



Article Deep and Wide Transfer Learning with Kernel Matching for Pooling Data from Electroencephalography and Psychological Questionnaires

Diego Fabian Collazos-Huertas *[®], Luisa Fernanda Velasquez-Martinez, Hernan Dario Perez-Nastar, Andres Marino Alvarez-Meza and German Castellanos-Dominguez [®]

Signal Processing and Recognition Group, Universidad Nacional de Colombia, Manizales 170001, Colombia; lfvelasquezm@unal.edu.co (L.F.V.-M.); hdperezn@unal.edu.co (H.D.P.-N.); amalvarezme@unal.edu.co (A.M.A.-M.); cgcastellanosd@unal.edu.co (G.C.-D.)

* Correspondence: dfcollazosh@unal.edu.co

Abstract: Motor imagery (MI) promotes motor learning and encourages brain-computer interface systems that entail electroencephalogram (EEG) decoding. However, a long period of training is required to master brain rhythms' self-regulation, resulting in users with MI inefficiency. We introduce a parameter-based approach of cross-subject transfer-learning to improve the performances of poorperforming individuals in MI-based BCI systems, pooling data from labeled EEG measurements and psychological questionnaires via kernel-embedding. To this end, a Deep and Wide neural network for MI classification is implemented to pre-train the network from the source domain. Then, the parameter layers are transferred to initialize the target network within a fine-tuning procedure to recompute the Multilayer Perceptron-based accuracy. To perform data-fusion combining categorical features with the real-valued features, we implement stepwise kernel-matching via Gaussian-embedding. Finally, the paired source-target sets are selected for evaluation purposes according to the inefficiency-based clustering by subjects to consider their influence on BCI motor skills, exploring two choosing strategies of the best-performing subjects (source space): single-subject and multiple-subjects. Validation results achieved for discriminant MI tasks demonstrate that the introduced Deep and Wide neural network presents competitive performance of accuracy even after the inclusion of questionnaire data.

Keywords: kernel-embedding; transfer learning; Deep and Wide network; motor imagery

1. Introduction

Motor imagery (MI) is related to the process of mentally generating a quasi-perceptual experience in the absence of any appropriate external stimuli [1]. MI practice promotes children's motor learning and has been suggested to provide benefits in enhancing the musicality of untrained children [2,3], in evaluating the screen-time and cognitive development [4], and improving attentional focus and rehabilitation [5–7], among others. MI-based brain-computer interface (BCI) systems often entail electroencephalogram (EEG)-decoding because of their ease of use, safety, high portability, relatively low cost, and, most importantly, high temporal resolution [8]. EEG is a non-invasive and portable neuroimaging technique that records brain electrical signals over the scalp, reflecting the synchronized oscillatory activity originating from the pyramidal cells of the sensorimotor cortex. However, evoked responses in frequency bands, besides the eliciting stimuli, depend upon every individual. In addition, in MI-based cognitive tasks, the evoked event-related de/synchronization of the sensorimotor area is perturbed by other background brain processes or even artifacts, seriously reducing the signal-to-noise ratio [9]. Hence, to generate steady evoked control patterns, long training must master brain rhythms' self-regulation. As a result, the percentage of users with MI inefficiency (or BCI-illiteracy) is high enough



Citation: Collazos-Huertas, D.F.; Velasquez-Martinez, L.F.; Perez-Nastar, H.D.; Alvarez-Meza, A.M.; Castellanos-Dominguez, G. Deep and Wide Transfer Learning with Kernel-Matching for Pooling Data from Electroencephalography and Psychological Questionnaires. *Sensors* 2021, *21*, 5105. https:// doi.org/10.3390/s21155105

Academic Editors: Natsue Yoshimura, Ludovico Minati, Masaki Nakanishi and Fernando E. Rosas

Received: 13 May 2021 Accepted: 9 July 2021 Published: 28 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). to limit this technology to lab environments even that MI research has been going for many years [10].

In practice, the MI ability can be assessed to determine to what extent a user engages in a mental representation of movements, mainly through self-report questionnaires developed explicitly for this purpose [11]. Yet, there is very little evidence stating a confident correlation between the classification accuracy and the questionnaire scores. Several reasons may account in this regard [12,13]: weak and ambiguous self-interpretation in understanding the questionnaire instructions, laboratory paradigms restricted to a narrow class of motor activity, timeline limitations guaranteeing consistent mental states, and difficulty in learning features from subjects with BCI-illiteracy, among others. Hence, although psychological assessment and questionnaires are probably the most accepted and validated methods in medical contexts [14], their inclusion in the automated prediction of the BCI skills remains very rare due to their disputed reliability and reproducibility [15]. For enhancing the predictive utility, the joint analysis of different imaging modalities is achieved, which may explain the discovered relationships between anatomical, functional, and electrophysiological properties of the brain [16,17]. Nonetheless, besides those issues that may arise by the questionary implementation, research endeavors of multimodal analysis pose a challenging problem in terms of combining categorical data with imaging measurements, facing the following restrictions [18,19]: Different spatial and temporal sampling rates, noninstantaneous and nonlinear coupling, low signal-to-noise ratios, a lack of interpretable results, and the optimal combination of individual modalities is still undetermined, as well as effective dimensionality reduction to enhance the discriminability of extracted multi-view features [20].

Another approach to improve BCI skills is to perform several training sessions in which participants learn how to modulate their sensorimotor rhythms appropriately, relying on the spatial specificity of MI-induced brain plasticity [21]. However, collecting extensive data is time-consuming and mentally exhausting during a prolonged recording session, deteriorating the measurement quality. To overcome this lack of subject-specific data, transfer learning-based approaches are increasingly integrated into MI systems using pre-existing information from other subjects (source domain) to facilitate the calibration for a new subject (target domain) through a set of shared features among individuals under the assumption of a unique data acquisition paradigm [22–24]. Therefore, to have the advantages of transfer learning in EEG signal analysis, strategies for individual difference matching and data requirement reduction are needed to fine-tune the model for the target subject [25]. For example, in [26], the authors use pre-trained models (e.g., VGG16 and Alex-net) as the starting point for approach-fitting. This strategy limits the amount of training data required to support the MI classification task. In this case, they compute the continuous wavelet transform from EEG signals to represent the time-series data into equivalent image representation that can be trained in deep networks. Similarly, Zhang et al. in [27] proposed five schemes for adaptation of a deep convolutional neural network-based EEG-BCI system for decoding MI. Specifically, each procedure fine-tunes a pre-trained model to enhance the evaluation performed on a target subject. Recently, approaches based on weighted instances [28] and domain adaptation [29] have been studied. In the first case, instance-based transfer learning is used to select the source domain data that is most similar to the target domain to assist the training of the target domain classification model. In the second case, researchers extend deep transfer learning techniques to the EEG multi-subject training case. In particular, they explore the possibility of applying maximum-mean discrepancy to align better distributions of features from individual feature extractors in an MI-based BCI system. Nonetheless, to extract sets of shared features among subjects with a similar distribution, there is a need to adequately handle two main limitations of subject-dependent and subject-independent training strategies: small-scale datasets and a significant difference in signals across subjects [30]. In fact, several issues remain as challenges to obtaining adequate consistency of the feature space and probability distribution of training and test data, avoiding negative transfer effects [31,32]: feature

extraction from available multimodal data effective enough to discriminate between MI tasks, and the choosing of transferable objects and transferability measures along with the assignation of their weights [33].

Here, we introduce a parameter-based approach of cross-subject transfer learning for improving poor-performing individuals in MI-based BCI systems, and pooling data from labeled EEG measurements and psychological questionnaires via kernel-embedding. For sharing the discovered model parameters, as presented in [34], an end-to-end Deep and Wide neural network for MI classification is implemented that is, firstly, fed by data from the whole trial set to pre-train the network from the source domain. Then, the layer parameter layers are transferred to initialize the target network within a fine-tuning procedure to recompute the Multilayer Perceptron-based accuracy. To perform data fusion combining categoricals with the real-valued features, we implement the stepwise kernel-matching via Gaussian embedding, resulting in similarity matrices that hold a relationship with the BCI inefficiency clusters. For evaluation purposes, the paired source-target sets are selected according to the inefficiency-based clustering by subjects to consider their influence on BCI motor skills, exploring two choosing strategies of the best-performing subjects (source space): Single-subject and multiple-subjects, as delivered in [35]. The validation results for discriminating MI tasks show that the proposed Deep and Wide neural network gives promising accuracy performance, even after including questionnaire data. Therefore, this deep learning framework with cross-subject transfer learning is a promising way to address small-scale data limitations from the best-performing subjects.

The remainder of this paper is as follows: Section 2 presents the materials and methods, Section 3 describes the experiments and the corresponding results, putting effort into their interpretation. Lastly, Section 4 highlights the conclusions and recommendations.

2. Materials and Methods

2.1. 2D Feature Representation of EEG Data

From the EEG database collected by an *C*-channel montage, we build a single matrix for the *n*-th trial $\{X_n \in \mathbb{R}^C \times T, \lambda_n \in \{0, 1\}^A\}_{n=1}^N$, that contains *T* time points at the sampling rate F_s . Along with the EEG data, we also create the one-hot output vector λ_n in $\Lambda \in \mathbb{N}$ labels. For evaluation in discriminating MI tasks, the proposed transfer learning model is assessed on a trial basis. That is, we extract the feature sets per trial $\{\hat{X}_n^r \in \mathbb{R}^C\}_{r=1}^R$, incorporating a pair of EEG-based feature representation approaches (R = 2): Continuous Wavelet Transform (CWT) and Common Spatial Patterns (CSP), as recommended for Deep and Wide learning frameworks in [36].

Further, the extracted multi-channel features (using CSP and CWT methods) are converted into a two-dimensional topographic interpolation $\mathbb{R}^C \to \mathbb{R}^{W \times H}$ to preserve their spatial interpretation, mapping into a two-dimensional circular view for every extracted trial feature set. As a result, we obtain the labeled 2D data $\{Y_n^z \in \mathbb{R}^{W \times H}, \lambda_n : n \in N\}$, where Y_n^z is a single-trial bi-domain *t*-*f* feature array, termed *topogram*, extracted from every *z*-th set. Of note, the triplet $z = \{r, \Delta_t, \Delta_f \text{ (with } z \in Z) \text{ indexes a topogram estimated for each included domain principle <math>r \in R$ at the time-segment $\Delta t \in T$, and within the frequency-band $\Delta f \in F$.

Besides, we estimate the local spatial patterns of relationships from the input topographic set through the square-shaped layer kernel arrangement $\{K_{i,l}^z \in \mathbb{R}^{P \times P}\}^{I_l, Z}$ (as in straightforward convolutional networks), where *P* holds the kernel size. Therefore, the number of kernels varies at each layer $i \in I_l$, so that the stepwise 2D-convolutional operation is performed over the input topogram, Y^z , as follows:

$$\hat{\mathbf{Y}}_{L}^{z} = (\varphi_{L}^{z} \circ \dots \circ \varphi_{1}^{z})(\mathbf{Y}^{z}), \tag{1}$$

where $\varphi_l^z(\hat{\mathbf{Y}}_{l-1}^z) = \gamma_l(\mathbf{K}_{i,l}^z \otimes \hat{\mathbf{Y}}_{l-1}^z + \mathbf{B}_{i,l}^z)$ is the convolutional layer, followed by a non-linear activation function $\gamma_l : \mathbb{R}^{W_l^z} \times H_l^z \to \mathbb{R}^{W_l^z} \times H_l^z$, $\hat{\mathbf{Y}}_l^z \in \mathbb{R}^{W_l^z \times H_l^z}$ is the resulting 2D feature map of the *l*-th layer (adjusting $\hat{\mathbf{Y}}_0^z = \mathbf{Y}^z$), and the arrangement $\mathbf{B}_{i,l}^z \in \mathbb{R}^{W_l^z \times H_l^z}$ denotes

the bias matrix. Notations \circ and \otimes stand, respectively, for the function composition and convolution operator.

2.2. Multi-Layer Perceptron Classifier Using 2D Feature Representation

In this stage, we employ the deep learning-based classifier function $\varphi : \mathbb{R}^{W \times H} \mapsto \Lambda$ developed through a Multilayer Perceptron (MLP) Neural Network that predicts the label probability vector $\tilde{v} \in \{0, 1\}^{\Lambda}$, as below [37]:

$$\tilde{\boldsymbol{v}} = \varphi(\boldsymbol{u}_0, \boldsymbol{\Theta}; \boldsymbol{\phi}_D^z \circ \cdots \circ \boldsymbol{\phi}_1^z), \tag{2a}$$

s.t.:
$$\Theta_0^* = \arg \min_{K_i^z, \lambda_d, B_i^z, \lambda_d} \{ \mathcal{L}(\tilde{v}_n, \lambda_n | \Theta); \forall n \in N \}$$
 (2b)

where $\phi_d(u_{d-1}) = \eta_d(A_d u_{d-1} + \alpha_d)$ is the fully-connected layer ruled by the non-linear activation function: $\eta_d : \mathbb{R}^{P'_d} \to \mathbb{R}^{P'_d}$, $P'_d \in \mathbb{N}$ is the number of hidden units at the *d*-th layer, $d = \{0, \ldots, D\}$ (d = 0 is the initial concatenation before the classification layer), $A_d \in \mathbb{R}^{P'_d} \times P'_{d-1}$ is the weighting matrix containing the connection weights between the preceding neurons and the hidden units P' of layer d, $\alpha_d \in \mathbb{R}^{P'_d}$ is the bias vector, and the $u_d \in \mathbb{R}^{P'_d}$ hidden layer vector holds the extracted spatial information encoded by the resulting 2D feature maps in the Q domain.

For computation at each layer, the hidden layer vector is iteratively updated by the rule (composition function-based approach of deep learning methods) $u_d = \phi_d(u_{d-1})$, for which the initial state vector is flattened by concatenating all matrix rows across z and I_l domains as $u_0 = [\operatorname{vec}(\hat{Y}_L^z) : \forall z \in Z]$. The input vector u_0 sizes $G = W'H'Z \sum_{l \in L} I_l$, holding W' < W, H < H'. Besides, the optimizing estimation framework of label adjustment estimates the training parameter set $\Theta_0 = \{K_{i,l}^z, A_d, b_{i,l}^z, \alpha_d\}$, fixing the loss function $\mathcal{L} : \mathbb{R}^A \times \mathbb{R}^A \to \mathbb{R}$ to calculate the gradients employed to update the weights and bias of the proposed Deep and Wide neural network through a certain number of training epochs. Remarkably, we refer to our method as Deep and Wide because of the inclusion of a set of different topograms (along time and frequency domains) from the extracted multi-channel features using CSP and CWT algorithms. A mini-batch-based gradient implements the solution, as commonly used in deep learning methods, equipped with automatic differentiation and back-propagation [38].

2.3. Transfer Learning with Added Questionnaire Data

In EEG analysis based on Deep Learning, for enhancing the classifier performance, transfer learning is a common approach to adjust a pre-trained neural network model equipped with the label probability vector \tilde{v} , aiming to provide a close domain distance measurement $\delta(\cdot, \cdot)\mathbb{R}^+$, lower than a given value $\epsilon \in \mathbb{R}^+$, between the paired domains to approximate the source $\mathcal{Y}^{(s)}$ to the target $\mathcal{Y}^{(t)}$ [24], as follows:

$$\delta(\mathfrak{Y}^{(s)}(\boldsymbol{Y},\boldsymbol{S}),\mathfrak{Y}^{(t)}(\boldsymbol{Y},\boldsymbol{S})|\tilde{\boldsymbol{v}}) \leq \epsilon$$
(3)

s.t.:
$$\Theta^* = \{K_{i,l}^{q*}, B_{i,l}^{q*}\}$$
 (4)

Here, we propose to conduct the transfer learning procedure to learn a target prediction function that is enhanced by the addition of the categorical assessments of a psychological questionnaire data matrix, *S*, along with the stepwise multi-space kernel-embedding, including EEG-based features, to perform the whole network parameter optimization in Equation (2b). Besides, for interpretation purposes, selecting the paired source–target sets is accomplished according to the inefficiency-based clustering of subjects.

Therefore, to combine the categorical data, *S*, with the real-valued feature map set extracted from EEG as exposed in Sections 2.1 and 2.2, *Y*, we compute the tensor product space between the corresponding kernel-matching representations, $\kappa_{\hat{u}}$ and κ_{S} , as suggested in [39]:

$$\bar{\kappa} = \kappa_{\hat{U}} \circ \kappa_{S}, \quad \bar{\kappa} \in \mathbb{R}^{J \times J} \tag{5}$$

where $J = \sum_{m=1}^{M} N_m$ (N_m holds the trials for the *m*-th subject), $\kappa_S \in \mathbb{R}^{J \times J}$ is the kernel matrix directly extracted from the questionnaire data $S \in \mathbb{R}^{J \times N_Q}$ (N_Q is the questionnaire vector length), $\kappa_{\hat{u}} \in \mathbb{R}^{J \times J}$ is the kernel topographic matrix estimated from the projected version $\hat{u} = uY^*$, with $\hat{u} \in \mathbb{R}^{J \times G'}$ (holding that G' < G), $u \in \mathbb{R}^{J \times G}$ $u \in \mathbb{R}^{J \times G}$ is the initial data matrix build by concatenating across the trial and subject sets all flattened vectors u_0^* , which are computed by adjusting the optimized parameters $\Theta^* = \{K_{i,l}^{q*}, B_{i,l}^{q*}\}$, and $Y^* \in \mathbb{R}^{G \times G'}$ is the projection matrix introduced to maximize the similarity between both estimated kernel-embeddings derived from the labeled EEG measurements of MI responses, namely, one from the one-hot label vectors, $\kappa_V \in \mathbb{R}^{J \times J}$, and another from the topographic features, $\kappa_U \in \mathbb{R}^{J \times J}$.

In particular, we match both estimated kernel-embeddings through the centered kernel alignment (CKA), as detailed in [40]:

$$\mathbf{Y}^* = \arg\max_{\mathbf{x}} \operatorname{CKA}(\kappa_{\boldsymbol{U}}, \kappa_{\boldsymbol{V}}) \tag{6}$$

where the kernel κ_V is obtained from the matrix of predicted label probabilities $V \in \mathbb{R}^{J \times \Lambda}$ build by concatenating across the trial and subject sets all label probability vectors \tilde{v}_{nn} .

3. Experimental Set-Up

Training of the proposed Deep and Wide neural network model for transfer learning to improve classification of MI responses, including EEG and questionnaire data, encompasses the following stages (see Figure 1): (i) Preprocessing and spatial filtering of EEG signals, followed by 2D features extracted from the input topogram set using the convolutional network (see Section 2.1). (ii) MLP classification applying the extracted 2D feature maps (see Section 2.2), (iii) Cross-subject transfer learning, including stepwise multi-space kernel-embedding of the real-valued and categorical variables (see Section 2.3). The paired source-target sets are selected according to the inefficiency-based clustering by subjects to consider their influence on BCI motor skills.

Nonetheless, the classifier performance can decrease since the extracted representation sets may still involve irrelevant and/or similar features. Therefore, for reducing the data complexity, we accomplish dimensionality reduction by evaluating a widely-used unsupervised feature extractor of Kernel PCA (KPCA) that provides a representation of data points' global structure [41].



Figure 1. Guideline of the proposed transfer learning approach, including Stepwise Kernel Matching to combine data from Electroencephalography and Psychological Questionnaires.

3.1. Database Description and Preprocessing

GigaScience (publicly available at http://gigadb.org/dataset/100295 (accessed on 9 July 2021)): This acquisition holds EEG data recorded by a BCI experimental paradigm of MI movement collected from 52 subjects (though only 50 is available). Data were acquired by a 10–10 *C*-electrode system C = 64 with 512 Hz sampling rates, collecting 100 individual

trials (each one lasting 7 s) in either task (left or right hand). The MI paradigm begins with a fixation cross presented on a black screen within 2 s. Next, a cue instruction appeared randomly on the screen for 3 s to ask each subject to imagine moving the fingers, starting to from the forefinger and proceeding to the little finger, touching each to their thumb. A blank screen was then shown at the beginning of a break period, lasting randomly between 4.1 and 4.8 s. For each MI class, these procedures were repeated 20 times within a single testing run.

GigaScience also collected subjective answers to physiological and psychological questionnaires (categorical data), intending to investigate the evidence on performance variations to work out strategies of subject-to-subject transfer in response to intersubject variability. To this end, all subjects were invited to fill out a questionnaire during three different phases of the MI paradigm timeline: before beginning the experiment (each subject answered $N_Q = 15$ questions); after every run within the experiment ($N_Q = 10$ questions were answered); and at the experiment's termination ($N_Q = 4$ answered questions, { $Q_i : i = 1, 2, 3, 4$ }).

As preprocessing, we filtered each raw channel $x_n^c \in \mathbb{R}^T$ within [8–30] Hz using a five-order Butterworth band-pass filter. Further, we carry out a bi-domain short-time feature extraction (i.e., CWT and CSP—see Section 2.1), as performed in [42]. In the former extraction, the wavelet coefficients are assumed to provide a compact representation pinpointing the EEG data energy distribution, yielding a time-frequency map in which the amplitudes of individual frequencies (rather than frequency bands) are represented. In the latter extraction, the goal of CSP is to employ a linear relationship to transfer a multichannel EEG dataset into a subspace with a lower dimension (i.e., latent source space), aiming to enhance the class separability by maximizing the labeled covariance in the latent space. In both extraction cases, we fix the sliding short-time window length parameter $\tau \in \mathbb{R}^+$ according to the accuracy achieved by the baseline Filter Bank CSP algorithm that is performed using the whole range of considered frequency bands. The sliding window is adjusted to $\tau = 2$ s with a step size of 1 s as an appropriate choice to extract $N_{\tau} = 5$ EEG segments, as performed in [43]. Since electrical brain activities provoked by MI tasks are commonly related to μ and β rhythms [44], the spectral range is split into the following bandwidths of interest: $\Delta f \in \{\mu \in [8-12], \beta \in [12-30]\}$ Hz. The CWT feature set is computed by the Complex Morlet function frequently applied in the spectral EEG analysis, fixing a scaling value to 32. Additionally, we set the number of CSP components as 3Λ ($\Lambda \in \mathbb{N}$ holds the number of MI tasks), utilizing a regularized sample covariance estimation.

3.2. MLP Classifier Performance Fed by 2D Features

At this stage, we carry out the extraction of 2D feature maps from the input topogram set using the convolutional network. Further, the 2D features extracted to feed the MLP-based classifier with the parameter tuning shown in Table 1, and the resulting layer-by-layer model architecture is illustrated in Figure 2. For implementation purposes, we apply the Adam algorithm using the optimizing procedure with fixed parameters: a learning rate of 1×10^{-3} , 200 training epochs, and a batch size of 256 samples. Additionally, the mean squared error (MSE) is chosen as the loss function $\mathcal{L}(:)$ in Equation (2b), that is, $\mathcal{L}(\tilde{v}_n, \lambda_n | \Theta) = \mathbb{E}\left\{(\tilde{v}_n - \lambda_n)^2\right\}$. For speeding the learning procedure, the Deep and Wide neural network framework is written in Python code (TensorFlow toolbox and Keras API) trained to employ multiple GPU devices at the Google Colaboratory. The codes are made available at a public GitHub repository (codes available at https://github.com/dfcollazosh/DWCNN_TL (accessed on 9 July 2021)).

As the performance measure, the classifier accuracy $A_c \in [0, 1]$ is computed by the expression: $A_c = (T_P + T_N)/(T_P + T_N + F_P + F_N)$, where T_P , T_N , F_P , and F_N are true-positives, true-negatives, false-positives, and false-negatives, respectively. In this case, we split the subject's dataset and built the training set using 90% of trials and the remaining 10% for the test set. Further, the individual training trial set is randomly partitioned by a stratified 10-fold cross-validation to generate a validation trial partition.

Table 1. Detailed *Deep and Wide* architecture of transfer learning. Layer FC8 accomplishes the regularization procedure using the *Elastic-Net* configuration, while layers FC8 and OU10 apply a kernel constraint adjusted to *max_norm*(1.). Notation $O = RN_{\Delta}N_{\tau}$, N_{Δ} denotes the number of filter banks, P'—the number of hidden units (neurons), C—the number of classes, and I_L stands for the amount of kernel filters at layer L. Notation $|| \cdot ||$ stands for the concatenation operator.

Layer	Assignment	Output Dimension	Activation	Mode
IN1	Input	$ 40 \times 40 $		
CN2	Convolution	$ 40 \times 40 \times 2 $	ReLu	<i>Padding</i> = SAME
				$Size = 3 \times 3$
				$Stride = 1 \times 1$
BN3	Batch-normalization	40 imes 40 imes 2		
MP4	Max-pooling	20 imes 20 imes 2		$Size = 2 \times 2$
				$Stride = 1 \times 1$
CT5	Concatenation	$ 20 imes 20 imes O \cdot I_L $		
FL6	Flatten	$20 \cdot 20 \cdot O \cdot I_L$		
BN7	Batch-normalization	$20 \cdot 20 \cdot O \cdot I_L$		
FC8	Fully-connected	P' imes 1	ReLu	Elastic-Net
				max_norm(1.)
BN9	Batch-normalization	$ P' \times 1 $		
OU10	Output	$ C \times 1 $	Softmax	max_norm(1.)



Figure 2. Scheme of the proposed Deep and Wide neural network architecture to support MI discrimination.

For the tested subject set, Figure 3 displays the results of accuracy that the MLPbased classifier produces if fed by just the 2D feature set extracted before. From the obtained accuracy values, we evaluate the performance to be considered as inadequate in brain–computer interface systems as detailed in [45]. Namely, we cluster the individual set into the following three groups with distinctive BCI skills: (i) Group of individuals performing the highest accuracy but with very low variability of neural responses (colored in green). (ii) A group that reaches superior classifier performance but with some response fluctuations (yellow color). (iii) A group that produces modest performance along with a high unevenness of responses (red color).





Figure 3. Partitions of individuals clustered by the MLP-based accuracy. Each subject performance is painted by this estimated BCI inefficiency partition: Group *I* (green), Group *II* (yellow), and Group *III* (red).

3.3. Performed Stepwise Multi-Space Kernel Matching

Algorithm 1 presents the procedures to complete the validation of the suggested transfer learning with multi-space kernel-embedding. We implement the Gaussian kernel to represent the available data because of its universal approximating ability and mathematical tractability. The length scale hyperparameter $\sigma \in \mathbb{R}^+$, ruling the variance of the described data, is adjusted to their median estimate. The following steps (3: and 4:) accomplish the pairwise kernel matching, firstly between the sets of EEG measurement U and label probability V. To this end, the CKA matching estimator is fed by the concatenated EEG features together with the predicted label probabilities to perform alignment across the whole subject set, empirically fixing the parameter G' to 50 according to the subjects' number in this experiment. In the second matching, we encode all the available categorical information about the psychological and physiological evaluation with the relevant feature set, resulting from CKA, by their projection onto a common matrix space representation, using the kernel/tensor product. Note that the projected data \hat{U} by CKA are also embedded. We also perform dimensionality reduction of the feature sets generated after stepwise-matching using Kernel Principal Component Analysis (KPCA) for evaluating the representational ability.

Further, we estimate the subject similarity matrix from the extracted feature sets, aiming to assess the domain distance between the source-target pairs, which are to be selected from different clusters of BCI inefficiency. Since the clustering of individuals relies on the ordered accuracy vector, we introduce the following neighboring similarity matrix $\bar{\Delta}_{\hat{c}}$ with pairwise metric elements computed from the matrices $\hat{\zeta} = \{\hat{\kappa}, \hat{\kappa}_{KPCA}\}$, as follows:

$$\bar{\Delta}_{\hat{\xi}}(m,m') = \operatorname{cov}(\operatorname{seq}(\Delta_{\hat{\xi}}(m,\forall m')),\operatorname{seq}(\Delta_{\hat{\xi}}(m',\forall m))), \quad \bar{\Delta}_{\hat{\xi}}(m,m') \in \bar{\mathbf{\Delta}}_{\hat{\xi}} \in \mathbb{R}^{M \times M} \quad (7)$$

$$\Delta_{\hat{\xi}}(m,m') = \sum_{\forall j \in J} |\hat{\xi}(m,j) - \hat{\xi}(m',j)|^2, \quad \Delta_{\hat{\xi}}(m,m') \in \mathbf{\Delta}_{\hat{\xi}} \in \mathbb{R}^{M \times M}$$

where notations $cov(\cdot, \cdot)$ and $seq(\Delta(m, \forall m')) \in \mathbb{R}^M$ stand for, respectively, the covariance operator and the sequence composed of all $\forall m' \in M$ elements of row *m* ranked in decreasing order of the achieved MLP-based accuracy. The rationale for applying the covariance over the ranked row vectors of $\mathbf{\Delta}_{\hat{\boldsymbol{\zeta}}}$ is to preserve the similarity information between neighboring subjects.

Algorithm 1 Validation procedure of the proposed approach for transfer learning with stepwise, multi-space kernel matching. [†] Dimensionality reduction is an optional procedure performed for comparison purposes.

Input data: EEG measurement U, predicted label probabilities V, questionary data $S, \forall m \in M$

- 1: INITIAL PARAMETER SET ESTIMATION Θ_0^* : Compute the baseline MLP-based accuracy from \boldsymbol{U} and \boldsymbol{V} by optimizing $\Theta = \{\boldsymbol{K}_{i,l}^q, \boldsymbol{A}_d, \boldsymbol{b}_{i,l}^q, \boldsymbol{\alpha}_d\}$
- 2: for $\forall m \in M, n \in N$ do
- 3: KERNEL MATCHING between EEG measurement U and labels V, Compute kernel-embedding of input data $\xi = {\kappa_{U}, \kappa_{S}, \kappa_{V}}$: $\kappa_{\xi} = \mathcal{N}_{\xi}(\mu_{\xi}, \sigma_{\xi})$ Compute Center Kernel Alignment between both spaces: $CKA(\kappa_{U}, \kappa_{V})$
- 4: KERNEL MATCHING on supervised EEG representation for the categorical data Compute kernel-embedding of projected data \hat{U} using $\kappa_{\hat{U}} = \mathcal{N}_{\hat{U}}(\mu_{\hat{U}}, \sigma_{\hat{U}})$ Compute tensor product, including the categorical data $\bar{\kappa} = \kappa_{\hat{U}} \circ \kappa_{S}, \bar{\kappa} \in \mathbb{R}^{J \times J}, J = NM$
- 5: end for
- DIMENSIONALITY REDUCTION[†] by Kernel Principal Components: $\bar{\kappa}_{KPCA} \in \mathbb{R}^{J \times J}$ 6: TRANSFER LEARNING OF PAIRED SOURCE-TARGET SUBJECTS: $\mathcal{Y}^{(s)}$ and $\mathcal{Y}^{(t)}$
 - Perform matrix reshaping $\mathbb{R}^{J \times J} \mapsto \mathbb{R}^{M \times J}$: $\hat{\xi} = {\{\hat{\kappa}, \hat{\kappa}_{KPCA}\}}$ Compute the neighboring similarity matrix of individuals: $\bar{\Delta}, \bar{\Delta}_{KPCA}$ Compute the intra-subject distance matrix through the domain distance measurement: $\bar{\delta}_{\hat{\epsilon}}(m) \in \mathbb{R}^+, \forall m \in M$

Select paired subjects for each transfer learning strategy evaluated:

- (*a*) One-source versus one-target, (*b*) multiple-source versus one-target
- Recompute the MLP-based accuracy of targets, initializing the parameter set as $\Theta = \Theta_0^*$, fixing the *P*' parameter according to the source subject.
- 7: **Output data**: Accuracy gain achieved by each individual target, according to the selection transfer learning strategy evaluated.

Figure 4 displays the similarity matrix performed by the tensor product $\Delta_{\hat{\zeta}}$ (left column), evidencing some of the relations between the clustered subjects, but depending on the evaluated questionary data. Thus, the collection Q_1 yields two groups, while Q_4 exhibits three partitions. Instead, Q_2 and Q_3 do not cluster the individuals precisely. After KPCA dimensionality reduction, however, the proximity assessments $\bar{\Delta}_{KPCA}$ tend to make the neighboring association more solid, resulting in clusters of subjects with more distinct feature representation, as shown in the middle column for each questionary.

Under the assumption that the closer the association between the paired source-target couples, the more effectively their cross-subject transfer learning is implemented, we estimate the marginal distance $\bar{\delta}_{\hat{\zeta}}(m) \in \mathbb{R}^+$ from either version $\Delta_{\hat{\zeta}}$, $\bar{\Delta}_{KPCA}$ by averaging the neighboring similarity of each subject over the whole set, as follows:

$$\bar{\delta}_{\hat{\xi}}(m) = \mathbb{E}\Big\{ |\bar{\Delta}_{\hat{\xi}}(m, m')| : \forall m' \in M \Big\},\tag{8}$$

where the notation $\mathbb{E}\{z : \forall \zeta\}$ stands for the expectation operator computed across the whole set $\{\zeta\}$.

The right column displays the values of marginal values $\delta_{\hat{\xi}}(m)$, showing that each individual is differently influenced by the stepwise multi-space kernel matching of electroencephalography to psychological questionnaires Q_i . These results are in agreement with the subject cluster properties evaluated above. Thus, Q_1 and Q_4 , having more discernible partitions, yield the feature representations that are more even in the subject set, while Q_2 and Q_3 provide irregular representations. One more aspect is the effect of dimensionality reduction that improves the representation of Q_1 and Q_2 cases. On the contrary, the use of KPCA tends to worsen the global similarity level of individuals.



Figure 4. Similarity matrix performed by the tensor product and computed domain marginal values $\bar{\delta}_{\xi}(m)$. The subjects are ranked in decreasing order of accuracy.

3.4. Estimation of Pre-Trained Weights for Cross-Subject Transfer Learning

The following step is to pair the representation learned on a *source* to be transferred to a given *target* subject. Starting from the subject partitions according to their BCI skills performed above in Section 3.2, we select the candidate sources (i.e., the source space $\mathcal{Y}^{(s)}(,)$) within the best-performing subjects (Group *I*), while the target space $\mathcal{Y}^{(t)}(,)$ becomes the worst-performing participants (Group *III*). Here, we validate two choosing strategies of subjects from the source space (Group *I*):

(a) *Single source-single-target*, when we select the subject of Group *I*, achieving the highest value of the domain distance measurement in Equation (9) computed as follows:

$$\max_{\forall m \in \text{Group } I} \bar{\delta}_{\hat{\xi}}(m; Q_i) \tag{9}$$

Once the *source-target* pairs are selected, the pre-trained weights are computed from each designed *source* subject to initialize the Deep and Wide neural network, rather than introducing a zero-valued starting iterate, and thus enabling a better convergence of the training algorithm. Note that the fulfilling condition in Equation (9) depends on Q_i , meaning distinct selected sources for each questionnaire data.

(b) Multiple sources-single-target when the selected subjects of Group I achieve the four highest domain distance values. In this case, the Deep and Wide initialization procedure applies the pre-trained weights estimated from the concatenation of the source topograms.

Figure 5 details the reached classification performance using the proposed transfer learning approach for either strategy of selecting the candidate sources through a radar diagram that includes all target subjects (axes). For comparison's sake, the graphical representation depicts, with a line colored in black, the MLP-based accuracy (see Figure 3) as a reference for assessing the performance classifier gain due to the applied transfer learning approach, the accuracy achieved by the features extracted by the tensor product (blue line) and KPCA (magenta line), $\hat{\kappa}$, $\hat{\kappa}_{KPCA}$, respectively.

The odd columns (first and third) present the *Single source-Single-target* diagrams, while the even ones are for the *Multiple sources-Single-target* strategy. In all cases of questionnaire data Q_i , the transfer learning with stepwise, multi-space kernel matching allows increasing, on average, the baseline classifier performance of the subjects belonging to Group *III* with modest accuracy and high unevenness of responses. Nevertheless, there are still some issues to be clarified. The accuracy gain performed by the *Single source-Single-target* strategy is lower than the one achieved by the latter approach, but the number of subjects that benefit from the transfer learning approach is higher. On the contrary, the presence of multiple sources halves the number of poor-performing subjects that are improved, though they produce accuracy gain values up to 25% (see subject #45). The next aspect of addressing is the contribution of categorical data in terms of classifier performance. The first two radars in the bottom row (labeled as EEG) present the accuracy improvement performed by the features extracted from the EEG measurements after CKA alignment (CKA(κ_u , κ_v)), underperforming the transfer learning adding questionnaires.

Topographic maps of representative subjects (computed with and without transfer learning) using just the feature map information, presenting the learned weights with assumed meaningful activity.

Regarding the dimensional reduction additionally considered, its delivered accuracy (outlined in magenta) strongly depends on the specific case of fused data Q_i . Thus, while Q_1 and Q_2 benefit from the KPCA procedure, Q_4 reduces the performance achieved. This result becomes evident in two bottom radars (3-th and 4-th) that depict the effect of transfer learning averaged across the data Q_i , showing that the classifier performance of almost each target individual can be enhanced by the proposed transfer learning approach for either strategy of selecting the candidate sources. However, there are a couple of subjects (# 38 and # 20) that did not have a positive impact.



Figure 5. Achieved accuracy by validated strategies of selecting source subjects from Group I. (**a**,**c**) *Single source-single-target*, (**b**,**d**) *Multiple sources-single-target*. Individual gain reports the average accuracy per subject of questionnaire data Q_i and EEG.

Lastly, the topographic maps shown in Figure 6 give a visual interpretation of the proposed transfer learning, which are reconstructed from the learned network weights according to the algorithm introduced in [37]. We compare the estimated approaches under the assumption that the discriminating power is directly proportional to the reconstructed weight value. Thus, the top row shows the topograms of the single-source strategy built from both bandwidths (β and μ) within different intervals of the neural response. As seen, the source selected (subject #3) performs a weight set with the spatial distribution related to the sensorimotor area, focusing their neural responses within the MI segment correctly. Next to S #3, we present the target's topograms that benefit the most from the transfer learning, holding weights with a spatial distribution that is a bit blurred. The effect of the single-source transfer learning approach is the reduction of the weight variability, as shown in the adjacent topograms. However, the source effectiveness to reduce the variability is limited in the case of the low-skilled target #38 that presents many contributing weights spread all over the scalp area. Moreover, the weights appear inside the two intervals (before cue-onset and the ending segment) at which the responses elicited by MI tasks are believed to vanish. As a result, the single-source strategy yields a negative accuracy gain of Target #38 (it drops from 70% to 65%).



Figure 6. Topographic maps of representative subjects with and without transfer learning using just feature map information, presenting the learned weights with meaningful activity reconstructed within both bandwidths (β and μ) across the whole signal length, N_{τ} .

Similar behavior is also observed in the second row, displaying the topograms of the multi-source strategy performed by the most benefitting (T#11) and the worst-achieving target (T#22), respectively. However, the inclusion of multiple sources leads to weights with a sparse distribution, as observed in the topograms of the selected subjects (S#3,14,41,28). This effect may explain the small number of targets improved by the multi-source strategy. In order to clarify this point, the bottom row displays the corresponding spatial distribution performed by the multi-source strategy when including the whole subject set of Group I, resulting in weights that are very weak and scattered. Moreover, compared with the first two rows, the all-subjects source approach of the bottom row makes the related transfer learning deliver the worst performance averaged across the target subject set.

4. Discussion and Concluding Remarks

Here, we introduce a cross-subject transfer learning approach for improving the classification accuracy of elicited neural responses, pooling data from labeled EEG measurements and psychological questionnaires through a stepwise multi-space kernel-embedding. For validation purposes, the transfer learning is implemented in a Deep and Wide framework, for which the source-target sets are paired according to the BCI inefficiency, showing that the classifier performance of almost each target individual can be enhanced using single or multiple sources.

From the evaluation results, the following aspects are to be highlighted:

Evaluated NN framework: The Deep and Wide learning framework is supplied by the 2D feature maps extracted to support the MLP-based classifier. As a result, Table 2 compares the bi-class accuracy of the *GigaScience* database achieved by several recently published approaches, which are outperformed by the learning algorithm with the proposed transfer learning method. Of note, the MSNN algorithm presented in [46] achieves a competitive classification accuracy on average, 82.6 (ours) vs. 81.0 (MSNN), but with a higher standard deviation in comparison with our proposal, 12.0 vs. 8.4. Besides, our method can include categorical data from questionnaires within the MI paradigm, which favors the interpretability concerning the studied subject from spatial, time, and frequency patterns from EEG data coupled with categorical physiological and psychological data.

Table 2. Comparison of bi-class accuracy achieved by state-of-the-art approaches in *GigaScience*. The best value is marked in bold. Notation * denotes Deep and Wide framework results with transfer learning (TL). CSP + FLDA: Common spatial patterns and Fisher linear discriminant analysis, LSTM + Optical: Long-short term memory network and optical predictor, SFCSP: Sparse filter-bank CSP, DCJNN: Deep CSP neural network with joint distribution adaptation, MINE+EEGnet: Mutual information neural estimation, MSNN: Multi-scale Neural Network.

Approach	A_c	Interpretability
CSP + FLDA [47]	67.60	_
LSTM + Optical [48]	68.2 ± 9.0	_
SFBCSP [49]	72.60	_
DCJNN [50]	76.50	\checkmark
MINE + EEGnet [51]	76.6 ± 12.48	\checkmark
MSNN [46]	81.0 ± 12.00	\checkmark
Proposal	79.5 ± 10.80	\checkmark
Proposal + TL *	$\textbf{82.6} \pm \ \textbf{8.40}$	\checkmark

Feature representation challenges and computational requirements: The bi-domain extraction is presented (CWT and CSP) to deal with the substantial intra-subject variability in patterns across trials. However, for improving their combination with categorical data, more compact feature representations can be explored, for instance, using connectivity metrics like in [52]. Besides, neural network architectures capturing the temporal dynamics local structures of the EEG time-series associated with the elicited MI responses could be helpful to upgrade our approach [53]. Moreover, it is well-known that deep learning approaches require considerable computational time when training the model. For clarity, a computational time experiment is carried out. Specifically, for the parameter setting of the FC8 layer, with regularization values *l*1 and *l*2 tuned by a grid search around [0.0005, 0.001, 0.005], and a number of neurons fixed through a grid search within P' = [100, 200, 300], the fitting time with and without our transfer learning approach is summarized in Table 3. As seen, the multi-source scheme requires more computation time per fold. Still, real-time BCI requirements can be satisfied once the model is trained, and a new instance must be predicted. In short, for a new subject, the following stages must be carried out: (i) Store the EEG and questionnaire information of the new and training subjects. (ii) Apply our transfer learning approach as exposed in Figure 1 to couple EEG and questionary psychological data for the new subject. (iii) Once the model is trained, new instances of the studied subject can be predicted as straightforward deep learning methods (in this stage, only the EEG data is required).

Table 3. Computational time experiments. The achieved training time (average) per fold and epoch is presented.

Approach	Time per Fold	Time per Training Epoch
Proposal (Single-source)	\sim 984 s	<1 s
Proposal (Multi-source (4))	$\sim 1663 \text{ s}$	<1 s
Proposal (Multi-source (all))	\sim 3176 s	$\sim 1 \mathrm{s}$
Proposal + TL	~341 s	<1 s

Multi-space kernel matching: To overcome the difficulties in utilizing data-fusion combining categorical with the real-valued features, we implement the stepwise kernel matching via Gaussian embedding. As a consequence, the obtained similarity matrices evidence the relationship with the BCI inefficiency clusters of subjects. Even though the association is highly influenced by each evaluated questionnaire data, this result becomes essential in light of previous reports stating that no statistically significant differences can be detected between questionary scores and EEG-based performance [54]. One more aspect is the effect of dimensionality reduction through kernel PCA that improves the representation, but to a certain extent (only in Q_1 and Q_2 cases). For tackling the differences in subjective criteria for predicting MI performance, however, two main issues need to be addressed: The use of more appropriate kernel-embedding for categorical scores [55] and dimensionality reduction approaches, providing representation of data points with a wide range of structures like *t*-Distributed Stochastic Neighbor Embedding [56].

Cross-subject transfer learning: We conduct the transfer learning to infer a target prediction function from the kernel spaces embedded before, selecting the paired source-target sets according to the Inefficiency-based clustering by subjects. Overall, the transfer learning with feature representations, combined with questionary data, allows for an increase of the baseline classifier accuracy of the worst-performing subjects. Nevertheless, source selection through a different method impacts the classifier performance; while the *Multiple-source-Single-target* strategy tends to produce accuracy improvements that are bigger than the *Single-source-Single-target*, and the number of the benefited targets declines. This result may point to future exploration of more effective transfer learning of BCI inefficiency devoted to bringing together, as much as possible, the source domain to each target space. This task also implies improving the similarity metric in Equation (7) proposed for comparing ordered-by-accuracy vectors of different BCI inefficiency clusters.

As future work, the authors plan to validate the cross-subject transfer learning approach in applications with the joint incorporation of two or more databases (cross-database), growing the tested number of individuals significantly. For instance, we plan to consider the dataset collected by the Department of Brain and Cognitive Engineering, Korea University in [57], since this set holds questionnaire data information about the physiological and psychological condition of subjects. As a result, we will obtain classification performances based on transfer learning at intra-subject and inter-dataset levels.

Author Contributions: Conceptualization, D.F.C.-H., A.M.A.-M. and G.C.-D.; methodology, D.F.C.-H. and H.D.P.-N.; validation, D.F.C.-H., H.D.P.-N. and L.F.V.-M.; data curation, D.F.C.-H., H.D.P.-N. and L.F.V.-M.; writing—original draft preparation, D.F.C.-H. and A.M.A.-M.; writing—review and editing, D.F.C.-H. and G.C.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This research manuscript is developed supported by "Convocatoria Doctorados Nacionales COLCIENCIAS 727 de 2015" and "Convocatoria Doctorados Nacionales COLCIENCIAS 785 de 2017" (Minciencias). Additionally, A.M. Álvarez-Meza thanks to the project: *Prototipo de interfaz cerebro-computador multimodal para la detección de patrones relevantes relacionados con trastornos de impulsividad*-Hermes 50835, funded by Universidad Nacional de Colombia.

Informed Consent Statement: No aplicable since this study uses duly anonymized public databases.

Data Availability Statement: The databases used in this study are public and can be found at the following links: *GigaScience*: http://gigadb.org/dataset/100295, accessed on 10 March 2021.

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

References

- 1. Ladda, A.; Lebon, F.; Lotze, M. Using motor imagery practice for improving motor performance—A review. *Brain Cogn.* **2021**, 150, 105705. [CrossRef] [PubMed]
- James, C.; Zuber, S.; Dupuis Lozeron, E.; Abdili, L.; Gervaise, D.; Kliegel, M. How Musicality, Cognition and Sensorimotor Skills Relate in Musically Untrained Children. Swiss J. Psychol. 2020, 79, 101–112. [CrossRef]
- 3. Basso, J.; Satyal, M.; Rugh, R. Dance on the Brain: Enhancing Intra- and Inter-Brain Synchrony. *Front. Hum. Neurosci.* **2021**, 14, 586. [CrossRef]
- Suggate, S.; Martzog, P. Screen-time influences children's mental imagery performance. Dev. Sci. 2020, 23, e12978. [CrossRef] [PubMed]
- Bahmani, M.; Babak, M.; Land, W.; Howard, J.; Diekfuss, J.; Abdollahipour, R. Children's motor imagery modality dominance modulates the role of attentional focus in motor skill learning. *Hum. Movem. Sci.* 2020, 75, 102742. [CrossRef]
- Souto, D.; Cruz, T.; Fontes, P.; Batista, R.; Haase, V. Motor Imagery Development in Children: Changes in Speed and Accuracy With Increasing Age. *Front. Pediatr.* 2020, *8*, 100. [CrossRef]

- 7. Simpson, T.; Ellison, P.; Carnegie, E.; Marchant, D. A systematic review of motivational and attentional variables on children's fundamental movement skill development: The OPTIMAL theory. *Int. Rev. Sport Exer. Psychol.* **2020**, 1–47. [CrossRef]
- 8. Singh, A.; Hussain, A.; Lal, S.; Guesgen, H. A Comprehensive Review on Critical Issues and Possible Solutions of Motor Imagery Based Electroencephalography brain–computer Interface. *Sensors* **2021**, *21*, 2173. [CrossRef] [PubMed]
- 9. Al-Saegh, A.; Dawwd, S.; Abdul-Jabbar, J. Deep learning for motor imagery EEG-based classification: A review. *Biomed. Signal Process. Control* 2021, 63, 102172. [CrossRef]
- 10. Thompson, M. Critiquing the Concept of BCI Illiteracy. Sci. Eng. Ethics 2019, 25, 1217–1233. [CrossRef] [PubMed]
- 11. McAvinue, L.; Robertson, I. Measuring motor imagery ability: A review. Eur. J. Cogn. Psychol. 2008, 20, 232–251. [CrossRef]
- Yoon, J.; Lee, M. Effective Correlates of Motor Imagery Performance based on Default Mode Network in Resting-State. In Proceedings of the 2020 8th International Winter Conference on brain–computer Interface (BCI), Gangwon, Korea, 26–28 February 2020; pp. 1–5.
- 13. Rimbert, S.; Gayraud, N.; Bougrain, L.; Clerc, M.; Fleck, S. Can a Subjective Questionnaire Be Used as brain–computer Interface Performance Predictor? *Front. Hum. Neurosci.* **2019**, *12*, 529. [CrossRef] [PubMed]
- 14. Vasilyev, A.; Liburkina, S.; Yakovlev, L.; Perepelkina, O.; Kaplan, A. Assessing motor imagery in brain–computer interface training: Psychological and neurophysiological correlates. *Neuropsychologia* **2017**, *97*, 56–65. [CrossRef]
- Seo, S.; Lee, M.; Williamson, J.; Lee, S. Changes in Fatigue and EEG Amplitude during a Longtime Use of brain-computer Interface. In Proceedings of the 2019 7th International Winter Conference on brain-computer Interface (BCI), Gangwon, Korea, 18–20 February 2019; pp. 1–3.
- Lioi, G.; Cury, C.; Perronnet, L.; Mano, M.; Bannier, E.; Lécuyer, A.; Barillot, C. Simultaneous MRI-EEG during a motor imagery neurofeedback task: An open access brain imaging dataset for multi-modal data integration. *bioRxiv* 2019, 2019, 862375.
- 17. Collet, C.; Hajj, M.E.; Chaker, R.; Bui-xuan, B.; Lehot, J.; Hoyek, N. Effect of motor imagery and actual practice on learning professional medical skills. *BMC Med. Educ.* **2021**, *21*, 59. [CrossRef]
- Dähne, S.; Bießmann, F.; Samek, W.; Haufe, S.; Goltz, D.; Gundlach, C.; Villringer, A.; Fazli, S.; Müller, K. Multivariate Machine Learning Methods for Fusing Multimodal Functional Neuroimaging Data. *Proc. IEEE* 2015, 103, 1507–1530. [CrossRef]
- 19. Dai, C.; Wang, Z.; Wei, L.; Chen, G.; Chen, B.; Zuo, F.; Li, Y. Combining early post-resuscitation EEG and HRV features improves the prognostic performance in cardiac arrest model of rats. *Am. J. Emerg. Med.* **2018**, *36*, 2242–2248. [CrossRef] [PubMed]
- 20. Xu, J.; Zheng, H.; Wang, J.; Li, D.; Fang, X. Recognition of EEG Signal Motor Imagery Intention Based on Deep Multi-View Feature Learning. *Sensors* 2020, 20, 3496. [CrossRef]
- 21. Zhuang, M.; Wu, Q.; Wan, F.; Hu, Y. State-of-the-art non-invasive brain–computer interface for neural rehabilitation: A review. J. Neurorestoratol. 2020, 8, 4. [CrossRef]
- 22. Wu, D.; Jiang, X.; Peng, R.; Kong, W.; Huang, J.; Zeng, Z. Transfer Learning for Motor Imagery Based brain–computer Interfaces: A Complete Pipeline. *arXiv* 2021, arXiv:eess.SP/2007.03746.
- 23. Zheng, M.; Yang, B.; Xie, Y. EEG classification across sessions and across subjects through transfer learning in motor imagery-based brain-machine interface system. *Med. Biol. Eng. Comput.* **2020**, *58*, 1515–1528. [CrossRef]
- 24. Wan, Z.; Yang, R.; Huang, M.; Zeng, N.; Liu, X. A review on transfer learning in EEG signal analysis. *Neurocomputing* **2021**, 421, 1–14. [CrossRef]
- Zhang, R.; Zong, Q.; Dou, L.; Zhao, X.; Tang, Y.; Li, Z. Hybrid deep neural network using transfer learning for EEG motor imagery decoding. *Biomed. Sig. Process. Control* 2021, 63, 102144. [CrossRef]
- Kant, P.; Laskar, S.; Hazarika, J.; Mahamune, R. CWT Based Transfer Learning for Motor Imagery Classification for Brain computer Interfaces. J. Neurosci. Methods 2020, 345, 108886. [CrossRef] [PubMed]
- 27. Zhang, K.; Robinson, N.; Lee, S.; Guan, C. Adaptive transfer learning for EEG motor imagery classification with deep Convolutional Neural Network. *Neural Netw.* **2021**, *136*, 1–10. [CrossRef] [PubMed]
- 28. Wei, X.; Ortega, P.; Faisal, A. Inter-subject Deep Transfer Learning for Motor Imagery EEG Decoding. *arXiv* 2021, arXiv:cs.LG/2103.0535.
- 29. Zheng, M.; Yang, B.; Gao, S.; Meng, X. Spatio-time-frequency joint sparse optimization with transfer learning in motor imagerybased brain-computer interface system. *Biomed. Signal Process. Control* **2021**, *68*, 102702. [CrossRef]
- 30. Zhang, K.; Xu, G.; Chen, L.; Tian, P.; Han, C.; Zhang, S.; Duan, N. Instance transfer subject-dependent strategy for motor imagery signal classification using deep convolutional neural networks. *Comput. Math. Methods Med.* 2020, 2020. [CrossRef] [PubMed]
- 31. Luo, W.; Zhang, J.; Feng, P.; Yu, D.; Wu, Z. A concise peephole model based transfer learning method for small sample temporal feature-based data-driven quality analysis. *Knowl. Based Syst.* **2020**, *195*, 105665. [CrossRef]
- 32. Zhang, K.; Xu, G.; Zheng, X.; Li, H.; Zhang, S.; Yu, Y.; Liang, R. Application of Transfer Learning in EEG Decoding Based on brain–computer Interfaces: A Review. *Sensors* 2020, 20, 6321. [CrossRef]
- Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In Artificial Neural Networks and Machine Learning—ICANN 2018; Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 270–279.
- 34. Zhao, D.; Tang, F.; Si, B.; Feng, X. Learning joint space-time-frequency features for EEG decoding on small labeled data. *Neural Netw.* **2019**, *114*, 67–77. [CrossRef] [PubMed]

- Parvan, M.; Ghiasi, A.R.; Rezaii, T.; Farzamnia, A. Transfer Learning based Motor Imagery Classification using Convolutional Neural Networks. In Proceedings of the 2019 27th Iranian Conference on Electrical Engineering (ICEE), Yazd, Iran, 30 April–2 May 2019; pp. 1825–1828.
- Collazos-Huertas, D.; Álvarez-Meza, A.; Acosta-Medina, C.; Castaño-Duque, G.; Castellanos-Dominguez, G. CNN-based framework using spatial dropping for enhanced interpretation of neural activity in motor imagery classification. *Brain Inf.* 2020, 7, 8. [CrossRef] [PubMed]
- Collazos-Huertas, D.; Álvarez-Meza, A.; Castellanos-Dominguez, G. Spatial interpretability of time-frequency relevance optimized in motor imagery discrimination using Deep and Wide networks. *Biomed. Sig. Process. Control* 2021, 68, 102626. [CrossRef]
- 38. Mammone, N.; Ieracitano, C.; Morabito, F. A deep CNN approach to decode motor preparation of upper limbs from time–frequency maps of EEG signals at source level. *Neural Netw.* **2020**, *124*, 357–372. [CrossRef]
- Song, L.; Fukumizu, K.; Gretton, A. kernel-embeddings of Conditional Distributions: A Unified Kernel Framework for Nonparametric Inference in Graphical Models. *IEEE Signal Process. Mag.* 2013, 30, 98–111. [CrossRef]
- 40. Alvarez-Meza, A.M.; Orozco-Gutierrez, A.; Castellanos-Dominguez, G. Kernel-Based Relevance Analysis with Enhanced Interpretability for Detection of Brain Activity Patterns. *Front. Neurosci.* **2017**, *11*, 550. [CrossRef]
- 41. You, Y.; Chen, W.; Zhang, T. Motor imagery EEG classification based on flexible analytic wavelet transform. *Biomed. Signal Process. Control* **2020**, *62*, 102069. [CrossRef]
- 42. Collazos-Huertas, D.; Caicedo-Acosta, J.; Castaño-Duque, G.A.; Acosta-Medina, C.D. Enhanced Multiple Instance Representation Using Time-Frequency Atoms in Motor Imagery Classification. *Front. Neurosci.* **2020**, *14*, 155. [CrossRef]
- Velasquez, L.; Caicedo, J.; Castellanos-Dominguez, G. Entropy-Based Estimation of Event-Related De/Synchronization in Motor Imagery Using Vector-Quantized Patterns. *Entropy* 2020, 22, 703. [CrossRef] [PubMed]
- 44. McFarland, D.; Miner, L.; Vaughan, T.; Wolpaw, J. Mu and Beta Rhythm Topographies During Motor Imagery and Actual Movements. *Brain Topogr.* **2004**, *12*, 177–186. [CrossRef]
- 45. Sannelli, C.; Vidaurre, C.; Muller, K.; 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. [CrossRef]
- Ko, W.; Jeon, E.; Jeong, S.; Suk, H. Multi-Scale Neural Network for EEG Representation Learning in BCI. 2020. Available online: http://xxx.lanl.gov/abs/2003.02657 (accessed on 9 July 2021).
- 47. Cho, H.; Ahn, M.; Ahn, S.; Kwon, M.; Jun, S. EEG datasets for motor imagery brain–computer interface. *GigaScience* 2017, *6*, gix034. [CrossRef] [PubMed]
- 48. Kumar, S.; Sharma, A.; Tsunoda, T. Brain wave classification using long short-term memory network based OPTICAL predictor. *Sci. Rep.* **2019**, *9*, 9153. [CrossRef] [PubMed]
- Zhang, Y.; Zhou, G.; Jin, J.; Wang, X.; Cichocki, A. Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface. J. Neurosci. Methods 2015, 255, 85–91. [CrossRef] [PubMed]
- Zhao, X.; Zhao, J.; Liu, C.; Cai, W. Deep Neural Network with Joint Distribution Matching for Cross-Subject Motor Imagery brain–computer Interfaces. *BioMed Res. Int.* 2020, 2020. [CrossRef]
- Jeon, E.; Ko, W.; Yoon, J.; Suk, H. Mutual Information-Driven Subject-Invariant and Class-Relevant Deep Representation Learning in BCI. 2020. Available online: http://xxx.lanl.gov/abs/1910.07747 (accessed on 9 July 2021).
- 52. Mirzaei, S.; Ghasemi, P. EEG motor imagery classification using dynamic connectivity patterns and convolutional autoencoder. *Biomed. Signal Process. Control* **2021**, *68*, 102584. [CrossRef]
- 53. Freer, D.; Yang, G. Data augmentation for self-paced motor imagery classification with C-LSTM. *J. Neural Eng.* **2020**, *17*, 016041. [CrossRef]
- 54. Lee, M.; Yoon, J.; Lee, S. Predicting Motor Imagery Performance From Resting-State EEG Using Dynamic Causal Modeling. *Front. Human Neurosci.* **2020**, *14*, 321. [CrossRef]
- 55. Cardona, L.; Vargas-Cardona, H.; Navarro, P.; Cardenas Peña, D.; Orozco Gutiérrez, A. Classification of Categorical Data Based on the Chi-Square Dissimilarity and t-SNE. *Computation* **2020**, *8*, 104. [CrossRef]
- 56. Anowar, F.; Sadaoui, S.; Selim, B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Comput. Sci. Rev.* **2021**, *40*, 100378. [CrossRef]
- 57. Lee, M.; Kwon, O.; Kim, Y.; Kim, H.; Lee, Y.; Williamson, J.; Fazli, S.; Lee, S. EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience* **2019**, *8*, giz002. [CrossRef] [PubMed]