UNIMORE
Potenzialità e innovazione nella ricerca biomedica: approcci interdisciplinari
26 giugno 2018 | Via del Pozzo 71 – Modena

Numerical methods for the identification of central biomarkers of fibromyalgia

Mathide Galinier[1,2], Giuseppe Pagnoni[3], Carlo Adolfo Porro[3], Marco Prato[1,2]

1. Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Italy
2. Gruppo Nazionale per il Calcolo Scientifico (GNCS) – Istituto Nazionale di Alta Matematica (INdAM), Italy
3. Dipartimento di Scienze Biomediche, Metaboliche e Neuroscienze, Università di Modena e Reggio Emilia, Italy

OASIS Optimization Algorithms and Software for Inverse problemS

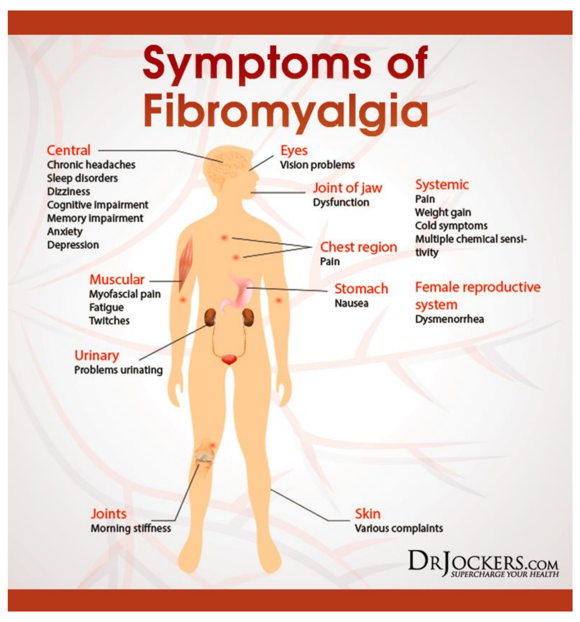
Abstract

This proposal aims to investigate the putative suitability of cerebral functional connectivity patterns obtained by non-invasive resting state Blood-Oxygenation-Level-Dependent (BOLD) functional magnetic resonance imaging (fMRI) as biomarkers for chronic pain states. In a preliminary analysis, we used a dataset of 27 fibromyalgic patients and 21 healthy controls (all females, age- and education-matched) in a machine-learning computational framework, aimed at automatic classification of subjects into the patient/control categories. The classification and predictive algorithm use as input functional connectivity matrices from a 120-node parcellation of the whole cerebral cortex, thus allowing an anatomically unbiased investigation and a reasonably high degree of spatial specificity. Different feature selection techniques are investigated with the aim of detecting which connections are mainly responsible for discriminating patients from healthy controls. The design of appropriate predictors based on this subset of relevant features is the final output of the project.

The dataset: resting-state BOLD-fMRI images

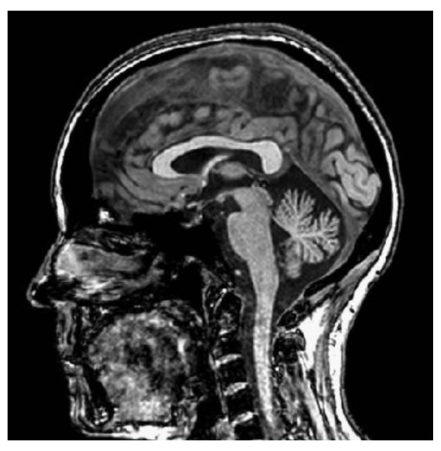
Few words about fibromyalgia

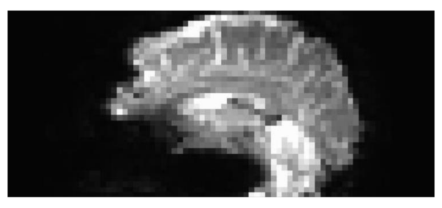
- ✓ One of the most common chronic pain disorders
- ✓ Debilitating pain syndrome associated with other symptoms such as fatigue, sleep disturbances, cognitive difficulties and stiffness
- ✓ Not fully understood
- ✓ Prevention, diagnosis and therapy very challenging



Acquisition and preprocessing of fMRI data

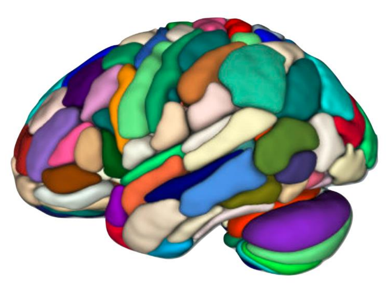
- ✓ Acquisition of an anatomical image (T1-weighted) and of a sequence of functional images by EPI (T2*-weighted)
- ✓ Preprocessing of the data (using AFNI):
  - ✓ Slice timing
  - ✓ Head motion correction
  - ✓ Alignment of EPI and anatomical images
  - ✓ Normalization of the anatomical image
  - ✓ Application of the normalization to EPI
  - ✓ Regression step to clean the signal





Parcellation and creation of computation matrix

- ✓ Use of an anatomical parcellation of the entire brain[1]: AAL2
- ✓ Computation of the first singular vector for each region of the parcellation
- ✓ Computation of the correlation matrix



- ✓ Our dataset:
  - ✓ 48 subjects : 21 controls, 27 patients with fibromyalgia
  - ✓ For each subject : a 120 x 120 correlation matrix

Feature selection approach

Neighbourhood component analysis (NCA)

- ✓ Learns the feature weights by using a diagonal adaptation of neighbourhood component analysis (NCA) with regularization[2]
- ✓ Training set:

$$S = \{(\boldsymbol{x}_i, y_i) \in \mathbb{R}^p \times \{1, \dots, C\} : \ i = 1, \dots, n\}$$

- ✓ Consider a randomized classifier which labels an example x using the label of its reference point $Ref(\boldsymbol{x}) \in S$, picked at random
- ✓ Assume that $p\big(Ref(\boldsymbol{x}) = \boldsymbol{x}_j \,|\, S\big) \propto \kappa\left(d_W(\boldsymbol{x}, \boldsymbol{x}_j)\right)$, being

$$d_W(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{r=1}^{p} w_r^2 |x_{ir} - x_{jr}| \quad ; \quad \kappa(z) = \exp\left(-\frac{z}{\sigma}\right) \quad ; \quad \sigma > 0$$

- ✓ Define the probability $\mathbb{P}\big(Ref(\boldsymbol{x}) = \boldsymbol{x}_j \,|\, S\big) = \frac{\kappa\left(d_W(\boldsymbol{x}, \boldsymbol{x}_j)\right)}{\sum_{k=1}^{n} \kappa\left(d_W(\boldsymbol{x}, \boldsymbol{x}_k)\right)}$
- ✓ Leave-one-out (LOO) probability of correct classification, i.e., the probability p_i that the randomized classifier correctly predict observation i using $S^{-i} \equiv S \setminus \{(\boldsymbol{x}_i, y_i)\}$ $\ (i = 1, \dots, n)$:

$$p_i = \sum_{j=1, j \neq i}^{n} \mathbb{P}\big(Ref(\boldsymbol{x}_i) = \boldsymbol{x}_j \,|\, S^{-i}\big) \times \delta_{ij} \quad ; \quad \delta_{ij} = \begin{cases} 1 & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases}$$

- ✓ Approximate LOO classification accuracy: $\xi(\boldsymbol{w}) = \sum_{i=1}^{n} p_i$
- ✓ In order to perform feature selection and alleviate overfitting, a Tikhonov regularization term is added

$$\xi(\boldsymbol{w}) = \sum_{i=1}^{n} p_i - \lambda \sum_{r=1}^{p} w_r^2$$

- ✓ Gradient ascent strategy with adaptive stepsize exploited to find the weights $\boldsymbol{w}^*$ which maximize the cost function ξ

Numerical tests

Dataset info

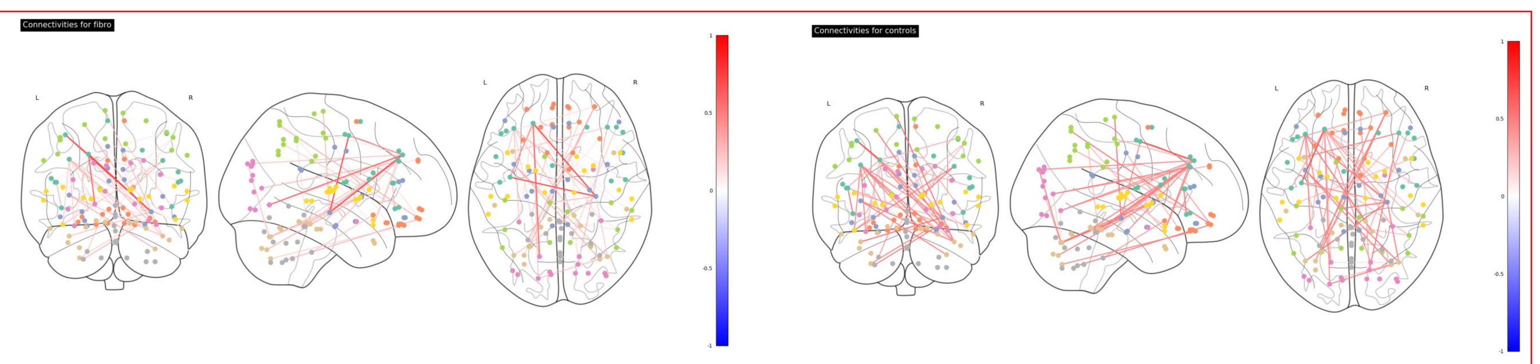
- ✓ Support Vector Machines (SVM) used to build the classifier
- ✓ Dataset divided into training set (33 subj, ~70%) and test set (15 subj, ~30%)
- ✓ Different random subdivision training/set performed to obtain stable results
- ✓ For each subject :
  - ✓ 7140 features (upper triangular part of the correlation matrix)
  - ✓ two labels (-1 if control, 1 if fibromyalgia-affected patient)

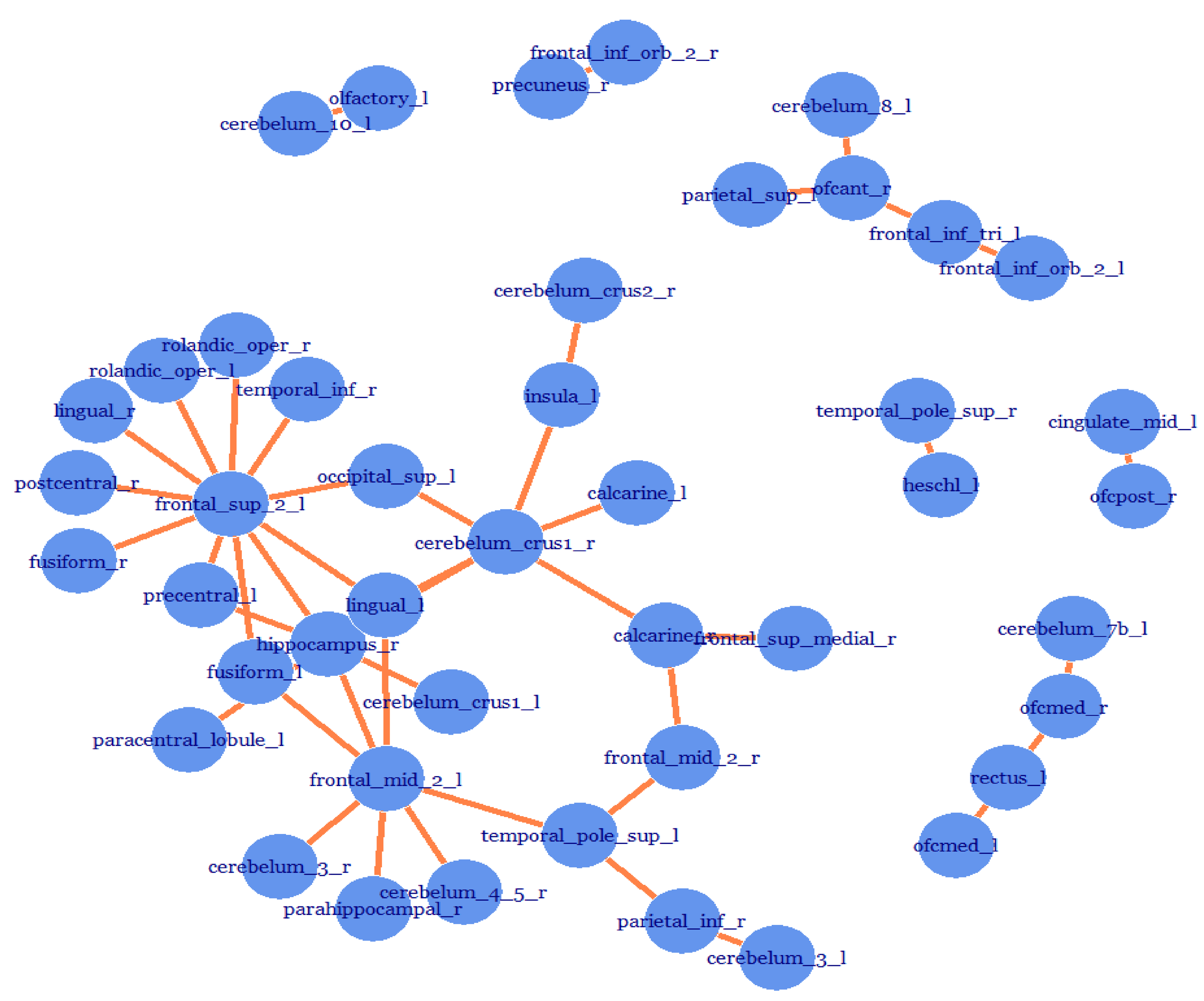
Looking for robust features and building the final classifier

- ✓ NCA feature selection applied for different values of σ and λ
- ✓ Features corresponding to the 150 highest weights intersected with those obtained with MATLAB kNN-based relieff function
- ✓ Resulting 44 features used as input to a Gaussian SVM → final prediction accuracy between 80% and 100% depending on the distribution training/test sets (average: **87.4%**)

Future work

- ✓ Compare/intersect results with other feature selection strategies
- ✓ Increase the number of examples by including patients/controls data from other databases
- ✓ Extend the analysis to predict individual clinical fibromyalgia scores
- ✓ Apply same approach to fMRI data collected in economic decision-making experiments





References

[1] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**(1), 273-289

[2] Yang, W., K. Wang, W. Zuo (2012). Neighborhood Component Feature Selection for High-Dimensional Data. *Journal of Computers* **7**(1), 161-168