

## RESEARCH ARTICLE

# Measuring Connectivity in Linear Multivariate Processes With Penalized Regression Techniques

YURI ANTONACCI<sup>1</sup>, (Member, IEEE), JLENIA TOPPI<sup>2,3</sup>, (Member, IEEE), ANTONIO PIETRABISSA<sup>2</sup>, (Senior Member, IEEE), ALESSANDRA ANZOLIN<sup>4,5</sup>, (Member, IEEE), AND LAURA ASTOLFI<sup>2,3</sup>, (Member, IEEE)

<sup>1</sup>Department of Engineering, University of Palermo, 90128 Palermo, Italy

<sup>2</sup>Department of Computer, Control and Management Engineering, University of Rome La Sapienza, 00185 Rome, Italy

<sup>3</sup>IRCCS Fondazione Santa Lucia, 00179 Rome, Italy

<sup>4</sup>Spaulding Rehabilitation Hospital, Harvard Medical School, Boston, MA 02129, USA

<sup>5</sup>Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA

Corresponding author: Yuri Antonacci (yuri.antonacci@unipa.it)

The work of Yuri Antonacci was supported by SiciliAn MicronanOTeCH Research And Innovation Center “SAMOTHRACE,” spoke 3-Università degli Studi di Palermo “S2-COMMs—Micro and Nanotechnologies for Smart and Sustainable Communities,” under Grant ECS\_0000022 (MUR, PNRR-M4C2). This work was supported in part by the Italian Ministry of University and Research—PRIN under Grant 20207S3NB8 and Grant 2022LB7WKJ, in part by the Sapienza University of Rome—Progetto di Ateneo 2022 under Grant RM1221816C8C757C and SEED-PNR 2021, and in part by Bit-Brain Award 2020 under Grant 2962-B2B.

**ABSTRACT** The evaluation of time and frequency domain measures of coupling and causality relies on the parametric representation of linear multivariate processes. The study of temporal dependencies among time series is based on the identification of a Vector Autoregressive model. This procedure is pursued through the definition of a regression problem solved by means of Ordinary Least Squares (OLS) estimator. However, its accuracy is strongly influenced by the lack of data points and a stable solution is not always guaranteed. To overcome this issue, it is possible to use penalized regression techniques. The aim of this work is to compare the behavior of OLS with different penalized regression methods used for a connectivity analysis in different experimental conditions. Bias, accuracy in the reconstruction of network structure and computational time were used for this purpose. Different penalized regressions were tested by means of simulated data implementing different ground-truth networks under different amounts of data samples available. Then, the approaches were applied to real electroencephalographic signals (EEG) recorded from a healthy volunteer performing a motor imagery task. Penalized regressions outperform OLS in simulation settings when few data samples are available. The application on real EEG data showed how it is possible to use features extracted from brain networks for discriminating between two tasks even in conditions of data paucity. Penalized regression techniques can be used for brain connectivity estimation and can be exploited for the computation of all the connectivity estimators based on linearity assumption overcoming the limitations imposed by the classical OLS.

**INDEX TERMS** Brain connectivity, linear regression, penalized regression analysis, multivariate autoregressive models, motor imagery.

The associate editor coordinating the review of this manuscript and approving it for publication was Luca Cassano.

## I. INTRODUCTION

In the framework of linear signal processing, Vector Autoregressive model (VAR) has been proved to be a robust and reliable tool for analyzing temporal dependencies among

time series in several research fields, ranging from economics to biomedical sciences. In neuroscience, these models are extensively used for evaluating brain connectivity in order to understand how different brain areas communicate [1]. With the term brain connectivity, we refer both to the concept of “functional connectivity”, that is the evaluation of statistical dependencies between different neuronal units, and “effective connectivity” that describes networks with directional effects, thereby requiring the generation of a mechanistic model of the cause-effect relationships [2]. Granger causality (GC) is a versatile tool for the analysis of the cause-effect relationships between different time series and was firstly formulated in the framework of bivariate autoregressive modeling, stating that a time series  $X$  G-causes another time series  $Y$  if the past of  $X$  contains information that helps to predict the future of  $Y$  above and beyond the information already contained in the past of  $Y$  [3]. To account for the influence of other time series, the bivariate formulation has been extended to the multivariate case by using multivariate autoregressive (MVAR) models, leading to the well-known conditional form of the GC [4]. The MVAR analysis allows the time and frequency domain representation of physiological time series to be derived directly from model coefficients and their frequency-domain representation [5]. Accordingly, several measures of connectivity have been introduced and applied in recent years, many of which are based on the concept of GC and rely on the identification of the MVAR model. The most used measures, able to quantify causality in the frequency domain, are the directed transfer function (DTF) [6], the directed coherence (DC) [7] and the partial directed coherence (PDC) with all its formulations [8], [9]. These measures are widely used for the analysis of the interactions between physiological time series and, more generally to characterize brain connectivity [10].

The computation of all these connectivity estimators is based on the identification procedure of an MVAR model which in turn includes the solution of a linear regression problem to estimate the matrix of autoregressive (AR) coefficients and the innovation variance. The Ordinary Least Squares (OLS) method finds the optimal solution, starting from the hypothesis of independence between regressors and residuals of the regression problem [1]. However, its accuracy is strongly influenced by the number of available data samples. In particular, the mean squared error (MSE) in the MVAR parameters estimation decreases when the amount of available data increases [11]. As a rule of thumb, it is necessary that the ratio between the number of data samples available and the number of parameters to be estimated (k-ratio) should be at least equal to 10 to ensure the accuracy of the OLS estimator [12], [13]. Otherwise, particularly when the k-ratio is close to one, the estimation problem becomes ill-posed and under-determined [12], [13]. In such a situation, the OLS does not guarantee the uniqueness of the solution, leading to an ambiguity in the MVAR

representation of the data. One possible solution to overcome this methodological limitation, is to use penalized regression techniques. The idea is to add a constraint (based on  $l_1$ -norm and/or  $l_2$ -norm) to regularize the OLS problem. The effect of  $l_2$ -norm is to shrink the MVAR parameters towards zero, while the effect of  $l_1$ -norm is to select only specific coefficients, setting the others to zero. Both  $l_2$ - and  $l_1$ -norm methods can improve the OLS accuracy when the k-ratio is lower than 10, with the effect of also reducing the mean squared error [13], [14]. Penalized regressions can be convex or non-convex depending on whether the constrained optimization problem they solve guarantees or not the existence of a global minimum [15]. In the literature they have been used for GC estimation [13], [16], [17], [18], [19], neuro-imaging data analysis [20] and brain connectivity analysis between EEG time series in a single trial or in real-time settings [21], [22]. Nevertheless, these applications show some limitations: the comparison between different penalized regressions has been done only for a single value of k-ratio; for a real-time application, only one method based on  $l_1$ -norm was used; convex and non-convex methods were directly compared even if it is well known that the latter do not achieve a stable solution. Although penalized regression techniques have been proposed in different studies, they are not commonly used to study brain connectivity from EEG signals. Therefore, this work is motivated by the fact that there are no available extensive studies that analyze the behavior of penalized regressions in both simulated and real EEG time series by also providing guidelines for their exploitation. Specifically, we have selected five penalized regression techniques that solve a linear convex optimization problem guaranteeing the existence of a solution: Ridge Regression (RR) [23], Least Absolute Shrinkage and Selecting Operator regression (LASSO) [24], Elastic-Net regression (E-NET) [25], fused LASSO regression (F-LASSO) [26] and Sparse group LASSO regression (SG-LASSO) [27]. We will aim to demonstrate by using both simulated and real EEG data, that it is possible, also in conditions of data paucity, to accurately reconstruct connectivity patterns describing networks of several coupled Gaussian systems exhibiting complex interactions, and to discriminate between two different experimental conditions by using features extracted from a brain connectivity analysis starting from short windows of EEG signals [21].

The remainder of this paper is organized as follows. The penalized regression techniques are presented in Sec. II. Sec. III presents a simulation study on EEG-like surrogate data, generated by reproducing ground-truth networks, for different values of the k-ratio. Sec. IV, presents an application of penalized regression techniques to real EEG data, recorded from a healthy subject performing an motor imagery (MI) task. In Sec. V the main features of the different regression techniques are discussed. In Sec. VI and VII the limitations of this study are outlined and the conclusions of the work are drawn.

II. METHODS

In this section, we first introduce the MVAR model identification procedure. Then, we describe the different penalized regressions by highlighting their distinctive features and the underlying principles of each presented technique.

A. MVAR MODEL IDENTIFICATION

Let us consider a dynamical system  $\mathcal{Y}$ , whose activity is mapped by a discrete-time stationary vector stochastic process composed of  $M$  real-valued zero-mean stationary scalar processes,  $Y = [Y_1 \dots Y_M] \in \mathbb{R}^{1 \times M}$ . Considering the time step  $n$  as the current time, the present and the past of the process are denoted as  $Y_n = [Y_{1,n} \dots Y_{M,n}] \in \mathbb{R}^{1 \times M}$  and  $Y_n^- = [Y_{n-1} Y_{n-2} \dots]$ , respectively. Moreover, assuming that  $Y$  is a Markov process of order  $p$ , its whole past history can be truncated using  $p$  time steps, i.e. using the  $Mp$ -dimensional vector  $Y_n^p$  so that  $Y_n^- \approx Y_n^p = [Y_{n-1} \dots Y_{n-p}] \in \mathbb{R}^{1 \times Mp}$ . Then, in the linear signal processing framework, the dynamics of  $Y$  can be completely described by the VAR model:

$$Y_n = \sum_{k=1}^p Y_{n-k} \mathbf{A}_k + U_n, \tag{1}$$

where  $\mathbf{A}_k \in \mathbb{R}^{M \times M}$  is the matrix containing the autoregressive (AR) coefficients, and  $U = [U_1 \dots U_M] \in \mathbb{R}^{1 \times M}$  is a zero-mean white process, denoted as innovations, with covariance matrix  $\Sigma \equiv \mathbb{E}[U_n^T U_n] \in \mathbb{R}^{M \times M}$  ( $\mathbb{E}$  is the expectation value).

Let us now consider a realization of the process  $Y$  involving  $N_s$  consecutive time steps, collected in the data matrix  $[y_1; \dots; y_{N_s}] \in \mathbb{R}^{N_s \times M}$ , where the operator “;” stands for row separation, so that the  $i^{th}$  row is a realization of  $Y_i$ , i.e.  $y_i = [y_{1,i} \dots y_{M,i}]$ ,  $i = 1, \dots, N_s$ , and the  $j^{th}$  column is the time series collecting all realizations of  $Y_j$ , i.e.  $[y_{j,1} \dots y_{j,N_s}]^T$ ,  $j = 1, \dots, M$ . OLS finds an optimal solution for the problem (1) by solving the following linear quadratic problem [1]:

$$\hat{\mathbf{A}} = \underset{\mathbf{A}}{\operatorname{argmin}} \| \mathbf{y} - \mathbf{y}^p \mathbf{A} \|_2^2, \tag{2}$$

where  $\mathbf{y} = [y_{p+1}; \dots; y_{N_s}] \in \mathbb{R}^{(N_s-p) \times M}$  is the matrix of the responses values,  $\mathbf{y}^p = [y_{p+1}^p; \dots; y_{N_s}^p] \in \mathbb{R}^{(N_s-p) \times Mp}$  is the matrix of the regressors and  $\mathbf{A} = [\mathbf{A}_1; \dots; \mathbf{A}_p] \in \mathbb{R}^{Mp \times M}$  is the matrix of coefficients. The problem has a solution in a closed form  $\hat{\mathbf{A}} = ([\mathbf{y}^p]^T \mathbf{y}^p)^{-1} [\mathbf{y}^p]^T \mathbf{y}$  for which the residual sum of squares is minimized. When  $N_s - p \leq Mp$  the OLS does not guarantee the uniqueness of the solution since the matrix  $([\mathbf{y}^p]^T \mathbf{y}^p)$  becomes singular [15].

B. PENALIZED LINEAR REGRESSION TECHNIQUES

From a mathematical point of view, regularizing the OLS problem means adding a constraint to the problem (2). In the Lagrangian form, the constraints are written in the cost function as additional weighted costs. In this scenario, it is possible to carry out separate regression analyses for each process composing the dynamical system  $\mathcal{Y}$ . In other words,

TABLE 1. Penalty terms associated with problem 3.

Method	Norm	Penalty term
OLS	-	-
RR	$l_2$	$\lambda \ A_j\ _2^2$
LASSO	$l_1$	$\lambda \ A_j\ _1$
E-NET	$l_1, l_2$	$\lambda_1 \ A_j\ _1 + \lambda_2 \ A_j\ _2^2$
F-LASSO	$l_1, l_2$	$\lambda_1 \ A_j\ _1 + \lambda_2 \sum_{i=1}^{Mp-1}  A_j(i+1) - A_j(i) $
SG-LASSO	$l_1, l_2$	$\lambda_1 \ A_j\ _1 + \lambda_2 \sum_{i=1}^g w_i^g \ A_j(G_i)\ _2$

it is possible to estimate separately each column  $A_j$  of  $\mathbf{A}$ . The multi-objective optimization problems reads as:

$$\hat{A}_j = \underset{A_j}{\operatorname{argmin}} (\|y_j - \mathbf{y}^p A_j\|_2^2 + \mathcal{F}_p(A_j)), \tag{3}$$

where  $y_j = [y_{j,p+1} \dots y_{j,N_s}]^T \in \mathbb{R}^{(N_s-p) \times 1}$  denotes the  $j$ -th column of  $\mathbf{y}$  containing the response vector for the specific response variable ( $j = 1, \dots, M$ ),  $A_j = [a_{j,1}(1), \dots, a_{j,M}(1), \dots, a_{j,1}(p), \dots, a_{j,M}(p)]^T \in \mathbb{R}^{Mp \times 1}$  is the vector coefficient corresponding to the response variable  $j$  and  $\mathcal{F}_p$  is a penalty function applied to  $A_j$ . As anticipated in Section I, in this work we analyze the linear penalty functions reported in Table 1.

RR is characterized by a  $l_2$ -norm based term that shrinks the estimated AR parameters towards zero, reducing the MSE [23]. The regularization parameter  $\lambda$  controls the amount of penalization to be applied, if  $\lambda = 0$ , RR reduces to OLS. The minimization of the RR functional leads to the closed form solution  $\hat{\mathbf{A}} = ([\mathbf{y}^p]^T \mathbf{y}^p + \lambda \mathbf{I})^{-1} [\mathbf{y}^p]^T \mathbf{y}$ , where  $\mathbf{I} \in \mathbb{R}^{Mp \times Mp}$  is the identity matrix. Besides the reduction of the MSE, it would be of interest to reduce the number of selected parameters (i.e., parameters where non-null values are imposed) especially when  $N_s < Mp$ . This result could be achieved by using a penalty term based on the  $l_0$ -norm, which in turn is equal to the number of non-zero elements in a vector [14]. Unfortunately, such a penalization function would render the optimization problem non-convex [28]. The effect of the  $l_0$ -norm can be approximated by using penalty terms based on the  $l_1$ -norm, such as the LASSO: the penalty term acts so as to shrink some coefficients and set others to 0 [24]. In the LASSO, the parameter  $\lambda$  sets the trade-off between the number of non-null coefficients selected in the matrix  $\mathbf{A}$  and residuals sum of squares (RSS).

Similarly, the E-NET simultaneously performs the automatic variable selection and the coefficient shrinkage [25]. The variable selection is regulated by the parameter  $\lambda_1$ , weighting the term based on the  $l_1$ -norm. The shrinkage of coefficients, instead, is regulated by the parameter  $\lambda_2$  weighting the term based on the  $l_2$ -norm. The E-NET coincides with RR if  $\lambda_1 = 0$  and with LASSO if  $\lambda_2 = 0$ . F-LASSO includes in its penalty function a further term with respect to LASSO, which computes the  $l_1$ -norm of the vector of the differences between the coefficients of successive predictors. This term is used to enforce smoothness along the predictors, i.e. along each column of the matrix  $\mathbf{A}$  of

the AR parameters. When  $\lambda_2=0$  F-LASSO coincides with LASSO [26].

Finally, SG-LASSO is a convex combination of the LASSO and the group LASSO penalties [27]. This procedure imposes a structural constraint on the AR coefficients matrix, in addition to the basic sparsity one ( $l_1$  term), to model the assumption that the predictors can be aggregated in groups of a given size. In the penalty term of SG-LASSO,  $g$  is the number of groups, pre-determined, and  $w_i^g$  denotes the weight for the  $i$ -th group (in this study, as specified in [29], the weight of each group was set to 1). As [16] pointed out, in a multivariate regression analysis,  $M$  groups of  $p$  elements for each column of  $\mathbf{A}$  can be used. The  $p$  elements of each group,  $[a_{ij}(1), \dots, a_{ij}(p)] \in \mathfrak{R}^{p \times 1}$  with  $(i, j) \in (1, \dots, M)$ , can be represented by one AR parameter from the time lag 1 to time lag  $p$ . In other words, the second penalty term of the SG-LASSO selects a subset of groups by setting to zero all the coefficients in some groups (sparsity between groups), whereas the first term encourages sparsity within each group.

Since the different rows of the matrices  $\mathbf{y}^p$  and  $\mathbf{y}_j$  can be considered as independent from each other, it is possible to estimate the optimal value of lambda ( $\lambda_{opt}$ ) by dividing them into training and test sets. In the present work, 50% of the rows were used for the training set and 50% for the testing set. Training and test sets are standardized and the optimal values for  $\lambda$  are selected by using a Generalized Cross Validation (GCV) criterion [13], [18], [20], [30]. This criterion comprises the computation of RSS defined as:  $\|\mathbf{y} - \mathbf{y}^p \hat{\mathbf{A}}\|^2$ , and an estimate of the degrees of freedom, which in turn, for LASSO based regressions, is represented by the number of non-zero coefficient in the matrix  $\hat{\mathbf{A}}$  [31].

### III. COMPARING DIFFERENT PENALIZED LINEAR REGRESSION TECHNIQUES

This section reports the theoretical design and the practical implementation of a simulation of multiple interacting stochastic processes, which is used as a benchmark to illustrate the properties of the penalized regression techniques analyzed in this work. Using this simulation we show how the different techniques behave in different conditions of data paucity. Their performances are then compared by using different performance parameters.

#### A. SIMULATION STUDY

Problem (3) was solved for the different penalty terms listed in Table 1. We used the SLEP package (MatLab<sup>®</sup>) [29] with the multivariate linear regression solved with accelerated gradient methods [32], [33] by exploiting the parallel computing toolbox of MatLab. The code used in this paper, for the model identification through penalized regression techniques, is collected in the S-MVAR MatLab toolbox available at <https://github.com/YuriAntonacci/S-MVAR>. Included in the toolbox, we provide a comprehensive script (Test\_Simulation.m) designed to test all penalized regressions, along with a detailed pipeline, providing a

practical means to assess their effectiveness collectively. All the MatLab functions are described in a dedicated documentation (Readme.pdf).

The simulation study included the following steps:

- 1) Generation of simulated datasets, fitting predefined ground-truth networks under different values of the k-ratio (0.5, 0.8, 1, 1.5, 2, 3). The k-ratio is computed as the number of available data samples  $N_s$  divided by the number of parameters to be estimated  $Mp$ .
- 2) Selection of the regularization parameters for each regression method by means of GCV criterion. The process was iterated 10 times for each  $\lambda$  randomly changing the training and testing sets.
- 3) Estimation of the AR parameters by using the six regression methods of interest, OLS, RR, LASSO, E-NET, F-LASSO and SG-LASSO.
- 4) Evaluation of the performances by comparing the estimated AR parameters with those imposed in the corresponding ground-truth network.

To increase the robustness of the statistical analysis, the entire procedure was repeated 50 times randomizing the data generation.

#### B. SIGNAL GENERATION

The simulated data-sets were generated according to different ground-truth networks by means of an MVAR model used as a generator filter [34]. The simulated multivariate time series ( $M = 10$ ) were generated as realizations of a VAR(10) process fed by Gaussian noise with variance equal to 1. Furthermore, an autoregressive component extracted from a real EEG signal was imposed in the model to reproduce its spectral properties [35]. The EEG signals used for this purpose were acquired from the  $C_z$  location of a healthy subject with sampling frequency equal to 200 Hz during the resting state. The simulated networks were randomly generated with a connection density of 15% (14 out of 90 possible connections). The matrix of AR parameters were automatically generated by the SEED-G software [35] by assigning randomly the lag in the range [1,10] and the coefficient value in the interval [-0.9, 0.9]. This value is imposed only at one lag among the possible  $p$  lags, randomly selected. The null connections are instead characterized by null coefficients at each lag. Under these constraints, 50 realizations of the VAR(10) process were generated for different values of the k-ratio parameter in the set {0.5, 0.8, 1, 1.5, 2, 3}, so that the length of the simulated time series was  $N_s = 50$  when  $k = 0.5$  and  $N_s = 300$  when  $k = 3$ .

#### C. PERFORMANCES EVALUATION

The performances of the investigated regression approaches were assessed in terms of accuracy in estimating the strength of the network link (i.e., values of MVAR coefficients) and in terms of ability to reconstruct the network structure. The error in the connection strength was evaluated separately for non-null and null links. Specifically, the values of the

theoretical AR matrix  $\mathbf{A}$  were compared with the estimated values stored in  $\hat{\mathbf{A}}$  using the Mean Absolute Percentage Error (MAPE) [36] if the theoretical value is different from zero, and the Root Mean Squared Error (RMSE), otherwise:

$$MAPE = \frac{100}{D} \sum_{d \in D} \left| \frac{\mathbf{A}_d - \hat{\mathbf{A}}_d}{\mathbf{A}_d} \right|, \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{n \in N} (\mathbf{A}_n - \hat{\mathbf{A}}_n)^2}{N}}, \quad (5)$$

where  $D$  and  $N$  represent the set of non-null and null elements, respectively. The distributions of MAPE and RMSE across the 50 iterations were then presented separately for each method.

The ability in reconstructing the network structure was tested only for the subset of regression methods that yield a sparse matrix as output (i.e., LASSO, E-NET, F-LASSO and SG-LASSO). In particular, the problem of comparing the two AR parameter matrices, representative of the estimated and theoretical network structure, can be seen as a binary classification task. The existence (class 1) or absence (class 0) of each estimated connection is assessed and compared to the underlying ground-truth structure. Performances were assessed computing the false positive rate (FPR) that measures the fraction of null links for which an AR coefficient different from zero was detected, the false negative rate (FNR) that measures the fraction of non-null links for which an AR coefficient equal to zero was detected, and the Area Under Curve (AUC) that summarized the information provided by FNR and FPR [37], [38]. These performance parameters were evaluated for each individual network and their distribution across the 50 simulated connectivity structures was then presented separately for each regression method.

The last considered performance parameter is the computational time (in seconds), required for the computation of the AR matrix  $\hat{\mathbf{A}}$ . In particular, for penalized regressions, the estimation process is divided into two different steps: 1) selection of the regularization parameters, taking  $T_{sel}$  seconds; 2) computation of  $\hat{\mathbf{A}}$  taking  $T_{comp}$  seconds. The process for the selection of the regularization parameters was performed by means of the GCV criterion within a range of 350 values for  $\lambda_1$  and 100 values for  $\lambda_2$  for each method and for each value of k-ratio (where applicable, see Table 1). To speed-up the entire process, the parallel computing toolbox implemented in MatLab<sup>®</sup> 2016a was used.

#### D. STATISTICAL ANALYSIS

In the present work, a repeated measures two-way ANOVA test was performed for each performance parameter (MAPE, RMSE, FNR, FPR and AUC) in order to evaluate the effect of the factors: k-ratio with the levels {0.5, 0.8, 1, 1.5, 2, 3} (factor K); type of regression method with the levels OLS, RR, LASSO, E-NET, F-LASSO, SG-LASSO (factor TYPE). FNR, FPR and AUC were not considered in the

TABLE 2. F-values of the two-way repeated measures ANOVA.

Factor	MAPE	RMSE	FNR	FPR	AUC
K	367.2**	324**	140.6**	23.7**	143.7**
TYPE	149.8**	1863**	85.3**	277.5**	129.1**
$K \times TYPE$	95.6**	321**	13.5**	10.1**	11.5**

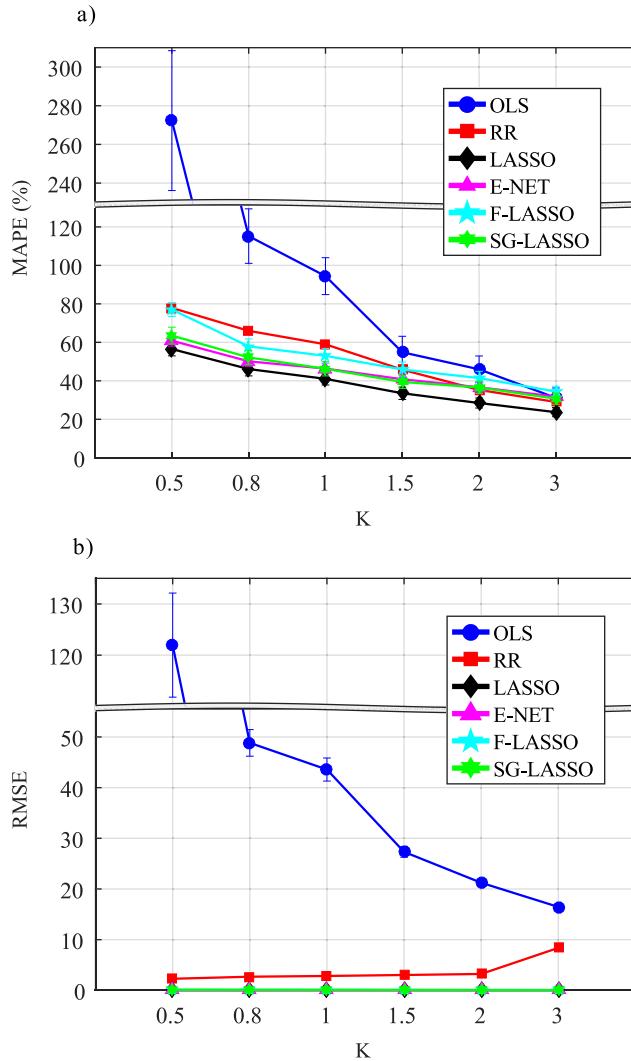
ANOVA for the methods OLS and RR because they do not produce sparse AR matrices. The Greenhouse-Geisser correction for the violation of the spherical hypothesis was used in all the analyses. Tukey's post-hoc test was used for testing the differences between sub-levels of ANOVA factors. Bonferroni-Holm correction was applied for multiple ANOVAs computed on different performance parameters.

#### E. RESULTS OF THE SIMULATION STUDY

The results of the two-way repeated measures ANOVAs computed separately for all the performance parameters are expressed in terms of F-values considering K and TYPE as within factors and reported in Table 2 (\*\* is associated with  $p < 10^{-5}$ ). The two-way ANOVA performed on MAPE and RMSE reveals a strong statistical influence of the main factors K, TYPE, and their interaction  $K \times TYPE$  on the two performance parameters. Fig. 1 reports the distribution of the parameters MAPE and RMSE as a function of the interaction factor  $K \times TYPE$ . The comparison of the six procedures shows remarkable differences between penalized regression techniques and the classical OLS. Specifically, MAPE decreases as the number of data samples available for the estimation procedure increases (Fig. 1.a). In particular, when  $K=3$ , there are no statistically significant differences between all methods, with a MAPE of approximately  $\sim 20\%$ . On the contrary, when  $K \leq 1$ , OLS shows significantly higher values of MAPE (with a sharp increase between 100% and 280%) compared to penalized regression methods where MAPE values remain below 80%. Tukey's post-hoc test reveals statistically significant differences between regression methods only for  $K = 0.5$ . In particular, MAPE computed using LASSO, was significantly lower than RR and F-LASSO.

The analysis of RMSE (Fig. 1.b) reveals overlapping trends that are very close to zero, obtained for the penalized regressions LASSO, F-LASSO, SG-LASSO, and E-NET. Only RR exhibits a higher mean value of RMSE when compared with the other penalized regressions. However, it performs better than OLS. Similarly to the results obtained for the MAPE parameter, OLS shows a sharp increase in RMSE as the data samples (factor K) decrease. In this case as well, there is a noticeable discontinuity in the RMSE trend, between  $K=1.5$  and  $K=1$ , as also highlighted in the analysis of MAPE.

The two-way ANOVA performed on FNR, FPR and AUC reveals a strong statistical effect of the main factors K and TYPE and of their interaction  $K \times TYPE$  (Table 2).



**FIGURE 1.** Performance parameters computed for non-null (MAPE, a) and null (RMSE, b) connections as a function of the factors  $K$  ( $k$ -ratio) and TYPE (regression methods). Both parameters are reported as mean value and 95% confidence interval computed across 50 realizations for each experimental condition (6 regression methods and 6 values of  $K$ ).

Fig.2 reports the distributions of the parameters FNR, FPR, and AUC as a function of the factor  $K$  for each regression approach performing a variable selection procedure (i.e., LASSO, E-NET, F-LASSO, SG-LASSO). All regression methods exhibit a decreasing number of false negatives as the number of available data samples increases (Fig. 2.a). However, under the most challenging condition of  $K = 0.5$ , F-LASSO, and SG-LASSO (azure line and green line, respectively) demonstrate the highest values of FNR, classifying roughly half of the theoretically non-zero links as null. LASSO and E-NET demonstrate better performance compared to all the other methods. No statistically significant differences are noticed between LASSO and E-NET, regardless of factor  $K$ , as highlighted by Tukey’s post-hoc tests. Also, LASSO and E-NET show false negative values ranging between zero ( $K = 3$ ) and 30% ( $K = 0.5$ ).

The analysis of false positives (Fig. 2.b) reveals that the number of links incorrectly classified as non-null remains stable and almost negligible for all methods. The only exception is represented by F-LASSO which on average exhibits higher values of FPR compared to the other methods.

The overall network reconstruction performance, summarized by the AUC parameter (Fig. 2.c), demonstrates that LASSO and E-NET yield the most effective results. Specifically, when compared with F-LASSO and SG-LASSO, both LASSO and E-NET exhibit a higher rate of correctly detected links in all analyzed conditions. SG-LASSO performs better than F-LASSO, which shows the worst performance. For  $K = 3$ , LASSO and E-NET achieve an average AUC value close to 1, indicating a nearly perfect reconstruction of the network structure. Tukey’s post-hoc test indicates that there are no statistically significant differences between LASSO and E-NET. However, both LASSO and E-NET perform significantly better (higher AUC) compared to SG-LASSO and F-LASSO, irrespective of factor  $K$ . Additionally, AUC computed for SG-LASSO is significantly higher than F-LASSO. These regression methods demonstrate the ability to reconstruct the network structure with a very high accuracy of approximately 0.75 even in the worst-case scenario where the number of data samples is half the number of coefficients to be estimated.

Table 3 reports the computational time (in seconds) required for the selection of the optimal value of lambda/s ( $T_{sel}$ ) and the following computation of the MVAR parameters ( $T_{comp}$ ). The two values are computed for each regression method and for each value of  $K$ . In the case of OLS, it was not possible to evaluate the  $T_{sel}$  parameter since it is not included in the set of penalized regression methods. Times were recorded on a PC with an IntelCore i7-6700 processor, clock speed 3.40 GHz, 8 Gb RAM DDR4 (1.33 MHz), Intel (R) HD Graphics 530, 1024 Mb dedicated VRAM. Both,  $T_{sel}$  and  $T_{comp}$  increase with  $K$ . For all the values of  $K$  analyzed, the methods based on one single regularization parameter (i.e., LASSO and RR) need a shorter time for the selection step if compared with those based on two regularization parameters (i.e., F-LASSO, E-NET, and SG-LASSO). Furthermore, among the methods based on  $l_1$ -norm, LASSO shows the lowest computational time required for the estimation process. The most time-consuming method is the SG-LASSO.

To summarize, all performance parameters are directly influenced by the amount of data samples available for the estimation process, exhibiting a direct proportionality. This means that the accuracy of the estimation process improves as the value of  $K$  increases. Even in cases where data is scarce, all the analyzed penalized regression methods demonstrate the ability to achieve good performance in both estimating the values of AR coefficients and reconstructing the network structure. Among the  $l_1$ -norm-based regression methods, LASSO and EL-NET show the best performances, with no statistically significant differences between them. Notably,

**TABLE 3.** Computational time required for each regression method.

Regression Method	$T_{sel}$						$T_{comp}$					
	K=0.5	K=0.8	K=1	K=1.5	K=2	K=3	K=0.5	K=0.8	K=1	K=1.5	K=2	K=3
OLS	-	-	-	-	-	-	0.03	0.02	0.03	0.05	0.05	0.06
RR	7.7	9.4	10.3	13.1	15.3	20.4	0.01	0.02	0.02	0.03	0.04	0.06
LASSO	6.7	10.8	12.9	19.2	23.8	35.7	0.02	0.03	0.03	0.04	0.05	0.06
E-NET	315	495	572	872	927	1400	0.03	0.03	0.04	0.04	0.05	0.09
F-LASSO	588	860	882	1200	1300	1700	0.05	0.06	0.06	0.07	0.08	0.09
SG-LASSO	870	1200	1500	1800	2000	2300	0.03	0.04	0.05	0.07	0.08	0.19

LASSO also exhibits a significantly shorter computational time when compared to EL-NET.

#### IV. TESTING OF LINEAR REGRESSION TECHNIQUES ON REAL EEG DATA

The study of human brain activity during motor imagery (MI) plays an important role in clinical neurology and neuroscience and for this reason the neuronal representation of MI and motor execution has been studied intensively for years using brain imaging techniques such as functional magnetic resonance [39], EEG [40], [41], [42] and positron emission tomography [43]. It is well established that adults present 8-12 Hz (i.e.  $\mu$ ) and 16-26 Hz (i.e.  $\beta$ ) rhythms in the EEG recorded from the scalp over the primary sensorimotor cortex. These rhythms show amplitude fluctuations that are synchronized to movement imagery [44]. These fluctuations are more prominent in the contralateral primary motor cortex and contralateral somatosensory regions as demonstrated in [45]. Given their clearly defined occurrence, the information derived from these oscillations and their localization on the scalp has also been used for developing brain computer interfaces [46]. Since MI involves a dynamic interplay between separate brain regions, it is reasonable to assume that brain connectivity can provide useful information for understating this phenomenon. For instance, in the context of brain computer interfaces, the feasibility of using brain connectivity as an additional feature for the discrimination between different mental tasks has been explored [21], [47]. However, the authors pointed out how the accuracy during the entire process of connectivity estimation drops dramatically when few data samples are available and it would be necessary to use penalized regressions. In such a way, the estimate of connectivity becomes more reliable with the resulting increase in classification accuracy. In this section, by using the EEG dataset previously used in [21], we demonstrate that it is possible: (i) to estimate brain connectivity using a MVAR model identified through penalized regressions, especially when a limited number of data samples are available for the estimation process; (ii) to demonstrate the discriminative power of features extracted directly from brain connectivity analysis in the context of a binary classification problem. All data were drawn from a publicly-available dataset (<https://github.com/scot-dev/scot-data/tree/main/scotdata>) and no new experiments were performed as a part of this work.

#### A. DATA DESCRIPTION AND PRE-PROCESSING

The EEG dataset made available by the authors of [21] consists of 45 EEG channels located over the scalp according to the international 10-20 system. The signals were recorded at a sampling rate of 300 Hz with three synchronized g.USBamp amplifiers (g.tec, Guger Technologies OEG, Graz, Austria) and pre-processed following the pipeline described in [21]. To test the regression methods on real EEG signals, a healthy subject (male, right handed), with no prior experience in brain computer interface (BCI) control based on motor imagery, was selected to record one session, consisting of 90 trials of right hand motor imagery (HAND) and 90 trials of foot motor imagery (FOOT). Further details about the experimental paradigm and the pre-processing process are available in [21].

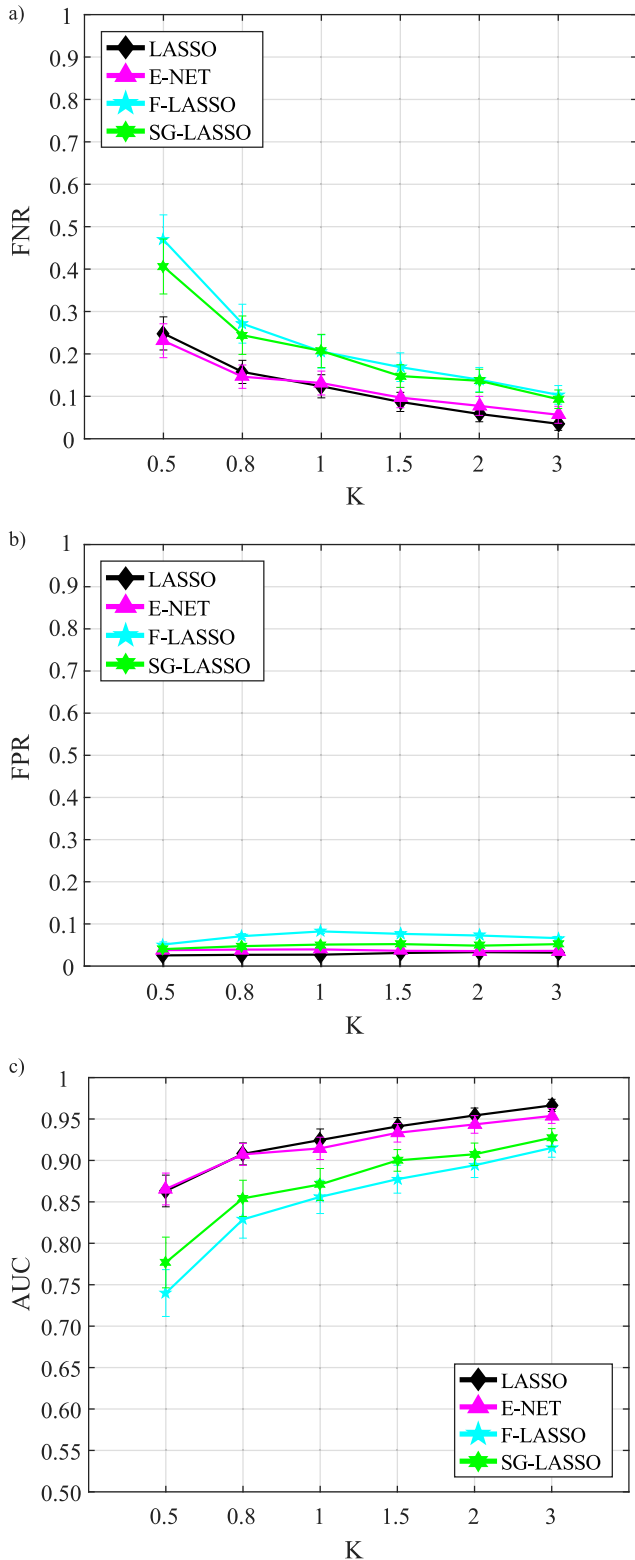
#### B. SINGLE-TRIAL CONNECTIVITY ESTIMATION

To reproduce the condition of the simulation study, 11 out of the 45 available channels were selected ( $C_5, C_3, C_1, C_2, C_4, C_6, CP_3, CP_4, C_z, CP_z, FP_z$ ) and, as suggested in [21], 100 samples between the third and the fourth second were selected to obtain a proper time window in which the task was correctly performed. Each of the eleven time series obtained from each trial and condition (HAND - FOOT) were interpreted as a realization of a VAR process whose matrix of parameters  $\hat{A}$  was estimated with the six different regression methods (i.e. OLS, RR, LASSO, EL-NET, F-LASSO, SG-LASSO). The order of the model, denoted as  $p$ , was estimated for every experimental condition and trial. This estimation was done using the Final Prediction Error (FPE) criterion [1]. Afterwards, the average value of the model order across trials and conditions was calculated, resulting in a value of 10. All the analyses were performed by identifying MVAR models of dimension  $Mp$ , where  $M=11$  and with the model order fixed to  $p = 10$ , which together with a time series of 100 points guarantees a connectivity estimation with a value of  $K$  close to 1 for each regression method.

To retrieve information about the connectivity between EEG signals in the frequency domain, the estimated AR parameters were then Fourier transformed [10].

#### C. CLASSIFICATION TASK

To verify (i) if penalized regressions can be used for estimating connectivity between EEG time series and, (ii) if the features derived from a connectivity analysis can



**FIGURE 2.** Distribution of FNR (a), FPR (b) and AUC (c) parameters considering the interaction  $K \times TYPE$ , expressed as mean value and 95% confidence interval of the parameter computed across 50 realizations for the four regression methods, performing variable selection, and for different values of  $K$ .

provide useful information even in data paucity conditions, a classification task was performed.

The classification HAND vs. FOOT was repeated by using as features the frequency version of estimated AR parameters (real part of the Fourier transform of the AR parameters) in two different frequency bands, typically related to the MI tasks [46]:  $\alpha$  (8-12 Hz) and  $\beta$  (13-30 Hz). By considering  $f_i$  as the frequency interval under investigation and  $M = 11$ , it was possible to extract  $M^2 f_i = 605$  features in  $\alpha$  band ( $f_i \in [8, 12]$ ) and 2178 in  $\beta$  band ( $f_i \in [13, 30]$ ).

In order to ensure a ratio of at least 5 between the number of training cases and the number of classifier parameters to be estimated, i.e., to avoid overfitting during the training phase, a sub-selection of the features was performed [48]. In particular, we selected the top 25 highest -ranked ones according to the associated t-value (independent samples t-test, HAND vs FOOT) for each feature in each frequency band [49]. Note that all the diagonal elements of the AR matrices were removed to maintain only information related to causal effects.

The data were divided into 70% for the training process (126 cases out of 180), 15% for validation (27 cases out of 180) and 15% for testing (27 cases out of 180). Each of the three data sets contained the same number of observations for each class (HAND-FOOT) to avoid class imbalance problem [50]. By using a holdout approach at each iteration, one different feed-forward neural network (FFNN) was trained, validated, and tested [48]. The structure chosen for the FFNN is one of the most widely used for classification purposes [51], [52] that includes one hidden layer with one neuron (sigmoid activation function) and an output layer with a softmax activation function [53]. The initial weights of the network were randomly generated and the training was performed with a gradient descent algorithm (with learning rate set to  $10^{-3}$ ), with cross-entropy used as the cost function [54]. The training process was stopped by means of the early stopping criterion [55]. The process was repeated 100 times for each considered frequency band. As a performance parameter, we computed the classification accuracy (ACC) on the test set [56].

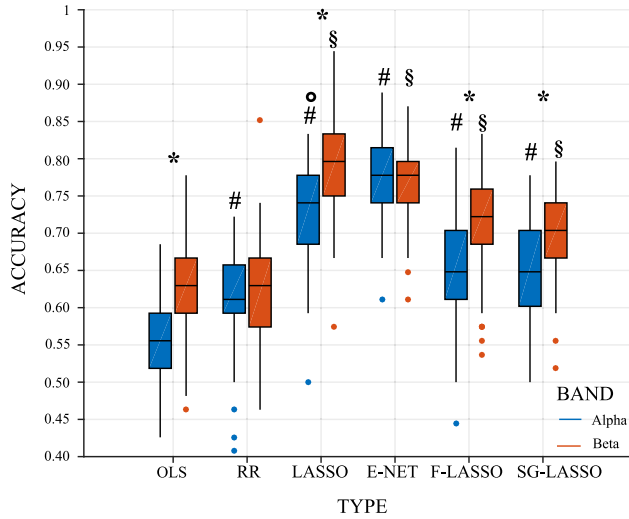
For this study, a two-way repeated measures ANOVA test on the classification accuracy (ACC) was performed in order to evaluate the performance of classification accuracy depending on the regression method (factor TYPE: OLS, RR, E-NET, LASSO, F-LASSO, SG-LASSO) and the frequency band (factor BAND:  $\alpha, \beta$ ). Moreover, to assess statistical differences between pairs of distributions independent samples t-tests were performed.

#### D. RESULTS ON REAL EEG DATA

Fig. 3 reports the distributions of the classification accuracy evaluated on the test set according to the interaction  $BAND \times TYPE$  for each regression method and for  $\alpha$  and  $\beta$  frequency band.

For all the regression methods, except RR and E-NET, ACC is significantly greater in  $\beta$  band (blue boxes) with respect to  $\alpha$  band (orange boxes). The analysis of  $\alpha$  band





**FIGURE 3.** Classification accuracy evaluated on the testing set for each regression method obtained for each frequency band ( $\alpha$  and  $\beta$ ). Box plots report the distribution of the classification accuracy across the 100 iterations by considering the interaction  $TYPE \times BAND$  ( $F=18.57$ ,  $p < 10^{-5}$ ). On each box, the central line indicated the median, the edges of the box indicate 25<sup>th</sup> and 75<sup>th</sup> percentiles, and outliers are marked with a circle. Statistically significant differences between pairs of distributions are marked with \* ( $\alpha$  vs  $\beta$ ), with # ( $\cdot$  vs OLS for  $\alpha$  band), with § ( $\cdot$  vs OLS for  $\beta$  band), and with o (LASSO vs E-NET in  $\alpha$  band).

**TABLE 4.** Computational time required for each regression method ( $K \sim 1$ ).

Regression Method	$T_{sel}$	$T_{comp}$
OLS	-	0.04
RR	6.9	0.05
LASSO	8.1	0.11
E-NET	719	0.15
F-LASSO	997	0.16
SG-LASSO	1978	0.18

reveals a statistically significant increment of the ACC if penalized regressions are used when compared with OLS (distributions marked with #); this is not the case for the  $\beta$  band in which RR is not significantly different from OLS. All the other regression methods (i.e. LASSO, E-NET, F-LASSO, and SG-LASSO) show a statistically significant increment of ACC with respect to that obtained with OLS (distribution marked with §). LASSO and E-NET show the highest values of ACC ( $\sim 0.8$ ) but only using LASSO it is possible to highlight a significant difference between  $\alpha$  and  $\beta$  frequency bands. On the other hand, OLS shows the worst performance in discriminating between HAND and FOOT with an average value of ACC  $\sim 0.55$  in  $\alpha$  band and  $\sim 0.65$  for  $\beta$  band.

Table 4 reports the time required for estimating the AR coefficients matrix  $\hat{A}$  for each regression method. Also in this case the process of cross-validation described in Section III-C was used for the selection of different optimal values of lambda/s with the same PC used for the simulation experiment. As a first observation, the computational times

required are comparable to those obtained in the simulation study (see Table 3) performed on  $M = 10$  EEG-like data for the case  $K = 1$ . Table 4 shows that OLS, which does not require the selection of a regularization parameter, is the fastest method. Among penalized regression methods, RR and LASSO show a difference of a few seconds for the selection of  $\lambda_{opt}$  ( $\sim 7s$ ,  $\sim 8s$  respectively). E-NET, F-LASSO, and SG-LASSO represent the slowest regression methods with several minutes needed for the selection of  $\lambda_1$  and  $\lambda_2$  ( $\sim 12 min$  for E-NET and  $\sim 32 min$  SG-LASSO).

## V. DISCUSSION

The present work compares the performances of different regression techniques for the study of connectivity estimates in the framework of linear signal processing. The behavior of the different regression techniques was validated both in a simulation setting and in a real-world scenario relevant to the study of brain connectivity from EEG signals.

### A. SIMULATION EXPERIMENT

The simulation study was designed to assess the performance of the classic OLS when compared with five different penalized regression techniques (RR, LASSO, E-NET, F-LASSO, and SG-LASSO) used in the identification procedure of the MVAR model. Performance parameters were evaluated in terms of the strength of the network links (values of the MVAR coefficients) as well as network structure (false positives and false negatives). Lastly, the computational time needed to solve all the linear problems was analyzed.

The accuracy in estimating the value of the MVAR coefficients was investigated across different k-ratio values by means of MAPE and RMSE used as performance parameters (Fig. 1). As expected, both performance parameters for all the regression methods show a tendency to increase as the k-ratio decreases. This tendency is more evident for OLS, as already documented testing different VAR identification approaches (e.g., the Levinson recursion for the solution of Yule-Walker equations) in the context of signal processing [12]. For  $K=1$  the uniqueness of the solution is not guaranteed for OLS as the matrix  $([y^p]^T y^p)$ , necessary for the solution of the problem (2), approaches singularity. As a consequence, OLS exhibits a strong bias, as reported in previous studies [13], [24], [57]. On the other hand, we document very low values of MAPE and RMSE for penalized regressions, even when  $K \leq 1$ , confirming their high tolerance to collinearity between regressors caused by the drastic reduction of data samples available [13], [58]. The trends obtained for MAPE highlight the best performances of LASSO and E-NET without any statistically significant differences between them.

The ability in the reconstruction of the network structure was investigated in terms of false negatives, false positives and AUC. The latter appears to be the best-suited indicator in terms of its capability to synthesize the similarity of two networks also in the condition of class imbalance, a typical condition in sparse networks, as those simulated in this work [37]. The regression methods tested here are those

based on  $l_1$ -norm which in turn induces sparsity within the matrix of estimated MVAR coefficients [14]. Focusing on Fig. 2, all the penalized regressions are characterized by a low percentage of false positives (FPR  $\sim 0.05$ ). Additionally, both false negatives and AUC values increase as  $K$  decreases. In particular, the maximum value of AUC ( $\sim 0.95$ ) is obtained by LASSO and E-NET regardless of the value of  $K$ . Even in the worst condition ( $K = 0.5$ ), these two methods reach high values of AUC (above 0.85). On the other hand, F-LASSO showed the worst performance for all the values of  $K$ , with AUC  $\sim 0.75$  when  $K$  is equal to 0.5. The results obtained in this study align with those reported in a previous work that utilized Ridge and LASSO regression to investigate brain connectivity estimated from fMRI data [20]. Specifically, the authors demonstrated how penalized regression techniques can effectively identify causal relationships, even when the number of nodes in the network exceeds the available data samples. The obtained AUC values ranged from 0.6 to 0.9. Moreover, these findings are consistent with our previous work, which compared only OLS and LASSO in the context of information dynamics analysis [13].

The observed differences between LASSO and E-NET when compared to other penalized regressions can potentially be explained through a methodological consideration. In their work, Zou and Hastie [25] highlighted that LASSO and EL-NET may not be the ideal approaches when dealing with structurally grouped regressors (referred to as  $\mathbf{y}^p$  in this study). Specifically, only SG-LASSO introduces a structural constraint to account for grouped variables, for instance when the predictors are represented by genes. In the study of connectivity, where prior information about potential connections between different brain areas is lacking, it is reasonable to assume that LASSO and EL-NET may exhibit superior performance in retrieving the theoretical network structure compared to SG-LASSO and F-LASSO.

Finally, the computational time required from each method to provide an estimate of the coefficient matrix  $\hat{\mathbf{A}}$  was computed for each regression method. Two different computational times were reported, one for the GCV procedure used for the selection of the penalization terms (depending on whether the methods present one or two parameters for regularization) and the other one for the MVAR parameters computation (Table 2). The total time required for the entire estimation procedure increases as the value of  $K$  increases. As expected, regression methods that incorporate two regularization parameters (E-NET, F-LASSO, and SG-LASSO) require more time because all combinations of the two parameters need to be tested through GCV. The less time-consuming regression methods are LASSO (among those based on  $l_1$  norm) and the E-NET regressions (among those based on a linear combination of  $l_1$  and  $l_2$ -norm). To limit the computational time, a smaller number of values could be tested for the different regularization parameters (i.e.  $\lambda_1, \lambda_2$ ). However, this could have an impact on the performance of the selected regression method since the optimal solution may not be in the range selected by the

operator. It is therefore recommended to use at least a range of 100 values for each regularization parameter.

A note should be made concerning the potential impact of cross-validation approach used to identify regularization parameters on the results of the simulation study. Cross-validation procedures are widely employed for selecting regularization parameters in penalized regressions, with their popularity attributed to intuitive appeal, ease of implementation, and the ability to provide a reliable estimate of expected prediction error [15]. In this study, we employed a generalized form of cross-validation, namely GCV, which not only considers prediction error but also incorporates considerations for the interpretability of the estimated model. This approach has demonstrated its reliability in prior studies [13], [18], [20], [59]. However, high variability in the results has been found according to the cross-validation method used for the analysis and the randomly assigned folds used in the training phase [60]. This model selection instability can lead to a lack of interpretability in the results of LASSO and LASSO-based regression analyses. Although various solutions have been introduced in the literature to address this issue, most of them are designed for LASSO regression [61]. Hence, additional research endeavors are essential to generalize these principles across all penalized regression techniques, underlining the absence of a unanimous consensus in the existing literature on the optimal procedures for selecting regularization parameters.

## B. APPLICATION OF LINEAR REGRESSION TECHNIQUES ON REAL EEG DATA

A well-established body of literature demonstrates an intrinsic balance between excitatory and inhibitory coupling among brain regions and between hemispheres occurring in the execution of a motor imagery task [62]. Based on this evidence, we selected the EEG time series recorded from a total of 11 channels including the left hemisphere (contralateral to the hand motor imagination), the right hemisphere (ipsilateral), and the midline. As a consequence, our EEG results are discussed in terms of primary motor cortex activity of the contralateral and ipsilateral hemispheres [63].

The simulation study showed that penalized regressions are reliable tools for estimating brain connectivity in conditions of data paucity where the classical OLS fails. Previous studies highlighted the possibility of discriminating different experimental conditions by using features related to brain connectivity, such as the estimated AR parameters and their frequency version [64], [65]. However, it was pointed out that, with the current methodology based on the OLS estimator, it is impossible to reach an appropriate accuracy and it is necessary to move towards penalized regression techniques [21]. The combined use of features derived from brain connectivity with neural network classifiers showed an increase in classification accuracy [66] with most of the results found in  $\alpha$  and  $\beta$  frequency bands [40], [65]. Interestingly, the results reported here show comparable classification accuracy even for different tasks and for a

different number of data samples available for the estimation procedure. In fact, as reported in Fig. 3, LASSO and E-NET attain the highest values of ACC in the  $\alpha$  and  $\beta$  frequency bands ( $\sim 0.8$ ). Furthermore, as proof of the unsuitability of OLS in a condition of strong data paucity ( $K \sim 1$ ), Fig. 3 showed the lowest values of ACC in both frequency bands.

## VI. FURTHER REMARKS AND LIMITATIONS

Despite the potentialities of the penalized regression techniques, there are some limitations in the performed analyses that should be taken into account. In the literature, it is well-known that conventional EEG-based computer-aided methods involve three basic steps: preprocessing of signals, feature extraction from preprocessed signals, and feature classification. The EEG data recording procedure is highly susceptible to different noise artifacts and several methodologies can be used to counteract this issue. Among these, multiscale principal component analysis (MSPCA) has proven to be a reliable tool for noise elimination from EEG signals through the combined use of wavelets and principal component analysis (PCA) [67], [68], [69]. In particular, the MSPCA works in two folds: first, it applies wavelet transforms to obtain deterministic attributes from EEG signals; second, it performs PCA on each wavelet coefficient [70]. Additionally, a novel empirical Fourier decomposition (EFD) framework was recently proposed for decomposing EEG signals into different modes after the denoising step achieved through MSPCA [68], [71]. As the main challenge in BCI applications is to achieve accurate classification results, preprocessing of EEG signals has a significant impact on classification performance, often yielding accuracy values close to 100% [67], [68], [69], [70], [72]. In the present study, we did not perform denoising of EEG signals using MSPCA and EFD, which likely affected the classification performance, resulting in an accuracy of approximately 80%. This is notably lower when compared to results obtained in recent studies that classified different motor imagery tasks [67], [68], [69], [70], [72]. However, it is well-documented in the literature how a simple filter with a non-zero phase response can impair brain connectivity analysis [73], highlighting the need for further research to evaluate the effects of these new signal processing techniques on brain connectivity analysis.

As for the feature selection procedure, this study utilized features directly derived from brain connectivity analysis. While previous studies have employed these features in EEG signal analysis [47], [64], exploring alternative feature sets has the potential to enhance classification accuracy. Recently, a novel category of features, known as geometrical features, has emerged to bolster the classification of various physiological and pathological states directly from EEG signals. Geometrical features pertain to a collection of quantitative measurements or characteristics extracted from the geometric representation of data in two- or three-dimensional space. In the literature, these features have found utility in the analysis of various biosignals, including

heart rate variability [74], electrocardiography [75], and EEG for the detection of different conditions such as depression [76] and seizures [77]. Previous studies have demonstrated that the incorporation of geometrical features can significantly boost classification accuracy, often by as much as 20%, when compared to features extracted from the time, frequency, or time-frequency domains. In summary, the use of geometrical features in motor imagery EEG data holds promise, particularly when motor imagery tasks involve spatial representations. However, the extent of this improvement hinges on several factors, including feature selection, data quality, classifier choice, and the nature of the motor imagery tasks. Therefore, careful feature selection and systematic experimentation are crucial for optimizing classification accuracy in this context.

Another important issue in the EEG-based brain-computer interfaces is the classification procedure. Our decision to use a NN as a classifier is motivated by the advancement in Artificial Intelligence. Indeed, various techniques have emerged and in this context NNs have received considerable attention. To analyze EEG signals, deep learning models have emerged to be more proficient in comparison to traditional techniques and other models [78] and, among the deep learning algorithms available in the literature, convolutional neural networks (CNNs) have proved to be the most suitable for EEG analysis [79]. Notably, CNNs have proved their ability to automate the classification, recognition, and identification processes for various conditions using EEG signals, extending beyond motor imagery to encompass areas such as sleep disorders [80], Alzheimer's disease [81], seizures [82], depression [83], and alcoholism [71], [84].

One final limitation of the present study is the relatively small sample size of subjects analyzed. While penalized regression techniques have shown promising results, further investigation on larger and more diverse datasets is required. Expanding the scope of our research to include a broader range of subjects will not only enhance the robustness of our findings but also provide a more comprehensive understanding of their practical applicability.

## VII. CONCLUSION AND IMPLICATION

The aim of this work was to evaluate the usefulness of penalized regression methods for brain connectivity estimation. The simulation study showed how LASSO and E-NET can estimate, with high accuracy, not only the value of AR coefficients but also the related connectivity structure even in conditions of data paucity in which OLS fails (e.g. when collinearity between regressors arises for the lack of data points).

The results consider the possibility of discriminating between two tasks through a classifier trained with features extracted from a brain connectivity analysis. LASSO and E-NET showed the best performances in terms of accuracy of classification. Notably, these results were obtained for  $K = 1$ , a condition in which OLS fails whereas E-NET and LASSO showed comparable performances when simulated EEG

signals were used. These findings suggest that, when OLS cannot be used, LASSO might represent the most suitable alternative, even when a limited amount of computational power is available.

If confirmed on a larger sample, the overall results pave the way for using sparse identification procedures for connectivity estimation in all those conditions in which few data points are available, such as in the estimation of brain networks at the single trial level as well as during real-time applications. Since all the linear connectivity estimators are based on the identification of an MVAR model, it is reasonable to assume that penalized regressions could be used and integrated for the computation of all the connectivity estimators in both time and frequency domains as already partially demonstrated [85], [86].

Several aspects deserve further investigation with a direct impact on brain connectivity estimation. First and foremost, an important research activity will be a systematic comparison of various brain connectivity estimators, through penalized regressions. This in-depth analysis will unravel the strengths and weaknesses inherent in each approach, significantly advancing our understanding of the modalities used for connectivity estimation. Equally important is the exploration of the feasibility of comparing these estimators with their non-linear counterparts, grounded in deep learning approaches. This exploration is key to effectively model the intricate relationships among brain signals as stated in previous works on this topic [18], [87], [88]. The findings from these comparative analyses will foster a deeper comprehension of the nuances associated with each approach. Given that the majority of the tools currently used to perform brain connectivity analysis rely on MatLab code, the development of Python-based code can enhance the accessibility and applicability of penalized regression techniques even in different research fields.

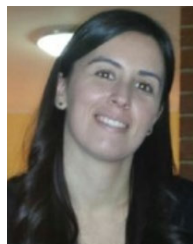
From an applicative point of view, it will be interesting to test the effects of EEG signal denoising on the estimation of brain connectivity. Finally, to validate our findings, penalized regression techniques will be tested on a larger experimental group, potentially providing the basis for further research in the direction of connectivity-based BCIs

## REFERENCES

- [1] H. Lütkepohl, *New Introduction To Multiple Time Series Analysis*. Cham, Switzerland: Springer, 2005.
- [2] G. Chiarion, L. Sparacino, Y. Antonacci, L. Faes, and L. Mesin, "Connectivity analysis in EEG data: A tutorial review of the state of the art and emerging trends," *Bioengineering*, vol. 10, no. 3, p. 372, Mar. 2023.
- [3] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, p. 424, Aug. 1969.
- [4] J. F. Geweke, "Measures of conditional linear dependence and feedback between time series," *J. Amer. Stat. Assoc.*, vol. 79, no. 388, p. 907, Dec. 1984.
- [5] P. L. Nunez and B. A. Cutillo, *Neocortical Dynamics and Human EEG Rhythms*. London, U.K.: Oxford Univ. Press, 1995.
- [6] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol. Cybern.*, vol. 85, no. 2, pp. 145–157, Aug. 2001.
- [7] L. A. Baccalá, K. Sameshima, G. Ballester, A. C. D. Valle, and C. Timo-Iaria, "Studying the interaction between brain structures via directed coherence and Granger causality," *Appl. Signal Process.*, vol. 5, no. 1, p. 40, 1998.
- [8] L. A. Baccalá and K. Sameshima, "Partial directed coherence: A new concept in neural structure determination," *Biol. Cybern.*, vol. 84, no. 6, pp. 463–474, May 2001.
- [9] G. Plomp, C. Quairiaux, C. M. Michel, and L. Astolfi, "The physiological plausibility of time-varying granger-causal modeling: Normalization and weighting by spectral power," *NeuroImage*, vol. 97, pp. 206–216, Aug. 2014.
- [10] L. Faes, S. Erla, and G. Nollo, "Measuring connectivity in linear multivariate processes: Definitions, interpretation, and practical analysis," *Comput. Math. Methods Med.*, vol. 2012, pp. 1–18, Mar. 2012.
- [11] M. H. Ahmad, R. Adnan, and N. Adnan, "A comparative study on some methods for handling multicollinearity problems," *Matematika*, vol. 22, pp. 109–119, Dec. 2006.
- [12] A. Schlögl and G. Supp, "Analyzing event-related EEG data with multivariate autoregressive parameters," *Prog. brain Res.*, vol. 159, pp. 135–147, Jan. 2006.
- [13] Y. Antonacci, L. Astolfi, G. Nollo, and L. Faes, "Information transfer in linear multivariate processes assessed through penalized regression techniques: Validation and application to physiological networks," *Entropy*, vol. 22, no. 7, p. 732, Jul. 2020.
- [14] I. Rish and G. Grabarnik, *Sparse Modeling: Theory, Algorithms, and Applications*. Boca Raton, FL, USA: CRC Press, 2014.
- [15] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Cham, Switzerland: Springer, 2009.
- [16] S. Haufe, K.-R. Müller, G. Nolte, and N. Krämer, "Sparse causal discovery in multivariate time series," in *Proc. Workshop Causality, Objectives Assessment NIPS*, Jun. 2010, pp. 97–106.
- [17] Y. Antonacci, L. Astolfi, and L. Faes, "Testing different methodologies for Granger causality estimation: A simulation study," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 940–944.
- [18] Y. Antonacci, L. Minati, L. Faes, R. Pernice, G. Nollo, J. Toppi, A. Pietrabissa, and L. Astolfi, "Estimation of Granger causality through artificial neural networks: Applications to physiological systems and chaotic electronic oscillators," *PeerJ Comput. Sci.*, vol. 7, p. e429, May 2021.
- [19] A. Arnold, Y. Liu, and N. Abe, "Temporal causal modeling with graphical Granger methods," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2007, pp. 66–75.
- [20] P. A. Valdés-Sosa, J. M. Sánchez-Bornot, A. Lage-Castellanos, M. Vega-Hernández, J. Bosch-Bayard, L. Melie-García, and E. Canales-Rodríguez, "Estimating brain functional connectivity with sparse multivariate autoregression," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 360, no. 1457, pp. 969–981, May 2005.
- [21] M. Billinger, C. Brunner, and G. R. Müller-Putz, "Single-trial connectivity estimation for classification of motor imagery data," *J. Neural Eng.*, vol. 10, no. 4, Aug. 2013, Art. no. 046006.
- [22] T. R. Mullen, C. A. E. Kothe, Y. M. Chi, A. Ojeda, T. Kerth, S. Makeig, T.-P. Jung, and G. Cauwenberghs, "Real-time neuroimaging and cognitive monitoring using wearable dry EEG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2553–2567, Nov. 2015.
- [23] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 42, no. 1, p. 80, Feb. 2000.
- [24] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Society: Ser. B (Methodol.)*, vol. 58, no. 1, pp. 267–288, Jan. 1996.
- [25] H. Zou and T. Hastie, "Regression shrinkage and selection via the elastic net, with applications to microarrays," *JR Stat. Soc. Ser. B*, vol. 67, pp. 20–301, 2003.
- [26] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 67, no. 1, pp. 91–108, Feb. 2005.
- [27] J. Friedman, T. Hastie, and R. Tibshirani, "A note on the group lasso and a sparse group lasso," 2010, *arXiv:1001.0736*.
- [28] O. P. Burdakov, C. Kanzow, and A. Schwartz, "Mathematical programs with cardinality constraints: Reformulation by complementarity-type conditions and a regularization method," *SIAM J. Optim.*, vol. 26, no. 1, pp. 397–425, Jan. 2016.

- [29] J. Liu, S. Ji, and J. Ye, "SLEP: Sparse learning with efficient projections," *Arizona State Univ.*, vol. 6, no. 491, p. 7, Aug. 2009.
- [30] R. J. Tibshirani and J. Taylor, "Degrees of freedom in lasso problems," *Ann. Statist.*, vol. 40, no. 2, pp. 1198–1232, Apr. 2012.
- [31] J. M. Sánchez-Bornot, E. Martínez-Montes, A. Lage-Castellanos, M. Vega-Hernández, and P. A. Valdés-Sosa, "Uncovering sparse brain effective connectivity: A voxel-based approach using penalized regression," *Statistica Sinica*, vol. 18, no. 4, pp. 1501–1518, Oct. 2008.
- [32] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.
- [33] Y. Nesterov, "Gradient methods for minimizing composite functions," *Math. Program.*, vol. 140, no. 1, pp. 125–161, Aug. 2013.
- [34] J. Toppi, D. Mattia, M. Riseti, R. Formisano, F. Babiloni, and L. Astolfi, "Testing the significance of connectivity networks: Comparison of different assessing procedures," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 12, pp. 2461–2473, Dec. 2016.
- [35] A. Anzolin, J. Toppi, M. Petti, F. Cincotti, and L. Astolfi, "SEED-G: Simulated EEG data generator for testing connectivity algorithms," *Sensors*, vol. 21, no. 11, p. 3632, May 2021.
- [36] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecasting*, vol. 32, no. 3, pp. 669–679, Jul. 2016.
- [37] J. Toppi, N. Sciaraffa, Y. Antonacci, A. Anzolin, S. Caschera, M. Petti, D. Mattia, and L. Astolfi, "Measuring the agreement between brain connectivity networks," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 68–71.
- [38] Y. Antonacci, J. Toppi, D. Mattia, A. Pietrabissa, and L. Astolfi, "Estimation of brain connectivity through artificial neural networks," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 636–639.
- [39] C. A. Porro, M. P. Francescato, V. Cettolo, M. E. Diamond, P. Baraldi, C. Zuiani, M. Bazzocchi, and P. E. D. Prampero, "Primary motor and sensory cortex activation during motor performance and motor imagery: A functional magnetic resonance imaging study," *J. Neurosci.*, vol. 16, no. 23, pp. 7688–7698, Dec. 1996.
- [40] F. Pichiorri, F. D. V. Fallani, F. Cincotti, F. Babiloni, M. Molinari, S. C. Kleih, C. Neuper, A. Kübler, and D. Mattia, "Sensorimotor rhythm-based brain-computer interface training: The impact on motor cortical responsiveness," *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025020.
- [41] I. Pirovano, A. Mastropietro, Y. Antonacci, C. Barà, E. Guanziroli, F. Molteni, L. Faes, and G. Rizzo, "Resting state EEG directed functional connectivity unveils changes in motor network organization in subacute stroke patients after rehabilitation," *Frontiers Physiol.*, vol. 13, Apr. 2022, Art. no. 862207.
- [42] I. Pirovano, Y. Antonacci, A. Mastropietro, C. Barà, L. Sparacino, E. Guanziroli, F. Molteni, M. Tettamanti, L. Faes, and G. Rizzo, "Rehabilitation modulates high-order interactions among large-scale brain networks in subacute stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 4549–4560, 2023.
- [43] J. Decety, D. Perani, M. Jeannerod, V. Bettinardi, B. Tadary, R. Woods, J. C. Mazziotta, and F. Fazio, "Mapping motor representations with positron emission tomography," *Nature*, vol. 371, no. 6498, pp. 600–602, Oct. 1994.
- [44] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topogr.*, vol. 12, no. 3, pp. 177–186, Feb. 2000.
- [45] S. Hétu, M. Grégoire, A. Saimpont, M.-P. Coll, F. Eugène, P.-E. Michon, and P. L. Jackson, "The neural network of motor imagery: An ALE meta-analysis," *Neurosci. Biobehavioral Rev.*, vol. 37, no. 5, pp. 930–949, Jun. 2013.
- [46] F. Pichiorri, G. Morone, M. Petti, J. Toppi, I. Pisotta, M. Molinari, S. Paolucci, M. Inghilleri, L. Astolfi, F. Cincotti, and D. Mattia, "Brain-computer interface boosts motor imagery practice during stroke recovery," *Ann. Neurol.*, vol. 77, no. 5, pp. 851–865, May 2015.
- [47] D. Rathee, H. Cecotti, and G. Prasad, "Single-trial effective brain connectivity patterns enhance discriminability of mental imagery tasks," *J. Neural Eng.*, vol. 14, no. 5, Oct. 2017, Art. no. 056005.
- [48] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surveys*, vol. 4, no. none, pp. 40–79, Jan. 2010.
- [49] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, *Introduction To Pattern Recognition: A MATLAB Approach*. New York, NY, USA: Academic, 2010.
- [50] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1–54, Dec. 2019.
- [51] N. J. Guliyev and V. E. Ismailov, "A single hidden layer feedforward network with only one neuron in the hidden layer can approximate any univariate function," *Neural Comput.*, vol. 28, no. 7, pp. 1289–1304, Jul. 2016.
- [52] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.
- [53] C. M. Bishop, *Pattern Recognition and Machine Learning*. Cham, Switzerland: Springer, 2006.
- [54] G. P. Zhang, "Neural networks for classification: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 30, no. 4, pp. 451–462, Nov. 2000.
- [55] L. Prechelt, "Automatic early stopping using cross validation: Quantifying the criteria," *Neural Netw.*, vol. 11, no. 4, pp. 761–767, Jun. 1998.
- [56] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognit. Lett.*, vol. 30, no. 1, pp. 27–38, Jan. 2009.
- [57] Y. Antonacci, J. Toppi, D. Mattia, A. Pietrabissa, and L. Astolfi, "Single-trial connectivity estimation through the least absolute shrinkage and selection operator," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 6422–6425.
- [58] M. Irfan, M. Javed, and M. A. Raza, "Comparison of shrinkage regression methods for remedy of multicollinearity problem," *Middle-East J. Sci. Res.*, vol. 14, no. 4, pp. 570–579, 2013.
- [59] O. Yamashita, A. Galka, T. Ozaki, R. Biscay, and P. Valde-Sosa, "Recursive penalized least squares solution for dynamical inverse problems of EEG generation," *Human Brain Mapping*, vol. 21, no. 4, pp. 221–235, Apr. 2004.
- [60] H. M. Bøvelstad, S. Nygård, H. L. Størvold, M. Aldrin, Ø. Borgan, A. Frigessi, and O. C. Lingjærde, "Predicting survival from microarray data—A comparative study," *Bioinformatics*, vol. 23, no. 16, pp. 2080–2087, Aug. 2007.
- [61] S. Roberts and G. Nowak, "Stabilizing the lasso against cross-validation variability," *Comput. Statist. Data Anal.*, vol. 70, pp. 198–211, Feb. 2014.
- [62] C. Gerloff, "Functional coupling and regional activation of human cortical motor areas during simple, internally paced and externally paced finger movements," *Brain*, vol. 121, no. 8, pp. 1513–1531, Aug. 1998.
- [63] N. Swindale, "Cortical cartography: What's in a map?" *Current Biol.*, vol. 11, no. 19, pp. R764–R767, Oct. 2001.
- [64] D. Vidaurre, C. Bielza, and P. Larrañaga, "Classification of neural signals from sparse autoregressive features," *Neurocomputing*, vol. 111, pp. 21–26, Jul. 2013.
- [65] M. Hamed, S.-H. Salleh, and A. M. Noor, "Electroencephalographic motor imagery brain connectivity analysis for BCI: A review," *Neural Comput.*, vol. 28, no. 6, pp. 999–1041, Jun. 2016.
- [66] N.-J. Huan and R. Palaniappan, "Neural network classification of autoregressive features from electroencephalogram signals for brain-computer interface design," *J. Neural Eng.*, vol. 1, no. 3, pp. 142–150, Sep. 2004.
- [67] M. T. Sadiq, X. Yu, Z. Yuan, M. Z. Aziz, S. Siuly, and W. Ding, "Toward the development of versatile brain-computer interfaces," *IEEE Trans. Artif. Intell.*, vol. 2, no. 4, pp. 314–328, Aug. 2021.
- [68] X. Yu, M. Z. Aziz, M. T. Sadiq, Z. Fan, and G. Xiao, "A new framework for automatic detection of motor and mental imagery EEG signals for robust BCI systems," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.
- [69] M. T. Sadiq, X. Yu, Z. Yuan, and M. Z. Aziz, "Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform," *Electron. Lett.*, vol. 56, no. 25, pp. 1367–1369, Dec. 2020.
- [70] M. T. Sadiq, X. Yu, Z. Yuan, M. Z. Aziz, N. U. Rehman, W. Ding, and G. Xiao, "Motor imagery BCI classification based on multivariate variational mode decomposition," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 5, pp. 1177–1189, Oct. 2022.
- [71] M. T. Sadiq, H. Akbari, S. Siuly, Y. Li, and P. Wen, "Alcoholic EEG signals recognition based on phase space dynamic and geometrical features," *Chaos, Solitons Fractals*, vol. 158, May 2022, Art. no. 112036.
- [72] M. T. Sadiq, X. Yu, Z. Yuan, F. Zeming, A. U. Rehman, I. Ullah, G. Li, and G. Xiao, "Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain-computer interfaces," *IEEE Access*, vol. 7, pp. 171431–171451, 2019.

- [73] L. Barnett and A. K. Seth, "The MVGC multivariate Granger causality toolbox: A new approach to granger-causal inference," *J. Neurosci. Methods*, vol. 223, pp. 50–68, Feb. 2014.
- [74] S. Behbahani, "Electrocardiogram and heart rate variability assessment in patients with common autoimmune diseases: A methodological review," *Türk Kardiyoloji Derneği Arşivi-Turkish Soc. Cardiology*, vol. 48, no. 3, pp. 1–16, Apr. 2020.
- [75] I. Nejadgholi, M. H. Moradi, and F. Abdolali, "Using phase space reconstruction for patient independent heartbeat classification in comparison with some benchmark methods," *Comput. Biol. Med.*, vol. 41, no. 6, pp. 411–419, Jun. 2011.
- [76] H. Akbari, M. T. Sadiq, M. Payan, S. S. Esmaili, H. Baghri, and H. Bagheri, "Depression detection based on geometrical features extracted from SODP shape of EEG signals and binary PSO," *Traitement du Signal*, vol. 38, no. 1, pp. 13–26, Feb. 2021.
- [77] H. Akbari, M. T. Sadiq, N. Jafari, J. Too, N. Mikaeilvand, A. Cicone, and S. Serra-Capizzano, "Recognizing seizure using Poincaré plot of EEG signals and graphical features in DWT domain," *Bratislava Med. J.*, vol. 124, no. 1, pp. 12–24, 2022.
- [78] S. Siuly, Y. Li, and Y. Zhang, "EEG signal analysis and classification," *IEEE Trans. Neural. Syst. Rehabil. Eng.*, vol. 11, pp. 141–144, 2016. [Online]. Available: <https://doi.org/10.1007/978-3-319-47653-7>
- [79] S. Rajwal and S. Aggarwal, "Convolutional neural network-based EEG signal analysis: A systematic review," *Arch. Comput. Methods Eng.*, vol. 30, no. 6, pp. 3585–3615, Jul. 2023.
- [80] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.
- [81] F. C. Morabito, M. Campolo, C. Ieracitano, J. M. Ebadi, L. Bonanno, A. Bramanti, S. Desalvo, N. Mammone, and P. Bramanti, "Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings," in *Proc. IEEE 2nd Int. Forum Res. Technol. Soc. Ind. Leveraging Better Tomorrow (RTSI)*, Sep. 2016, pp. 1–6.
- [82] A. O'Shea, G. Lightbody, G. Boylan, and A. Temko, "Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture," *Neural Netw.*, vol. 123, pp. 12–25, Mar. 2020.
- [83] Z. Wan, J. Huang, H. Zhang, H. Zhou, J. Yang, and N. Zhong, "HybridEEGNet: A convolutional neural network for EEG feature learning and depression discrimination," *IEEE Access*, vol. 8, pp. 30332–30342, 2020.
- [84] L. Farsi, S. Siuly, E. Kabir, and H. Wang, "Classification of alcoholic EEG signals using a deep learning method," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3552–3560, Feb. 2021.
- [85] Y. Antonacci, L. Minati, D. Nuzzi, G. Mijatovic, R. Pernice, D. Marinazzo, S. Stramaglia, and L. Faes, "Measuring high-order interactions in rhythmic processes through multivariate spectral information decomposition," *IEEE Access*, vol. 9, pp. 149486–149505, 2021.
- [86] E. Siggiridou and D. Kugiumtzis, "Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1759–1773, Apr. 2016.
- [87] N. Talebi, A. M. Nasrabadi, and I. Mohammad-Rezazadeh, "Estimation of effective connectivity using multi-layer perceptron artificial neural network," *Cognit. Neurodynamics*, vol. 12, no. 1, pp. 21–42, Feb. 2018.
- [88] A. Montalto, S. Stramaglia, L. Faes, G. Tessitore, R. Prevete, and D. Marinazzo, "Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality," *Neural Netw.*, vol. 71, pp. 159–171, Nov. 2015.



**JLENIA TOPPI** (Member, IEEE) received the bachelor's degree (summa cum laude) in clinical engineering and the master's degree (summa cum laude) in biomedical engineering from the University of Rome La Sapienza, in 2006 and 2009, respectively, and the Ph.D. degree (Hons.) in biomedical engineering from the University of Bologna (Alma Mater Studiorum). She is currently an Associate Professor with the Department of Computer, Control and Management Engineering, University of Rome La Sapienza. Since 2010, she has carried out research activity on healthy and pathological individuals at the Neuroelectrical Imaging and Brain Computer Interface Laboratory, IRCCS Fondazione Santa Lucia, Rome, Italy. Her expertise is in the fields of signal processing, mathematical modeling of biological systems, EEG, neuroelectrical imaging, connectivity estimation, hyperscanning, statistical assessment, and graph theory. She has participated in several national and international research projects, funded by Italian Ministry of Education, 7th Framework Program and Horizon 2020 of the European Commission. Her research interests include the development and implementation of new approaches for biomedical signal processing, with a special focus on neuroelectrical data (electroencephalography, EEG) to reconstruct the brain circuits at the basis of cognitive processes and social cognition. She is an Editor of the *Computational and Mathematical Methods in Medicine*. She serves as a reviewer for several peer-reviewed journals.



**ANTONIO PIETRABISSA** (Senior Member, IEEE) received the M.Sc. degree in electronics engineering and the Ph.D. degree in systems engineering from the University of Rome La Sapienza, in 2000 and 2004, respectively. He is currently an Associate Professor with the Department of Computer, Control, and Management Engineering "Antonio Ruberti" (DIAG), University of Rome La Sapienza, where he teaches automatic control and process automation. Since 2000, he has been participating in more than 30 EU and national research projects, playing the role of scientific coordinator of the projects 5G-ALLSTARS, on 5G communications, funded within the H2020 Europe-South Korea Cooperation, ARIES, on fire emergency prevention, funded by ESA, DAAS, on safe communications in explosive risk areas, funded by MIUR-FILAS, and FedMedAI, on medical applications of federated learning, funded by Regione Lazio (IT). He is the author of more than 60 journal articles and 85 conference papers. His research interests include the application of systems and control theory to the analysis and control of networks. He serves as an Associate Editor for *Control Engineering Practice* (Elsevier) and *IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING*.



**ALESSANDRA ANZOLIN** (Member, IEEE) was born in Taranto, in 1988. She received the Ph.D. degree in bioengineering from the University of Rome La Sapienza. She started her research activity with the Neuro-Electrical Imaging and BCI Laboratory, IRCCS Fondazione Santa Lucia, and the Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG), University of Rome La Sapienza. She is currently a Postdoctoral Fellow with the Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Harvard Medical School, where she is investigating the neural mechanisms at the basis of therapeutic alliance and how it contribute to chronic pain treatment and management. Her appointments include the Athinoula A. Martinos Center for Biomedical Imaging (Massachusetts General Hospital) where she also conducts research on non-invasive brain stimulation.



**YURI ANTONACCI** (Member, IEEE) was born in Rome, Italy, in 1989. He received the M.Sc. degree (cum laude) in biomedical engineering and the Ph.D. degree in bioengineering from the University of Rome "La Sapienza," Italy, in 2016 and 2021, respectively. From 2021 to 2023, he was a Postdoctoral Research Fellow with the Department of Physics and Chemistry "Emilio Segre" and the Department of Engineering, University of Palermo, where he is currently an Assistant Professor. His main research interests include time series analysis, digital signal processing, and system modeling applied to physiological data.



**LAURA ASTOLFI** (Member, IEEE) received the M.S. degree in electronic engineering from the Sapienza University of Rome and the Ph.D. degree in biomedical engineering from the University of Bologna Alma Mater Studiorum. She is currently a Full Professor with the Department of Computer, Control and Management Engineering, Sapienza University of Rome, where she is Head of the Bioengineering and Bioinformatics Laboratory. Her research focuses on the development and application of computational methods and models to study brain activity and functional connectivity in humans. She has served as Chair of the Technical Committee on Biomedical Signal Processing and as a member of the Administrative Committee of the IEEE Engineering in Medicine and

Biology Society (EMBS). She is a member of the Scientific Board of the International Society for Brain Electromagnetic Topography (ISBET) and a Fellow of the European Alliance for Medical and Biological Engineering and Science (EAMBES). She has participated and/or been PI in several national (Ministry of Health, Ministry of University, private foundations), European (7th FP, Horizon2020, Horizon Europe) and U.S. (NSF and NIH) funded research projects. She has received several national and international awards for her scientific activities and has been a keynote speaker and invited lecturer at more than 40 international conferences and schools. She is an Associate Editor for *Medical & Biological Engineering & Computing* and *IEEE Reviews in Biomedical Engineering*. She has authored 178 papers with more than 11000 citations and an H-index of 52.

...