Article

A Discriminative Multi-Output Gaussian Processes Scheme for Brain Electrical Activity Analysis

Cristian Torres-Valencia [1,*,†] , **Álvaro Orozco** [1,†], **David Cárdenas-Peña** [1,†],
Andrés Álvarez-Meza [2,†] **and Mauricio Álvarez** [3,†]

[1] Automatics Research Group, Engineering Faculty, Universidad Tecnológica de Pereira, Pereira 660003, Colombia; aaog@utp.edu.co (Á.O.); dcardenasp@utp.edu.co (D.C.-P.)

[2] Signal Processing and Recognition Group, Universidad Nacional de Colombia, sede Manizales, Manizales 170001, Colombia; amalvarezme@unal.edu.co

[3] Computer Science Department, University of Sheffield, Sheffield S102TN, UK; mauricio.alvarez@sheffield.ac.uk

* Correspondence: cristian.torres@utp.edu.co; Tel.: +57-6-313-7300 (ext. 7680)

† These authors contributed equally to this work.

check for updates

Abstract: The study of brain electrical activity (BEA) from different cognitive conditions has attracted a lot of interest in the last decade due to the high number of possible applications that could be generated from it. In this work, a discriminative framework for BEA via electroencephalography (EEG) is proposed based on multi-output Gaussian Processes (MOGPs) with a specialized spectral kernel. First, a signal segmentation stage is executed, and the channels from the EEG are used as the model outputs. Then, a novel covariance function within the MOGP known as the multispectral mixture kernel (MOSM) allows us to find and quantify the relationships between different channels. Several MOGPs are trained from different conditions grouped in bi-class problems, and the discrimination is performed based on the likelihood score of the test signals against all the models. Finally, the mean likelihood is computed to predict the correspondence of new inputs with each class's existing models. Results show that this framework allows us to model the EEG signals adequately using generative models and allows analyzing the relationships between channels of the EEG for a particular condition. At the same time, the set of trained MOGPs is well suited to discriminate new input data.

Keywords: brain electrical activity; spectral mixture kernel; gaussian processes; electroencephalography

---

1. Introduction

From the neuroscience perspective, each physiological or cognitive process will produce a particular pattern of electrical interactions linking neurons from different brain regions. Therefore, the electrical response related to the interactions between neuron assemblies allows studying the brain function. Under such a perspective, the neuroscientists have found brain electrical activity (BEA) patterns associated with motor functions, cognitive processes, and neuropathologies [1]. However, the wide range of possible mental conditions hampers the BEA analysis and poses very challenging tasks [2]. In particular, capturing BEA by placing a set of electrodes over the scalp, known as electroencephalography (EEG), gathers and amplifies currents reflected in the brain cortex from all the possible brain sources, yielding a mixture of latent activity sources at each channel. For discovering such a latent activity, the literature considers different types of EEG analyses ranging from time and spectral domain processing [3,4], through connectivity measures between channels [5], to the complex network analysis [6].

Due to the non-stationarity behavior of EEG data, the classical temporal analyses cannot decode differences among several mental states or conditions. Besides, several studies demonstrate that

the variations in the EEG oscillatory patterns play a fundamental role in the maintenance of brain functions and the identification of different neural conditions [7]. Hence, spectral decomposition approaches look for relevant information at the brain rhythms (namely alpha, beta, theta, delta, and gamma). For instance, performing working memory or creative tasks evokes discriminative oscillatory patterns [8,9]. In addition, neuropathologies, such as Alzheimer's disease, cause abnormal cortical neural synchronization at resting-state rhythms [10]. The general spectral analysis procedure consists of a channel-wise frequency band splitting, followed by an identification of relevant time intervals, usually supervised by a specialist. The subsequent stages compute descriptors from the time-frequency representation. The power spectral density (PSD) and spectral entropy are among the most considered descriptors due to their straightforward interpretation [11,12]. Although EEG frequency analysis has proven to extract useful information for brain function understanding, the channel-wise approach still lacks the interpretability of several cognitive processes because each channel only holds a reflected version of BEA from a neural assembly [13]. Besides, the low spatial resolution of EEG restricts the extracted information about the behavior of some brain regions involved in high-level cognitive or physiological processes [4].

In recent years, connectivity analysis techniques account for channel interdependencies to enhance EEG spatial resolution through functional relationships captured in the BEA. Metrics from several domains attempt to quantify different properties of the BEA. Specifically, the coherence (COH), phase value index (PVI), and phase-locking value (PLV) capture pair-wise channel dependencies in the frequency domain. Moreover, the latter two gained attention due to their non-linear capability for unraveling latent connectivity patterns, which have proven valuable for applications, such as motor imagery and emotion recognition [14,15]. Nevertheless, the selection of the connectivity measure is not entirely straightforward. Their dependence on an estimated cross-spectrum and subject data variability yield to low generalization capability in a wide range of scenarios [16,17].

The above issues demand reliable brain connectivity approaches to automatically identify the relevant spectral information from the inherent EEG uncertainty. In this regard, a probabilistic modeling framework, such as Gaussian Processes (GPs), along with an appropriate covariance function possesses the capability for characterizing the latent processes from BEA data [18,19]. Moreover, the extension to vector-valued processes or multi-output GPs (MOGPs) adjusts the probabilistic model to map inputs into a multidimensional output space as an EEG channel array [20,21]. Former GP applications to EEG analysis in stress detection and cognitive stimulus recognition demonstrated the potential of GPs for BEA data modeling [22,23]. Furthermore, a recently proposed covariance function, known as the multi-output spectral mixture (MOSM) analyzes dependencies with spectral information for multidimensional output processes [24]. The proposal novelty relies on the PSD design, ensuring the frequency constraints for real-valued signals. Additionally, the inverse Fourier transform of the MOSM covariance results in a temporal kernel with positive definiteness conditions that are harder to accomplish during kernel design.

In this work, we propose the extension of MOGPs with an MOSM kernel for BEA discrimination, termed MOSM-GP. To this end, we learn an MOSM-GP for each EEG trial in a training set to model and quantify channel relationship patterns associated with particular EEG conditions. Then, we implement a likelihood measure to label new trial data into a specific class. The proposed framework of discriminative MOSM-GP, termed DMOSM-GP, is tested in two publicly available EEG datasets acquired under emotional [25] and motor imagery [26] conditions. The attained results show that the likelihood measure of testing data on the trained MOSM-GPs corresponds to a reliable discriminative index.

The manuscript organization is as follows: Section 2.2 describes the theoretical background of MOSM-GPs and introduces the proposed framework. Section 3 presents the attained results on both EEG modeling and classification performance. Finally, Section 4 concludes the works with the most relevant findings.

2. Material and Methods

2.1. EEG Databases

This work considers two publicly available EEG datasets for testing the proposed DMOSM-GP methodology. The two datasets are widely used in the development of neuroscience techniques as they hold challenging cognitive conditions. Contained EEG data allows testing the quantification of channel dependencies at the frequency level by the DMOSM-GP for binary classification experiments.

DEAP dataset "A Database for Emotional Analysis using Physiological data" (DEAP) contains EEG data from 32 subjects acquired under 40 emotion elicitation experiments, with one-minute recordings at 128 Hz and 32 channels distributed over the scalp [25]. Each participant rated the emotional stimulus at the end of a video, following two dimensions: arousal and valence. Other scores, such as dominance and liking, were also reported, although arousal and valence are the most prominent dimensions for affective computing works. These dimensions characterize a more extensive range of emotions than just the classical categorical description in six basic emotions. For our experiments, we consider the classification valence dimension as either low (ranging from one to five) or high (from five to nine) valence.

MI dataset The "Brain Computer Interface (BCI) competition 2008-Graz data set A" contains motor imagery experiments from nine subjects performing four specific tasks involving movements of hands and feet [26] under the motor imagery (MI) paradigm. The set of 22 EEG channels was band-pass filtered between 0.5 Hz and 100 Hz, and further down-sampled at 128 Hz. Each subject performed two experiment sessions, consisting of six runs and 48 six second-long trials per run and task. From the BCI dataset, we select the left- and right-hand movement imagination tasks for evaluating the proposed discriminative framework in a subject-wise scheme.

2.2. Cross-Spectral Estimation from Kernel Mixtures

The communication between neuron cells is the basis of every neuronal processing task. The electrical impulses result in every possible cognitive or physiological condition, such as behavior, sensation, thoughts, and emotions [27]. Due to equally measuring normal and abnormal BEA, EEG is considered a well-suited neuroimaging technique for diagnosis, treatment, and clinical procedures across several neurological pathologies [1]. Equation (1) presents the mathematical representation of the BEA from an EEG of C channels holding T time instants [28].

$$\chi = \{t \in \mathbb{R}^T, X \in \mathbb{R}^{C \times T}\}, \tag{1}$$

with t as the time sample positions of the recordings, and $X = \{x_i\}_{i=1}^C$ holding the brain electrical responses measured by the EEG array at channel i. To quantify the spectral content between channels, the introduced cross-spectrum estimation relies on specific covariance functions. The Cramer's theorem states that a family of integrable functions $\{\kappa_{ij}(\tau)\}_{i,j=1}^C$ are the covariance functions of a weakly-stationary stochastic process if and only if they admit the following representation:

$$\kappa_{ij}(\tau) = \int_{\mathbb{R}^n} e^{\iota \omega^\top \tau} S_{ij}(\omega) d\omega, \tag{2}$$

being ι the imaginary unit, each $S_{ij}(\omega)$ an integrate complex-valued function $S_{ij} : \mathbb{R} \to \mathbb{C}$ that is also positive definite, and i, j the indices of two EEG channels. This relationship between covariance functions κ_{ij} in the time domain with argument $\tau \in \mathbb{R}$ and their corresponding spectral density S_{ij} with arguments ω in the Fourier domain allows designing a desired spectral density and obtaining a covariance function [24].

Now, a family $S = \{S_{ij}\}_{i,j=1}^C \in \mathbb{R}^{C \times C}$ of positive-definite complex-valued functions can be used as cross-spectral densities for multi-output data [24]. These functions are designed by including specific parameters that allow physical interpretation of the obtained covariance kernel regarding the input

data. Moreover, complex-valued and positive-definite matrices can be decomposed in the form $\mathbf{S}(\omega) = \mathbf{R}^H(\omega)\mathbf{R}(\omega)$ where $\mathbf{R}(\omega) \in \mathbb{R}^{Q \times C}$, Q represents the rank of decomposition, and $(\cdot)^H$ denotes the Hermitian operator. Since Fourier transforms and multiplications of squared exponential (SE) functions are also SE, the autocovariance function $R_i(\omega)$ of i-th channel is modeled as the complex-valued SE in Equation (3).

$$R_i(\omega) = w_i \exp\left(-\frac{1}{4}\frac{(\omega - \mu_i)^2}{\sigma_i^2}\right) \exp\left(-\iota(\theta_i \omega + \phi_i)\right), \tag{3}$$

with $w_i, \phi_i, \mu_i, \theta_i \in \mathbb{R}$ and $\sigma_i \in \mathbb{R}^+$. With such a choice of functions, the cross-spectral density between channels i and j is given by Equation (4) with covariance $\sigma_{ij} \in \mathbb{R}^+$, mean $\mu_{ij} \in \mathbb{R}$, magnitude $w_{ij} \in \mathbb{R}$, delay $\theta_{ij} \in \mathbb{R}$, and phase $\phi_{ij} \in \mathbb{R}$.

$$S_{ij}(\omega) = w_{ij} \exp\left(-\frac{1}{2}\frac{(\omega - \mu_{ij})^2}{\sigma_{ij}} + \iota(\theta_{ij}\omega + \phi_{ij})\right). \tag{4}$$

Finally, in order to guarantee that the model is restricted to real-valued stochastic processes, the spectral density is reassigned to become symmetric with respect to ω by $S_{ij}(\omega) \to \frac{1}{2}(S_{ij}(\omega) + S_{ij}(-\omega))$. Then, the inverse Fourier transform of the resulting cross-spectral density becomes the corresponding temporal domain real-valued kernel; the kernel and the symmetric version of the spectral density are presented in Equations (5) and (6), respectively.

$$\kappa_{ij}(\tau) = \alpha_{ij} \exp\left(-\frac{1}{2}(\tau + \theta_{ij})^2 \sigma_{ij}\right) \cos\left((\tau + \theta_{ij})\mu_{ij} + \phi_{ij}\right), \tag{5}$$

$$S_{ij}(\omega) = \frac{w_{ij}}{2} \exp\left(-\frac{1}{2}\frac{(\omega - \mu_{ij})^2}{\sigma_{ij}} + \iota(\theta_{ij}\omega + \phi_{ij})\right)$$
$$+ \frac{w_{ij}}{2} \exp\left(-\frac{1}{2}\frac{(\omega + \mu_{ij})^2}{\sigma_{ij}} + \iota(-\theta_{ij}\omega + \phi_{ij})\right), \tag{6}$$

where the term $\alpha_{ij} = w_{ij}\sqrt{2\pi}|\sigma_{ij}|^{1/2}$ absorbs the constant resulting from the inverse Fourier transform. Equation (5) allows computing the real-valued autocovariances ($i = j$) and cross-covariances ($i \neq j$) with negatively and positively correlated channels through the magnitude parameter $\alpha_{ij} \in \mathbb{R}$; delayed channels through the $\theta_{ij} \neq 0$ delay parameter, and channels out-of-phase through the $\phi_{ij} \neq 0$ phase parameter. Moreover, increasing the rank of decomposition Q corresponds to considering more components in the multiple-output spectral mixture (MOSM) kernel as shown in Equations (7) and (8).

$$\kappa_{ij}(\tau) = \sum_{q=1}^{Q} \alpha_{ij}^{(q)} exp\left(-\frac{1}{2}\frac{(\tau + \theta_{ij}^{(q)})^2}{\sigma_{ij}^{(q)}}\right) \cos\left((\tau + \theta_{ij}^{(q)})\mu_{ij}^{(q)} + \phi_{ij}^{(q)}\right), \tag{7}$$

$$S_{ij}(\omega) = \sum_{q=1}^{Q} \frac{w_{ij}^{(q)}}{2} \exp\left(-\frac{1}{2}\frac{(\omega - \mu_{ij}^{(q)})^2}{\sigma_{ij}^{(q)}} + \iota(\theta_{ij}^{(q)}\omega + \phi_{ij}^{(q)})\right)$$
$$+ \sum_{q=1}^{Q} \frac{w_{ij}^{(q)}}{2} \exp\left(-\frac{1}{2}\frac{(\omega + \mu_{ij}^{(q)})^2}{\sigma_{ij}^{(q)}} + \iota(-\theta_{ij}^{(q)}\omega + \phi_{ij}^{(q)})\right), \tag{8}$$

denoting the superindex (q) the q−th spectral component. Then, MOSM effectively computes autocovariance and cross-covariances through the spectral-mixture of positive-definite kernels from

the Fourier transform of spectral functions $S_{ij}(\omega)$. In practice, the adjustment of the cross-spectrum parameters should be performed on the evidence of the EEG data.

2.3. Multi-Output Spectral Mixture Gaussian Process

Given an EEG trial, the Gaussian Process (GP) probabilistic framework computes the mixture parameters by maximizing the data likelihood as the cost function. A Gaussian Process (GP) is a real-valued stochastic process $(f(t))$ over a input set t, such that for any finite subset of inputs $t \in \{1, \ldots, T\}$, the random variables $f(t)$ are jointly Gaussian [20]. Additionally, the GP is uniquely determined by its mean function $m(t) := E_t(f(t))$, typically assumed $m(t) = 0$ and its covariance function $\kappa(t, t) := cov(f(t), f(t')) \in \mathbb{R}^{T \times T}$ known as the kernel.

Then, the multivariate extension of GPs is derived by assembling C different scalar-valued stochastic processes, one for each EEG channel. Any finite collection of values across all such processes are jointly Gaussian, termed multiple-output Gaussian Process (MOGP). This extension results in a vector-valued process $f \sim \mathcal{GP}(m, K)$, where $m(t) \in \mathbb{R}^{TC}$ is a concatenated vector from the mean vectors associated to the outputs and $K \in \mathbb{R}^{TC \times TC}$ a block partitioned matrix of the form [20]:

$$K(X, X) = \begin{bmatrix} K(X_1, X_1) & \cdots & K(X_1, X_C) \\ K(X_2, X_1) & \cdots & K(X_2, X_C) \\ \cdots & \cdots & \cdots \\ K(X_C, X_1) & \cdots & K(X_C, X_C) \end{bmatrix}, \tag{9}$$

where each block $K(X_i, X_j)$ is a $T \times T$ matrix denoting the covariance between output channels i, j. Furthermore, a multivariate kernel $K(t, t')$ is stationary if $K(t, t') = K(t - t')$. In this case, the kernel becomes $\kappa_{ij}(t, t') = \kappa_{ij}(\tau)$ if substituting $\tau = t - t'$. Therefore, we use the MOSM kernel in Equation (7) as the covariance function to be implemented within the MOGP. By defining such a process as the MOSM kernel, the model adjustment to the data is performed by maximizing the data log-probability. Since the observations in the multiouput case are jointly Gaussian, they are concatenated into the vector $y = [x_1^\top, x_2^\top \cdots, x_C^\top]^\top \in \mathbb{R}^{CT}$ the channel observed value. Then, the negative log-likelihood (NLL) can be expressed as in Equation (10).

$$-\log p(y|t, \Theta) = \frac{CT}{2} \log 2\pi + \frac{1}{2} \log |K| + \frac{1}{2} y^\top K^{-1} y, \tag{10}$$

with $\Theta = \{w_i^{(q)}, \mu_i^{(q)}, \sigma_i^{(q)}, \theta_i^{(q)}, \phi_i^{(q)}, \sigma_i^2\}_{i=1, q=1}^{C, Q}$ holding the complete set of parameters. As a result, minimization of NLL with respect to Θ designs an spectral kernel quantifying the EEG channel relationships at automatically tuned frequency bands.

2.4. Discriminative Scheme Using MOSM-GP

Let a set of N labeled BEA trials $\{\chi_n, l_n\}_{n=1}^N$, each of them belonging either class A or B, that is, $l_n \in \{A, B\}$. In the case of DEAP dataset, classes correspond to low and high valence, while, for the MI dataset, left- and right-hand movement imagination are considered. As stated in Section 2.3, a single MOSM-GP models each BEA trial as the stochastic process $f_n^{(l)}$, resulting in N_A and N_B MOSM-GPs for classes A and B, respectively. Furthermore, on the evidence of a new BEA trial X_*, the marginal likelihood for each learned MOSM-GP is computed as:

$$p(f_n^{(l)}(X_*)|X_n, f_n^{(l)}, X_*) = \mathcal{N}(f_n^{(l)}(X_*), K(X_*, X_*)). \tag{11}$$

By evaluating the marginal likelihoods on all training MOSM-GPs, the new BEA trial label is estimated as follows:

$$l_* = \arg\max_{l \in \{A, B\}} E\left\{p(f_n^{(l)}(X_*)|X, f_n^{(l)}, X_*) : l_n = l\right\}, \tag{12}$$

where $E\{\cdot : l_n = l\}$ denotes the expectation operator over training trials belonging to class l. Figure 1 illustrates the proposed discriminative MOSM-GP framework, termed DMOSM-GP.
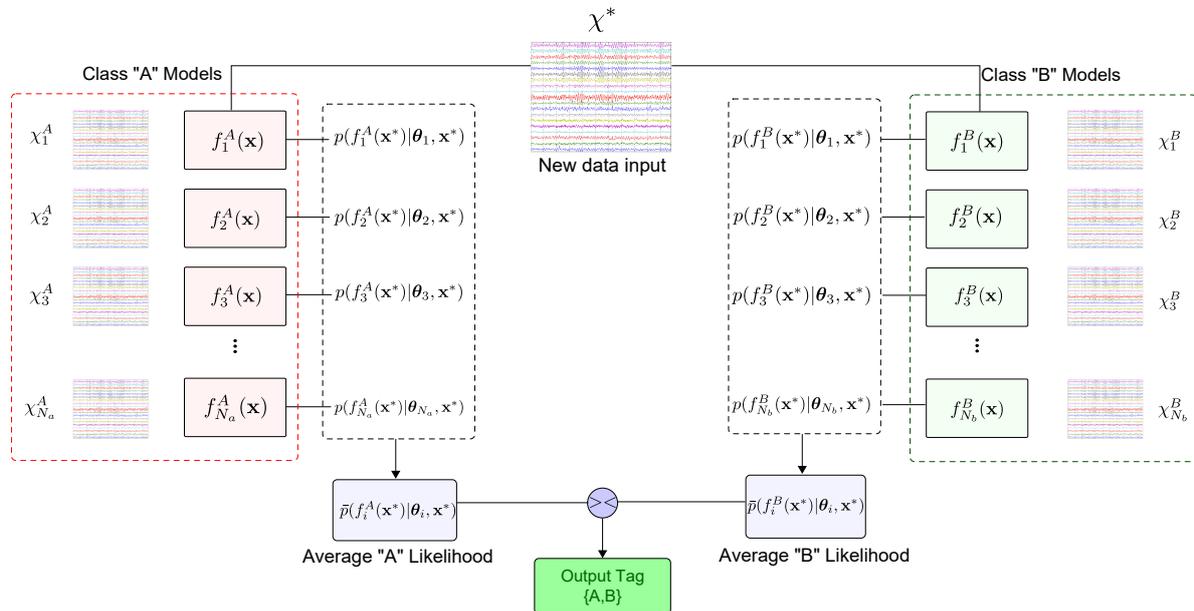

Figure 1. Proposed framework for discriminative multispectral mixture kernel (MOSM).

2.5. Implementation Details

Before the model training stage, a channel selection is carried out to reduce the training computational cost. For the MI dataset, channels are selected based on the evidence that body movement triggers neural activity in the opposite brain hemisphere within the sensorimotor area. Regarding the DEAP dataset, channels related to brain regions more likely to participate in affective states are considered. Figure 2 depicts the subset of selected channels for both EEG datasets. Regarding the DMOSM-GP free parameter, the rank of decomposition is chosen from a grid search within the range $Q \in \{1, \ldots, 10\}$ to minimize the mean absolute error (MAE) of the model prediction against the original EEG data. The GPflow framework is employed for the model definition [29], and the kernel function is optimized via the minimization of NLL cost function using the autograd library. Finally, for the statistical significance assessment of the classification performance, an 10-fold cross-validation scheme was applied.
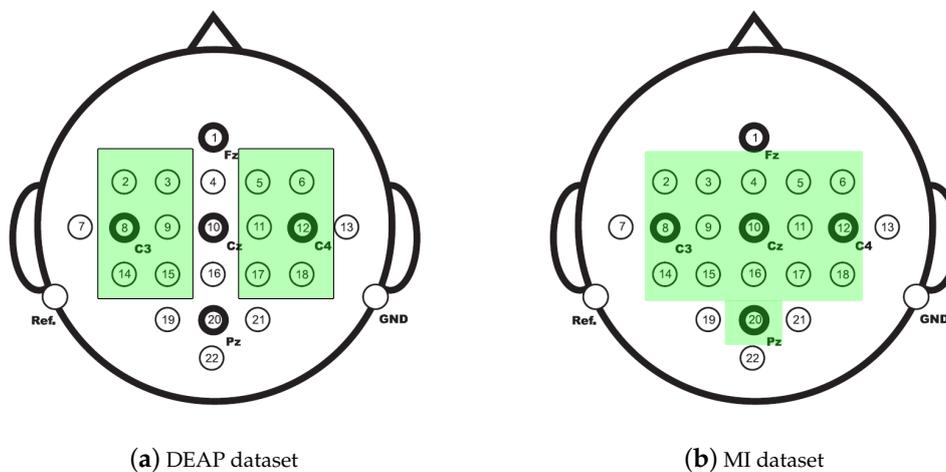

(**a**) DEAP dataset　　　　　　　　　　　　　　(**b**) MI dataset

Figure 2. Positioning scheme of electrodes for acquisition in both datasets. Green boxes highlight the selected channels for testing the proposed methodology.

3. Results

3.1. Parameter Tuning and Spectral Modeling

To tune the rank of decomposition Q, we evaluate the MOSM-GP performance for modeling the selected output channels from the data posterior distribution at specific temporal locations. Moreover, the mean absolute error (MAE) quantifies the difference between the target and predicted outputs as a function of the rank. Figure 3 presents the mean MAE across the GP outputs against the number of spectral components defined for the MOSM kernel. As a first insight, the MAE values evidence that the MOSM-GP effectively reconstruct EEG recordings. Nonetheless, the error increases over six spectral components, due to a large number of kernel parameters to be tuned, which in turn increases the computational complexity without providing relevant information to the probabilistic model. On the contrary, a single component lacks the complexity to account for the brain activity changes. Therefore, a rank of decomposition between three and five benefits the most the MAE performance, implying a balance between model complexity and generalization capability. Consequently, for the remaining of the work, we selected three spectral mixtures as the optimal Q for testing the MOSM-GP scheme. For the purpose of visualization, Figure 4 exemplifies the MOSM-GP output for channels FP1 and AF3 using $Q = 3$. As seen, the posterior MOSM distribution suitably models EEG data at all time locations with bounded deviations.

An analysis of the spectral information quantified by the MOSM-GP is carried out using Equation (6) that decomposes the spectral content shared by two channels into Q terms. Figure 5 plots the component-wise spectral distribution between channels FC2 and FC6 in a trial from left- (Figure 5a) and right-hand movement (Figure 5b). Attained spectra prove that each component automatically fits a particular frequency band. Moreover, the component magnitude $\alpha_{ij}^{(q)}$ highlights the dominant component from each trial as the most discriminative frequency band. As expected, such frequencies lay around 15 Hz to 35 Hz for MI experiments, being associated with an activation of alpha and beta rhythms.
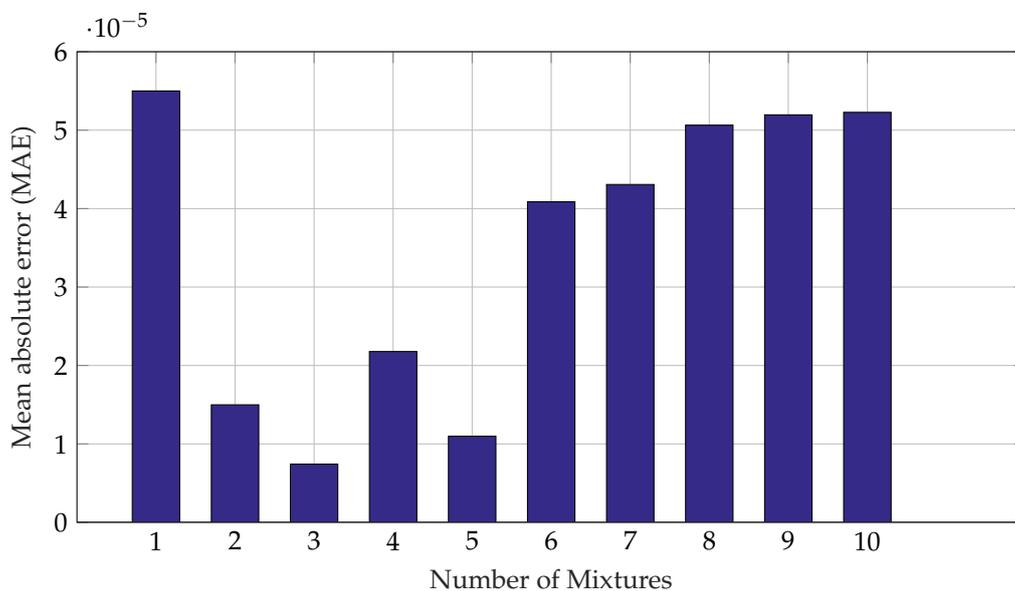

Figure 3. Mean absolute error (MAE) values for the data reconstruction of electroencephalography (EEG) channels after training a MOSM-Gaussian Process (GP) compared to the original recordings.

Later, Equation (8) is used to compute the cross-spectrum visualizing the computed covariances by the MOSM kernel for every channel pair. Since there is valuable information on the analysis of the cross-covariances between channel pairs, a complete trial visualization of the quantified spectral information is presented in Figure 6a,b for left- and right-hand movement trials, respectively. Each one

of the horizontal axis sections, corresponds to a particular channel i and its MOSM PSD against all the channels. The vertical axis represents the frequency bin at which the connectivity is assessed. Then, lighter green colors are related with strong spectral densities shared between i, j channels, while darker blue colors are related with lower interdependencies. Despite most of the strong spectral density being constrained to the $[15, 40]$ Hz band, there is clear evidence that not all the channels are synchronized at this frequency when performing MI activity. Particularly for Figure 6a, there are channels sections that seems to be uncorrelated in the complete spectrum for this particular task. For example, channels Cz, C2, and CP6 seem to have low frequency dependencies shared with the rest of the EEG array. On the other hand, channel CP2 presents strong connections with most of the channels. Furthermore, for the opposite MI task, there exist variations on the captured frequency relationships among the EEG array. C2 and C4 result as the most highly correlated the other channels when performing the MI task.
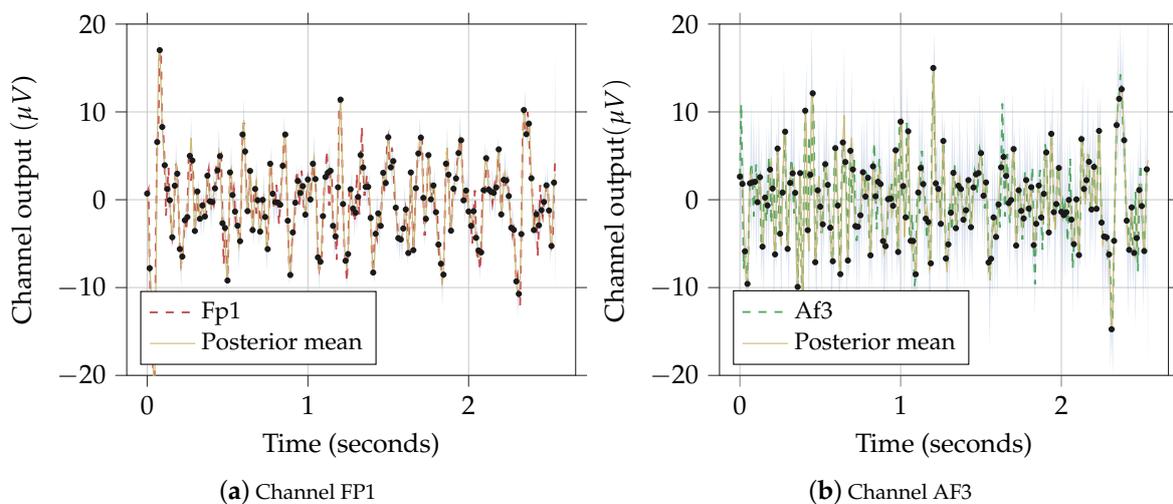

(**a**) Channel FP1
(**b**) Channel AF3

Figure 4. Channels FP1 and AF3 from EEG recordings adjusted by an MOSM with three spectral mixtures and its corresponding model posterior along time.
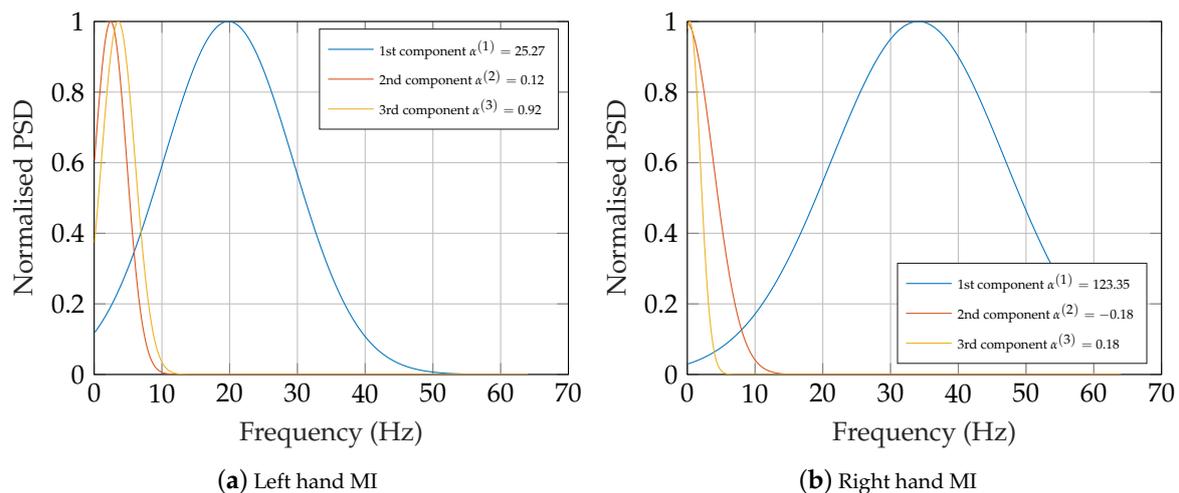

(**a**) Left hand MI
(**b**) Right hand MI

Figure 5. Normalized power spectral density $S_{ij}(\omega)$ for a given pair of channels (*FC2-FC6*) from MI database for two opposite conditions. Top row for left-hand movement condition and low row for right-hand movement condition. Figure 5a,b are the spectral content per spectral component.
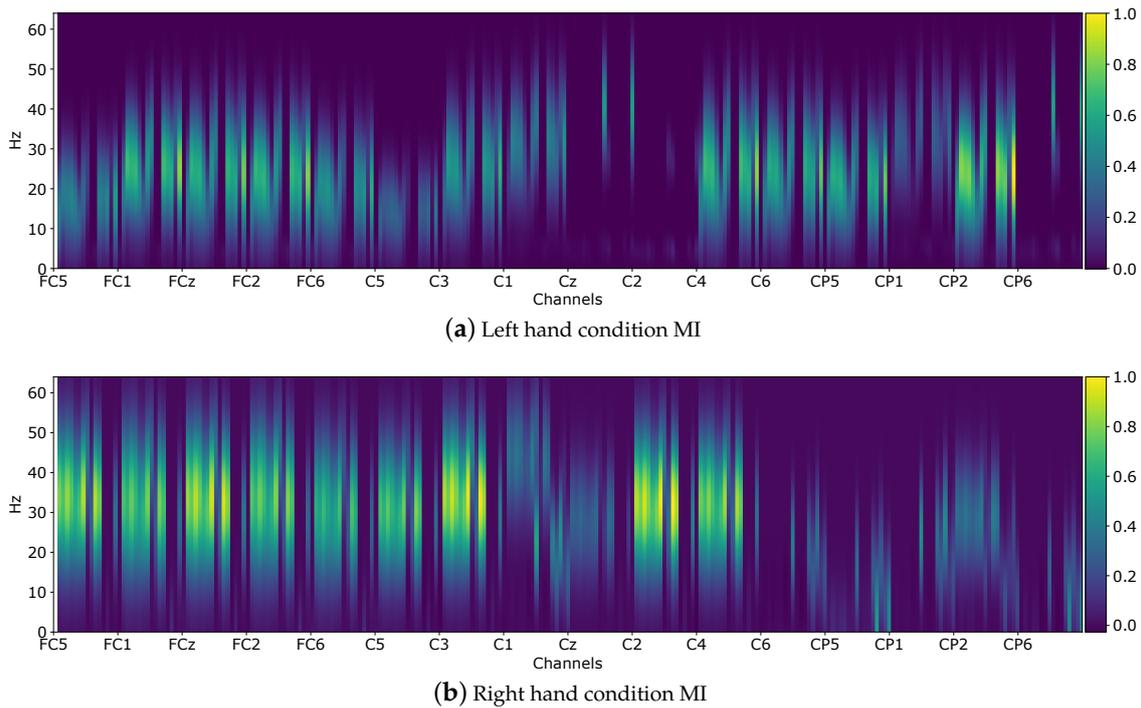
(**a**) Left hand condition MI

(**b**) Right hand condition MI

Figure 6. Quantified power spectral interdependencies for two particular conditions of the motor imagery dataset. The power spectral density (PSD) has been normalized between 0–1.

3.2. Discriminative MOSM-GP

For the DMOSM-GP framework, each EEG trial is trimmed into two-seconds lasting time series to find the desired spectral relationships. Table 1 presents the absolute value of the average likelihood as the class dependency measure, depicted in Equation (12). The first column corresponds to the test condition level regarding the emotional content (valence for this experiment), columns two and three are the values obtained for the model testing, and fourth is the valid tag of the corresponding test signal. Finally, the fifth column is the predicted tag from the magnitude of the mean test likelihood. Similarly, in Table 2, the results for the DMOSM-GP strategy over the BCI database are presented. The first column corresponds to the movement associated with the experiment. Columns two and three are the average likelihoods obtained by testing the new input against each class's models. The fourth column is associated with the resulting tag. Each row in Tables 1 and 2 corresponds to a particular trial, for emotional or motor imagery experiments.

From this test on complete datasets, it can be evidenced that the prediction regarding the likelihood of the trained model with the test signals is promising. The accuracy of the test data is around 73.3% for the DEAP database and about 87.33 for the BCI database. Specifically, for comparison purposes against works using the DEAP dataset, a selection of 9 subjects is performed. This selection is based on the evidence of an uncertainty test performed in [17], where the authors concluded that the subject itself did not adequately tag some experiments of the DEAP database. The scheme of implicit tagging used in this particular database allows the subjects to rate their affective result, so some of the acquired signals seem not correctly related to the emotional tests. The results reported in Table 3 show the accuracy resulting from the cross-validated DMOSM-GP scheme. The subject IDs reported are the same in the database, and some high accuracy (state-of-art comparable) ones were achieved around 78%, which is the case of subject 18.

On the other hand, for the BCI database, the complete set of 9 subjects is employed, and Table 4 reports the accuracy of classification within 5 folds of DMOSM-GP. Some higher accuracy was obtained for the MI database in the subject-dependent experiments around 87% for subject ID 2. In general terms, the results associated with the BCI database are higher than the DEAP database, and it can

be related directly with the condition that is tested. In the case of emotional experiments, there is a high degree of subjectivity in the elicitation of the states. In contrast, in the case of motor imagery, the conditions are somehow more consistently replicated.

Table 1. DMOSM-GP Test for brain electrical activity (BEA) analysis of emotional conditions of two classes. The mean absolute value of the likelihood from the test signals against the trained models is included—"A Database for Emotional Analysis using Physiological data" (DEAP) database. The boldface indicates the class with the highest log-likelihood.

Level	Low Models	High Models	True	Predicted
9.0	138.6293	**143.0221**	A	A
9.0	10.8743	**14.7149**	A	A
9.0	**103.9478**	100.7107	A	B
9.0	299.0507	**304.0222**	A	A
8.81	299.7143	**305.0394**	A	A
8.56	363.5983	**369.2760**	A	A
9.0	348.1999	**353.7941**	A	A
9.0	38.8484	**42.9119**	A	A
1.0	**193.7571**	191.0369	B	B
1.0	**145.3716**	142.2964	B	B
1.0	400.7168	**406.6640**	B	A
1.0	**36.0769**	32.4378	B	B
1.1	48.8835	**53.0558**	B	A
1.0	132.5064	**136.9884**	B	A
1.08	**83.4116**	80.2945	B	B
One-Fold Average Accuracy 73.35 $\pm$ 5.64%				

Table 2. DMOSM-GP Test for BEA analysis of emotional conditions of two classes. The mean absolute value of the likelihood from the test signals against the trained models is included—MI dataset. The boldface indicates the class with the highest log-likelihood.

Side	Left Models	Right Models	Predicted
L	**163.8711**	159.3530	L
L	**181.6294**	176.9928	L
L	29.7624	**35.1477**	R
L	**183.1083**	179.0200	L
L	**128.3101**	123.8268	L
L	**24.0204**	20.2636	L
L	**142.6790**	138.4443	L
L	**11.2064**	7.0390	L
R	59.3072	**63.8494**	R
R	**111.6419**	95.6175	L
R	191.7138	**196.6392**	R
R	193.1533	**196.7049**	R
R	35.7346	**39.8927**	R
R	194.2656	**198.3379**	R
R	54.5305	**59.0000**	R
One-Fold Average Accuracy 87.33 $\pm$ 3.68			

Table 3. Discriminative test for subjects from the DEAP database using the DMOSM-GP scheme, the test are developed within the valence emotional dimension.

Subject ID	Accuracy	Subject ID	Accuracy
17	0.65 ± 4.32	2	0.70 ± 6.19
18	0.78 ± 8.12	16	0.50 ± 5.28
22	0.60 ± 3.49	19	0.65 ± 5.66
29	0.65 ± 4.28	31	0.50 ± 6.77
30	0.70 ± 5.28	**Average**	**0.64 ± 0.09**

Table 4. Discriminative test for subjects from the MI database using the DMOSM-GP scheme.

Subject ID	Accuracy	Subject ID	Accuracy
1	0.82 ± 3.45	6	0.85 ± 3.18
2	0.87 ± 4.16	7	0.80 ± 3.73
3	0.85 ± 2.56	8	0.82 ± 4.21
4	0.77 ± 4.56	9	0.79 ± 2.66
5	0.82 ± 4.03	**Average**	**0.82 ± 0.03**

3.3. Results Comparison

The results obtained from the DMOSM-GP strategy are compared with some state-of-the-art results in terms of classification accuracy. However, since the proposed methodology uses generative models for discriminative purposes, it has to be stated that the comparison should be made on a few additional items than the classification accuracy. As can be seen in Tables 5 and 6, the average classification results obtained by the DMOSM-GP strategy are competitive among the state-of-art works. It has to be noticed that this work does not implement a preprocessing or feature extraction stage but uses the acquired data to train the MOSM-GPs and perform the discriminative task.

Table 5. Mean emotion classification results [%] for selected DEAP subjects, only mean value for comparison consistency against reported works in Reference [17].

Reference	Approach	Accuracy
Koelstra et al. [25]	PSD-SVM	57.50
Soleymani et al. [3]	PSD-SVM	62.00
Gupta et al. [30]	PSD-HJORT-SVM	60.00
Padilla-Buritica et al. [17]	MSP-SVM	55.76
Daimi and Saha [31]	Wavelet-SVM	65.00
Torres-Valencia et al. [28]	RFCV-KNN	65.73
Pan et.al. [11]	PSD-LORSAL	63.29
Pan et.al. [11]	DE-LORSAL	77.17
This work	DMOSM-GP	64.35

Table 6. Mean emotion classification results [%] for MI dataset, only mean value for comparison consistency against reported works.

Reference	Approach	Accuracy
Qureshi et al [32]	ICA-ELM	94.29
Li et al. [33]	CSP-SVM	78.78
Liang et al. [34]	PDC-MEMD	70.22
Elastuy and Eldawlatly [35]	DBN	73.44
Gómez et al. [36]	CSP-SVM	81.41
This work	DMOSM-GP	82.11

4. Discussion and Concluding Remarks

Brain information processing is a complex task that is not yet entirely mapped and understood. Despite the previous knowledge about brain regions interactions in motor or emotional means, works that allow improving the interpretability of results in different BEA scenarios would lead to the development of more precise frameworks for analysis, diagnosis, and treatment of mental pathologies, among other tasks. In this work, we proposed a framework for discriminate BEA using raw data with the support of a spectral kernel that identifies relationships between channels on behalf of a probabilistic methodology of multi-output Gaussian process. One of the essential remarks of this framework is the capability of learning EEG connectivity patterns by estimating raw data spectral components without a feature extraction stage. Further, this proposal of generative models working directly with EEG data has the advantage of adjusting the model relying on the optimization of kernel hyperparameters just from the channels information in a data-driven framework.

The results presented in Section 3 evidenced that introducing the MOSM kernel to MOGPs becomes a reliable tool for BEA modeling, due to the spectral designing properties. It is well known that EEG channels share complex frequency relationships that can be exploited using the design of a covariance function in that particular domain. Regarding this property, the posterior distribution over the data measured by the MAE in Figure 3 shows an adequate adjustment of the model on the original data. It also allows us to conclude that the inclusion of more spectral components into the covariance function benefits the model adjusting to the data.

Moreover, the identification of the spectral relationships between channels, performed by the MOSM kernel, allows gaining a better understanding of the latent functional connectivity between brain regions. As Figure 6 evidenced, the MOSM-GP identifies representative frequency bands for the cognitive process. The positive linear relationships are grouped among channels from the same hemisphere, with strong specific dependencies between channels, like $P3 - P7$, and $Fc6 - Af4$. The negative relationships are explained by lower PSD amplitudes at Cz, $C2$, and $CP6$ for the left-hand MI task, and $C6, CP1, CP6$ for the right-hand MI task. All these interactions quantified by the MOSM kernel in terms of higher or lower values of the PSD can be directly related with the activations of neural cells in different regions of the brain related with emotions (amygdala, hippocampus, and frontal cortex) or related with motor activities (primary motor cortex and posterior parietal cortex). Nevertheless, further studies must be completed from the evidence of these relationships to an accurate source reconstruction before determining the specific brain region of the acquired neural activity.

Finally, the discriminative results regarding the emotional and motor imagery conditions conclude that probabilistic models can be efficiently employed as a classification tool for EEG data. In this case, the probability distribution of tested data against the trained models directly becomes a classification algorithm by following a direct comparison of the mean likelihood value between the models from two classes. Despite lacking a feature extraction stage, the proposed DMOSM-GP produces discriminative information from the data. Moreover, the total accuracy of the subject-dependent and condition-dependent tests is comparable with state-of-art works, as Tables 5 and 6 illustrate.

One of the drawbacks of this framework is the computational complexity of training a considerable number of MOSM-GP models. In addition, increasing the number of MOSM kernel mixtures and the size of BEA data will derive into an exponential growth of the training time. Further improvements of this methodology will be directed towards using lighter versions of probabilistic models, such as the sparse GPs aiming at solving more difficult supervised learning tasks from EEG data as multilabel classification or regression.

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

Abbreviations

The following abbreviations are used in this manuscript:

BEA	Brain Electrical Activity
EEG	Electroencephalography
BCI	Brain-Computer Interface
GP	Gaussian Process
MOGP	Multi-output Gaussian Process
MI	Motor Imagery
MAE	Mean Absolute Error

References

1. Sanei, S. *Adaptive Processing of Brain Signals*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
2. Fadlallah, B.; Seth, S.; Keil, A.; Principe, J. Quantifying Cognitive State From EEG Using Dependence Measures. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2773–2781. [CrossRef]
3. Soleymani, M.; Pantic, M.; Pun, T. Multimodal Emotion Recognition in Response to Videos. *IEEE Trans. Affect. Comput.* **2012**, *3*, 211–223. [CrossRef]
4. Torres-Valencia, C.; Álvarez-López, M.; Orozco-Gutiérrez, Á. SVM-based feature selection methods for emotion recognition from multimodal data. *J. Multimodal User Interfaces* **2017**, *11*, 9–23. [CrossRef]
5. Horwitz, B. The elusive concept of brain connectivity. *NeuroImage* **2003**, *19*, 466–470. [CrossRef]
6. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **2010**, *52*, 1059–1069. [CrossRef] [PubMed]
7. Babiloni, C.; Barry, R.J.; Başar, E.; Blinowska, K.J.; Cichocki, A.; Drinkenburg, W.H.; Klimesch, W.; Knight, R.T.; da Silva, F.L.; Nunez, P.; et al. International Federation of Clinical Neurophysiology (IFCN)–EEG research workgroup: Recommendations on frequency and topographic analysis of resting state EEG rhythms. Part 1: Applications in clinical research studies. *Clin. Neurophysiol.* **2020**, *131*, 285–307. [CrossRef] [PubMed]
8. Rominger, C.; Papousek, I.; Perchtold, C.M.; Weber, B.; Weiss, E.M.; Fink, A. The creative brain in the figural domain: Distinct patterns of EEG alpha power during idea generation and idea elaboration. *Neuropsychologia* **2018**, *118*, 13–19. [CrossRef]
9. Akiyama, M.; Tero, A.; Kawasaki, M.; Nishiura, Y.; Yamaguchi, Y. Theta-alpha EEG phase distributions in the frontal area for dissociation of visual and auditory working memory. *Sci. Rep.* **2017**, *7*, 42776. [CrossRef] [PubMed]
10. Babiloni, C.; Del Percio, C.; Lizio, R.; Noce, G.; Lopez, S.; Soricelli, A.; Ferri, R.; Pascarelli, M.T.; Catania, V.; Nobili, F.; et al. Abnormalities of resting state cortical EEG rhythms in subjects with mild cognitive impairment due to Alzheimer's and Lewy body diseases. *J. Alzheimers Dis.* **2018**, *62*, 247–268. [CrossRef]
11. Pan, C.; Shi, C.; Mu, H.; Li, J.; Gao, X. EEG-Based Emotion Recognition Using Logistic Regression with Gaussian Kernel and Laplacian Prior and Investigation of Critical Frequency Bands. *Appl. Sci.* **2020**, *10*, 1619. [CrossRef]
12. Wang, Q.; Li, Y.; Liu, X. Analysis of feature fatigue EEG signals based on wavelet entropy. *Int. J. Pattern Recognit. Artif. Intell.* **2018**, *32*, 1854023. [CrossRef]
13. Friston, K.J. Functional and effective connectivity: A review. *Brain Connect.* **2011**, *1*, 13–36. [CrossRef] [PubMed]
14. Jian, W.; Chen, M.; McFarland, D.J. EEG based zero-phase phase-locking value (PLV) and effects of spatial filtering during actual movement. *Brain Res. Bull.* **2017**, *130*, 156–164. [CrossRef] [PubMed]

15. Wang, Z.; Tong, Y.; Heng, X. Phase-locking value based graph convolutional neural networks for emotion recognition. *IEEE Access* **2019**, *7*, 93711–93722. [CrossRef]

16. Bakhshali, M.A.; Ebrahimi-Moghadam, A.; Khademi, M.; Moghimi, S. Coherence-based correntropy spectral density: A novel coherence measure for functional connectivity of EEG signals. *Measurement* **2019**, *140*, 354–364. [CrossRef]

17. Padilla-Buritica, J.I.; Martinez-Vargas, J.D.; Castellanos-Dominguez, G. Emotion Discrimination Using Spatially Compact Regions of Interest Extracted from Imaging EEG Activity. *Front. Comp. Neurosci.* **2016**, *10*, 55. [CrossRef] [PubMed]

18. Aghaei, A.S.; Mahanta, M.S.; Plataniotis, K.N. Separable Common Spatio-Spectral Patterns for Motor Imagery BCI Systems. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 15–29. [CrossRef]

19. Shahidi Zandi, A.; Tafreshi, R.; Javidan, M.; Dumont, G.A. Predicting Epileptic Seizures in Scalp EEG Based on a Variational Bayesian Gaussian Mixture Model of Zero-Crossing Intervals. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 1401–1413. [CrossRef]

20. Alvarez, M.A.; Rosasco, L.; Lawrence, N.D. Kernels for vector-valued functions: A review. *Found. Trends Mach. Learn.* **2012**, *4*, 195–266. [CrossRef]

21. Gómez-González, S.; Álvarez, M.A.; García, H.F.; Ríos, J.I.; Orozco, A.A. Discriminative Training for Convolved Multiple-Output Gaussian Processes. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*; Pardo, A., Kittler, J., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 595–602.

22. Desai, R.; Porob, P.; Rebelo, P.; Edla, D.R.; Bablani, A. EEG Data Classification for Mental State Analysis Using Wavelet Packet Transform and Gaussian Process Classifier. *Wirel. Pers. Commun.* **2020**, 1–21. [CrossRef]

23. Remes, S.; Heinonen, M.; Kaski, S. Latent Correlation Gaussian Processes. *Stat* **2017**, *1050*, 27.

24. Parra, G.; Tobar, F. Spectral mixture kernels for multi-output Gaussian processes. *In Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 6681–6690.

25. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31. [CrossRef]

26. Brunner, C.; Leeb, R.; Müller-Putz, G.; Schlögl, A.; Pfurtscheller, G. BCI Competition 2008–Graz data set A. In *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces)*; Graz University of Technology: Graz, Austria, 2008; Volume 16.

27. Alotaiby, T.; El-Samie, F.E.A.; Alshebeili, S.A.; Ahmad, I. A review of channel selection algorithms for EEG signal processing. *EURASIP J. Adv. Signal Process.* **2015**, *2015*, 66. [CrossRef]

28. Torres-Valencia, C.; Alvarez-Meza, A.; Orozco-Gutierrez, A. Emotion Assessment Based on Functional Connectivity Variability and Relevance Analysis. In Proceedings of the Natural and Artificial Computation for Biomedicine and Neuroscience: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2017, Corunna, Spain, 19–23 June 2017; Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Toledo Moreo, J., Adeli, H., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 353–362.

29. De G. Matthews, A.G.; van der Wilk, M.; Nickson, T.; Fujii, K.; Boukouvalas, A.; Le'on-Villagr'a, P.; Ghahramani, Z.; Hensman, J. GPflow: A Gaussian process library using TensorFlow. *J. Mach. Learn. Res.* **2017**, *18*, 1–6.

30. Gupta, R.; ur Rehman Laghari, K.; Falk, T.H. Relevance Vector Classifier Decision Fusion and EEG Graph-theoretic Features for Automatic Affective State Characterization. *Neurocomputing* **2016**, *174*, 875–884. [CrossRef]

31. Daimi, S.N.; Saha, G. Classification of emotions induced by music videos and correlation with participants' rating. *Expert Syst. Appl.* **2014**, *41*, 6057–6065. [CrossRef]

32. Qureshi, M.N.I.; Cho, D.; Lee, B. EEG classification for motor imagery BCI using phase-only features extracted by independent component analysis. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, Korea, 11–15 July 2017; pp. 2097–2100. [CrossRef]

33. Li, D.; Zhang, H.; Khan, M.S.; Mi, F. A self-adaptive frequency selection common spatial pattern and least squares twin support vector machine for motor imagery electroencephalography recognition. *Biomed. Signal Process. Control* **2018**, *41*, 222–232. [CrossRef]

34. Liang, S.; Choi, K.S.; Qin, J.; Wang, Q.; Pang, W.M.; Heng, P.A. Discrimination of motor imagery tasks via information flow pattern of brain connectivity. *Technol. Health Care* **2016**, *24*, S795–S801. [CrossRef]

35. Elasuty, B.; Eldawlatly, S. Dynamic Bayesian Networks for EEG motor imagery feature extraction. In Proceedings of the 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 22–24 April 2015; pp. 170–173.

36. Gómez, V.; Álvarez, A.; Herrera, P.; Castellanos, G.; Orozco, A. Short Time EEG Connectivity Features to Support Interpretability of MI Discrimination. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*; Vera-Rodriguez, R., Fierrez, J., Morales, A., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 699–706.