Neurocomputing 510 (2022) 159–171

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A hypothesis-driven method based on machine learning for neuroimaging data analysis

J.M. Gorriz [a,b,c,*], R. Martín-Clemente [e], C.G. Puntonet [d], A. Ortiz [f], J. Ramírez [a], SiPBA group [a], J. Suckling [b]

[a] DaSCI Institute, University of Granada, Spain
[b] Dpt. of Psychiatry, University of Cambridge, UK
[c] ibs.Granada, Granada, Spain
[d] Dpt. Computer Architecture and Technology, University of Granada, Spain
[e] Dpt. Signal Theory and Communications, University of Seville, Spain
[f] Dpt. Communication Engineering, University of Málaga, Spain

ARTICLE INFO

Article history:
Received 15 June 2022
Revised 18 July 2022
Accepted 3 September 2022
Available online 9 September 2022

Keywords:
General Linear Model
Linear Regression Model
Support Vector Regression
permutation tests
Magnetic Resonance Imaging
Random Field Theory

ABSTRACT

There remains an open question about the usefulness and the interpretation of machine learning (ML) approaches for discrimination of spatial patterns of brain images between samples or activation states. In the last few decades, these approaches have limited their operation to feature extraction and linear classification tasks for between-group inference. In this context, statistical inference is assessed by randomly permuting image labels or by the use of random effect models that consider between-subject variability. These multivariate ML-based statistical pipelines, whilst potentially more effective for detecting activations than hypotheses-driven methods, have lost their mathematical elegance, ease of interpretation, and spatial localization of the ubiquitous General linear Model (GLM). Recently, the estimation of the conventional GLM parameters has been demonstrated to be connected to an univariate classification task when the design matrix in the GLM is expressed as a binary indicator matrix. In this paper we explore the complete connection between the univariate GLM and ML-based *regressions*. To this purpose we derive a refined statistical test with the GLM based on the parameters obtained by a linear Support Vector Regression (SVR) in the *inverse* problem (SVR-iGLM). Subsequently, random field theory (RFT) is employed for assessing statistical significance following a conventional GLM benchmark. Experimental results demonstrate how parameter estimations derived from each model (mainly GLM and SVR) result in different experimental design estimates that are significantly related to the predefined functional task. Moreover, using real data from a multisite initiative the proposed ML-based inference demonstrates statistical power and the control of false positives, outperforming the regular GLM.

1. Introduction

Whole-brain analyses in neuroimaging (NI), comprising a large number of independent statistical tests, have been traditionally conducted with classical statistics, either hypothesis testing or Bayesian inference, and the univariate General Linear Model (GLM) [8]. These hypothesis-driven methods gained their popularity due to the ease of interpretation and function localization across experimental designs [9]. However, they usually rely on assumptions that are frequently violated; e.g. homogeneity, Gaussianity, etc. [30] and, consequentially, inflated type I error rates have become problematic and a key contributor to the replication crisis [8,23].

Furthermore, technological advances are increasing spatial and temporal resolutions as well as the range of available measurements of anatomy and physiology; a true exemplar of the *curse of dimensionality* [1]. In this context, analyses of contemporary large image repositories retain the difficulties associated with small sample sizes. One of them is the inflated false-positives observed across experiments as a consequence of the multiple comparison problem. This problem is partly solved by *over conservative* approaches, such as Bonferroni or Random Field Theory (RFT) corrections [7].

One promising solution for the aforementioned problems in NI is machine learning (ML) [11,12]. ML provides us with high-dimensional relationships between datasets that are empirically established based on data-driven methods [36]. Estimating depen-

* Corresponding author at: DaSCI Institute, University of Granada, Spain.
E-mail address: gorriz@ugr.es (J.M. Gorriz).

https://doi.org/10.1016/j.neucom.2022.09.001
0925-2312/

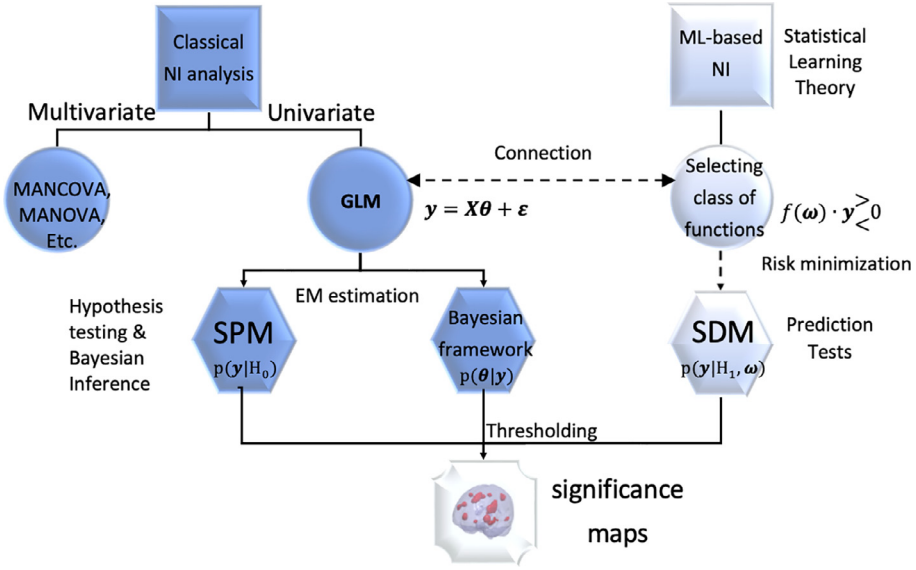


Fig. 1. NI analysis and methods. An univariate hypothesis-driven method based on the regression parameters obtained by ML (dash line) is proposed. Note the different meaning among the activation maps prior to probability thresholding, e.g. by means of a multiple comparison correction (RFT, Bonferroni, etc.) in classical approaches or selecting accuracies above random performance in SDM. In particular, SPM obtained by hypothesis testing indicates the likelihood of the observation given the null hypothesis H_0 (data likelihood) whilst the Bayesian framework considers the posterior probability map of the observation (effect likelihood). SDM and current multivariate ML approaches for image classification provide the (class-membership) probability of the observation given the alternative hypothesis H_1 (prevalence). MANCOVA: Multivariate analysis of covariance, MANOVA: Multivariate analysis of variance.

dencies in regression or classification tasks using statistical learning theory (SLT), unlike classical statistics, characterizes the actual relationships or effects with a limited dataset. NI in particular has embraced ML as a technology to deliver diagnostic and prognostic classification of neurological and psychiatric disorders [4,19,42,21,20]. Nevertheless, the mainstay of NI studies are observational and mechanistic, seeking to identify regional between-group differences in brain structure and function.

1.1. Related Works

Data-driven analysis methods based on ML [27,39,40,33,13] have demonstrated their ability for detecting activations in fMRI data, outperforming conventional hypothesis-driven approaches, i.e. the standard GLM inference based on random effect models (see Fig. 1). The core idea on these agnostic (model-free) approaches is to perform an accurate feature extraction based on a fixed-complexity MLE classifier, e.g. linear support vector machine (SVM), between predefined groups. They all share the same characteristic processing pipeline of a data-driven multivariate approach that enhances detection ability within a classification task.

Consequently, the Statistical Parametric Maps (SPM) derived from the GLM parameters are replaced by the spatial discriminance map (SDM) [1] obtained by ML [37,31]; or some other specific feature extracted at the training stage, e.g. distance to the separating hyperplane [39], replaces the explanatory variables in the GLM, previous to the statistical inference stage. These data-driven maps are then deployed in conventional pipelines (hybrid approaches) for statistical inference. However, some approaches depart from the p-value-based frequentist, Bayesian or permutation analyses, and introduce the concept of the probability of the *worst case* in neuroimaging [13].

Despite the popularity of ML as a solution for a wide range of complex problems, there remains an open question about its usefulness for statistical inference. Mainly, what is the statistical significance of its performance in a classification/regression task? Moreover, several works in NI [37,38,11] have revealed the shortcomings of predictive inference methods based on ML, e.g. cross-validation failure results in poor estimates. Another example is the inclusion of covariates and nuisance variables [2] in the analysis using ML. This procedure, that is massively used in the conventional NI analysis, is usually overlooked when using ML tools [19,42,21,20,25]. At most, the aforementioned hybrid approaches make use of a fitting process in the "ML space" in combination with the classical GLM, which includes these variables [17].

Efforts with ML around these issues are increasing with continuous output variables ([5], with remarks in [29]) rather than the more typical categorical classifications. However, the classification task is just a particular case of the regression problem with discrete labels, thus exploring general ML methods for linear regression, such as SV regression (SVR), is currently relevant in addition to its use as a simple extension of SVM [42].

Recently, a connection between both domains, that is, the standard GLM and classification tasks by ML, has been established [14] using *binary experimental design matrices*. The aim was to formalize the relationship that has been evidenced in several neuroimaging applications in the extant literature. In this paper, we show a complete and novel connection between the classical GLM, including random effect models, and the ML framework in the estimation of optimum regression parameters.

1.2. Aims

Overall, the main achievement of the article is the formalization of the use of ML in the context of hypothesis-driven neuroimaging statistical inference by deriving its formal relationship to the GLM; the method almost ubiquitously used for mass univariate testing of imaging datasets. Outputs from both approaches are then compared in a variety of commonly encountered applications. In doing so, we leverage one of the most commonly used packages for mass

[1] mainly based on accuracy or prevalence, that is, the proportion of a population who have a specific characteristic or effect versus control subjects

[2] they are usually referred to as confounds, i.e. variables that may blur the effect that is being sought.

univariate testing: SPM [8], and within it the methods for spatially extended statistics and p-value corrections for multiple comparisons. In summary, in this paper:

- We formalize the use of ML within the univariate GLM pipeline for statistical inference,
- We give an explanation on why hybrid approaches, massively used in this context, are actually working, e.g. the main regressor (experimental condition) in the design matrix could be replaced by a data-driven property [39].
- The connection is established by selecting the SVR in the inverse domain (SVR-iGLM) that allows us to include all the covariates (age, sex, etc.) in the analysis at once.

Subsequent analyses based on frequentist inference and permutation testing are carried out to calculate the level of significance in between-group testing of SPMs with a multiple comparisons correction.

2. Theory

2.1. The General Linear Model and its statistical framework

The GLM [9] is defined for a single observation level, e.g. in a between-subject comparison, as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y}$ is the $N \times 1$ observation vector with units of time, signal change, volume, etc., $\boldsymbol{\epsilon}$ is the $N \times 1$ vector of errors that is assumed to be Gaussian distributed, $\mathbf{X}$ is the $N \times M$ matrix containing the explanatory variables or constraints, and $\boldsymbol{\theta}$ is the $M \times 1$ vector of parameters explaining the observations $\mathbf{y}$. Note that: i) for a hierarchical observation model each level requires prior estimation at the previous levels; and ii) in terms of ML, $\mathbf{X}$ are the multidimensional labels or regressors acting on the observations $\mathbf{y}$. In the classic GLM, $\boldsymbol{\theta}$ is usually estimated by a maximum likelihood criterion based on the Gaussianity assumption and is given by:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^t\mathbf{C}_\epsilon^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{C}_\epsilon^{-1}\mathbf{y} \tag{2}$$

where $\mathbf{C}_\epsilon$ is the covariance matrix of errors. Inferences on this estimate[3] determine the components of $\boldsymbol{\theta}$, and the relationship between classical GLM and ML-based prevalence inferences can be obtained using a linear compound specified by a contrast weight vector $\mathbf{c}$, and writing a T statistic as:

$$T = \frac{\mathbf{c}^t\hat{\boldsymbol{\theta}}}{\sqrt{\mathbf{c}^t Cov(\hat{\boldsymbol{\theta}})\mathbf{c}}} \tag{3}$$

where $Cov(\hat{\boldsymbol{\theta}}) = \left(\mathbf{X}^t\mathbf{C}_\epsilon^{-1}\mathbf{X}\right)^{-1}$. This T statistic gives us the probability of observing the ML estimation under H_0, and when it is small enough, e.g. $p < 0.05$, the linear compound is considered significantly different from zero. As an example, given a set of two parameters in $\boldsymbol{\theta} = [\theta_1, \theta_2]^t$, if we select $\mathbf{c} = [1\ -1]$ we are assessing how large is the first parameter with respect to the second; i.e. the difference $\theta_1 - \theta_2$. Thus, if the T statistic suggests a small probability, the contrast is statistically significant with observations generated from different sources.

A similar procedure could be established based on a Bayesian estimation and inference to handle complex hierarchical observational models. This framework would be based on the Expectation Maximization (EM) algorithm for parameter estimation, along with known priors and *a priori* probability models, with the aim of evaluating the posterior probability (ppm). By thresholding the ppm, relationships between this and the frequentist approach can be established for both their similarities (statistical power) and differences (specificity) [9].

[3] Here, we refer to voxelwise inference since we use a threshold u to classify voxels i as "active" if $T_i \geqslant u$. Clusterwise inference uses a cluster-forming threshold to define contiguous suprathreshold regions [28].

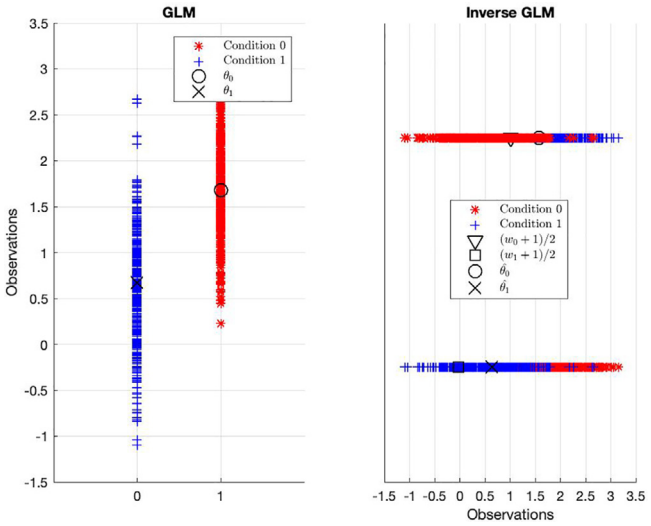


Fig. 2. Illustration example of the connection between GLM and iGLM. On the left observations in $\mathbf{y}$ and the best parameters of the GLM (θ_1, θ_2) given the experimental conditions (blue and red) are shown. On the right, given the set of observations, we show the parameters (ω_1, ω_2) that better explain the classes (red and blue) and the equivalent GLM parameters $(\hat{\theta}_1, \hat{\theta}_2)$.

2.2. Machine learning and the inverse GLM

The GLM can be interpreted as the inverse problem of regressing the observations onto the conditions (see Fig. 2). Instead of assuming the model in Eq. 1, the inverse GLM is defined as:

$$\mathbf{X} = \mathbf{y}\boldsymbol{\omega} + \hat{\boldsymbol{\epsilon}} \tag{4}$$

where $\boldsymbol{\omega}$ is a set of $(1 \times M)$ parameters that best explains the design matrix given the observations, and $\hat{\boldsymbol{\epsilon}}$ is noise with unknown pdf. We can readily see that:

$$(\mathbf{X} - \hat{\boldsymbol{\epsilon}})\boldsymbol{\omega}^t(\boldsymbol{\omega}\boldsymbol{\omega}^t)^{-1} = \mathbf{y} \tag{5}$$

where we assume that the inverse of the norm $(\boldsymbol{\omega}\boldsymbol{\omega}^t)^{-1}$ exists. After some manipulations we finally get:

$$\mathbf{y} = \mathbf{X}\tilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\epsilon}} \tag{6}$$

where we define:

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\omega}^t(\boldsymbol{\omega}\boldsymbol{\omega}^t)^{-1}; \quad \tilde{\boldsymbol{\epsilon}} = -\hat{\boldsymbol{\epsilon}}\boldsymbol{\omega}^t(\boldsymbol{\omega}\boldsymbol{\omega}^t)^{-1} \tag{7}$$

Therefore, solving the multiple regression problem in equation 4, e.g. using a ML approach, is equivalent to estimating the parameters of the GLM. If we carefully examine Eq. 7, each column c of the design matrix ($c = 1, \ldots, M$ including covariates and nuisance variables) can be described by:

$$\begin{bmatrix} x_{1,c} \\ \vdots \\ x_{N,c} \end{bmatrix} = \mathbf{y}\omega_c + \hat{\boldsymbol{\epsilon}}_c \tag{8}$$

The set of parameters $\boldsymbol{\omega}$ could be determined by a classical approach such as Parametric Empirical Bayes (PEB), assuming a Gaussian model for the noise $\hat{\boldsymbol{\epsilon}}$. However, we prefer to use the ML approach based on the regularized risk minimization and linear

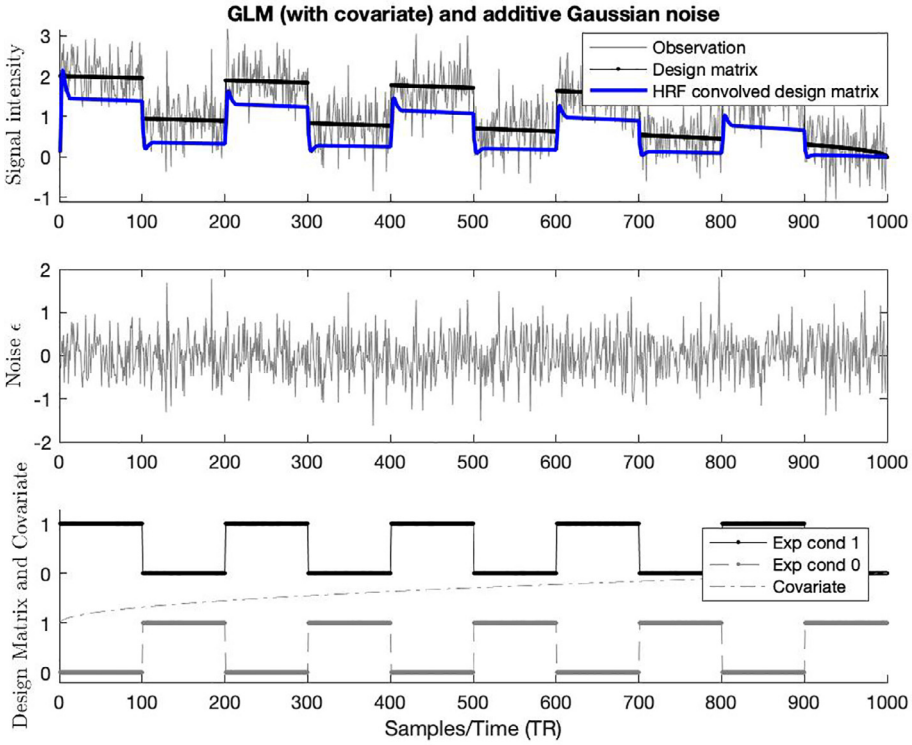


Fig. 3. Simulated data with noisy observations example with $\boldsymbol{\theta} = [1 \quad 0 \quad 1]^t$, CNR= 1 and $N = 1000$. We evaluated the GLM in Eq. 1 given the experimental conditions (main regressors) with a covariate and the Gaussian noise assumption.

regressors. For the reader's convenience, further mathematical discussion is available on 7.

Thus, given the set of observations $\mathbf{y}$ and the experimental conditions and covariates in $\mathbf{X}$, we regress M independent linear equations as:

$$\mathbf{X} = \mathbf{y}\boldsymbol{w} + \mathbf{B} = \hat{\mathbf{y}}\hat{\boldsymbol{w}} \tag{9}$$

where $\hat{\boldsymbol{w}} = \begin{bmatrix} \boldsymbol{\omega} \\ \mathbf{B} \end{bmatrix}$, the $N \times M$ matrix $\mathbf{B} = [\mathbf{b}^t \ldots \mathbf{b}^t]^t$ and $\mathbf{b}$ is the $1 \times M$ vector of biases. In view of Eq. 8, we realistically assume under this approach that the bias does not depend on the experimental condition or covariate realization.

Once the set of parameters and biases $\{\mathbf{w}, \mathbf{b}\}$ are estimated by a suitable procedure (see Appendix), such as least squares (LS) or SVR, we can calculate the observation $\mathbf{y}$ by simply inverting Eq. 9 as:

$$\mathbf{y}_{est} = (\mathbf{X} - \mathbf{B})\boldsymbol{w}^t(\boldsymbol{w}\boldsymbol{w}^t)^{-1} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\theta}} \tag{10}$$

where $\tilde{\boldsymbol{\theta}}$ is the ML estimated vector of parameters that combines the weight of multiple predictors to explain the observation, and $\tilde{\mathbf{X}} = (\mathbf{X} - \mathbf{B})$ is the adjusted design matrix. Instabilities that could arise from the scalar inversion in Eq. 10, that depend on the selected ML algorithm, are easily solved by bounding its value between -1 and 1.

It is worth mentioning here that our model does not use any data reduction techniques, e.g. principal component analysis (PCA), preserving the function localization of the univariate GLM. However, a similar description of the method can be given in terms of the difference of signed distances by replacing the main reference function in the GLM by $\frac{\mathbf{y}\boldsymbol{w}}{||\boldsymbol{w}||^2}$. In this way we a give an explanation of the multivariate approach proposed in [39] in terms of the solution in the iGLM.

2.3. Support Vector Regression

In general, the core idea of SVR [32] is to do (non-) linear regression in a feature space $\mathscr{F}$:

$$f(\mathbf{y}) = \boldsymbol{w} \cdot \Phi(\mathbf{y}) + b, \quad \Phi : \mathbb{R}^n \to \mathscr{F}, \quad \boldsymbol{w} \in \mathscr{F} \tag{11}$$

where $\cdot$denotes dot product and, in the case of the linear regression, Φ is simply the identity function. We determine $\boldsymbol{w}$ from the data by minimizing the sum of the empirical risk, e.g. ϵ-insensitive loss function [2], and a complexity term proportional to its norm:

$$R_{reg}(f) = \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N}|x_i - f(\mathbf{y}_i)|_\epsilon \tag{12}$$

This minimization can be transformed into a uniquely solvable quadratic programming problem [34] that provides the vector of parameters $\boldsymbol{w}$ in terms of the samples or support vectors:

$$\boldsymbol{w} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\Phi(\mathbf{y}_i) \tag{13}$$

where α_i, α_i^* are Lagrange multipliers; that is, the solutions of the quadratic programming problem. Finally, the bias term b can be computed by determining the prediction error on the margin $\delta_i = f(\mathbf{y}_i) - x_i = \epsilon \text{sign}(\alpha_i - \alpha_i^*)$ and taking the average of differences as $b = < f(\mathbf{y}_i) - \boldsymbol{w} \cdot \Phi(\mathbf{y}_i) >$.

2.4. Model Equivalence

After performing the regression in the iGLM domain by SVR and assuming that the explanatory matrix $\mathbf{X}$ contains two experimental conditions, i.e. indicator variables that refer to class membership, and a set of continuous covariates or nuisance variables, then we have $\mathbf{X} = [\mathbf{X}_b\mathbf{X}_c]$ and the observations can be estimated, following Eq. 10, by:

$$\mathbf{y}_{est} = \mathbf{X}_b\tilde{\theta}_b + \mathbf{X}_c\tilde{\theta}_c - \mathbf{B}\tilde{\theta} \quad (14)$$

where $\tilde{\theta} = [\tilde{\theta}_b^t \tilde{\theta}_c^t]^t$. Eq. 14 can be used to approximate a set of parameters that best explain the observation vector with maximum and minimum influence of covariates and nuisance variables ranged in the interval $[0, 1]$ as (see an example in Fig. 2)

$$\tilde{\theta}_{est} = \tilde{\theta}_b + [0, 1] \cdot \tilde{\theta}_c - \mathbf{B}\tilde{\theta} \quad (15)$$

3. Materials and Methods

3.1. Synthetic data

We generated synthetic data with the aim of modeling different scenarios in an fMRI time-series analysis with inserted activations using a block-design (baseline and task) paradigm [39]. For this purpose, we simulated one dimensional observation vectors $\mathbf{y}$ (Eq. 1) with different contrast to noise ratios (CNR: $0.25, 0.5, 0.75, 1$) and sample sizes (N:100–1000).

The design matrix $\mathbf{X}$ was the canonical HRF convolved boxcar function for fMRI simulated data. This matrix contained an exponentially decaying function $f(t) = (1 - \frac{t}{N})^{0.5}$ representing habituation during the fMRI task, or a covariate to simulate the effect of age when spatially testing brain activations, and as shown in Fig. 3. An N-dimensional Gaussian noise vector $\mathbf{v}$ was randomly drawn with zero mean and controlled variance (σ_X^2/CNR). Finally, a vector of observations was constructed by adding the noise to the design matrix with ideal parameters $\theta = [1 \quad 0 \quad cv]^t$, where cv is a constant that modulates the exponentially decreasing covariate in the time-series.

The use of synthetic data allowed us to estimated the noise covariance matrix by averaging a set of 100 noise realizations as an ideal comparison with the GLM benchmark. This is useful to simulate different performances of the ReML estimation that depends upon the observation vector $\mathbf{y}$ and some parametrization of the covariance components. The latter is the procedure in SPM12 that obtains the noise covariance matrix $\mathbf{C}_\epsilon$ in Eq. 3.

3.2. A structural MRI dataset including covariates: ADNI

Data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI database contains T1-weighted structural MRI scans acquired at 1.5 T and 3.0 T from patients with Alzheimer's disease (AD), Mild Cognitive Impairment (MCI), and cognitively normal controls (NC) at multiple time points. Here we only included structural MRI collected at 1.5T. The original database contains more than 1000 T1-weighted MRI images in total, although for this study only the first MRI examination of each participant was included, resulting in 417 structural images in the sample. Demographic data is summarized in Table 1.

The dataset was processed using the standard neuroimaging methods and protocols implemented in the SPM software (www.fil.ion.ucl.ac.uk/spm/), including registration to MNI space by spatial normalization and segmentation, to generate maps of grey matter (GM) volume [8].

Following the recommendation of the National Institute on Aging and the Alzheimer's Association for the use of imaging biomarkers [22], we considered the group comparison NC vs. AD for establishing a clear framework for comparing statistical paradigms. Age, sex and intracraneal volume (ICV) were included as covariates in between-group modelling as representing the most common set in the extant literature [17]. All the covariates were standardized with zero mean and standard deviation equal to one.

We selected two regions of interest using the 116-area automated anatomical labeling template [35]: one relevant area in AD, the bilateral hippocampus (denoted $Hippocampus_L$ and $Hippocampus_R$ for left and right hemispheres, respectively) with 2559 voxels; and the cerebellum ($Cerebelum9_L$ and $Cerebelum9_R$ of the atlas), which is considered not relevant to AD pathology, with 3027 voxels.

3.3. Statistical Analyses

To assess the performance of the methods presented in this paper, we used the benchmark proposed in SPM12 for statistical inference; that is, the GLM and RFT with FWE correction and $p = 0.05$ for second-level statistical inference.

First, we estimated the best set of parameters by regressing the design matrix, or the observations, using several configurations: i) the ideal ML method with synthetic data, where $\mathbf{C}_\epsilon$ is estimated from noise realizations, ii) Restricted (Re) ML, iii) LS, and iv) SVR. Then, we connected the ML-based estimates (LS and SVR) with the corresponding set of parameters in the GLM space, as shown in Eq. 15. To compare the inferences of each regression method, we assumed the same noise model and evaluated the T statistic in Eq. 3 on the set of parameters. Finally, we thresholded the resulting T-maps, e.g. derived from the SVR-iGLM method, by a detection threshold based on RFT.

A permutation analysis was also adopted to provide an alternative statistical inference based on a non-parametric approach [27]. By randomly permuting the experimental conditions 1000 times, we calculated the non-parametric T statistic based on the contrast of the estimates, thus avoiding the estimation of the denominator of Eq. 3).The probability of observed contrast was then calculated relative to the distribution of permuted contrasts representing the null distribution. If this value was less than a selected threshold, e.g. $p = 0.05$, then we rejected the null hypothesis.

4. Experimental results

4.1. Synthetic data: Estimating functional tasks

With the synthetic data we estimated the parameters by all the methods followed by permutation inference. We additionally plotted the non-normalized statistic (contrast) and evaluated the set of parameters, as a classifier, in the label domain, $\boldsymbol{w} \cdot \mathbf{y} \gtrless 0$.

The observations were artificially drawn from the same Gaussian sample distribution (Fig. 3) with varying sample size (i.e. time points) and CNR. We regressed both explanatory variables and observations to obtain the experimental parameters for each model $\boldsymbol{\theta}, \mathbf{w}$ given the simulated covariate. All these estimations were employed to calculate the regressed observed variables using Eqs. 1 and 10, given the explanatory matrix and the estimated parameters, and illustrated in Figs. 4 and 5. Only for extremely

Table 1
Demographic details of the MRI ADNI dataset, with group means and their standard deviations

Status	N	Age	Sex (M/F)	ICV($\times 10^5$)	MMSE
NC	229	75.9±5.0	119/110	15.3± 1.6	29.1±1.0
AD	188	75.3±7.5	99/89	15.5± 1.8	23.2±2.2

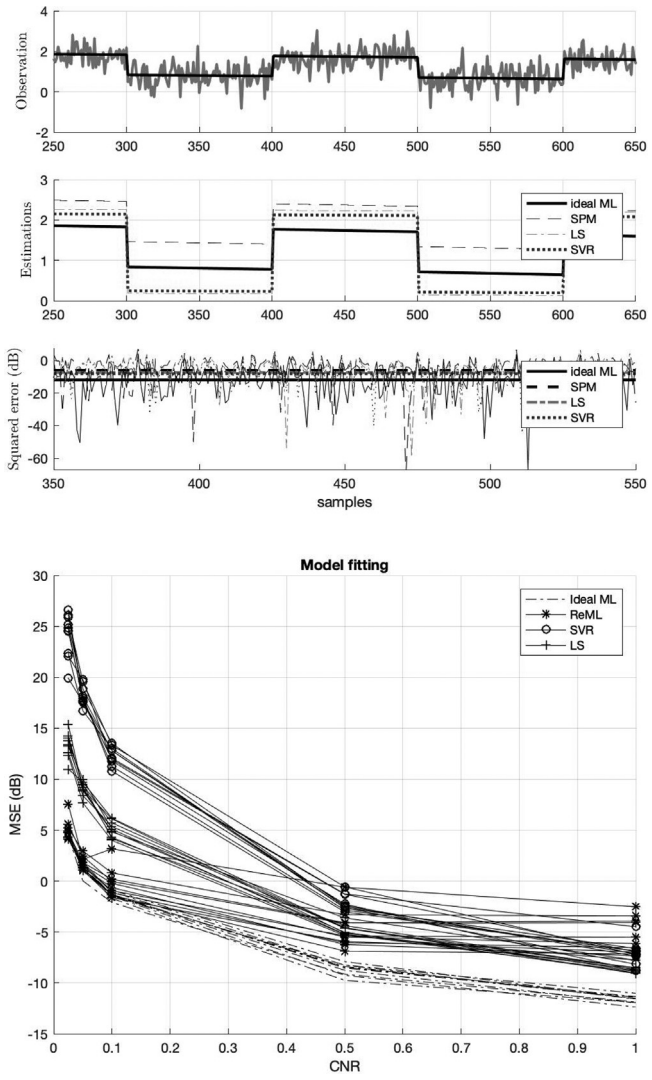


Fig. 4. Top: Example estimation with $N = 1000$, CNR= 1. Bottom: Averaged mean squared error (MSE) with different activation CNR and sample sizes (CNR: $0.25, 0.5, 0.75, 1$ and N:100,200,...1000).

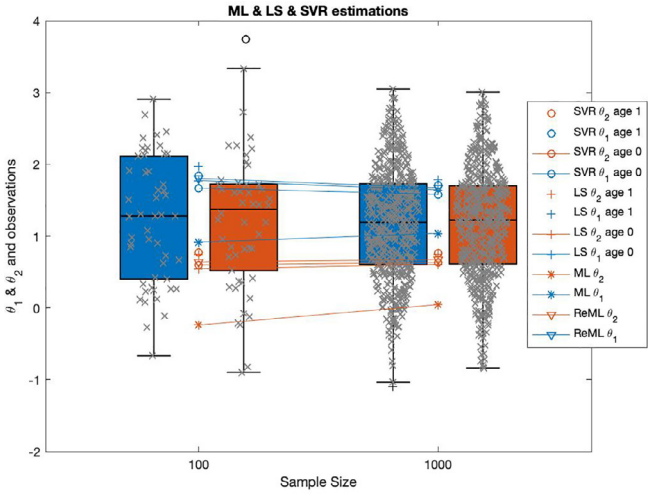


Fig. 5. Distribution of observations (in grey) separated by condition and estimations $\boldsymbol{\theta}$ (symbols in blue and red) for all the analysed methods including the covariate effect as shown in Eq. 15.

noisy observations (CNR<0.5) and considerably high sample sizes did the ReML (and the ideal ML) outperform ML approaches in MSE, even under the Gaussian assumption (Fig. 4). However, in the label domain, the set of parameters derived from the ML approaches provided the highest classification accuracy in the associated classification task, recapitulating prior findings [13]. In summary, the ML parameter fitting in the GLM does not imply an accurate regression in the iGLM.

The p-value of the estimated contrast (difference between parameters $\theta_1 - \theta_2$) was less than $p = 0.05$ for all methods. The probability of observation was $p = 1/1001$ in all cases when testing the null hypothesis. Thus, no false positive were detected during the simulated task, although the SVR estimation based inference provided results closest to the nominal false positive rate, whilst the remainder were over-conservative (Fig. 6).

4.2. Empirical data: a case-control design with the ADNI Dataset

In this section we show the inference derived from the two methodologies in each domain. We regressed on the observations and on the design matrix including covariates for age, sex and ICV (Table 1). Then, we constructed the spatially extended statistical processes, generating maps of significance, using GM estimated from the MRI ADNI dataset[13]. We compared the SPM (a two-sample T-statistic similar to Eq. 3), where significance is first individually assessed at each voxel, and then on clusters with $p = 0.05$ FWE corrected based on RFT. The univariate test based on the inverse GLM in Section 2.2 was also conducted.

In short, we will assess the detection ability (true positive detection) and the control of type I error (false positives) of the methods in a semi-controlled environment. To this purpose we selected two baseline regions in late AD, the hippocampus, that is primarily and generally affected in early AD; and the cerebellum, that remains unaffected. Finally, we will explore the detection ability within an Omnibus test (whole-brain analysis) using the multiple comparison correction provided by RFT and the optimum threshold derived from the baseline regions.

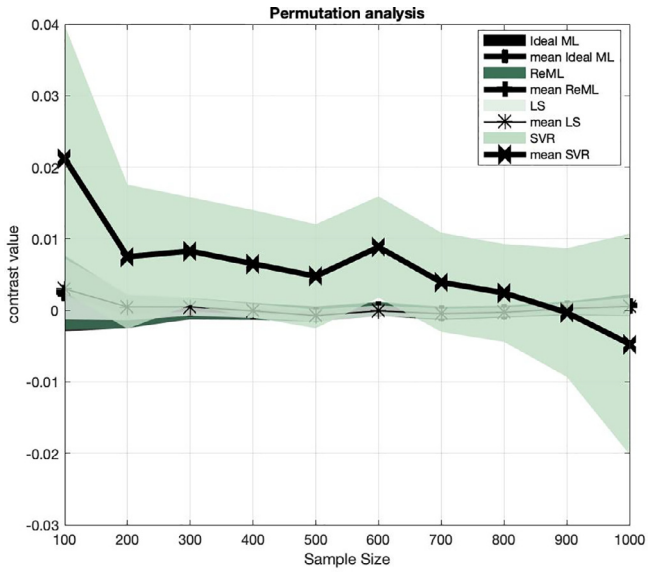


Fig. 6. Shadeplots (average and standard deviation) showing contrast estimations $(\theta_1 - \theta_2)$ for each method under the null hypothesis. They were calculated from 1000 random permutations and CNR= 1.

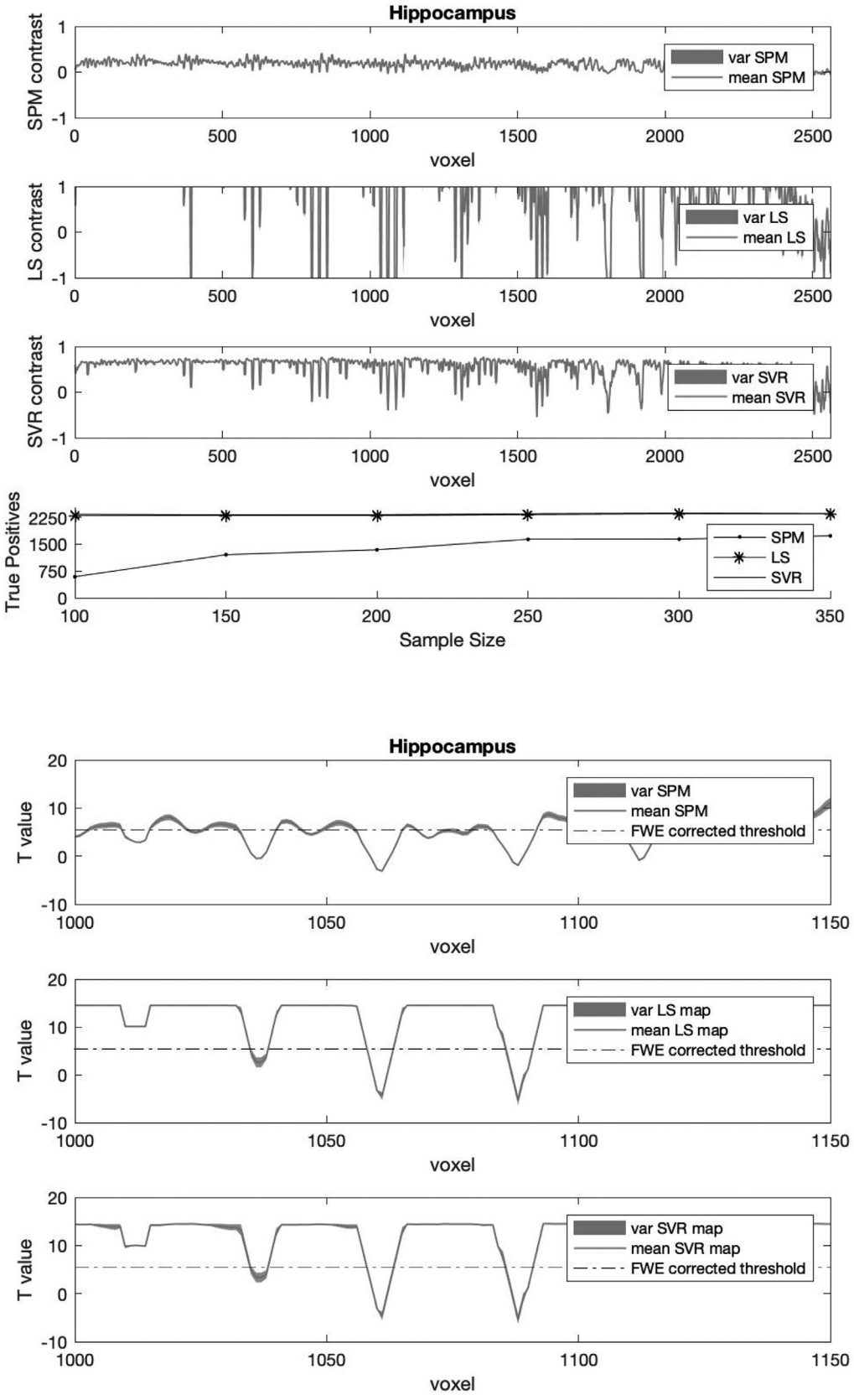


Fig. 7. Contrasts and T-statistics in the hippocampus with RFT FWE-corrected threshold at $p = 0.05$ and varying sample sizes. Note that the ML approaches are upper bounded for visualization purposes and the threshold plotted $T_{thres} = 5.2758$ is that obtained for $N = 350$.

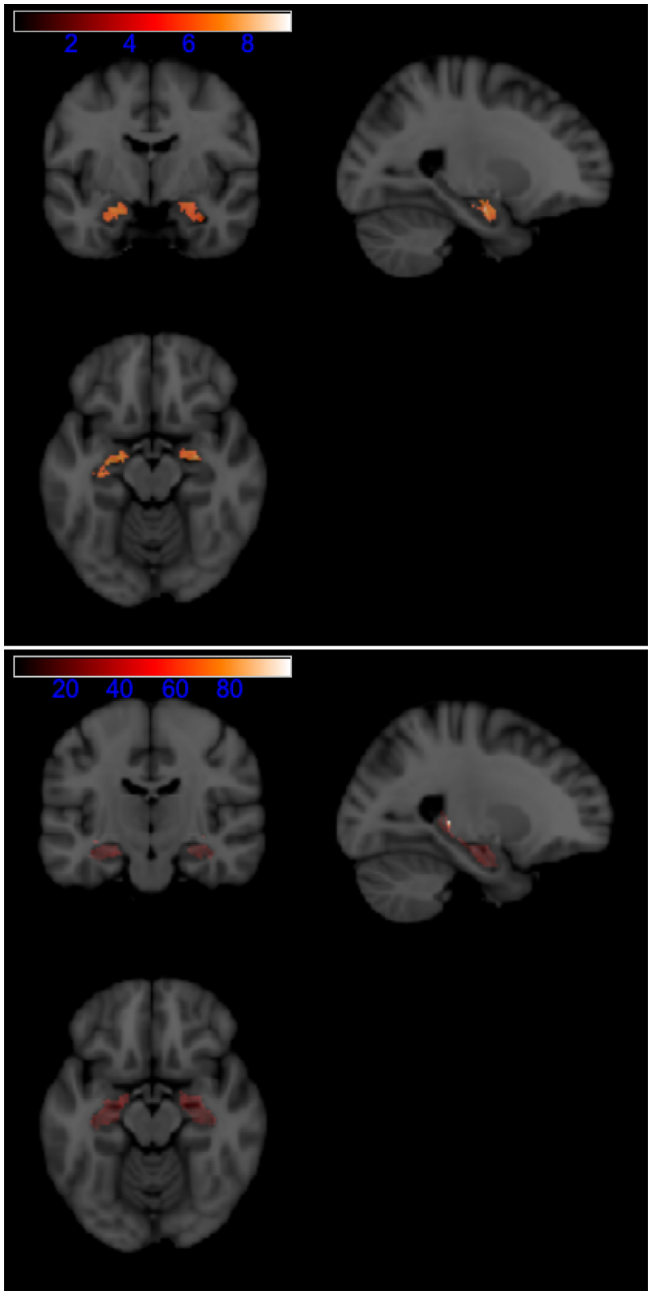


Fig. 8. Group-level statistical analysis results in the hippocampus of the ADNI MRI data. The univariate analysis was conducted using the standard GLM (top) and SVR-RFT approach (bottom). Note T-maps scales for comparison. The T-maps were thresholded at $u > 5.5704$ (equivalent to an uncorrected $p < 1.1802e^{-7}$) with sample size $N = 100$.

4.2.1. Hippocampal analysis to assess statistical power

First, we performed the univariate statistical analysis on GM in the left and right hippocamppal ROIs, which are typically the earliest sites of atrophy associated with AD [18]. Thus, the group comparison (advanced stage of the disease) AD vs. NC can be considered as a true positive (TP) region. We compared the GLM contrast with that obtained by the ML-based approach using the same parametrization of the noise covariance in Eq. 3, the standard approach in SPM12. SPMs were collected voxelwise and the inference method, based on a two-sample T-test with the linear compound $c = [1, -1, 0, 0, 0]$, was thresholded by means of the RFT FWE rate correction at $p = 0.05$, as previously.

Figs. 7 and 8 show the contrasts (shadeplots with increasing sample size) derived after model estimation, and the T-statistics obtained by the use of the same normalization term in Eq. 3. In this simulation, we employed the T-scores obtained by the standard SPM approach to bound the values obtained by our method. We readily see the higher statistical power ($T > T_{thres} = 5.2758$ for $N = 350$) of the ML methods and the strong dependence of the standard SPM on the sample size. Both methods converged as sample size increased. The LS approach was more unstable with fewer significant voxels than the SVR-based procedure ($N = 100, 2314 \, vs 2341; N = 150, 2300 \, vs. 2324; N = 200, 2300 \, vs. 2329$; $N = 250, 2334 \, vs. 2355; N = 300, 2355 \, vs. 2372$); $N = 350, 2344 \, vs. 2360$).

4.2.2. Type I error control in a putatively-null cerebellar region

We repeated the experiments in cerebellar ROIs (Cerebellum 9 left and right [35]) to evaluate the ability of the ML-based inference methods to control false positive rates using the same inference strategy used for thresholding the T-maps; i.e. FWE correction based on RFT in the standard GLM. Fig. 9 illustrates the over-conservative voxelwise inference with FWE correction based on RFT, although the proposed methods based on ML provided a significant number of tests closer to the expected value, i.e. $p = 0.05$. In other words, by chance the number of FPs should be around 5% of the total number of voxels. Parametric voxelwise inference is known to be valid but conservative, often falling below the nominal rate [6].

4.2.3. Global analysis in contrast images with limited sample size

Fig. 10 shows the vector of parameters in the selected axial slice Z = 47 derived from ReML and SVR in the analysis of the whole brain. The ML approach implemented in SPM finds very few significant relationships between the covariate effects (age, sex and ICV) and the observations, compared with the experimental conditions. On the contrary, the SVR yields stronger connections, mainly for the sex covariate. From these parameter images, contrast images are then derived and, subsequently, inference on their size relative to the estimate of their standard error is made in large target regions, as depicted in Figs. 7 and 9.

Selecting the aforementioned target regions and plotting the t-score histograms we can approximately compute the optimum threshold at a given level of significance α using the Neyman-Pearson lemma, e.g. at $\alpha = 0.05$, SVR: $t_{thresh} = 10.03$ and SPM: $T_{thresh} = 2.03$. By comparing the latter value with the one used in previous sections (RFT correction) we can readily see the over-conservative nature of the standard voxelwise inference (Fig. 11). Extending these thresholds to the whole volume we obtain the activation maps for the SVR and standard SPM approaches as shown in Fig. 12. (See Fig. 13).

5. Discussion

In the context of neuroimaging statistical inference, there is an increasing trend to incorporate exploratory methods into well-established GLM-based data analysis. Not only data preprocessing techniques, such as independent or principal component analysis (PCA) [26], but also multivariate ML approaches have been widely used in classification tasks to replace the predefined design matrix in the regular GLM pipeline [39] or to provide novel statistical maps of prevalence [27,13]. Existing multivariate approaches based on ML, including stages for smoothing or orthogonal decompositions, such as PCA [27,39,40,13], have provided promising results in (f) MRI-data analysis where there is a trade-off between sensitivity and computational cost.

In this paper we present a novel univariate methodology for (f) MRI image analysis based on the optimum performance in limited

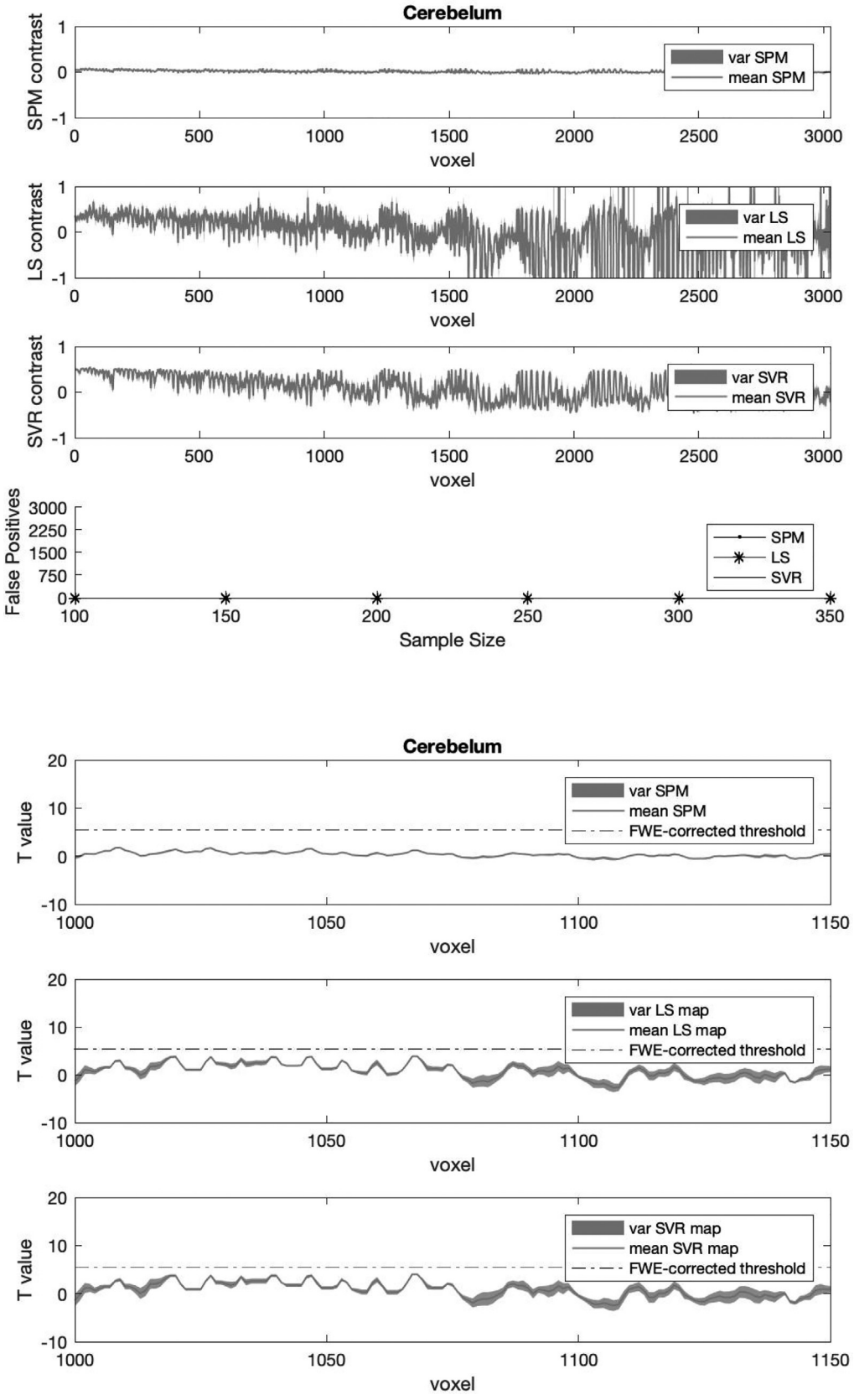


Fig. 9. Contrasts and T-statistics for the cerebellum with FWE-corrected threshold $t_{thres} = 5.2758$ at $p = 0.05$.

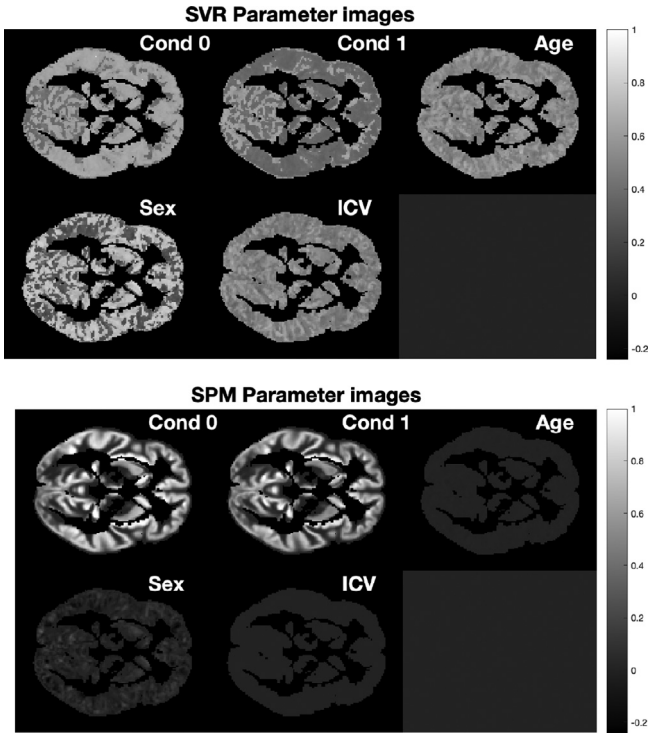


Fig. 10. Parameter images (axial slice Z = 47) derived from ReML (SPM) and SVR estimations for $N = 100$. Note that Cond 0 and Cond 1 denote NC and AD, respectively.

sample sizes of SVR over classical approaches. It departs from the use of the typical GLM frameworks based on classical estimations, or hybrid approaches that combine GLM with ML, and proposes a complete voxelwise inference method based on SVR. This methodology was demonstrated to provide optimum generalization ability in the case of regression estimation by the construction of regularizing functions in ill-posed problems [36].

The inverse GLM-based method (SVR-iGLM) is an univariate approach that preserves the function localization and better interpretation of classical approaches and incorporates the aforementioned advantages. We explored its performance and compared it with the regular GLM inference. As shown in previous approaches we demonstrated its higher detection ability where its regional sensitivity was controlled with a common p-value correction. In the whole volume analysis we employed the optimum Neyman-Pearson threshold obtained from baseline regions.

5.1. Covarying in neuroimaging

Although covarying for data variables in neuroimaging is routine when adjusting the model for confounding or nuisance factors [17], data-driven approaches have limited their operation to extracting whole brain SDMs as a description of the different responses to experimental conditions. Nevertheless, SDM extraction is not possible for multiple conditions/states in fMRI data due to the strictly use of data-driven properties [39] in a classification task.

Additional covariates could be partly avoided by the selection of balanced groups, sometimes a complex task, with the aim of reducing the impact of predictors that are not relevant to the research question at hand [24]. Moreover, group-balanced selection procedures further decrease sample sizes. The proposed ML approach, linked to the regular GLM, processes all the covariates at once, and combines their effect when estimating the observation or response variable. This effect is quantitatively determined by the definition of the equivalent vector of parameters $\tilde{\theta}$.

To our knowledge, this is the first work that formally extends the use of ML from classification to regression permitting the introduction of covariates, where they naturally weight the estimated parameters for statistical inference. Most importantly, the equations given in the theoretical section clearly and simply demon-

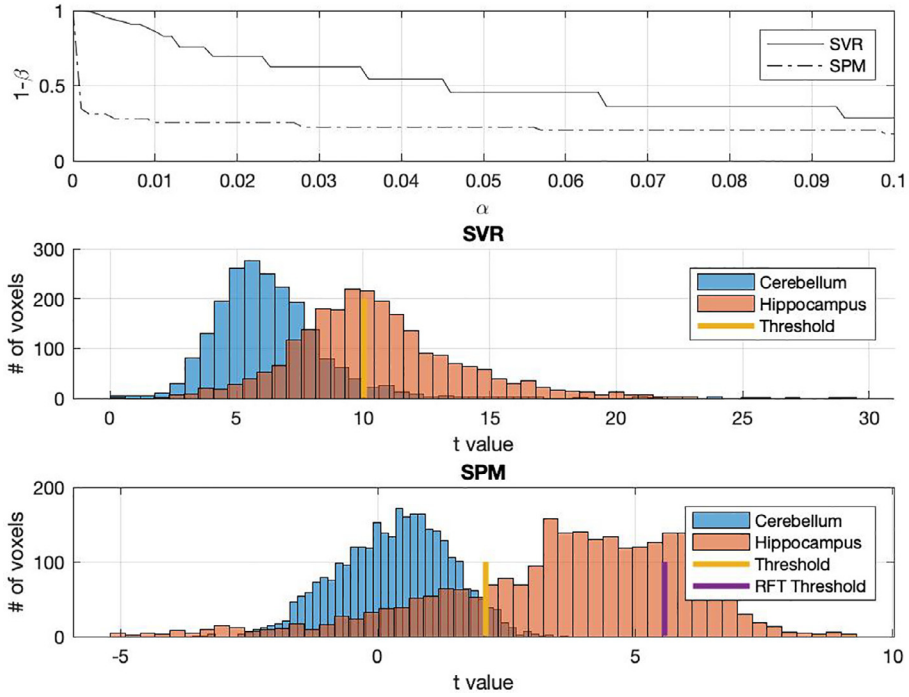
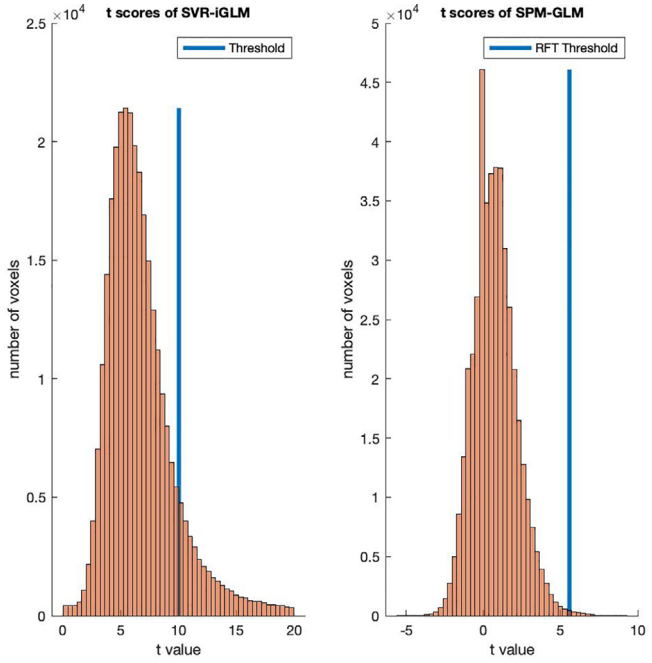


Fig. 11. Power analysis using the optimum Neyman-Pearson threshold. We selected a level of significance of $\alpha = 0.05$ that resulted in a threshold less conservative than the RFT correction.



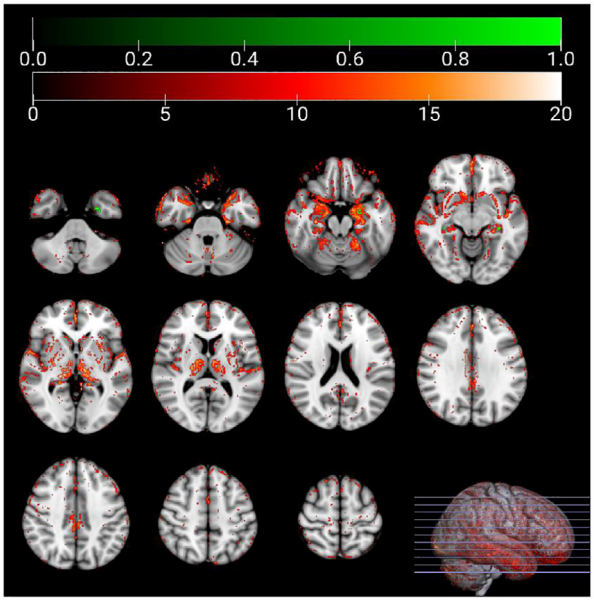


Fig. 12. Global T-scores and activation maps of the whole image volume, obtained by the Neyman-Pearson threshold at $\alpha = 0.05$ and the standard SPM FWE $p = 0.05$ correction.

strate how these weighted parameters obtained by the ML are connected to the parameters from GLM, and thus how statistical inference can be introduced. A main conclusion is that inference in the GLM is equivalent to a set of independent multiple regressions in the inverse space.

5.2. Regression techniques in ML

We tested the SVR estimation within a permutation inference and the regular GLM framework with FWE correction based on

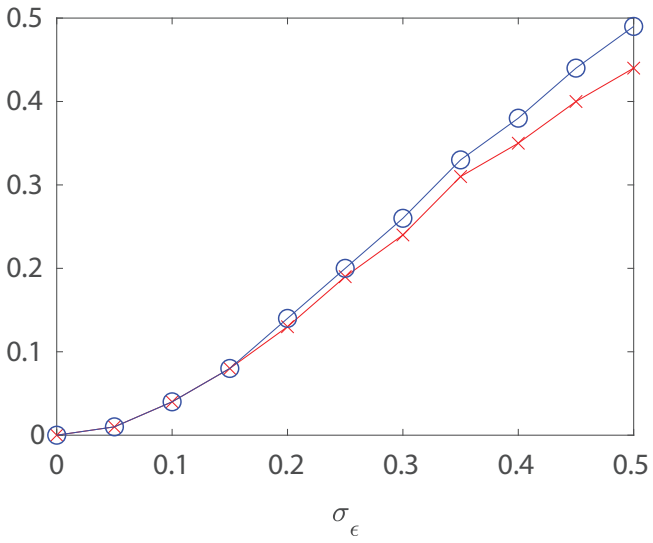


Fig. 13. Absolute error committed in the approximation of the optimal value $w = 1/(\theta_1 - \theta_2)$, as a function of σ_ε, by using the L1-approach (in red) and the MSE approach (in blue).

RFT. With synthetic data, the SVR approach resulted in a similar mean squared error as the GLM based on ReML estimations for CNR$>$ 0.5, whilst with real data the ML contrasts showed greater variability between conditions. The univariate SVR approach had improved discrimination compared to the GLM based inference, aligning with previous results in the extant literature that employed multivariate ML approaches to classify experimental conditions under the GLM. The proposed approach provides: i) inference results closer to the nominal values, i.e. $p = 0.05$, although over-conservative behaviour prevails; and ii) robust performance with increasing sample size.

Several limitations are known using LS linear regression models (LRM) for estimating the vector of parameters [13,15] due to instabilities of the algorithm, including effects of outliers, heteroskedasticity, etc. However, as shown in the experimental section, there was a strong correspondence between GLM and LS-LRM in the synthetic data analysis. Unfortunately, univariate ML approaches increase the computational burden of the analysis since they perform multiple LRMs independently at each voxel in the image. This is a drawback for fMRI analyses where the number of scans and voxels is inflated relative to structural images. Nevertheless, this issue is usually solved by the use of spatial dimension reduction and data representation techniques, such as Partial Least Squares (PLS) or PCA [10,40] in multivariate frameworks.

Another common caveat of the univariate ML approaches in this scenario is overfitting [32], a well-known problem in pattern recognition. Here, the regressor, once fitted, conforms to the specific samples in the training data with the consequence that its generalization ability or sensitivity to unseen data is reduced. The selected SVR employs the concept of maximizing the margin with a regularized term and a low-complexity model to reduce the risk of overfitting. Nevertheless, the fitting process could be affected by misregistration [30] or spatially incoherent activations [40]. The use of smoothing kernels in univariate approaches, such as SVR-iGLM, could potentially reduce this effect by taking spatial correlations into account, but could conversely worsen sensitivity in incoherent signals in multivariate approaches.

5.3. Data-driven approaches vs models

Most hypothesis-driven neuroimaging analyses depend on specified models when proposing a statistic and fitting parameters

of the GLM. This is a major advantage when data (nature's mechanism) is drawn assuming a Gaussian distribution (model's mechanism), and the inference drawn from the experiment may be misleading [3]. In the synthetic example, we assumed a known covariance matrix and a set of noise realizations for the formulation of the GLM. This experimental setup is imperfect in neuroimaging applications and the statistics following on from the best guess can fluctuate around the ideal value [13]. Frequentist and Bayesian analyses are strongly grounded on model selection and parameter fitting stages where, in complex scenarios with a limited sample size, heuristics are the common solution [41].

Finally, limited samples sizes and the selection/estimation of any specific model remains an issue in neuroimaging. This problem potentially deteriorates if the model, and the interaction between model parameters, becomes too complex for an accurate posterior probability estimation or a feasible numerical computation of Bayes rule. Given the relationship between the GLM and ML-based regression, we propose a conventional statistical inference based on the optimum estimations derived from ML from limited amounts of data [36,16]. The SVR-iGLM is not limited to linear regression since the main regressor in the design matrix could be replaced by another non-linear function; a common approach in fMRI data modelling. Moreover, SVR-iGLM could be incorporated in novel statistical tests, e.g. the P-tests [29], to highlight between-group differences in patterns of imaging-derived measures.

6. Conclusions

We addressed the open question on the usefulness and the interpretation of ML approaches for obtaining spatial patterns from brain imaging data that can discriminate between samples or brain states. We followed the natural path for using the ML framework by regressing observations onto conditions in a supervised learning manner, including a set of covariates. We thus explored the complete connection between the univariate GLM and ML *regressions* by deriving a refined statistical test based on the parameters obtained by a SVR in the *inverse* problem. Experimental results demonstrated how parameter estimations derived from ML and common statistical inference procedures provide a novel technique with good statistical power and control of false positives to obtain activation maps.

CRediT authorship contribution statement

J.M. Gorriz: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Project administration, Data curation. **R. Martín-Clemente:** Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft, Project administration. **C.G. Puntonet:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - review & editing. **A. Ortiz:** Methodology, Validation, Formal analysis, Investigation. **J. Ramírez:** Methodology, Validation, Formal analysis, Investigation, Methodology, Validation, Formal analysis, Investigation. **J. Suckling:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the MCIN/AEI/10.13039/501100011033/ and FEDER "Una manera de hacer Europa" under the RTI2018-098913-B100 project, by the Consejería de Economía, Innovación, Ciencia y Empleo (Junta de Andalucía) and FEDER under CV20-45250, A-TIC-080-UGR18, B-TIC-586-UGR20 and P20-00525 projects, as well as the research project ACACIA (Ref. US-1264994 US/JUNTA/FEDER, UE) awarded by FEDER and Junta de Andalucía (Consejería de Economía, Conocimiento, Empresas y Universidad).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.neucom.2022.09.001.

References

[1] R.E. Bellman, et al. Dynamic Programming. Courier Dover Publications. (2003) ISBN 978-0-486-42809-3.
[2] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (2) (1998) 121–167.
[3] L. Breiman, Statistical Modeling: The Two Cultures, Statistical Science 16 (3) (2001) 199–231.
[4] D. Bzdok, Classical Statistics and Statistical Learning in Imaging Neuroscience, Front. Neurosci. 06 (October 2017).
[5] J.R. Cohen, et al. Decoding continuous behavioral variables from neuroimaging data. Front. Neurosci. 5. 2011.
[6] A. Eklund, et al. Cluster failure: Inflated false positives for fMRI. Proceedings of the National Academy of Sciences Jul 2016, 113 (28) 7900–7905.
[7] R.S.J. Frackowiak, et al. Human Brain Function (Second Edition). Chap. 44. Introduction to Random Field Theory. 867–879, ISBN 978-0-12-264841-0 Academic Press, 2004.
[8] K.J. Friston et al., Statistical Parametric Maps in functional imaging: A general linear approach Hum, Brain Mapp. 2 (1995) 189–210.
[9] K.J. Friston et al., Classical and Bayesian inference in neuroimaging: theory, NeuroImage 16 (2) (2002) 465–483.
[10] J.M. Górriz, et al. A Machine Learning Approach to Reveal the NeuroPhenotypes of Autisms. International journal of neural systems, 1850058. 2019.
[11] J.M. Górriz et al., On the computation of distribution-free performance bounds: Application to small sample sizes in neuroimaging, Pattern Recognition 93 (2019) 1–13.
[12] J.M. Gorriz et al., Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications, Neurocomputing, 14(October) 410 (2020) 237–270.
[13] J.M. Gorriz et al., Statistical Agnostic Mapping: A framework in neuroimaging based on concentration inequalities, Information Fusion 66 (February 2021) 198–212.
[14] J.M. Gorriz et al., A connection between pattern classification by machine learning and statistical inference with the General Linear Model, IEEE Journal of Biomedical and Health Informatics (2021).
[15] T. Hastie, et al. The elements of statistical learning theory. Data Mining inference and prediction. Ed Springer. isbn 0-387-95284-5. 2001.
[16] D. Haussler, Decision theoretic generalizations of the PAC model for neural net and other learning applications, Information and Computation 100 (1) (1992) 78–150.
[17] C.S. Hyatt et al. The quandary of covarying: A brief review and empirical examination of covariate use in structural neuroimaging studies on psychological variables. Neuroimage 205, 116225.
[18] K.A. Johnson et al., Brain imaging in Alzheimer disease, Cold Spring Harb Perspect Med. 2 (4) (2012) a006213.
[19] I.A. Illan et al., Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging, Medical Physics (2012).
[20] N. Zeng et al., A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease, Neural Comput & Applic (2021).
[21] N. Zeng, A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease, Neurocomputing 320 (2018) 195–202.
[22] C.C. Jack Jr et al., NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease, Alzheimers Dement. 14 (4) (2018) 535–562.
[23] S. Noble et al., Cluster failure or power failure?, Evaluating sensitivity in cluster-level inference NeuroImage 209 (2020) 116468.
[24] M. Leming et al., Ensemble Deep Learning on Large, Mixed-Site fMRI Datasets in Autism and Other Tasks, M Leming, International Journal of Neural Systems. 30 (07) (2020) 2050012.
[25] F.J. Martinez et al., Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders. IEEE J Biomed Health, Inform. (2019), Jun 17.

[26] M.J. McKeown et al., Independent component analysis of functional MRI: what is signal and what is noise?, Curr Opin Neurobiol 13 (5) (2003) 620–629.
[27] J. Mouro-Miranda et al., Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data, NeuroImage 28 (2005) 980–995.
[28] T.E. Nichols, Multiple testing corrections, nonparametric methods, and random field theory, NeuroImage 62 (2012) 811–815.
[29] P.T. Reiss, et al. Cross-validation and hypothesis testing in neuroimaging: an irenic comment on the exchange between Friston and Lindquist et al. Neuroimage. 2015 August 1; 116: 248–254.
[30] J.D. Rosenblatt et al., Revisiting multi-subject random effects in fMRI: Advocating prevalence estimation, NeuroImage 84 (2014) 113–121.
[31] J.D. Rosenblatt et al., Better-than-chance classification for signal detection, Biostatistics (2016).
[32] B. Schölkopf et al., Learning with Kernels, MIT Press, 2002.
[33] D.V. Smith et al., Decoding the anatomical network of spatial attention, Proc Natl Acad Sci USA 110 (2013) 1518–1523.
[34] A. Smola et al., Convex cost functions for support vector regression, in: International Conference on Artificial Neural Networks, 1998.
[35] N. Tzourio-Mazoyer et al., Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single subject brain, Neuroimage 15 (2002) 273–289.
[36] V. Vapnik. Estimation dependencies based on Empirical Data. Springer-Verlach. 1982 ISBN 0-387-90733-5.
[37] G. Varoquaux, Cross-validation failure: Small sample sizes lead to large error bars, NeuroImage 180 (2018) 68–77.
[38] K. Gorgen et al., The same analysis approach: Practical protection against the pitfalls of novel neuroimaging analysis methods, NeuroImage 180 (2018) 19–30.
[39] Z. Wang, A hybrid SVM-GLM approach for fMRI data analysis, Neuroimage 46 (3) (2009) 608–615.
[40] Z. Wang et al., Support vector machine learning-based fMRI data group analysis, NeuroImage 36 (4) (2007) 1139–1151.
[41] M.K. Woolrich et al., Bayesian analysis of neuroimaging data in FSL, NeuroImage 45 (2009) S173–S186.
[42] Y. Zhang et al., Multivariate lesion-symptom mapping using support vector regression, Hum Brain Mapp. 35 (12) (2014) 5861–5876.

Juan Manuel Górriz is a Full Professor at the University of Granada, Spain and a Visiting Professor at the University of Cambridge, UK. He received his M.Sc. in Physics in 2000 and his M.Sc. in Electronic Engineering in 2001, both from the University of Granada, Spain. He then earned the Ph.D. degree in Electronic Engineering in 2003 and the Ph.D. degree in Physics in 2006, from the Universities of Cádiz and Granada, Spain, respectively. He has already been the supervisor of 14 Ph.D. students and has published more than 200 journal papers that have received more than 10K cites, with an h-index equal to 52. He actually leads the group in "Signal Processing and Biomedical Applications" (SIPBA) at the Data Science and Computational Intelligence Institute (DASCI-UGR). He has received awards and honors, such as the 2015 Spanish National Engineering Academy Medal (Young Researchers), the 2008 ASI Andalusian Prize for the best research project "DENCLASES" in neuroscience, full membership in Cambridge Neuroscience and Ellis Network, the Social Council Award at the UGR (Young Researchers), etc. He has developed several short and long stays as visiting Professor in top centers around the world, such as University of Cambridge (UK), University of Regensburg (Germany), University of Liege (Belgium), NorthEastern University (USA), Infineon-Siemens AG, Munich (Germany), etc.

Rubén Martín-Clemente (Member, IEEE) received the M.Eng. degree in telecommunications engineering and the Ph.D. degree (Hons.) in telecommunications engineering from the University of Seville, Seville, Spain, in 1996 and 2000, respectively. He was a Visiting Researcher at the University of Regensburg, Regensburg, Germany, in 2001 and 2009 and with the University of Nice, France, in 2015, 2016, and 2018, respectively. He is currently the Head of the Department of Signal Theory and Communications, University of Seville. He has authored or coauthored numerous publications on his research interests. His research interests include signal processing and machine learning with an emphasis on independent component analysis and its application to biomedical problems.

Carlos G. Puntonet received the Ph.D. degree in electronics from University of Granada, Granada, Spain, in 1994.He is currently a Full Professor with the "Departamento de Arquitectura y Tecnología de Computadores," at the University of Granada. His research interests lie in the fields of signal processing, linear and nonlinear independent component analysis, artificial neural networks, and optimization methods. He has coauthored more than 300 technical journals and conference papers in these areas. He has served as a reviewer for several international journals and conferences.

Andrés Ortiz received the B.Sc. degree in Physics (1998), the M.Sc. degree in electronics (2000) and the Ph.D. degree in Computer Science in 2008 all from the University of Granada. He then earned the Ph.D. degree from the University of Cádiz, Spain (2012). From 2000 to 2005 he was working as Networks and Systems Engineer with Telefónica, Madrid, Spain, where his work areas were high performance computing and network performance analysis. Currently he is an Associate Professor at the Department of Communication Engineering of the University of Malaga. His research interests include intelligence systems, statistical signal processing with biomedical applications and high performance computing. He has coauthored more than 120 technical journal and conference papers and leaded different research projects in these areas. He currently leads the "Biomedical signal processing, computational intelligence and communications security" (BioSiP) group at the University of Málaga, and co-directs the CINEMA Computational Neuroscience Laboratory. He has served as editor and reviewer for several international journals and conferences.

Javier Ramírez received the M.A.Sc. degree in Electronic Engineering in 1998, and the Ph.D degree in Electronic Engineering in 2001, all from the University of Granada. Since 2012, he is Professor at the Department of Signal Theory Networking and Communications of the University of Granada (Spain). His research interest includes signal processing and biomedical applications including brain image processing, robust speech recognition, speech enhancement, voice activity detection, seismic signal processing and implementation of high performance digital signal processing systems. He has coauthored more than 400 technical journals and conference papers in these areas. He has served as a reviewer for several international journals and conferences.

J. Suckling has over 250 peer-reviewed articles in medical image processing and psychiatric neuroimaging, including large-scale multicentre studies of autism spectrum condition, psychosis, drug addiction, and affective disorders. He has been involved in imaging research for over 25 years, initially with Positron Emission Tomography in the monitoring of tumors during chemotherapy, and in the implementation of image reconstruction techniques for over scanner geometries. His current interests include the neurobiological and neurodevelopmental bases for child and adolescent psychiatry with specific involvement of new neuroimaging techniques as primary outcome variables for studies of brain development and maturation, characterization of psychopathologies, and for clinical trials of investigational medical products.