64 Ag/AgCl electrodes placed according to the 10/10 international system, sampled at 512 Hz, and stored in *.mat format. Actual left-hand and right-hand movements and six types of noise (blinking eyes, eyeball movement up/down, eyeball movement left/right, head movement, jaw clenching, and resting state) were also collected, aside from the MI recordings.

Figure 2 provides an overview of the DBI (four-class problem) and DBII (binary-class problem) montage and paradigm used for MI classification.

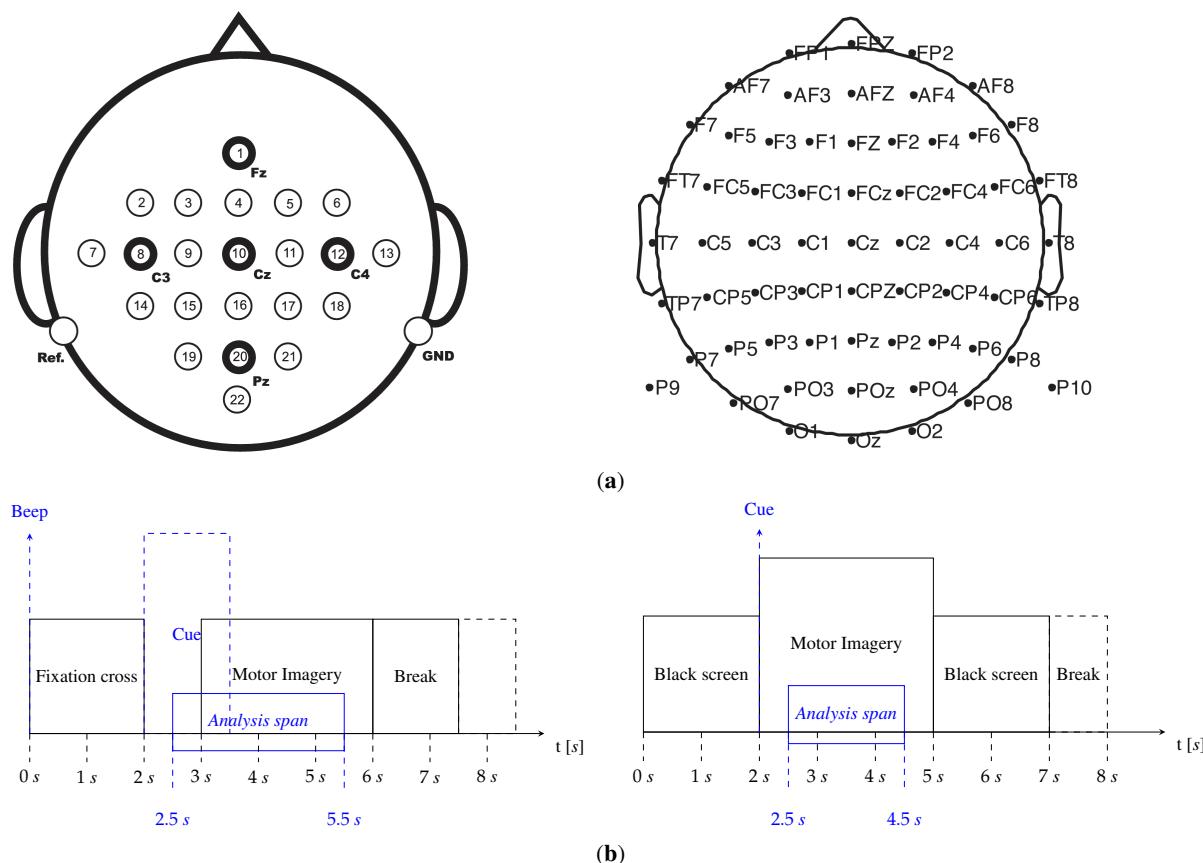


Figure 2. The EEG-MI databases examined: DBI (BCI Competition four-class task) and DBII (GigaScience binary task), displayed in the left and right columns, respectively. The top row shows the EEG montages, while the bottom row presents the MI paradigm tested.

3.2. KREEGNet Training Details and Assessment

The training of our KREEGNet comprises three stages: i) preprocessing of EEG records, ii) fine-tuning the network hyperparameters for improved classification performance, and iii) interpreting functional connectivity by identifying relevant patterns learned during deep learning training.

To initiate EEG preprocessing, a custom database loader module, see <https://github.com/UN-GCPDS/python-gcpds.databases> (accessed on 8 April 2023), was utilized to load the recordings. Only EEG channels were considered, and the signals were scaled to μ V to ensure suitability for analysis. Any trials marked as bad were rejected. A fifth-order Butterworth bandpass filter was applied to all channels within the $[4, 40]$ Hz range, where MI activity is observed [60]. Additionally, each channel's signal was clipped within the post-cue onset

time window, retaining only information from the MI task. For DBI, the time window was $0.5 - 3.5s$, while for DBII, it was $0.5 - 2.5s$. Then, to ensure the network parameters remained consistent, each channel's signal was downsampled in both databases from $256Hz$ for DBI and $512Hz$ for DBII to $128Hz$. Our preprocessing step is similar to the one described by the authors in [61].

Next, to ensure a reliable evaluation of our model, we employed the stratified shuffle split 5-fold 80 – 20 scheme within each subject's data. This process involved shuffling the data and selecting 80% for training while holding out the remaining 20% for testing. This procedure was repeated five times. Model performance was evaluated using accuracy, Cohen's kappa, and the area under the curve ROC [59]. An exhaustive search strategy for hyperparameter tuning was implemented, and the mean accuracy score across the folds was used to evaluate each hyperparameter's performance. In order to train our model, we formulated the loss function as a combination of the cross-entropy (CE) and the CKA-based regularization, with each component weighted accordingly. The CE component served as a guide for the model to perform the classification task effectively. On the other hand, the CKA component played a role in mitigating overfitting by considering the spatial information of the FCs computed in the GFC layer. The contribution of each term in the cost function was defined as $(1 - \lambda)$ for the CE component and λ for the CKA component (see Eq ??). The value of λ , a hyperparameter, was searched within the set $\{0, 0.2, 0.4, 0.6, 0.8\}$. We employed the Adam optimizer with an initial learning rate of $1e - 3$ to optimize the network parameters. Additionally, a callback mechanism was implemented to decrease the learning rate by 10 when the loss function no longer exhibited improvement. The KREEGNet was trained for 500 epochs, utilizing all available samples in the training set.

The experiments conducted in this study were performed using Python version 3.8 in both Google Collaboratory and Kaggle environments. We employed TensorFlow version 2.8.2 to construct models, define losses, create custom layers, and implement training strategies. To ensure reproducibility and facilitate further analysis and experimentation, we consistently saved the model weights and performance scores. For those interested in reproducing the training of our KREEGNet, we have provided a Kaggle notebook accessible at the following link: <https://www.kaggle.com/mateotobonhenao/rcka-eegnet-training> (accessed on 8 April 2023). This resource contains all the necessary details and code to replicate our training procedure.

3.3. Method Comparison

To assess the efficacy of our KREEGNet, we conducted a comprehensive analysis of its classification performance and the discriminability of estimated FCs. Additionally, we categorized subjects into groups (for DBII) based on their classification performance to gain insights into the impact of our proposal against four classical end-to-end deep learning models that incorporate both temporal and spatial information from EEG signals using stacked 1D convolutions. The first model, the baseline EEGNet [61], utilizes separable convolutions to reduce parameters while maintaining performance similar to traditional convolutional layers. In addition, it includes a depthwise convolution layer to capture spatial information and a fully connected layer with softmax activation for classification. The second model, Shallowconvnet [63], is a simpler architecture consisting of a single convolutional layer followed by non-linear activation, batch normalization, and pooling layers. Despite its simplicity, it effectively classifies EEG signals. The third model, Deepconvnet [63], is a deeper

architecture comprising five convolutional layers, followed by non-linear activation, batch normalization, and pooling layers. Although it performs well in EEG signal classification, it is computationally more expensive than Shallowconvnet and EEGNet. Finally, we consider the TCFusionnet proposed in [64]. This model consists of three main components: a temporal component that learns various bandpass frequencies, a depth-wise separable convolution that extracts spatial features for each temporal filter, and a Temporal Convolutional Network (TCN) block that captures temporal features. These features are combined to generate comprehensive feature maps, which are then classified into different MI classes using a dense layer with softmax activation. The Kaggle notebook available at (<https://www.kaggle.com/mateotobonhenao/dl-methods-comparison> - accessed on 8 April 2023) contains the code necessary to assess the MI classification effectiveness of the aforementioned deep learning models. Besides, the following GitHub repository holds the complete codes related to our experiments (<https://github.com/mtbonh/RCKA-EEGNet> - accessed on 8 April 2023).

4. Results and Discussion

4.1. Baseline EEGNet vs. KREEGNet: Subject and Group-Level Results

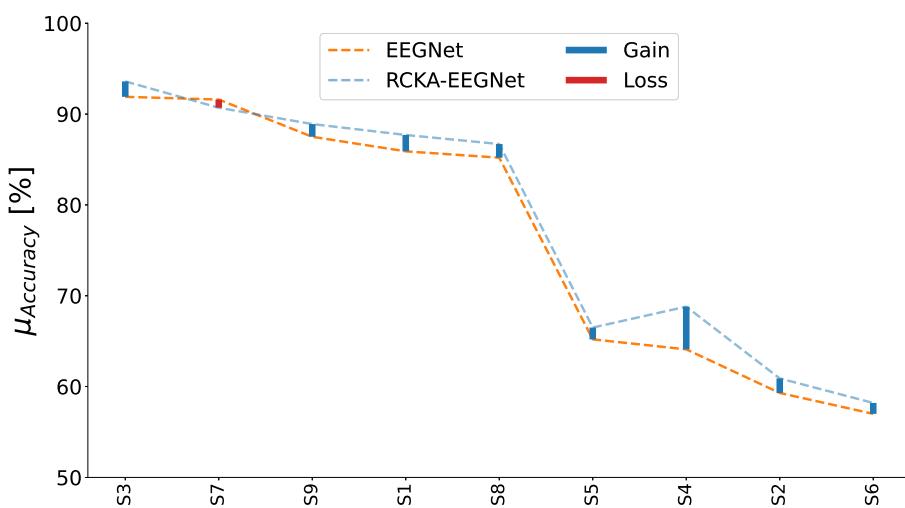
We conduct a comparative analysis of KREEGNet with the widely recognized benchmark, EEGNet, for both DBI and DBII in the context of binary MI classification tasks, explicitly focusing on distinguishing between left and right-hand imagery movements. A subject-specific examination is executed across both databases, while the group-level analysis is limited solely to DBII due to DBI's composition of a mere nine subjects. We construct a scoring matrix for robust validation with rows equivalent to the dataset's subject count—50 for DBII—and six columns representing accuracy, Cohen's kappa, the area under the ROC curve scores, and their corresponding standard deviations. To maintain the principle of 'the higher, the better' and restrict all column values within the $[0, 1]$ range in the scoring matrix, we substitute the standard deviation with its complement and normalize Cohen's kappa by adding one and dividing by two. Following that, we utilize this scoring matrix and the k-means clustering algorithm [59], setting k to three, to train a model that categorizes subject results based on the benchmark model EEGNet into three groups: top performers (GI), average performers (GII), and low performers (GIII). Subsequently, our KREEGNet's subject analysis results are clustered using the trained k-means and the score matrix. The ultimate goal is to examine and discern how subject classification shifts between the EEGNet and the KREEGNet-based groups [60].

Figure 3a and 3b present a comparative accuracy analysis of subject-specific and group-level analysis. The dotted orange line in the figures corresponds to the EEGNet; in contrast, the dotted blue line illustrates the proposed KREEGNet. The blue and red bars in the figures indicate the impact of employing the KREEGNet on individual subject accuracy. Specifically, the blue bars denote improvements in accuracy, while the red bars indicate decreases. These visual cues provide valuable insights into the performance enhancements achieved by our approach across specific subjects. Moreover, in the context of DBII, the figure's background incorporates bars with low opacity in opal green, lemon yellow, and salmon pink. These color-coded backgrounds denote the grouping of subjects into top-performing, average-performing, and low-performing subjects.

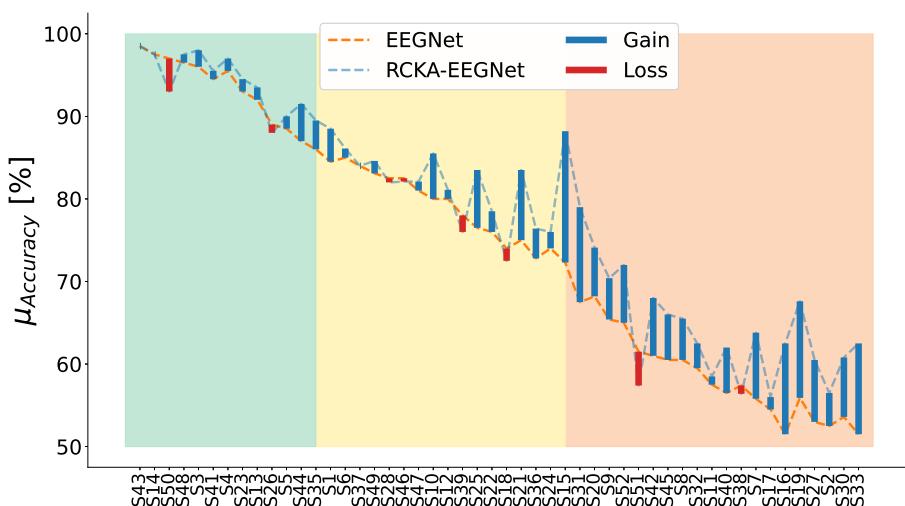
Our KREEGNet model's performance regarding DBI reveals a subject-dependent average accuracy of 78.0%, surpassing the baseline EEGNet by 1.6%. Notably, out of all the subjects, only Subject seven (S7) experienced a marginal decrease in performance, with a decline of less than 1%. Conversely, the remaining subjects demonstrated improvements in accuracy. Subject four (S4) was particularly impressive, exhibiting a remarkable performance increase of 4.7%, showcasing the effectiveness of our KREEGNet model in enhancing subject-specific analysis by coding relevant functional connections among channels within an end-to-end regularized network.

For DBII, the EEGNet and KREEGNet models achieved subject-dependent average accuracies of 74.4% and 77.9%, indicating an improvement of 3.5% for our proposal. The standard deviations for EEGNet and KREEGNet were 14.9% and 13.2%, respectively, suggesting that our approach resulted in less variability among subjects' performance. Interestingly, the accuracy of KREEGNet varied across the subjects, with three scenarios emerging from the results. Firstly, eight subjects showed a decrease in accuracy, with only three experiencing a reduction of 2% or more. Secondly, two subjects did not show any change in accuracy. Lastly, the remaining subjects demonstrated an increase in accuracy, with nineteen of them experiencing an increase of more than 5%.

Now, the impact of our method on the performance of different subject groups in DBII was substantial. In the case of Group GIII, the KREEGNet outperformed the baseline in all but two instances, with a remarkable increase of over 5% observed in fourteen cases. As for Group GII, four subjects experienced a minor decrease of less than 2%, while one remained unchanged. On the other hand, twelve subjects showed a performance improvement, with half achieving an increase of over 3%. Of particular note is Subject 15, which exhibited an impressive performance boost of 16%, highlighting the strong influence of our CKA-based regularizer on specific individuals. In Group G III, only two subjects witnessed a decrease in accuracy, while nine subjects demonstrated improved performance, including two with increases exceeding 3%. So then, our strategy yielded significant performance enhancements for most subjects across all groups, with a notable benefit observed in the poorly performing subject group.



(a) Average Accuracy: EEGNet 76.4%, KREEGNet A 78.0%.



(b) Average Accuracy: EEGNet 74.4%, KREEGNet A 77.9%.

Figure 3. EEGNet vs. KREEGNet comparison results. The top row demonstrates the subject-specific analysis for DBI, while the lower row exhibits the group-level evaluation for DBII (KREEGNet gain: GI 1.0%, GII 2.9%, and GIII 5.7%). The reported mean accuracy corresponds to a binary MI classification of left versus right-hand movement. Subjects have been organized following their EEGNet performance. The blue bars signify an enhanced performance achieved by our proposed KREEGNet, whereas the red bars highlight instances of reduced performance. The backdrop for the DBII results visually represents the group membership, with top performers in GI, average performers in GII, and low performers in GIII.

Similarly, Figure 4 presents the categorization of the subject group and the influence of the KREEGNet. The initial row displays the arrangement of subjects as per the results of EEGNet, while the final row illustrates the shift or constancy of each subject's group derived from the KREEGNet outcomes. For example, in GIII, our approach promoted four subjects to GII. Likewise, two individuals were elevated from GII to GI. Importantly, no individual experienced an in-group demotion status, underlining the equal or superior performance of KREEGNet compared to the standard EEGNet.

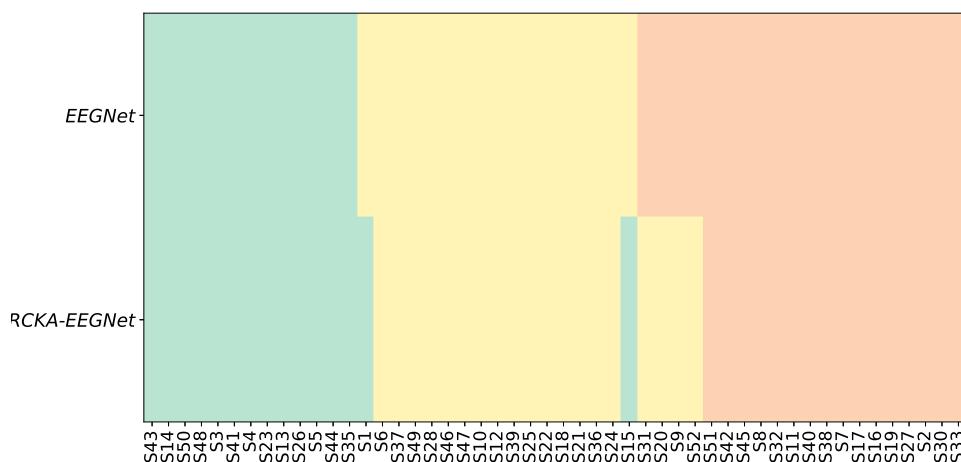


Figure 4. KREEGNet subject group enhancement (Baseline: EEGNet). Note that green, yellow, and red represent top, average, and low performance regarding the average accuracy along subjects. First row: The arrangement of subjects according to EEGNet classification. Second row: Alterations in subject group affiliations based on the results of KREEGNet.

Subsequently, we scrutinized the complex behavior of the hyperparameters λ and γ across different subject groups in DBII. λ symbolizes the importance given to the CKA-based regularizer in the cost function of KREEGNet, contributing to enhancing the network's classification capabilities. Conversely, γ sets the bandwidth scale for the Gaussian kernel employed in the GFC layer that calculates the FCs. By investigating the dynamics of these hyperparameters, we seek to understand their influence on performance and the GFC layer's FC estimation. Figure 5a presents a boxplot depicting the statistical distribution of the λ hyperparameter among the subject groups, with the background boxes denoting group membership. Firstly, most tend to possess lower λ values in GI, specifically below 0.6. This is attributed to the fact that subjects within this group display more evident MI patterns, readily captured by the standard EEGNet model. Secondly, GII exhibits a more evenly distributed set of values, with half of the subjects presenting λ values exceeding 0.3. This could imply that some subjects at this stage demonstrate noisy MI patterns that heighten the risk of overfitting the training data, thereby reducing the classification performance. Lastly, for GIII, λ values are predominantly higher. Precisely, half of the subjects in this group have λ values above 0.5, with the majority of the remainder having values ranging between 0.4 and 0.5. The latter suggests that most of the subjects' data in this group present noisy patterns. Nevertheless, the CKA-based regularizer, working on the FCs computed by the GFC layer, aids in eliminating this unwanted effect, leading to improved classification performance.

In the same way, Figure 5b displays the boxplot of the γ hyperparameter among different subject groups. This bandwidth filters the relationships between channels, suggesting that channels with higher noise levels have lower bandwidth values to circumvent unwarranted connections. The findings imply that subjects in GIII require more filtering through the γ parameter, hinting that these individuals typically have higher noise in their MI patterns. Our CKA-based regularizer and the GFC layer contribute to the reduction of these noises, thereby enhancing classification performance. Notably, our results demonstrate an inverse linear relationship between the fixed λ and γ values. Specifically, subjects with good performance, i.e., those in G I and some in G II,

exhibit lower values of λ and higher values of γ , indicating a low contribution of the CKA-based regularizer and that the bandwidth of the GFC layer is more flexible in filtering out the relationship between channels. This suggests that the MI patterns for these subjects are cleaner and less affected by noise. Contrariwise, subjects with poor performance, i.e., those in G III, exhibit higher values of λ and lower values of γ , indicating that the CKA-based regularizer contributes more to the cost function to reduce the effect of overfitting due to the presence of noisy. Additionally, γ shrinks the value of the bandwidth in the GFC layer to be more rigid in filtering out the relationship between channels, thereby avoiding spurious connectivities. These findings highlight our KREEGNet's importance in optimizing the performance and interpretability of EEG-based MI tasks.

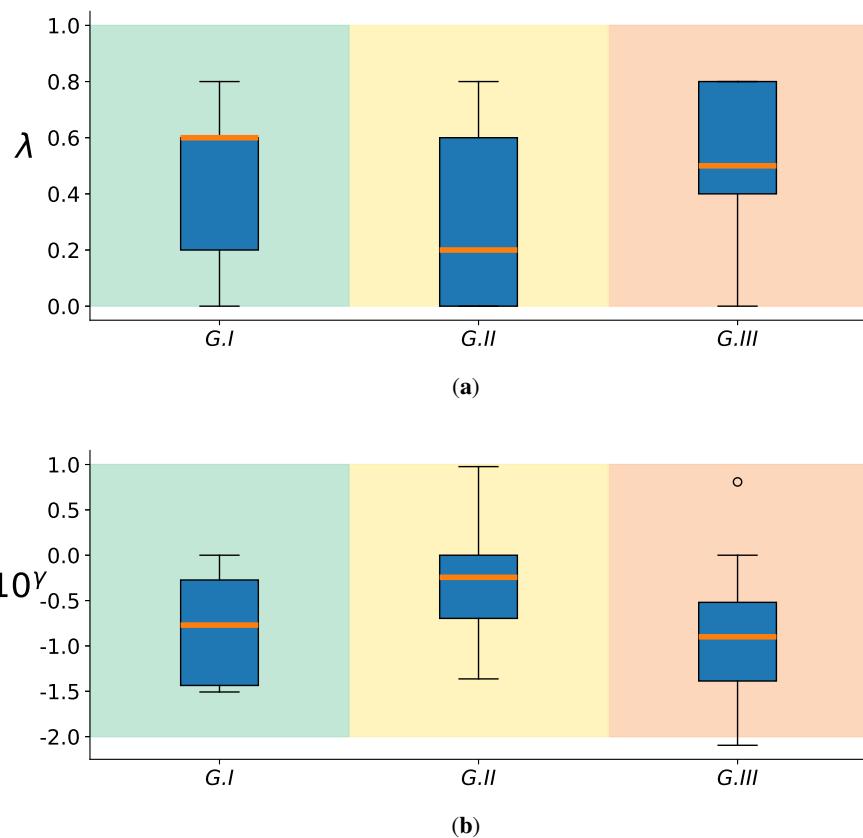


Figure 5. Analysis of KREEGNet hyperparameters at the group level for DBII. Boxplot diagrams are provided for the tuned λ and γ values in relation to the top (G.I), average (G.II), and low (G.III) performing subjects.

4.2. Relevance Analysis Results

We evaluated the FC variations across subjects, focusing on determining which connections significantly influence the ability to distinguish between the MI classes. Acknowledging that a strong correlation in the FC matrix does not automatically translate into enhanced class distinction is essential. In this endeavor, we utilized the Kolmogorov-Smirnov (KS) statistic [65], a tool that quantifies the disparity between the class distributions for each FC. Our KS-based connectivity pruning is as follows:

- We categorized each connection's trials for an individual based on the label, forming the right and left sample sets.

- Following this, we calculated the KS statistic for the connectivity between each pair of EEG channels along the training set trials. A KS value nearing 1 signifies a high level of distinguishability for the connectivity between two channels, whereas a value approaching 0 suggests a low level of separability.
- Moreover, we utilized the maximum operator across the estimated feature maps to establish a KS statistic matrix. This matrix denotes the class-separability of each connectivity.
- In order to illustrate the variations in each KS statistic matrix across subjects and groups, we depicted each matrix of KS statistic values on a two-dimensional scatter representation. Both dimensions were calculated employing the widely accepted *t*-SNE algorithm [66].
- Lastly, to fully comprehend the key connectivities and channels involved in the MI classification, we used topoplots from the KS statistic matrix.

Figures 6 and 7 depict the *t*-SNE 2D projections of the KS statistic matrices of each subject for DBI and DBII, respectively. In particular, the color-coded outer square of Figure 7 represents the group affiliation (GI, GII, and GIII). This visual representation enhances our comprehension of the significant connectivity patterns in the MI classification task. Figure 6 depicts the optimal performing subjects at the bottom, intermediate performers towards the left-middle, and the poorly performing ones at the top-left. Notably, the KS statistic matrices of high-performing subjects are more distinct, except for subject 7. This finding suggests that the FCs estimated by the GFC layer hold more significance in the MI classification. On the contrary, intermediate and poor performers show sparse KS matrices, implying their data has a higher noise level, which results in erroneous FCs that overlap with MI class distributions.

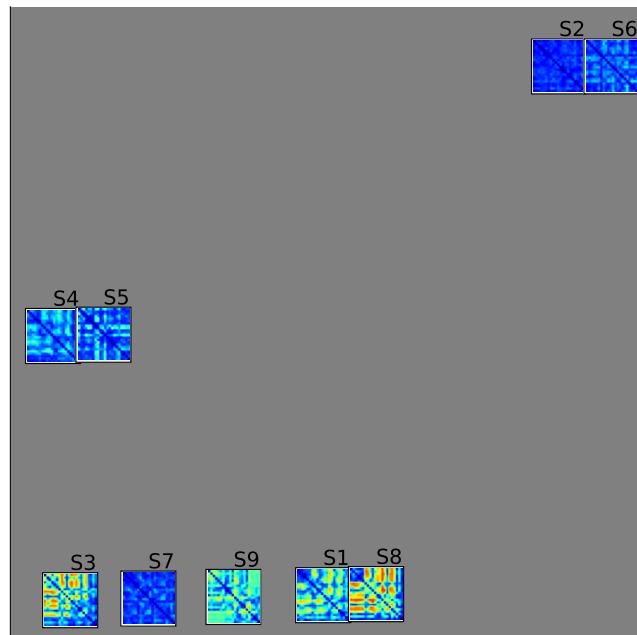


Figure 6. DBI-2D *t*-SNE projection of KS-based pruned FC matrices utilizing our KREEGNet. A gradation of colors ranging from blue to red represents a continuum from low to high separability.

Likewise, Figure 7 shows how G III exhibits sparse KS statistic matrices in the bottom-right corner, indicating that the FCs estimated are not discriminative among classes. This observation can be explained by the fact that the γ parameter took lower values for this particular group of subjects, which tend to produce sparse matrices regardless of MI classes. In contrast, the subjects in G I in the top-left tend to have more fired KS statistics, with a notable concentration over the MI area. Finally, G II reveals more erratic behavior, with the subjects near G III. The latter may be attributed to individual differences in brain activity during the MI tasks.

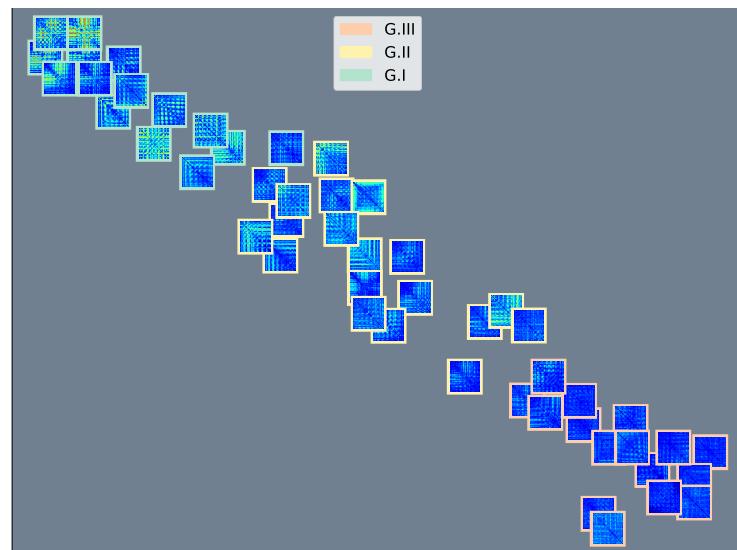


Figure 7. DBII-2D *t*-SNE projection of KS-based pruned FC matrices utilizing our KREEGNet. A gradation of colors ranging from blue to red represents a continuum from low to high separability. Outer boxes indicate subject group belongingness: green G I, yellow G II, and red G III.

In order to evaluate the informational dynamics of pruned FCs, we utilized quadratic Rényi's entropy, computed over the KS statistic matrices [67]. Our observations suggested that sparse KS statistic matrices corresponded with higher noise levels, whereas the KS matrices that had been freed up corresponded to lower noise levels. These statements are corroborated by Figures 8a and 8b. Furthermore, our findings align with the previously stated remarks. Specifically, the subjects that perform poorly in DBI tend to display higher entropy values, which is also true for the subjects classified under G III in DBII.

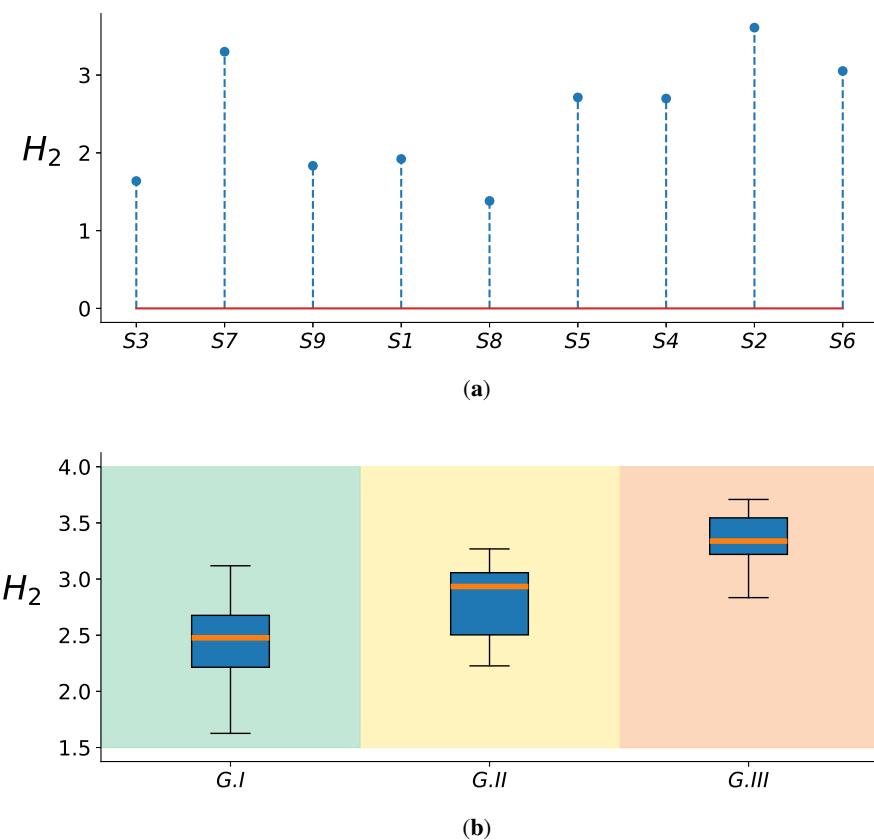


Figure 8. Renyi's entropy-based retained information within the estimated functional connectivity matrices (H_2 stands for quadratic entropy value). Top: DBI results sorted regarding the classification performance. Bottom: DBII results where the background codes the group membership (best, medium, and poor-performing subjects. Boxplot representation is used to present the retained information within each group.

Next, the topoplots in Figures 9a and 9b show the distribution of relevant connectivities and channels. Relevant subjects are selected for visualization purposes in DBI. Meanwhile, the centroid of each group in DBII is employed. The results indicate that the sensorimotor area is the most critical region for both databases. It suggests that our KREEGNet effectively improves classification performance and model interpretability by incorporating a CKA-based regularizer and a *GFC* layer.

For DBI, we analyzed S3, S4, and S6 to represent high-performing, intermediate, and low-performing subjects. Notably, S3 exhibited a higher number of relevant FCs compared to S4 and S6. Furthermore, the FCs of S3 and S4 are thickened in the central-brain region, consistent with the MI paradigm. However, S6 displayed a concentration of FCs in a single channel in the left-central region. Concerning DBII, the analysis of connectivities and channels in G I subjects revealed that the primary areas of interaction during MI tasks are located in the left-right central regions. This finding suggests that these subjects exhibit more distinct and reliable patterns of MI activity. The subjects belonging to G II displayed a pattern of connectivities and channels in the right-central brain region. However, a diffuse pattern was observed in the left hemisphere, covering some posterior and brain regions not strongly associated with MI activity. This diffuse pattern may be attributed to noise-induced EEG features, affecting classification performance. Finally, for the subjects in G III, the connectivities are concentrated in the central region of both hemispheres, which aligns with the MI paradigm. Similar to G II, the main channels

are located in the right-left central brain areas, but robust patterns are observed in the left-posterior and frontal areas, highlighting noisy behavior.

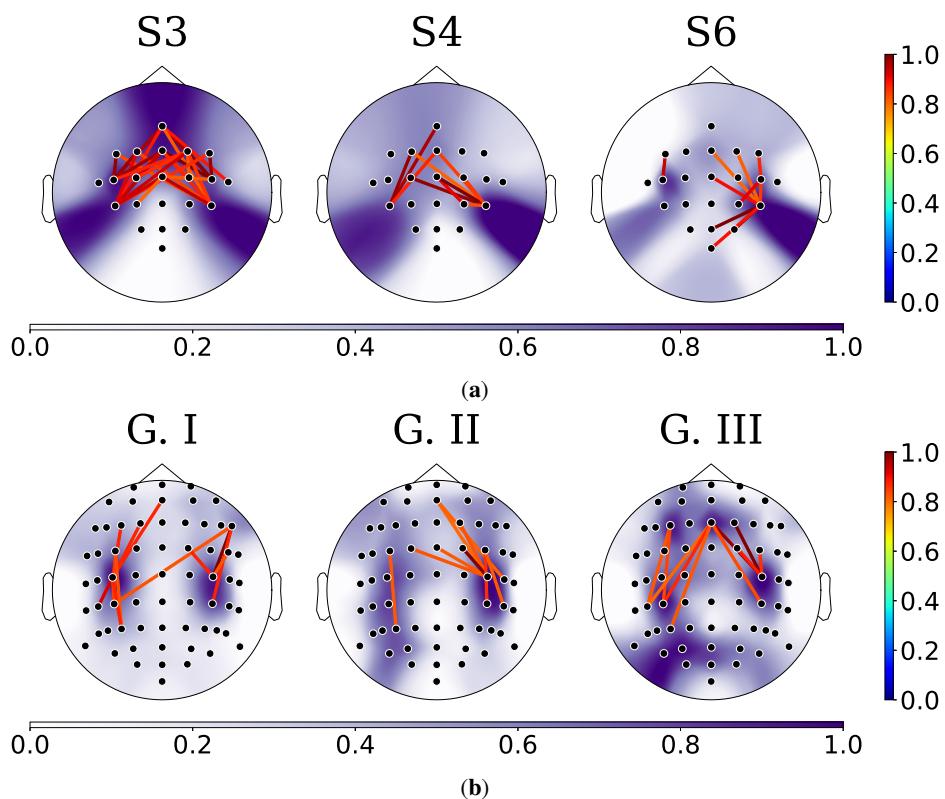


Figure 9. Visual outcomes of the topographical maps (DBI and DBII results). The top row illustrates the results related to significant subjects for the DBI. The bottom row displays group-oriented visualizations for the DBII. Only those connections that hold a value surpassing the 95th percentile are highlighted. The backdrop of these visualizations corresponds to the normalized cumulative connection strength across channels, which is projected onto the topographical map.

4.3. Method Comparison Results: Binary and Multi-Class MI Classification

The classification performance of the deep learning models discussed in Section 3.3 for DBI and DBII are presented in Tables 1 and 2, respectively. The results indicate that the DeepConvenet model performs the worst for both databases, while our proposed KREEGNet achieves the highest MI classification results. Notably, the Shallowconvnet, EEGNet, and TCFusionnet networks conduct similarly in both databases. Our KREEGNet attains outstanding results in all classification measures for DBII, demonstrating its superior performance. Although our model also achieves the best results for DBI, the difference in performance compared to other models is less significant. This can be attributed to the fact that DBI has fewer channels, with most of them concentrated in the central brain area, which limits the effect of the estimated FC by the GFC layer and the CKA-based regularizer. Then, only interactions between channels located in the same brain region are considered, reducing the diversity of information.

Table 1. Multi-class MI classification results for DBI. Average Accuracy, Kappa, and AUC are displayed \pm the standard deviation.

Approach	Accuracy	Kappa	AUC
Deepconvnet [63]	55.5 ± 24.3	40.6 ± 32.4	78.1 ± 20.4
Shallowconvnet [63]	74.9 ± 13.3	66.7 ± 17.7	91.6 ± 7.2
EEGNet [61]	76.4 ± 14.6	68.6 ± 19.5	92.5 ± 0.71
TCFusionnet [64]	77.3 ± 13.4	69.7 ± 17.9	92.6 ± 0.68
KREEGNet (ours)	78.0 ± 14.1	70.7 ± 18.8	92.6 ± 0.7

Table 2. Binary MI classification results for DBII. Average Accuracy, Kappa, and AUC are displayed \pm the standard deviation.

Approach	Accuracy	Kappa	AUC
Deepconvnet [63]	61.9 ± 12.4	23.6 ± 24.9	66.0 ± 16.0
Shallowconvnet [63]	72.5 ± 14.1	44.6 ± 28.3	77.9 ± 15.3
TCFusionnet [64]	73.9 ± 14.8	48.0 ± 30.0	80.0 ± 16.3
EEGNet [61]	74.4 ± 14.9	48.6 ± 29.8	79.6 ± 16.4
KREEGNet (ours)	77.9 ± 13.2	55.7 ± 26.5	82.5 ± 14.5

5. Conclusions

We introduced a novel deep learning approach for EEG-based Motor Imagery classification, named Kernel-based Regularized EEGNet grounded on Gaussian Functional Connectivity and Centered Kernel Alignment (KREEGNet). Our proposal addresses the issues of intra-subject variability caused by noisy EEG recordings and the absence of spatial interpretability in end-to-end frameworks for MI classification. Specifically, we amplified the widely recognized EEGNet architecture with a kernel-based layer designed to encode discriminant functional connectivities through a Gaussian similarity layer. Furthermore, the regularizer rooted in Centered Kernel Alignment seeks to minimize the overfitting effect caused by noise in the EEG data of subjects, thereby enhancing the performance of Motor Imagery classification.

The experimental outcomes obtained from binary and multi-class EEG-based Motor Imagery classification databases revealed the superior performance of our KREEGNet compared to the baseline EEGNet and other state-of-the-art deep-learning models. We further delved into the interpretability of our model at both the subject-dependent and group levels, using classification performance measures and pruned functional connectivities specific to our KREEGNet. In classifying subjects into three groups based on their performance, we showcased the capacity of KREEGNet to improve the MI classification performance, particularly among those subjects with previously poor performance. This highlights the fact that these individuals are most significantly impacted by noise.

In summary, the proposed KREEGNet effectively resolves issues like intra-subject variability attributed to noise in EEG data and the absence of spatial interpretability in deep learning models used for motor imagery classification. These insights will aid in developing brain-computer interfaces that are more accurate and interpretable, expanding their potential for diverse applications.

In future research, we aim to augment our KREEGNet to achieve end-to-end functional connectivity estimation via graph convolutional networks [68]. Additionally, we intend to investigate causal connectivity

rooted in information-theoretic learning for deep-learning-based estimations [69]. Lastly, we consider conducting subject-independent experiments and testing transformer networks [70].

Author Contributions: Conceptualization, A.A.-M., and G.C.-D.; methodology, M.T.-H., A.A.-M., and G.C.-D.; software, M.T.-H.; validation, M.T.-H. and A.A.-M.; formal analysis, A.A.-M., and G.C.-D.; investigation, M.T.-H.; data curation, M.T.-H.; writing original draft preparation, M.T.-H., G.C.-D.; writing—review and editing, A.A.-M. and G.C.-D.; visualization, M.T.-H.; supervision, A.A.-M. and G.C.-D. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://www.bbci.de/competition/iv/> and <http://gigadb.org/dataset/100295>.

Funding: Under grants provided by the project: "Sistema prototípico de procesamiento de bioseñales en unidades de cuidado intensivo neonatal utilizando aprendizaje de máquina - Fase 1: Validación en ambiente simulado" - HERMES 55063, funded by Universidad Nacional de Colombia.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Venkatachalam, K.; Devipriya, A.; Maniraj, J.; Sivaram, M.; Ambikapathy, A.; Iraj, S.A. A Novel Method of motor imagery classification using eeg signal. *Artificial intelligence in medicine* **2020**, *103*, 101787.
2. Dai, G.; Zhou, J.; Huang, J.; Wang, N. HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification. *Journal of neural engineering* **2020**, *17*, 016025.
3. Gaur, P.; McCreadie, K.; Pachori, R.B.; Wang, H.; Prasad, G. An automatic subject specific channel selection method for enhancing motor imagery classification in EEG-BCI using correlation. *Biomedical Signal Processing and Control* **2021**, *68*, 102574.
4. Khan, M.A.; Das, R.; Iversen, H.K.; Puthusserypady, S. Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application. *Computers in Biology and Medicine* **2020**, *123*, 103843.
5. Kanna, R.K.; Vasuki, R. Classification of Brain Signals Using Classifiers for Automated Wheelchair Application. *International Journal of Modern Agriculture* **2021**, *10*, 2426–2431.
6. Miao, M.; Hu, W.; Yin, H.; Zhang, K. Spatial-frequency feature learning and classification of motor imagery EEG based on deep convolution neural network. *Computational and mathematical methods in medicine* **2020**, *2020*.
7. Padfield, N.; Zabalza, J.; Zhao, H.; Masero, V.; Ren, J. EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges. *Sensors* **2019**, *19*, 1423.
8. Kaur, J.; Kaur, A. A review on analysis of EEG signals. In Proceedings of the 2015 International Conference on Advances in Computer Engineering and Applications. IEEE, 2015, pp. 957–960.
9. Ghosh, P.; Mazumder, A.; Bhattacharyya, S.; Tibarewala, D.N.; Hayashibe, M. Functional connectivity analysis of motor imagery EEG signal for brain-computer interfacing application. In Proceedings of the 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, 2015, pp. 210–213.
10. Han, C.H.; Kim, Y.W.; Kim, D.Y.; Kim, S.H.; Nenadic, Z.; Im, C.H. Electroencephalography-based endogenous brain–computer interface for online communication with a completely locked-in patient. *Journal of neuroengineering and rehabilitation* **2019**, *16*, 1–13.
11. Collazos-Huertas, D.F.; Álvarez-Meza, A.M.; Castellanos-Dominguez, G. Spatial interpretability of time-frequency relevance optimized in motor imagery discrimination using Deep&Wide networks. *Biomedical Signal Processing and Control* **2021**, *68*, 102626.
12. Saha, S.; Baumert, M. Intra-and inter-subject variability in EEG-based sensorimotor brain computer interface: a review. *Frontiers in computational neuroscience* **2020**, *13*, 87.
13. Pérez-Velasco, S.; Santamaria-Vazquez, E.; Martínez-Cagigal, V.; Marcos-Martínez, D.; Hornero, R. EEGSym: Overcoming Inter-Subject Variability in Motor Imagery Based BCIs With Deep Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2022**, *30*, 1766–1775.

14. Seghier, M.L.; Price, C.J. Interpreting and utilising intersubject variability in brain function. *Trends in cognitive sciences* **2018**, *22*, 517–530.
15. Sannelli, C.; Vidaurre, C.; Müller, K.R.; Blankertz, B. A large scale screening study with a SMR-based BCI: Categorization of BCI users and differences in their SMR activity. *PLoS One* **2019**, *14*, e0207351.
16. Vidaurre, C.; Blankertz, B. Towards a cure for BCI illiteracy. *Brain topography* **2010**, *23*, 194–198.
17. Caicedo-Acosta, J.; Castaño, G.A.; Acosta-Medina, C.; Alvarez-Meza, A.; Castellanos-Dominguez, G. Deep Neural Regression Prediction of Motor Imagery Skills Using EEG Functional Connectivity Indicators. *Sensors* **2021**, *21*, 1932.
18. LK Jaya Shree, B. Automatic Detection of EEG as Biomarker using Deep Learning: A review. *Annals of the Romanian Society for Cell Biology* **2021**, pp. 6502–6511.
19. Shoka, A.; Dessouky, M.; El-Sherbeny, A.; El-Sayed, A. Literature review on EEG preprocessing, feature extraction, and classifications techniques. *Menoufia J. Electron. Eng. Res* **2019**, *28*, 292–299.
20. Salami, A.; Andreu-Perez, J.; Gillmeister, H. EEG-ITNet: An explainable inception temporal convolutional network for motor imagery classification. *IEEE Access* **2022**, *10*, 36672–36685.
21. Somers, B.; Francart, T.; Bertrand, A. A generic EEG artifact removal algorithm based on the multi-channel Wiener filter. *Journal of neural engineering* **2018**, *15*, 036007.
22. Kwon, M.; Han, S.; Kim, K.; Jun, S.C. Super-resolution for improving EEG spatial resolution using deep convolutional neural network—feasibility study. *Sensors* **2019**, *19*, 5317.
23. Kotte, S.; Dabbakuti, J.K. Methods for removal of artifacts from EEG signal: A review. In Proceedings of the Journal of Physics: Conference Series. IOP Publishing, 2020, Vol. 1706, p. 012093.
24. Singh, A.; Hussain, A.A.; Lal, S.; Guesgen, H.W. A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface. *Sensors* **2021**, *21*, 2173.
25. Stergiadis, C.; Kostaridou, V.D.; Klados, M.A. Which BSS method separates better the EEG Signals? A comparison of five different algorithms. *Biomedical Signal Processing and Control* **2022**, *72*, 103292.
26. Rashid, M.; Sulaiman, N.; PP Abdul Majeed, A.; Musa, R.M.; Bari, B.S.; Khatun, S.; et al. Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review. *Frontiers in neurorobotics* **2020**, *14*, 25.
27. Uribe, L.F.S.; Stefano Filho, C.A.; de Oliveira, V.A.; da Silva Costa, T.B.; Rodrigues, P.G.; Soriano, D.C.; Boccato, L.; Castellano, G.; Attux, R. A correntropy-based classifier for motor imagery brain-computer interfaces. *Biomedical Physics & Engineering Express* **2019**, *5*, 065026.
28. Mridha, M.; Das, S.C.; Kabir, M.M.; Lima, A.A.; Islam, M.; Watanobe, Y.; et al. Brain-Computer Interface: Advancement and Challenges. *Sensors* **2021**, *21*, 5746.
29. Xu, B.; Zhang, L.; Song, A.; Wu, C.; Li, W.; Zhang, D.; Xu, G.; Li, H.; Zeng, H. Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification. *IEEE Access* **2018**, *7*, 6084–6093.
30. Tobón-Henao, M.; Álvarez-Meza, A.; Castellanos-Domínguez, G. Subject-dependent artifact removal for enhancing motor imagery classifier performance under poor skills. *Sensors* **2022**, *22*, 5771.
31. dos Santos, E.M.; Cassani, R.; Falk, T.H.; Fraga, F.J. Improved motor imagery brain-computer interface performance via adaptive modulation filtering and two-stage classification. *Biomedical Signal Processing and Control* **2020**, *57*, 101812.
32. Rajabioun, M. Motor imagery classification by active source dynamics. *Biomedical Signal Processing and Control* **2020**, *61*, 102028.
33. Taran, S.; Bajaj, V. Motor imagery tasks-based EEG signals classification using tunable-Q wavelet transform. *Neural Computing and Applications* **2019**, *31*, 6925–6932.
34. Sadiq, M.T.; Yu, X.; Yuan, Z.; Fan, Z.; Rehman, A.U.; Li, G.; Xiao, G. Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform. *IEEE access* **2019**, *7*, 127678–127692.
35. Zhang, R.; Xiao, X.; Liu, Z.; Jiang, W.; Li, J.; Cao, Y.; Ren, J.; Jiang, D.; Cui, L. A new motor imagery EEG classification method FB-TRCSP+ RF based on CSP and random forest. *IEEE Access* **2018**, *6*, 44944–44950.
36. Hsu, W.Y. Improving classification accuracy of motor imagery EEG using genetic feature selection. *Clinical EEG and neuroscience* **2014**, *45*, 163–168.
37. Bashashati, A.; Fatourechi, M.; Ward, R.K.; Birch, G.E. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering* **2007**, *4*, R32.
38. Van Erp, J.; Lotte, F.; Tangermann, M. Brain-computer interfaces: beyond medical applications. *Computer* **2012**, *45*, 26–34.

39. Altaheri, H.; Muhammad, G.; Alsulaiman, M.; Amin, S.U.; Altuwaijri, G.A.; Abdul, W.; Bencherif, M.A.; Faisal, M. Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review. *Neural Computing and Applications* **2021**, pp. 1–42.

40. Sun, B.; Liu, Z.; Wu, Z.; Mu, C.; Li, T. Graph Convolution Neural Network based End-to-end Channel Selection and Classification for Motor Imagery Brain-computer Interfaces. *IEEE Transactions on Industrial Informatics* **2022**.

41. Ma, Y.; Bian, D.; Xu, D.; Zou, W.; Wang, J.; Hu, N. A Spatio-Temporal Interactive Attention Network for Motor Imagery EEG Decoding. In Proceedings of the 2022 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2022, pp. 1–6.

42. Song, Y.; Jia, X.; Yang, L.; Xie, L. Transformer-based spatial-temporal feature learning for EEG decoding. *arXiv preprint arXiv:2106.11170* **2021**.

43. He, Y.; Lu, Z.; Wang, J.; Shi, J. A channel attention based MLP-Mixer network for motor imagery decoding with EEG. In Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022, pp. 1291–1295.

44. Song, Y.; Wang, D.; Yue, K.; Zheng, N.; Shen, Z.J.M. EEG-based motor imagery classification with deep multi-task learning. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019, pp. 1–8.

45. Berton, L.; Valverde-Rebaza, J.; de Andrade Lopes, A. Link prediction in graph construction for supervised and semi-supervised learning. In Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN). IEEE, 2015, pp. 1–8.

46. Kong, Q.; Wu, Y.; Yuan, C.; Wang, Y. Ct-cad: Context-aware transformers for end-to-end chest abnormality detection on x-rays. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2021, pp. 1385–1388.

47. Yang, L.; Song, Y.; Ma, K.; Xie, L. Motor imagery EEG decoding method based on a discriminative feature learning strategy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2021**, *29*, 368–379.

48. Phunruangsakao, C.; Achancaray, D.; Hayashibe, M. Deep adversarial domain adaptation with few-shot learning for motor-imagery brain-computer interface. *IEEE Access* **2022**, *10*, 57255–57265.

49. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-adversarial training of neural networks. *The journal of machine learning research* **2016**, *17*, 2096–2030.

50. Halme, H.L.; Parkkonen, L. Across-subject offline decoding of motor imagery from MEG and EEG. *Scientific reports* **2018**, *8*, 10087.

51. Marquand, A.F.; Brammer, M.; Williams, S.C.; Doyle, O.M. Bayesian multi-task learning for decoding multi-subject neuroimaging data. *NeuroImage* **2014**, *92*, 298–311.

52. Roy, S.; Chowdhury, A.; McCreadie, K.; Prasad, G. Deep learning based inter-subject continuous decoding of motor imagery for practical brain-computer interfaces. *Frontiers in Neuroscience* **2020**, *14*, 918.

53. Huang, Y.C.; Chang, J.R.; Chen, L.F.; Chen, Y.S. Deep neural network with attention mechanism for classification of motor imagery EEG. In Proceedings of the 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, 2019, pp. 1130–1133.

54. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **2014**, *15*, 1929–1958.

55. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International conference on machine learning. pmlr, 2015, pp. 448–456.

56. Cai, D.; He, X.; Han, J.; Huang, T.S. Graph regularized nonnegative matrix factorization for data representation. *IEEE transactions on pattern analysis and machine intelligence* **2010**, *33*, 1548–1560.

57. Cortes, C.; Mohri, M.; Rostamizadeh, A. Algorithms for learning kernels based on centered alignment. *The Journal of Machine Learning Research* **2012**, *13*, 795–828.

58. Alvarez-Meza, A.M.; Orozco-Gutiérrez, A.; Castellanos-Dominguez, G. Kernel-based relevance analysis with enhanced interpretability for detection of brain activity patterns. *Frontiers in neuroscience* **2017**, *11*, 550.

59. Géron, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*; " O'Reilly Media, Inc.", 2022.

60. García-Murillo, D.G.; Álvarez-Meza, A.M.; Castellanos-Dominguez, C.G. KCS-FCnet: Kernel Cross-Spectral Functional Connectivity Network for EEG-Based Motor Imagery Classification. *Diagnostics* **2023**, *13*, 1122.

61. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering* **2018**, *15*, 056013.

62. Zhang, A.; Lipton, Z.C.; Li, M.; Smola, A.J. Dive into deep learning. *arXiv preprint arXiv:2106.11342* **2021**.

63. Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping* **2017**, *38*, 5391–5420.
64. Musallam, Y.K.; AlFassam, N.I.; Muhammad, G.; Amin, S.U.; Alsulaiman, M.; Abdul, W.; Altaheri, H.; Bencherif, M.A.; Algabri, M. Electroencephalography-based motor imagery classification using temporal convolutional network fusion. *Biomedical Signal Processing and Control* **2021**, *69*, 102826.
65. Berger, V.W.; Zhou, Y. Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online* **2014**.
66. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **2008**, *9*.
67. Giraldo, L.G.S.; Rao, M.; Principe, J.C. Measures of entropy from data using infinitely divisible kernels. *IEEE Transactions on Information Theory* **2014**, *61*, 535–548.
68. Hou, Y.; Jia, S.; Lun, X.; Hao, Z.; Shi, Y.; Li, Y.; Zeng, R.; Lv, J. GCNs-net: a graph convolutional neural network approach for decoding time-resolved eeg motor imagery signals. *IEEE Transactions on Neural Networks and Learning Systems* **2022**.
69. De La Pava Panche, I.; Gómez-Orozco, V.; Álvarez-Meza, A.; Cárdenas-Peña, D.; Orozco-Gutiérrez, Á. Estimating Directed Phase-Amplitude Interactions from EEG Data through Kernel-Based Phase Transfer Entropy. *Applied Sciences* **2021**, *11*, 9803.
70. Song, Y.; Zheng, Q.; Wang, Q.; Gao, X.; Heng, P.A. Global Adaptive Transformer for Cross-Subject Enhanced EEG Classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2023**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.