Neurocomputing 273 (2018) 373–384

Contents lists available at ScienceDirect

Neurocomputing

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

A novel and effective fMRI decoding approach based on sliced inverse regression and its application to pain prediction

Y.H. Tu[a,b,c,1], Z.N. Fu[a,b,1], A. Tan[b], G. Huang[a], L. Hu[d], Y.S. Hung[b], Z.G. Zhang[a,*]

[a] School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
[b] Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China
[c] Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA, USA
[d] Institute of Psychology, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 11 August 2016
Revised 9 May 2017
Accepted 19 July 2017
Available online 18 August 2017

Communicated by Wei Wu

Keywords:
fMRI decoding
Dimension reduction
Sliced inverse regression
Pain prediction

ABSTRACT

Dimension reduction is essential in fMRI decoding, but the complex relationship between fMRI data and class labels is often unknown or not well modeled so that the most effective dimension reduction (e.d.r.) directions can hardly be identified. In the present study, we introduce a novel fMRI decoding approach based on an effective and general dimension reduction method, namely sliced inverse regression (SIR), which can exploit class information for estimating e.d.r. directions even when the relationship between fMRI data and class labels is not explicitly known. We incorporate singular value decomposition (SVD) into SIR to overcome SIR's limitation in dealing with ultra-high-dimensional data, and integrate SVD-SIR into a pattern classifier to enable quantification of the contributions of fMRI voxels to class labels. The resultant new SIR decoding analysis (SIR-DA) approach is capable of decoding behavioral responses and identifying predictive fMRI patterns. Simulation results showed that SIR-DA can more accurately detect e.d.r. directions and achieve higher classification accuracy than decoding approaches based on conventional dimension reduction methods. Further, we applied SIR-DA on real-world pain-evoked fMRI data to decode the level of pain perception and showed that SIR-DA can achieve higher accuracy in pain prediction than conventional methods. These results suggest that SIR-DA is an effective data-driven technique to decode behavioral or cognitive states from fMRI data and to uncover unknown brain patterns associated with behavior or cognitive responses.

1. Introduction

Decoding brain states from functional magnetic resonance imaging (fMRI) data using machine learning approaches has gained popularity [1,2]. Because of the high dimensionality and multicollinearity of fMRI data, effective dimension reduction is crucial for fMRI decoding. Subspace-projection-based dimension reduction techniques, such as principle component analysis (PCA) [3] and partial least squares (PLS) [4], are often used as a pre-processing step in fMRI decoding. On the other hand, sparsity-enhancing regularization techniques, such as L_1 or L_2 norm-regularized linear regression, can shrink small regression coefficients to zero to achieve automatic feature selection [5]. These two types of dimension reduction techniques could be combined for higher effectiveness in dimension reduction and one example is LASSO principle component regression (LASSO-PCR) used in [6] for fMRI-based pain prediction.

Existing dimension reduction techniques in fMRI decoding normally over-simplified the complex relationship between fMRI and class labels as a linear function and some even did not make use of class information. As clearly revealed by many studies, the linear assumption about the relationship between neural activities and behavioral responses is problematic. For example, pain perception was nonlinearly correlated with the magnitudes of laser-evoked potentials and fMRI responses in medial prefrontal cortex (mPFC) [7,8]. Consequently, linear dimension reduction methods may fail to select the most predictive features and could degrade prediction accuracy when the relationship between fMRI and class labels is complex and possibly even unknown.

In the present work, we aim to introduce a new fMRI decoding approach based on a powerful dimension reduction method, namely sliced inverse regression (SIR). SIR was originally introduced to estimate the effective dimension reduction (e.d.r.) direction along which the reduced data can well explain class labels regardless of linear or nonlinear relationship between data and labels

* Corresponding author.
E-mail address: zgzhang@szu.edu.cn (Z.G. Zhang).
[1] These authors contribute equally.

http://dx.doi.org/10.1016/j.neucom.2017.07.045
0925-2312/

Table 1
Detailed procedure for SIR.

Input	$\mathbf{x} \in \mathbf{R}^{n \times p}$: a matrix of fMRI data with n samples and p predictors; $y \in \mathbf{R}^n$: a vector of n samples of observations.
Output	$\hat{\beta}_k$: e.d.r. directions.
Step 1	Standardize $\mathbf{x}$ using the sample mean $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$ and the sample covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x}} = \sum_{i=1}^{n} \frac{1}{n-1}(\mathbf{x}_i - \bar{\mathbf{x}})^T(\mathbf{x}_i - \bar{\mathbf{x}})$, where $\mathbf{x}_i$is the ith row of $\mathbf{x}$.
Step 2	Divide the range of y into M slices, $G_1, \cdots, G_M$, and calculate the proportion of y_i that falls into the mth slice G_m, $m = 1, \cdots, M$, as $\hat{p}_m = \frac{1}{n}\sum_{i=1}^{n} \xi_m(y_i)$, where $\xi_m(y_i)$ equals 1 or 0 depending on whether y_i falls into the mth slice or not.
Step 3	For each slice, calculate the sliced mean $\bar{\mathbf{x}}_m = \frac{1}{n\hat{p}_m}\sum_{i \in G_m} \mathbf{x}_i$ and weighted covariance $\hat{\mathbf{\Sigma}}_W = \sum_{m=1}^{M} \hat{p}_m(\bar{\mathbf{x}}_m - \bar{\mathbf{x}})^T(\bar{\mathbf{x}}_m - \bar{\mathbf{x}})$.
Step 4	Solve the generalized eigen-decomposition problem: $\hat{\mathbf{\Sigma}}_W\hat{\beta}_k = \hat{\lambda}_k\hat{\mathbf{\Sigma}}_{\mathbf{x}}\hat{\beta}_k$, and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K$, where $\hat{\beta}_k$ denotes SIR directions.

[9]. Because of its generality and effectiveness, SIR has been used in bioinformatics [10,11] and economics [12]. But, SIR can only deal with a dataset where the number of features is smaller than the number of samples, which is obviously not the case of fMRI data, so it cannot been used for fMRI decoding directly. To overcome this problem, we proposed to use singular value decomposition (SVD) prior to SIR to transform fMRI data into a low-dimensional orthogonal basis, which has a dimension equal to or lower than the number of samples. We further integrate SVD-SIR into a classifier or a prediction model to build a new and general fMRI decoding analysis approach (namely, SIR-DA) which can decode behavior responses from SIR-reduced features and identify predictive fMRI patterns.

The effectiveness of SIR-DA was evaluated on simulated fMRI data under a variety of testing conditions (combinations of different linear or nonlinear fMRI-class relationships and different noise levels). Simulation results show that SIR outperformed other projection-based dimension reduction techniques (PCA and PLS) in terms of finding the correct e.d.r. directions, and SIR-DA provided better prediction accuracy than projection-based as well as regularization-based decoding analyses (PCA-DA, PLS-DA, LASSO-PCR and elastic net).

Further, we applied the proposed SIR-DA approach on real-world laser-evoked fMRI data, with the aim to predict the level of subjective pain perception elicited by laser stimulation. Results show that SIR-DA can identify brain patterns both linearly and nonlinearly correlated with pain perception, and achieved significantly higher pain prediction accuracy than linear decoding techniques.

2. Material and methods

2.1. Sliced inverse regression

SIR is based on a generalized linear model and it assumes that the response variables $y \in \mathbf{R}^n$ depend on the predictors $\mathbf{x} \in \mathbf{R}^{n \times p}$ $(n \geq p)$ via K linear combinations of $\mathbf{x}$ as:

$$y = f(\mathbf{x}\beta_1, \mathbf{x}\beta_2, ..., \mathbf{x}\beta_K, e), \tag{1}$$

where $f(\cdot)$ is an unknown function, β_k, $k = 1, \cdots, K$, are unknown p-dimensional projection vectors, and ε is zero-mean random noise independent of $\mathbf{x}$. If $K < p$, the dimension of predictors can be effectively reduced from p to K. Any linear combination defined by β_k is an e.d.r. direction, and the linear subspace B spanned by all β_k's is called the e.d.r. space. The basic principle of the SIR algorithm (shown in Table 1) is to estimate a crude inverse regression curve $E(\mathbf{x}|y)$, which will fall into the e.d.r. space if $\mathbf{x}$ is standardized to have zero mean and identity covariance [9]. The inverse regression curve $E(\mathbf{x}|y)$ can be estimated by partitioning $\mathbf{x}$ into M slices (each of which has the same number of samples, generally) according to the values of y and calculating the mean values of $\mathbf{x}$ within the slices. Then, a generalized eigen-decomposition is applied to the slice mean values to locate the most important K-dimensional subspace for tracking the inverse regression curve. It is important to note that, SIR can only work for cases where $n \geq p$, because the sample covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x}}$ will be singular when $n < p$ and consequently the solution to the eigen-decomposition problem will not be unique.

2.2. Sliced inverse regression-decoding analysis

Suppose in an fMRI decoding problem, we have behavioral data or cognitive parameters $\boldsymbol{\psi} \in \mathbf{R}^n$ with n samples and fMRI data $\mathbf{\Phi} \in \mathbf{R}^{n \times p}$ with p voxels for each sample. According to (1), an SIR model, $\psi = f(\mathbf{\Phi}\beta_1, \mathbf{\Phi}\beta_2, ..., \mathbf{\Phi}\beta_K, \varepsilon)$, can be used to reduce the dimension of $\mathbf{\Phi}$. But, the number of voxels (tens of thousands) is normally much larger than the number of samples (a few dozens or hundreds), so SIR cannot be directly applied to $\boldsymbol{\psi}$ and $\mathbf{\Phi}$. To address this problem, we proposed to reduce the number of columns (predictors) of $\mathbf{\Phi}$ from p to q $(q \leq n)$, which can be achieved by SVD prior to SIR. More precisely, we perform SVD on $\mathbf{\Phi}$ as $\mathbf{\Phi} = \mathbf{U\Sigma V}^T$ to obtain the principle components $\mathbf{U\Sigma} \in \mathbf{R}^{n \times n}$ and the column-orthogonal matrix $\mathbf{V} \in \mathbf{R}^{p \times n}$. To keep as much information from $\mathbf{\Phi}$ as possible, we retain all principle components to construct the dimension-reduced fMRI data matrix as $\tilde{\mathbf{\Phi}} = \mathbf{U\Sigma} \in \mathbf{R}^{n \times n}$. Then, the new SIR model reads $\psi = f(\tilde{\mathbf{\Phi}}\tilde{\beta}_1, \tilde{\mathbf{\Phi}}\tilde{\beta}_2, ..., \tilde{\mathbf{\Phi}}\tilde{\beta}_K, \varepsilon)$, where $\tilde{\beta} = [\tilde{\beta}_1, \tilde{\beta}_2, ..., \tilde{\beta}_K] \in \mathbf{R}^{n \times K}$ are the e.d.r. directions in the dimension-reduced space. Note that in the proposed dimension reduction process, the number of predictors is reduced from p to K (first from p to n by SVD and finally from n to K by SIR), yielding the feature set $\mathbf{S} = \tilde{\mathbf{\Phi}}\tilde{\beta} \in \mathbf{R}^{n \times K}$. Subsequently, a pattern classifier with $\boldsymbol{\psi}$ as the response variables and $\mathbf{S}$ as predictors is used for fMRI decoding.

Since SIR is based on a generalized linear model, it can be followed by any classifier that is also based on a generalized linear model, enabling the contribution (i.e., magnitudes of coefficients) of each feature to the classification to be reconstructed. Suppose a generalized linear classifier $\psi = C(\mathbf{S}a)$ with a weighting vector $\alpha \in \mathbf{R}^K$ is adopted, where the contribution of each variable in $\mathbf{S}$ to predict $\boldsymbol{\psi}$ is represented by the corresponding element in the weighting vector a [13]. Considering that $\mathbf{S} = \tilde{\mathbf{\Phi}}\tilde{\beta}$ and $\tilde{\mathbf{\Phi}} = \mathbf{\Phi V}$, one gets

$$\psi = C(\mathbf{S}a) = C(\tilde{\mathbf{\Phi}}\tilde{\beta}a) = C(\mathbf{\Phi V}\tilde{\beta}a) \tag{2}$$

Hence, the weighting vector denoting the contribution of each variable in the original fMRI data matrix $\mathbf{\Phi}$ to predict $\boldsymbol{\psi}$ is $\theta = \mathbf{V}\tilde{\beta}\alpha$. A summary of the SIR-DA approach, including all above procedures is given in Table 2.

Cross-validation (CV) is often used in a classifier to determine parameters and to evaluate performance. The use of CV in the framework of SIR-DA is illustrated in Fig. 1. In each fold of CV, the e.d.r. directions are firstly obtained from all training trials (with labels) and applied on both training and test trials to obtain dimension-reduced feature sets. Next, still in the same fold of CV, a classifier or a prediction model is trained by the dimension-reduced training data and then applied on the dimension-reduced test data to predict the class labels.

2.3. Complexity and parameter selection

Computational complexity and the selection of parameters are two practical issues when applying the SIR-DA in an fMRI study.

Table 2
Procedures of SIR-DA.

Input	$\mathbf{\Phi} \in \mathbf{R}^{n \times p}$: fMRI data with n samples and p voxels; $\boldsymbol{\psi} \in \mathbf{R}^{n}$: n samples of behavioral data or cognitive parameters.
Output	$\tilde{\beta} = [\tilde{\beta}_1, \tilde{\beta}_2, ..., \tilde{\beta}_K] \in \mathbf{R}^{n \times K}$: e.d.r. directions; $\theta \in \mathbf{R}^{p}$: regression weights for whole brain fMRI voxels.
SVD	Perform SVD on $\mathbf{\Phi}$ as $\mathbf{\Phi} = \mathbf{U\Sigma V}^T$ to obtain the principle components $\tilde{\mathbf{\Phi}} = \mathbf{U\Sigma} \in \mathbf{R}^{n \times n}$ and the column-orthogonal matrix $\mathbf{V} \in \mathbf{R}^{p \times n}$.
SIR	Perform SIR on $\tilde{\mathbf{\Phi}}$, resulting in the e.d.r. directions $\tilde{\beta}$ and SIR-derived feature set $\mathbf{S} = \tilde{\mathbf{\Phi}}\tilde{\beta} \in \mathbf{R}^{n \times K}$.
Classification	Construct a classifier $\psi = C(\mathbf{S}_t\alpha; t = 1, \cdots, n)$ with $\alpha \in \mathbf{R}^{K}$ to estimate classification accuracy and the contribution of each variable in the reduced space to predict $\boldsymbol{\psi}$.
Reconstruction	Transform α back to the original space by multiplying $\tilde{\beta}$ and $\mathbf{V}$, resulting in the weights for whole brain voxels $\theta = \mathbf{V}\tilde{\beta}\alpha$.

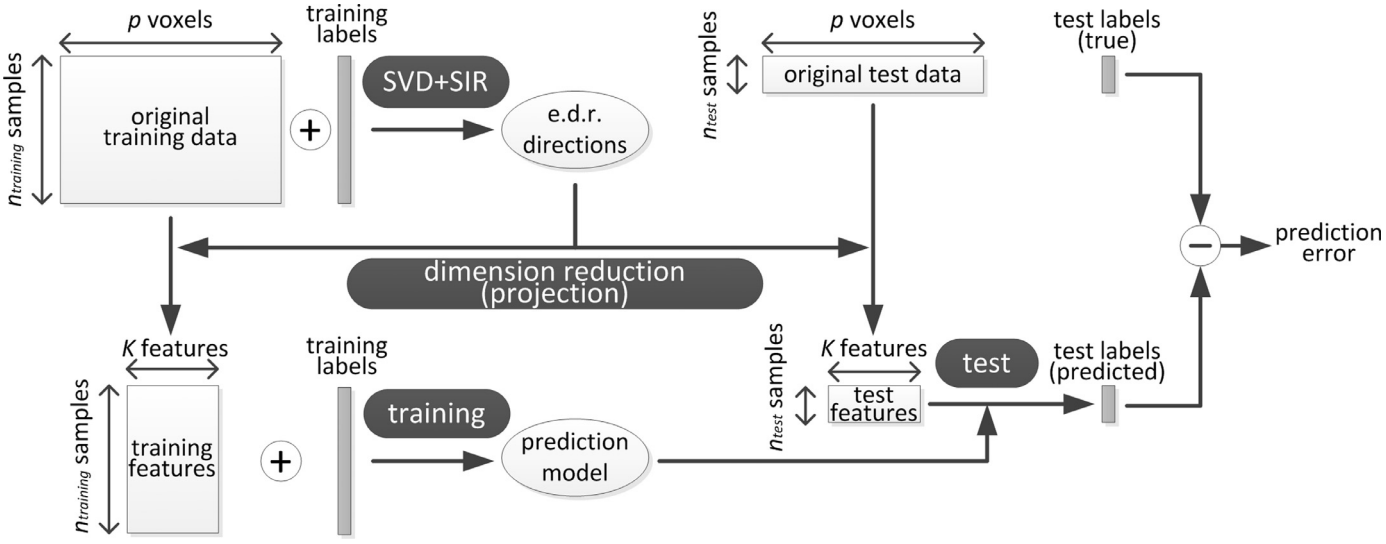


Fig 1. Flowchart of proposed SIR-DA pipeline in each fold of CV.

We first consider the arithmetic complexity of SVD. Since SVD is performed on $\mathbf{\Phi} \in \mathbf{R}^{n \times p}$, $n << p$, it has a complexity of $O(p^3)$. SIR is performed on the square matrix $\tilde{\mathbf{\Phi}} = \mathbf{U\Sigma} \in \mathbf{R}^{n \times n}$, and the most demanding operation is eigenvalue decomposition which has a complexity of $O(n^3)$. Therefore, SVD-SIR still has a complexity of $O(p^3)$.

It is necessary to specify the number of slices M and the number of projection vectors K in SIR-DA for practical implementation. The number of slices, M, affects the tradeoff between reliability of slice mean estimation and adequacy in the number of slice means for robust inverse regression curve estimation. More specifically, a larger M (more slices) will lead to more slice means for building a more precise inverse regression curve, but it will increase the estimation variance in the slice means due to the smaller sample size of each slice. On the contrary, a smaller M (fewer slices) will lead to smaller estimation variance in the slice means, but the inverse regression curve may not be very accurately estimated because of fewer number of slice means. The optimal M can be obtained by CV from training data. The literature and our simulation results suggested that M would not largely affect the output estimates if it is sufficiently large ($M > 4$) [14]. The number of e.d.r. directions, K, can also be trained by CV. In our application, we determined the value of K using the output eigenvalues of the generalized eigen-decomposition in SIR (Step 4 in Table 1). More precisely, the first leading K eigenvalues should explain the majority of data variance.

3. Simulation

3.1. Subjects and data acquisition

In this section, a resting state fMRI dataset collected from 32 healthy subjects (20 females) aged 22.1 ± 2.0 years (mean ± SD, range = 19–24 years) were used to generate synthetic fMRI activations. All subjects gave their written informed consent and were paid for their participation. The local ethics committee approved the procedures.

The fMRI data were acquired using a Siemens 3.0 Tesla Trio scanner with a standard head coil at the Key Laboratory of Cognition and Personality (Ministry of Education) of the Southwest University (China). A whole-brain gradient-echo, echo-planar-imaging sequence was used for functional scanning with a repetition time (TR) of 1500 ms (29 ms echo time, 25 contiguous 5.0 mm thick slices with 0.5 mm thick inter-slice gaps, 3 × 3 mm in-plane resolution, field of view 192 × 192 mm, matrix 64 × 64, flip angle = 90°). During the scan, subjects were instructed to rest with their eyes closed and in total 330 images were collected for each subject. For spatial normalization and localization, a high-resolution, T1-weighted structural image (1 mm^3 isotropic voxel MPRAGE) was acquired after functional imaging.

The fMRI data were preprocessed and analyzed using SPM8 (Wellcome Trust Center for Neuroimaging, London, UK). Images were slice-time corrected, motion corrected, spatially smoothed using a Gaussian kernel of 8 mm full width at half maximum (FWHM = 8 mm) and normalized to the Montreal Neurological Institute (MNI) space by matching grey matter [15].

3.2. Simulated data

Six linear and nonlinear models between variables $\mathbf{\Phi}$ (fMRI data) and $\boldsymbol{\psi}$ (response variables) were constructed and they were summarized in Table 3. fMRI data were synthesized by Monte-Carlo simulation. In each Monte-Carlo run, we randomly selected 100 images from the total 330 images of one subject to simulate 100 trials of this subject, and added a zero mean Gaussian noise

Table 3
Functions used in simulation.

Relationship type	Description
Linear ($K=1$)	$\psi = (\Phi\beta_1) + e$
Cubic ($K=1$)	$\psi = (\Phi\beta_1)^3 + e$
Exponential ($K=1$)	$\psi = \exp(\Phi\beta_1) + e$
Linear ($K=2$)	$\psi = (\Phi\beta_1 + \Phi\beta_2) + e$
Cubic ($K=2$)	$\psi = [(\Phi\beta_1)^3 + (\Phi\beta_2)^3] + e$
Exponential ($K=2$)	$\psi = [\exp(\Phi\beta_1) + \exp(\Phi\beta_2)] + e$

to the trial. For each subject, we run above Monte-Carlo testing for 100 times and then different methods (SIR, PCA, PLS) were compared on these 100 runs. Two projections vectors β_1 and β_2 in Table 3 were binary masks, both of which contains either "1" for specified region of interest or "0" elsewhere. In our stimulation, β_1 and β_2 were used to define fMRI activation patterns in two regions, the part of the insular cortex (Brodmann area [BA] 13) and the primary auditory cortex (A1, BA42/BA43), respectively. The relationships between $\boldsymbol{\psi}$ and $\boldsymbol{\Phi}\beta_1$ or $\boldsymbol{\Phi}\beta_2$ in different models are given in Supplementary Material Fig. 1.

In SIR-DA, the number of slices was set as $M = 10$ and the number of e.d.r. directions set to $K = 1$ or 2 depending on the number of β's in the model of Table 3. SIR-derived feature sets were fitted in a support vector regression (SVR) model with linear kernels and the regression weights were reconstructed in the original space to indicate the contribution of each voxel in predicting $\boldsymbol{\psi}$. A point-by-point one-sample t-test against zero with Bonferroni correction was performed on reconstructed whole-brain regression weights across subjects to detect the significantly activated regions. We tested the performance of SIR-DA under different levels of signal-to-noise ratio (SNR). Here, SNR is defined as:

$$\mathrm{SNR} = \frac{\sigma^2(A)}{\sigma^2(e)} \tag{3}$$

where A is the synthetic activation and e is the additive random noise generated from normal distribution with zero mean. By adjusting the variance of the additive noise e, we can vary the SNR.

It should be noted that, SNR defined in this study is used to describe the "cleanness" of response variables, not the "cleanness" of fMRI signals. Thus, SNR used in our study is different from the conventional use of SNR in fMRI research, which normally denotes the ratio between the magnitudes of fMRI activities and of additive interference/noise [16]. We use such a definition of SNR because we focus on the classification problem, in which the noise is not only added to features but also to responses (similar definitions of SNR are popular in pattern recognition). In our simulation, SNR is set to 1, 2, 5, and 10. Particularly, an SNR of 1 should be a small value because it means the additive noise is as large as the true response variables. In our pain study (Section 4), we approximated the variance of noise as the variance of prediction error (the difference between predicted pain rating and true pain rating), which is a commonly-used method to estimate the amount of additive noise in estimation theory [17], and the approximated SNR is 2.66 ± 0.85 (for 32 subjects).

Compared to the conventional simulation approach which adds noise to synthetic fMRI time-series, we added noise to the response variables in the fMRI decoding model. We adopted the current simulation approach because of two reasons. First, theoretically, SIR regresses explanatory variables $\mathbf{x}$ against response variables y. The reliability of the estimates of sliced means and inverse regression curve are largely influenced by the slicing according to the values of y. So, to examine the performance of SIR, it is common to add noise on the response variables [9,14]. Second, in fMRI decoding, response variables may not be correctly decoded from fMRI data solely, because other factors, such as brain structures and genetics, could also contribute to response variables but they may not be encoded by fMRI recording. For example, in our problem of pain prediction, fMRI data alone cannot completely explain the intra- and inter-subject variability in pain experience [18,19]. In our simulation approach, the noise term is used to describe all possible predication biases which are not related to fMRI data but caused by other factors.

3.3. Performance evaluation

To assess the accuracy of an estimated e.d.r. direction, the most direct criterion is based on some distance between the estimated e.d.r. direction $\hat{\beta}_k$ and the true e.d.r. direction β_k. However, distance-based criterion is not applicable due to the change of scale or affine transformation of $\boldsymbol{\Phi}$. Here, we used the absolute correlation (AC), $|corr(\boldsymbol{\Phi}\beta_1, \tilde{\boldsymbol{\Phi}}\hat{\beta}_1)|$, as the performance measure when $K = 1$, and the squared canonical correlation (SCC), $[canoncorr(\boldsymbol{\Phi}\beta, \tilde{\boldsymbol{\Phi}}\hat{\beta})]^2$with $\beta = [\beta_1, \beta_2]$ and $\hat{\beta} = [\hat{\beta}_1, \hat{\beta}_2]$, when $K = 2$ [9,20]. A higher AC or SCC indicates a more accurate estimate of the e.d.r. direction.

The prediction accuracy of SIR-DA was measured by the mean absolute error (MAE),

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|\psi_t - \hat{\psi}_t\right| \tag{4}$$

where $\boldsymbol{\psi}_t$ and $\hat{\psi}_t$ are the real and predicted values of the tth trial and T is the total number of trials for each subject. The MAE values were estimated using 5-fold CV: the e.d.r. directions and the prediction model were trained using 80 trials from each subject and tested on the remaining 20 trials of the subject.

To compare the performance of estimation of e.d.r. directions(s), conventional projection-based dimension reduction methods including PCA and PLS (regularization-based methods do not estimate e.d.r. directions), were also applied on the same $\boldsymbol{\psi}$ and $\boldsymbol{\Phi}$ generated by the models in Table 3. Reduced feature sets were extracted by maximizing the variance and covariance using PCA and PLS respectively. For PCA, we kept the principle components accounting for 90% of the total variance of the data. PLS was computed by nonlinear iterative partial least squares (NIPALS) algorithm, and the number of latent components was estimated using coefficient of determination [21]. Further, to compare the performance of prediction accuracy, PCA-DA, PLS-DA and regularization-based methods including LASSO-PCR [6] and elastic net [22] were applied on the models in Table 3. For LASSO and elastic net, the entire regularization path (from zero activate variables to least squares solution) were obtained and the optimal penalties were selected using Bayesian Information Criterion (BIC).

The performances of different methods in dimension reduction and classification were tested on these 100 sets of simulated data and compared using paired t-test with Bootstrap hypothesis tests [23]. Specifically, the statistical comparisons between different approaches were carried out as follows:

1. Calculate the statistics t using paired t-test from all Monte-Carlo simulation samples,
2. Select 20 samples with replacement from the Monte-Carlo simulation samples in each bootstrap,
3. Calculate the statistics t_i^* using paired t-test from each bootstrap run, $i = 1, ..., 1000$,
4. Get p value by estimating $p = \min(2k, 2-2k)$, where $k = \frac{1}{N}\sum_{i=1}^{N} 1\{t_i^* \le t\}, N = 1000$.

Lastly, we also assessed the performance of SIR-DA with different number of slices (M). We repeated the above simulation procedure with M ranging from 3 (being the least number of slices means to build a curve) to 20 (with a minimal of five samples in

each slice to estimate slice means). Performance of dimension reduction (in terms of AC or SCC) and prediction accuracy (in terms of MAE) were summarized in Figs. 4 and 5.

4. FMRI pain prediction

4.1. Experimental design

The analyses of this study were performed on the dataset collected from the same 32 healthy subjects as in the simulation study. Prior to data collection, we delivered a small number of laser pulses with different energies to familiarize the subjects with the stimulation. During fMRI data collection we delivered ten laser pulses at each of the four stimulus energies (E1: 2.5 J, E2: 3 J, E3: 3.5 J, E4: 4 J), making a total of 40 pulses. The order of stimuli at different energies was pseudorandomized. The inter-stimulus interval (ISI) varied randomly from 27 to 33 s (rectangular distribution). A visual cue delivered between 15 and 18 s after each stimulus prompted the subject to rate the intensity of pain perception elicited by the laser stimulus, using a visual analogue scale (VAS) ranging from 0 (corresponding to "no pain") to 10 (corresponding to "pain as bad as it could be").

4.2. General linear model analysis

The fMRI data were preprocessed following the standard procedure suggested by SPM8 (see preprocessing in simulation) and filtered using a high-pass filter with a cutoff frequency of 1/128 Hz. To obtain the brain activations from the conventional linear method, single-subject fMRI data were analyzed using a general linear model (GLM) approach [24]. BOLD responses were modeled as a series of events using a stick function and ratings were included as a parametric modulator of each stimulus, which were then convolved with a canonical hemodynamic response function (HRF). Group-level statistical analyses were carried out using a random effects analysis with the one-sample *t*-test as implemented in SPM8. To account for multiple comparisons, the significance level (*p* value) was corrected using a false discovery rate (FDR) in the whole-brain exploratory analysis (Heller et al., 2006). Trials were sorted according to the subjective pain rating and 10 trials were categorized into each of the four pain levels (R1-R4; R1 denoted trials with minimal pain perception, while R4 denoted trials with maximum pain perception).

4.3. SIR-DA analysis

SIR-DA ($M=4$, $K=2$) was applied on whole-brain fMRI data at the third image after stimulus onset, when the brain had the largest responses to the pain stimuli (as shown in Fig. 6). The number of slices, M, was set to 4 because there are 4 levels of laser stimulation, and we also used CV to make sure $M=4$ is an appropriate selection. In our case, the first eigenvalue can explain $59.1 \pm 7.0\%$ of the total variance, and the second one can explain $32.3 \pm 5.7\%$. Since first two eigenvalues can explain more than 90% of the total variance, we chose K as 2 here. Regression weights were reconstructed in the original space by fitting the pain rating and SIR-derived feature set. To identify the brain regions that significantly contributed to pain perception, a point-by-point one-sample *t*-test against zero was performed on the estimated regression weights across subjects. The significance level was corrected using an FDR to account for multiple comparisons. PCA-DA, PLS-DA, LASSO-PCR and elastic net were also applied on the same dataset to compare the performance (See Supplementary Figs. 2–5 for details).

4.4. Pain prediction

We performed a binary pain prediction, which used supported vector machine (SVM) with linear kernels to classify high-pain trials ($VAS \geq 5$) and low-pain trials ($VAS < 5$), and a continuous pain prediction which used SVR with linear kernels to predict the level of pain perception within the range between 0 and 10 for each trial. Both binary and continuous predictions were performed using 5-fold CV strategy and at the within-subject level. The prediction performances were measured by classification accuracy and MAE between reported pain intensity and predicted pain intensity. We also compared the prediction performance of SIR-DA with SVM (without dimension reduction), PCA-DA, PLS-DA, LASSO-PCR and elastic net using one-way repeated-measures ANOVA. When the main effect was significant, post hoc pairwise comparisons were performed.

5. Results

5.1. Simulations results

5.1.1. Accuracy of e.d.r. direction

Fig. 2 summarizes the performance of dimension reduction measured by AC and SCC for $K=1$ and $K=2$. For linear models ($K=1$ and $K=2$), PLS-DA had the best performance at different noise levels. SIR-DA performed significantly worse than PLS-DA when the noise was large (SNR = 1 for $K=1$, $p=0.012$; SNR = 1 and 2 for $K=2$, $p=0.010$ and $p=0.008$ respectively), but had comparable results with PLS-DA at higher levels of SNR (SNR = 2, 5 and 10 for $K=1$, SNR = 5, and 10 for $K=2$; $p>0.05$). For nonlinear models, SIR-DA performed significantly better than PLS-DA under most conditions ($p<0.05$) except for Cubic/Exp, $K=1$, and SNR = 1. As an unsupervised method, PCA-DA cannot accurately detect e.d.r. direction(s) under all testing conditions. Here, we only report SCC values obtained from the first component of SIR-DA and PLS-DA when $K=2$ because they already captured the major contribution of insula and A1 so that the second direction can be ignored without much performance loss [14].

5.1.2. Mean absolute error

The prediction performances of different methods measured by MAE are summarized in Fig. 3. For linear models ($K=1$ and $K=2$), SIR-DA showed comparable performance to PLS-DA, LASSO-PCR and elastic net at higher SNR (SNR = 5 and 10). In contrast, SIR-DA had significantly better performance than other linear methods on cubic and exponential models (SNR = 2, 5 and 10; $p<0.05$). As an unsupervised method, PCA-DA always had the worst performance among all three methods under test. For better illustration, we only showed the pair-wise statistical comparisons between SIR-DA and other methods in Fig. 3. We also conducted simulation in the canonical way and compared the results of two simulation approaches. Details of the comparison can be referred to Supplementary Fig. 7.

5.1.3. Performance of parameter selection

Fig. 4 summarizes the influence of M (the number of slices) on the performance of dimension reduction (in terms of AC or SCC) in all six models and under four SNR conditions. In most circumstances, SIR-DA had stable performance within a wide range of number of slices. Particularly, for SNR = 1, the performance decreased with increasing M. It is possible that less samples (more slices) within each slice may lead to unreliable estimation of slice means when the noise is large, and consequently hampered the performance. Fig. 5 summarizes the influence of M on MAE of prediction using SIR-DA in six models and under four SNR conditions. In all circumstances, MAE was decreased with increasing M, and

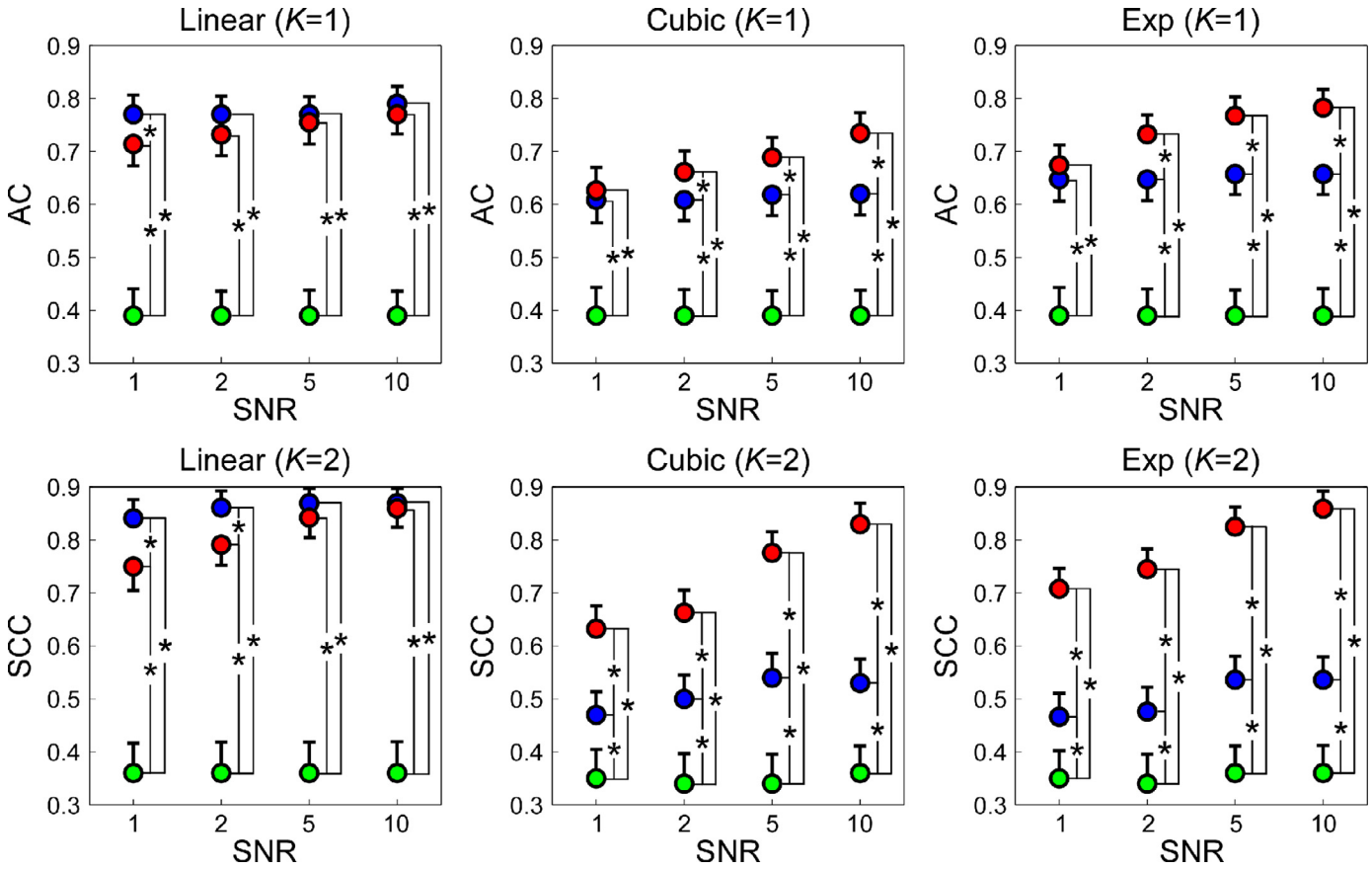


Fig 2. Performance of dimension reduction measured by AC and SCC. Each circle represents the mean value from 100 Monte Carlo simulations and the error bar represents standard deviation (STD). Significant differences between SIR (red dots), PLS (blue dots) and PCA (green dots) in different scenarios are indicated using asterisks ($p < 0.05$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tended to be stable for large number of slices ($M > 10$). Based on Figs. 4 and 5, we conclude that the performance of SIR-DA is not sensitive to the selection of M. We can also see from Figs. 4 and 5 that, varying M may lead to a minor tradeoff between dimension reduction and decoding when SNR = 1. This may be due to two reasons: (1) different degrees of influences of noise on dimension reduction and decoding; (2) the influence of M on the performance of SIR, but it is still difficult to quantify the influences of noise and the parameter M. We also provided the performance of SIR-DA in terms of dimension reduction and decoding using different number of slices ($M = 5$ and $M = 15$) in the Supplementary Tables S1–S4.

5.2. Real data results

5.2.1. Psychophysics

Laser stimuli of the four energies elicited graded subjective pain intensities (E1: 2.9 ± 1.5; E2: 3.8 ± 1.7, E3: 5.7 ± 1.6, and E4: 6.9 ± 1.5). One-way repeated-measures ANOVA revealed that the intensity of pain was significantly modulated by stimulus intensity ($F_{(3,93)} = 163.51$, $p < 0.001$). Post-hoc pairwise comparisons revealed that the intensity of pain was significantly larger at higher stimulus energies ($p < 0.001$ for all comparisons).

5.2.2. Pain-related brain patterns

The identified pain-related brain patterns ($P_{FDR} < 0.05$) using GLM and SIR-DA are summarized in Fig. 6. In the top left panel, GLM detected that pain perception was positively modulated by BOLD signals in cerebellum (CB), insula, thalamus, anterior cingulate cortex (ACC), primary somatosensory cortex (S1), secondary somatosensory cortex (S2), mid-cingulate cortex (MCC) and supplementary motor area (SMA). Group level BOLD time courses in some representative regions (ACC, S1 and insula) at different painful levels are displayed in the bottom left panel. We checked the relationships between BOLD magnitudes at peak scan (TR = 3) in ACC/S1/insula and pain perception by a linear fitting. P values from the linear fitting (ACC: $p = 0.009$; S1: $p = 0.013$; insula: $p = 0.015$) indicate possible linear relationships between pain perception and BOLD responses at these regions (in the linear regression model, the independent variables are the mean values of BOLD magnitudes for different pain levels and the dependent variables are pain ratings). In contrast, SIR-DA detected that pain perception was positively activated in CB, insula, thalamus, ACC, S1, S2, MCC and SMA, while negatively activated in mPFC, posterior cingulate cortex (PCC) and bilateral angular gyrus (Fig. 6, *top right panel*). Group-level BOLD signal time courses in ACC, S1, and mPFC at different painful levels are displayed in the bottom right panel of Fig. 6. P values from the linear fitting (ACC: $p = 0.012$; S1: $p = 0.016$; mPFC: $p > 0.05$) indicates possible linear relationships between pain perception and BOLD responses at ACC and S1 and the nonlinear relationship between pain perception and BOLD responses at mPFC.

5.2.3. Pain prediction

Fig. 7 summarizes the prediction performance by the six methods. For binary prediction, SVM, PCA-DA, PLS-DA, LASSO-PCR, elastic net and SIR-DA achieved classification accuracies of $71.52 \pm 2.48\%$, $70.98 \pm 2.52\%$, $74.06 \pm 2.31\%$, $73.39 \pm 2.00\%$, $73.05 \pm 2.40\%$ and $77.61 \pm 1.69\%$ (mean ± standard error of the mean [SEM]) respectively. One-way repeated measures ANOVA

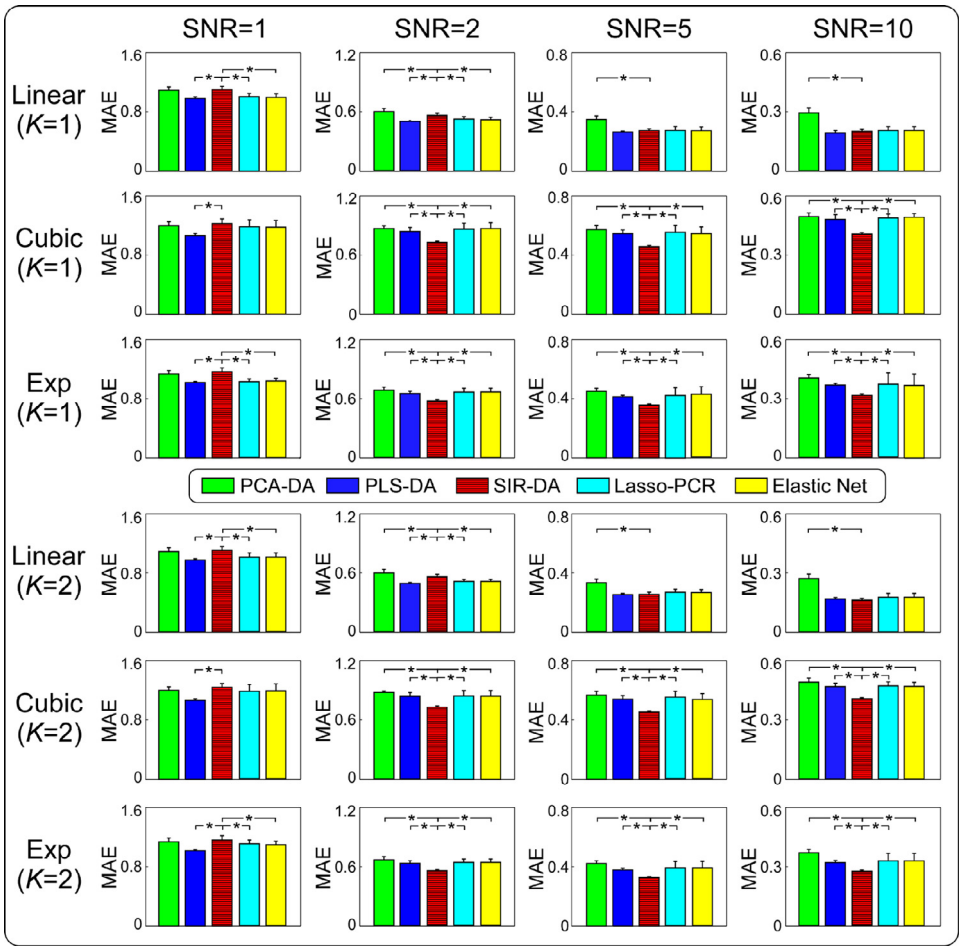


Fig 3. The prediction performances of different methods measured by MAE. Each color bar represents the mean value from 100 Monte Carlo simulations and the error bar represents STD. Significant differences between SIR-DA, PLS-DA and PCA-DA in different scenarios are indicated using asterisks ($p < 0.05$). For linear models, SIR-DA showed comparable results to PLS-DA at higher SNR. In contrast, SIR-DA had significantly better performance than PLS-DA on cubic and exponential models (SNR = 2, 5 and 10; $p < 0.001$; paired t-test). As an unsupervised method, PCA-DA always had inferior performance than the other two supervised methods.

revealed significant difference in classification accuracy between the six methods ($F_{(5,155)} = 5.33$, $p = 0.002$). Post hoc paired t-test showed that our proposed approach, SIR-DA, provided the best prediction performance. For continuous prediction, SVR, PCA-DA, PLS-DA, LASSO-PCR, elastic net and SIR-DA had MAE of 1.87 ± 0.12, 1.88 ± 0.11 1.73 ± 0.10, 1.75 ± 0.12, 1.76 ± 0.11 and 1.62 ± 0.09, respectively. One-way repeated measures ANOVA also revealed significant difference in MAE between six methods ($F_{(5,155)} = 12.22$, $p < 0.001$). Post hoc paired t-test showed that SIR provided the best prediction performance

6. Discussion

Dimension reduction is important in fMRI decoding for extracting an essential set of meaningful neural features to predict mental states or behavior responses from high-dimensional fMRI data. Conventional dimension reduction methods used in fMRI decoding do not use class labels or simply assume the relationship between class labels and fMRI data is linear, so they cannot always detect the most e.d.r. directions. To address this problem, we introduced the SIR-DA approach to accomplish supervised dimension reduction even when the complex relationship between fMRI data and class labels is not explicitly known. Results on simulated data and real data showed that the SIR-DA can more correctly estimate the e.d.r. direction(s) for dimension reduction and more accurately predict behavior responses for both linear and nonlinear fMRI prediction models.

6.1. Significance of SIR-DA

The relationship between neural activities and behavioral responses is complicated, and so dimension reduction method based on unjustified linear assumption could be sub-optimal. Nonlinear effects in the brain are actually very common. Such nonlinearities are believed to arise from both vascular response [25] and neuronal activity [26], and may be present between tasks and neural activations, between neural activity and BOLD response, or between tasks/behavior response and BOLD response [27]. Some studies have mentioned that the linear assumption for dimension reduction has some limitations, such as creating prediction error and causing unreliable brain activation [28,29], and nonlinear methods were suggested in a few studies [30,31]. In the present study, a new and general SIR-DA approach is developed to deal with the nonlinearity between behavior response and BOLD response.

The proposed SIR-DA approach is very useful because it allows the relationship between behavior responses and brain activity to be implicitly modeled. Due to the complex neurovascular mechanism, it is often difficult to formulate a clear hypothesis to model the relationship between behavior responses and brain activity. Our results showed that SIR-DA can well handle various types of

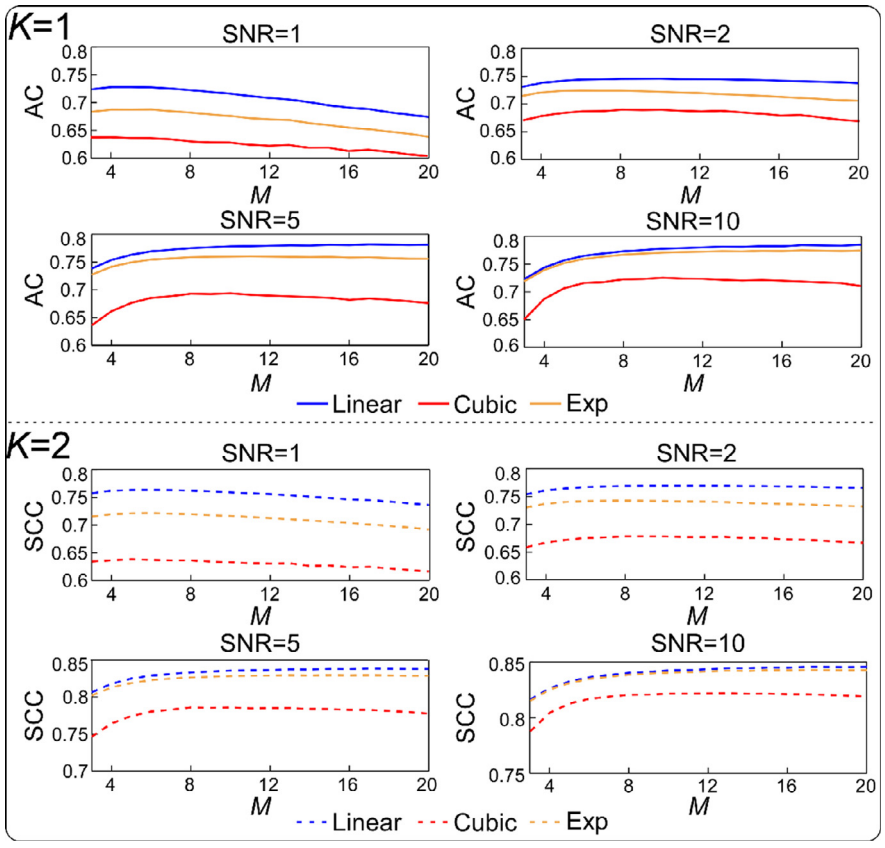


Fig 4. The performance of dimension reduction with SIR-DA using different numbers of slices in all six models and under four SNR conditions. In most circumstances, SIR-DA had stable performance within a wide range of numbers of slices.

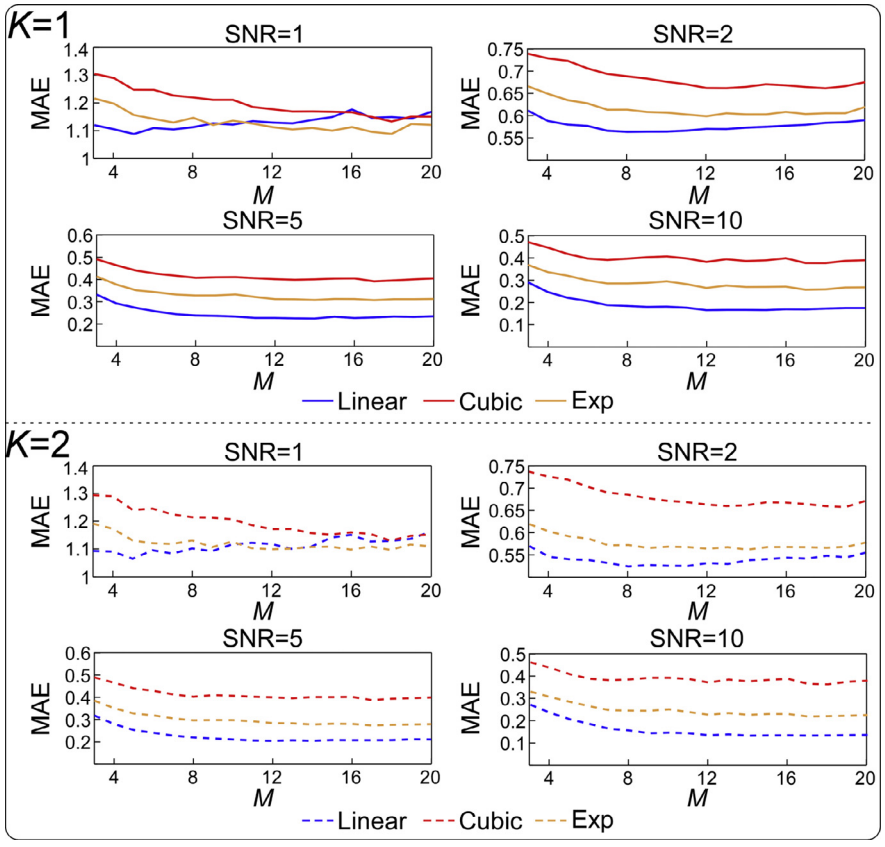


Fig 5. MAE of prediction using SIR-DA in six models and under four SNR settings. In all simulation circumstances, MAE was generally reduced with an increase in M, and tended to be stable for larger number of slices ($M > 10$).

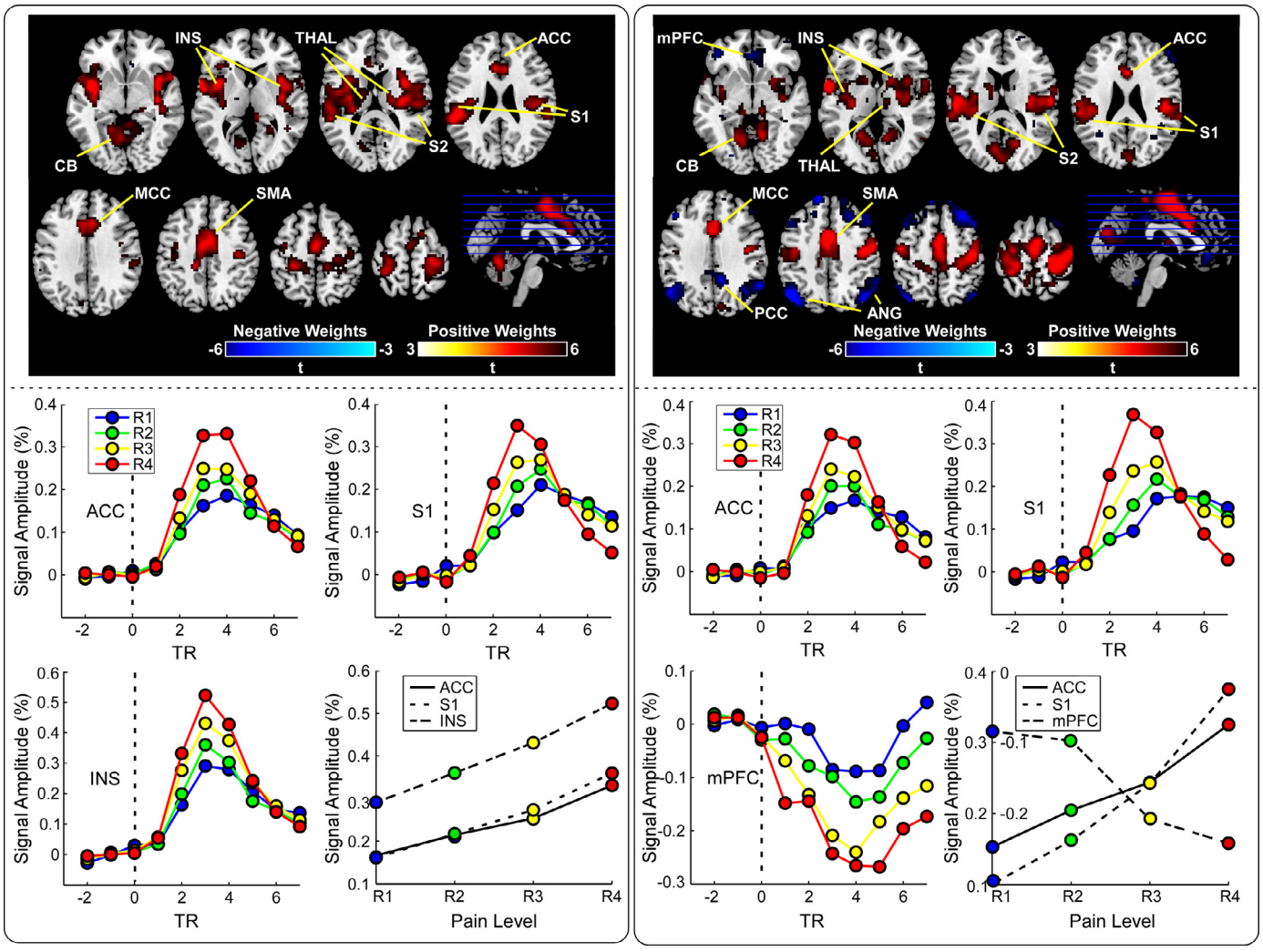


Fig 6. Pain-related brain regions and BOLD responses. *Top left panel*: Pain related brain regions detected by GLM ($P_{FDR} = 0.05$). BOLD responses in cerebrum (CB), insula (INS), thalamus (THAL), anterior cingulate cortex (ACC), primary somatosensory cortex (S1), secondary somatosensory cortex (S2), medial cingulate cortex (MCC) and supplementary motor area (SMA) were positively correlated with pain perception. *Bottom left panel:* Time series of BOLD responses in ACC, S1, INS and the relationships between BOLD responses at peak time (TR = 3) in these regions and pain intensity. *Top right panel*: Pain related brain regions detected by SIR-DA ($P_{FDR} = 0.05$). BOLD responses in CB, INS, THAL, ACC, S1, S2, MCC and SMA were positively correlated with pain perception, while in medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC) and angular gyrus (ANG) were negatively correlated with pain perception. *Bottom right panel*: Time series of BOLD responses in ACC, S1, mPFC and the relationship between BOLD responses at peak time (TR = 3) in these regions and pain intensity. The time courses of PCC and ANG were shown in Supplementary Fig. 6.

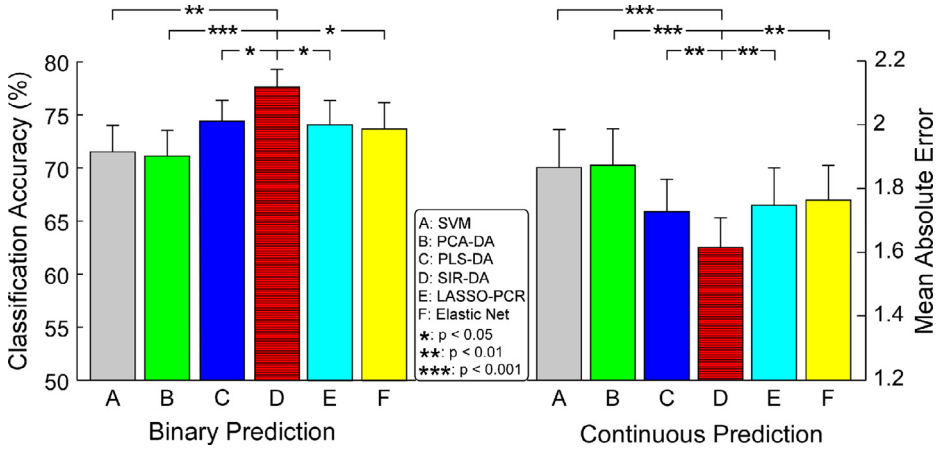


Fig 7. Performance of six methods in pain prediction. Each color bar represents the mean prediction performance across all subjects and the error bar represents SEM. Significant differences between SIR-DA and other methods are indicated by asterisks.

relationships (linear and nonlinear) between fMRI data and class labels. Therefore, this study provides a new data-driven approach for uncovering unknown brain patterns associated with behavior or cognitive responses.

Other supervised nonlinear dimension reduction methods also exist, such as kernel PLS (KPLS) [32], and they model nonlinear relationships in a different way. SIR assumes the relationship between features (in the original space as well as in the new reduced space) and class labels is nonlinear, but the transformation linking the reduced space with the original space is linear; while KPLS links the new feature space and the original space with a nonlinear transformation through an inner product operation, but it still assumes that the relationship between features in the new feature space and class labels is linear (so that the relationship between features in the original space and classes is still nonlinear). SIR has the advantage of being more suitable for interpreting the fMRI decoding result, because it enables the estimation of the contributing weight of each voxel in the original space from that in the reduced space since the relationship between these two spaces is linear. But it is hard for kernel-based methods to construct contributing weights of original fMRI voxels because new features are related nonlinearly to original features through inner product.

Because SIR can efficiently reduce the dimensionality under nonlinear relationship, it is possible to incorporate a nonlinear classifier into SIR-DA to boost its performance. However, the performance of a nonlinear classifier varies for it depends on how well its model matches the nonlinearity of the fMRI-class relationship. Searching for a most suitable nonlinear classifier for a specific application is outside the scope of this study, so we did not perform a more comprehensive comparison between different classifiers in this study.

6.2. Implementation of SIR-DA

In contrast to many sophisticated nonlinear dimension reduction techniques, such as projection pursuit [33], mutual information maximization [34] and kernel dimension reduction [35], SIR is very simple to implement. Two parameters, namely the number of slices and the number of e.d.r. directions, need to be prescribed for SIR-DA. It has been shown that the estimation of e.d.r. directions is not very sensitive to the choice of the number of slices [9,14]. For the number of e.d.r. directions, chi-squared test and BIC were suggested to determine the dimension [9,14],. In simulation results, we have shown that the performance of SIR-DA was stable within a wide range of the number of slices. But the number of slices may still affect the tradeoff between reliability of slice mean estimation and adequacy in the number of slice means for robust inverse regression curve estimation (as explained in Section 2.3). For the experimental results on real data, we empirically selected the number of slices as the levels of laser stimulation ($M = 4$), and set the number of e.d.r. directions as 2 since the first leading two eigenvectors had captured more than 90% variance of the data.

SIR is also computationally inexpensive. The computational time to find e.d.r. directions for one subject was around 0.58 s (averaged from 1000 simulations; dimension of fMRI data: $40 \times 71{,}450$ [trials $\times$ voxels]; PC configuration: Intel Core i7-2600 CPU @ 3.4 GHz and 16 GB RAM; software: MATLAB 7.1). Thus, SIR has the potential to be used for real-time neuroimaging data analysis.

SIR can achieve its maximum performance in finding the e.d.r. when brain-behavior relationships are monotonic, which holds true in most cases of fMRI research. The performance of SIR would be degraded when brain-behavior relationships are not monotonic [36]. Possible solutions (improved versions of SIR for non-monotonic relationships) have been reported in some studies, such as, second moment based sliced average variance estimate (SAVE) [36] and principle Hessian directions (pHd) [37]. These improved versions of SIR could be considered for fMRI decoding in future studies.

6.3. SIR-DA for pain prediction

The possibility of pain prediction using neuroimaging techniques including EEG and fMRI has been widely explored in recent years [6,38–41]. Due to the high spatial resolution, whole-brain fMRI technique could not only reveal spatially distributed patterns of pain perception, but also provide a relatively high prediction accuracy [6]. Machine learning approaches have been adopted in fMRI-based pain prediction studies, because it can reveal important brain patterns (spatial pattern in fMRI) without sophisticated feature extraction processes. However, the performance of fMRI decoding has been hampered by the ultra-high dimension of prediction variables. Including all whole-brain voxels in a classifier (e.g., SVM) is not only computationally expensive but also has the risk of overfitting. Therefore, dimension reduction techniques should be applied to extract most predictive patterns or feature sets before prediction.

However, the relationship between pain perception and brain patterns may not be linear. In the present study, SIR-DA and other linear methods identified a series of patterns of BOLD responses related to pain perception. Among them, insula, thalamus, ACC, S1, S2, MCC and SMA have been widely reported and they are commonly referred to as 'pain matrix' [42]. Notably, compared to linear methods, SIR-DA found the contribution of default model network (DMN), including mPFC, PCC and angular gyrus, in the prediction of pain. Such negative relationship between BOLD activity in DMN and behavior responses have been widely studied, and confirmed the indispensable role of DMN in the modulation of perception [8,43–45]. More importantly, SIR-derived feature sets provided the best pain prediction performance (Fig. 6), indicating that conventional linear methods may not fully detect the contribution from brain patterns nonlinearly correlated with pain perception.

FMRI-based pain prediction has been reported in a few studies. Brodersen et al., showed a prediction accuracy of 61% using a near-threshold experimental design [40], and Marquand et al., reported results for classifying three different levels of painful stimuli with accuracies ranging from 68.3% to 91.7% (the highest accuracy was obtained from classifying nonpainful and maximal tolerance painful stimuli) [38]. Although it is not appropriate to directly compare the prediction accuracy of this study with others because of different pain modalities and experimental settings, the prediction accuracy of 77.61% for binary prediction and MAE of 1.62 for continuous prediction achieved in this study are considered satisfactory. As a future development, the predication accuracy could be further improved by incorporating other pain-related variables (such as demographic parameters, resting-state neuroimaging data, etc.) into the prediction model, with an aim that, ultimately, the fMRI-based pain prediction method could be used in clinical practices.

Acknowledgment

We thank Prof. L.X. Zhu from Department of Mathematics, The Hong Kong Baptist University, for his inspiration and valuable suggestions.

Funding

This work was supported by National Natural Science Foundation of China (61640002), Shenzhen Peacock Plan (NO.KQTD2016053112051497) and a Hong Kong RGC GRF Grant (HKU 785913 M).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neucom.2017.07.045.

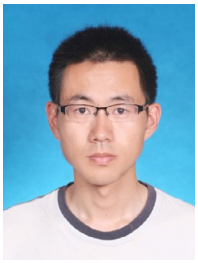
References

[1] K.A. Norman, S.M. Polyn, G.J. Detre, J.V. Haxby, Beyond mind-reading: multi-voxel pattern analysis of fMRI data, Trends Cognit. Sci. 10 (2006) 424–430.
[2] J.V. Haxby, Multivariate pattern analysis of fMRI: the early beginnings, Neuroimage 62 (2012) 852–855.
[3] I. Jolliffe, Principal Component Analysis, Wiley Online Library, 2002.
[4] J. Sui, T. Adali, Q. Yu, J. Chen, V.D. Calhoun, A review of multivariate methods for multimodal fusion of brain imaging data, J. Neurosci. Methods 204 (2012) 68–81.
[5] Y. Li, P. Namburi, Z. Yu, C. Guan, J. Feng, Z. Gu, Voxel selection in fMRI data analysis based on sparse representation, IEEE Trans. Biomed. Eng. 56 (2009) 2439–2451.
[6] T.D. Wager, L.Y. Atlas, M.A. Lindquist, M. Roy, C.-W. Woo, E. Kross, An fMRI-based neurologic signature of physical pain, New Engl. J. Med. 368 (2013) 1388–1397.
[7] L. Hu, M. Cai, P. Xiao, F. Luo, G. Iannetti, Human brain responses to concomitant stimulation of Aδ and C nociceptors, J. Neurosci. 34 (2014) 11439–11451.
[8] M.L. Loggia, R.R. Edwards, J. Kim, M.G. Vangel, A.D. Wasan, R.L. Gollub, R.E. Harris, K. Park, V. Napadow, Disentangling linear and nonlinear brain responses to evoked deep tissue pain, Pain 153 (2012) 2140–2151.
[9] K.-C. Li, Sliced inverse regression for dimension reduction, J. Am. Stat. Assoc. 86 (1991) 316–327.
[10] J.J. Dai, L. Lieu, D. Rocke, Dimension reduction for classification with gene expression microarray data, Stat. Appl. Genet. Mol. Biol. 5 (2006) Article 6.
[11] L. Li, Dimension reduction for high-dimensional data, Stat. Method Mol. Biol. 620 (2010) 417–434.
[12] P.A. Naik, M.R. Hagerty, C.-L. Tsai, A new dimension reduction approach for data-rich marketing environments: Sliced inverse regression, J Mark. Res. 37 (2000) 88–101.
[13] F. Pereira, T. Mitchell, M. Botvinick, Machine learning classifiers and fMRI: a tutorial overview, Neuroimage 45 (2009) S199–S209.
[14] L. Zhu, B. Miao, H. Peng, On sliced inverse regression with high-dimensional covariates, J. Am. Stat. Assoc. 101 (2006) 630–643.
[15] J. Ashburner, K.J. Friston, Unified segmentation, Neuroimage 26 (2005) 839–851.
[16] M. Welvaert, Y. Rosseel, On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data, PloS One 8 (2013) e77089.
[17] Z.G. Zhang, S.C. Chan, Recursive parametric frequency/spectrum estimation for nonstationary signals with impulsive components using variable forgetting factor, IEEE Trans. Instrum. Meas. 62 (2013) 3251–3264.
[18] Y. Tu, Z. Zhang, A. Tan, W. Peng, Y.S. Hung, M. Moayedi, G.D. Iannetti, L. Hu, Alpha and gamma oscillation amplitudes synergistically predict the perception of forthcoming nociceptive stimuli, Hum. Brain Mapp. 37 (2016) 501–514.
[19] R.C. Coghill, J.G. McHaffie, Y.-F. Yen, Neural correlates of interindividual differences in the subjective experience of pain, Proc. Natl. Acad. Sci. USA 100 (2003) 8538–8542.
[20] R.J. Carroll, K.-C. Li, Measurement error regression with unknown link: dimension reduction and data visualization, J. Am. Stat. Assoc. 87 (1992) 1040–1050.
[21] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.
[22] G.A. Cecchi, L. Huang, J.A. Hashmi, M. Baliki, M.V. Centeno, I. Rish, A.V. Apkarian, Predictive dynamics of human pain perception, PLoS Comput. Biol. 8 (2012) e1002719.
[23] F. Konietschke, M. Pauly, Bootstrapping and permuting paired *t*-test type statistics, Stat. Comput. 24 (2014) 283–296.
[24] K.J. Friston, A.P. Holmes, K.J. Worsley, J.P. Poline, C.D. Frith, R.S. Frackowiak, Statistical parametric maps in functional imaging: a general linear approach, Hum. Brain Mapp. 2 (1994) 189–210.
[25] R.B. Buxton, E.C. Wong, L.R. Frank, Dynamics of blood flow and oxygenation changes during brain activation: the balloon model, Magn. Reson. Med. 39 (1998) 855–864.
[26] N.K. Logothetis, J. Pauls, M. Augath, T. Trinath, A. Oeltermann, Neurophysiological investigation of the basis of the fMRI signal, Nature 412 (2001) 150–157.
[27] B. He, L. Yang, C. Wilke, H. Yuan, Electrophysiological imaging of brain activity and connectivity—challenges and opportunities, IEEE Trans. Biomed. Eng. 58 (2011) 1918–1931.
[28] K.J. Friston, O. Josephs, G. Rees, R. Turner, Nonlinear event-related responses in fMRI, Magn. Reson. Med. 39 (1998) 41–52.
[29] T.D. Wager, A. Vazquez, L. Hernandez, D.C. Noll, Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies, Neuroimage 25 (2005) 206–218.
[30] K. Friston, J. Phillips, D. Chawla, C. Buchel, Nonlinear PCA: characterizing interactions between modes of brain activity, Philos. Trans. R. Soc. B 355 (2000) 135–146.
[31] M.J. McKeown, T.J. Sejnowski, Independent component analysis of fMRI data: examining the assumptions, Hum. Brain Mapp 6 (1998) 368–372.
[32] R. Rosipal, Kernel partial least squares for nonlinear regression and discrimination, Neural Netw. World 13 (2003) 291–300.
[33] N. Vlassis, Y. Motomura, B. Kröse, Supervised dimension reduction of intrinsically low-dimensional data, Neural Comput. 14 (2002) 191–215.
[34] K. Torkkola, Feature extraction by non parametric mutual information maximization, J. Mach. Learn. Res. 3 (2003) 1415–1438.
[35] K. Fukumizu, F.R. Bach, M.I. Jordan, Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces, J. Mach. Learn. Res. 5 (2004) 73–99.
[36] R.D. Cook, S. Weisberg, Sliced inverse regression for dimension reduction: comment, J. Am. Stat. Assoc. 86 (1991) 328–332.
[37] K.-C. Li, On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma, J. Am. Stat. Assoc. 87 (1992) 1025–1039.
[38] A. Marquand, M. Howard, M. Brammer, C. Chu, S. Coen, J. Mourão-Miranda, Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes, Neuroimage 49 (2010) 2178–2189.
[39] E. Schulz, A. Zherdin, L. Tiemann, C. Plant, M. Ploner, Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data, Cereb. Cortex 22 (2011) 1118–1123.
[40] K.H. Brodersen, K. Wiech, E.I. Lomakina, C.-s. Lin, J.M. Buhmann, U. Bingel, M. Ploner, K.E. Stephan, I. Tracey, Decoding the perception of pain from fMRI using multivariate pattern analysis, Neuroimage 63 (2012) 1162–1170.
[41] G. Huang, P. Xiao, Y. Hung, G.D. Iannetti, Z. Zhang, L. Hu, A novel approach to predict subjective pain perception from single-trial laser-evoked potentials, Neuroimage 81 (2013) 283–293.
[42] A. Mouraux, A. Diukova, M.C. Lee, R.G. Wise, G.D. Iannetti, A multisensory investigation of the functional significance of the "pain matrix", Neuroimage 54 (2011) 2237–2249.
[43] M.F. Mason, M.I. Norton, J.D. Van Horn, D.M. Wegner, S.T. Grafton, C.N. Macrae, Wandering minds: the default network and stimulus-independent thought, Science 315 (2007) 393–395.
[44] M.N. Baliki, P.Y. Geha, A.V. Apkarian, D.R. Chialvo, Beyond feeling: chronic pain hurts the brain, disrupting the default-mode network dynamics, J. Neurosci. 28 (2008) 1398–1403.
[45] M.L. Loggia, J. Kim, R.L. Gollub, M.G. Vangel, I. Kirsch, J. Kong, A.D. Wasan, V. Napadow, Default mode network connectivity encodes clinical pain: an arterial spin labeling study, Pain 154 (2013) 24–33.

Yiheng Tu received M.S. and Ph.D. degree in the Department of Electrical and Electronic Engineering at the University of Hong Kong, China, in 2012 and 2016 respectively. Now he is a research fellow in Massachusetts General Hospital and Harvard Medical School, and a visiting research fellow at Shenzhen University. His research interests include biomedical signal processing, statistical machine learning and neural engineering.

Zening Fu received Ph.D. degree in the Department of Electrical and Electronic Engineering at the University of Hong Kong, China, in 2016. He is now a postdoctoral fellow in Mind Research Network, Albuquerque, NW, USA. His research interests include biomedical signal processing and machine learning.

Ao Tan received B.S. degree in Automation from Sun Yat-Sen University in 2012. He got the Ph.D. degree from Department of Electrical and Electronic Engineering at the University of Hong Kong, in 2017. His research interests include biomedical signal processing and statistical machine learning.



Gan Huang received his Bachelor and Master degree from Department of Mathematics at Southeast University, China, in 2005 and 2008, respectively. He got the Ph.D. degree from Shanghai Jiao Tong University, China, in 2013. From 2013 to 2016, he was a postdoctorate researcher in Institute of Neuroscience, Universite catholique de Louvain, Belgium. He is currently an assistant professor in Shenzhen University, China. His research interests include biomedical signal processing, brain-computer interface and neuroprosthetics.

Li Hu received his B.Eng and M.Sc from Tianjin University, China, in 2005 and 2007 respectively. He got the Ph.D. degree from the University of Hong Kong, in 2010. In 2010 to 2011, he was a visiting research scholar in University College London. He has worked in Southwest University, China from 2011 to 2016, before joined Institute of Psychology, Chinese Academy of Sciences, where he is currently a professor. His research interests include biomedical signal processing, neural engineering and pain neuroscience.

Y. S. Hung received his B.Sc. (Eng) in Electrical Engineering and B.Sc. in Mathematics, both from the University of Hong Kong, and M.Phil. and Ph.D. degrees from the University of Cambridge. He has worked at the University of Cambridge and the University of Surrey in U.K. before he joined the University of Hong Kong, where he is currently a professor at the Department of Electrical and Electronic Engineering. His research interests include biomedical engineering, computer vision and control systems.

Zhiguo Zhang obtained his Ph.D. degree in the Department of Electrical and Electronic Engineering at the University of Hong Kong in 2008. He was a Research Assistant Professor at the University of Hong Kong from 2012 to 2014 and an Assistant Professor at Nanyang Technological University from 2015 to 2016. He is now a Professor with the School of Biomedical Engineering, Health Science Center at Shenzhen University, Shenzhen, China. His research interests include biomedical signal processing, medical image analysis, neural engineering, and computational neuroscience.