Neuroimaging as an intermediate phenotype to bridge the gap between clinic and genetic: Machine learning methods

Edouard Duchesnay

Edith Le Floch, Vincent Frouin, Bertrand Thirion, JB Poline

NeuroSpin, LNAO, CEA Saclay
Overview

Genetic (DNA array)  Neuroimaging (MRI/PET)  Final phenotype
Clinic/behavioural

Principle
Imagery as an intermediate phenotype

Example of exploratory data strategy
Step 1: Neuroimaging to clinic
→ identify neuroimaging-based intermediary phenotypes
Step 2: Genetic to neuroimaging-based intermediary phenotypes
→ identify genetic markers (link to pathways etc.)

Problem
Vast amount of biological measurements:
- Neuroimaging (~10^6 voxels), DNA array (~10^6 SNPs)
→ Spurious associations between: genetic x imaging x clinic
→ Poor reproducibility (multiple comparison/over-fitting issues)
Outline

Neuroimaging (PET) \rightarrow Final phenotype Clinic/behavioural

Application to autism

Genetic (DNA array) \rightarrow Neuroimaging (fMRI)

Application: asymmetries in language processing
Neuroimaging to predict the clinical status

Neuroimaging (PET) → Final phenotype Clinic/behavioural

Application to autism

Genetic (DNA array) → Neuroimaging (fMRI)

Application: asymmetries in language processing
Brains variability

- Find the brain variability associated to a clinical trait
- How to compare brains?

Ideal scenario:
1) Neuroimaging measurements (anatomical/diffusion/functional-MRI)
2) Re-align brains: remove non-specific variability
3) Do Machine learning on the specific variability
Features extraction strategies

Volume of Interest (VOI or ROI)
- Manually defined
- Template-based

Structure identification
- Automatically defined
- Manually defined
=> **Structural data**

Brain warping
- Warp brain toward a common template (coordinate system)
=> **Iconic data**
ML in neuroimaging

**Classical voxel-wise analysis**
- Analyse each brain locus **Independently**: Measure the brain/clinic association
  - Similar to (GWAS) Genome-wide association study
  - **Overlook** brain-brain interaction
  - **No individual** classification
  - Multiple comparison issues

**Multivariate classification**
- Find a global mapping from the brain to the clinic
  - Improved sensitivity
  - Consider brain-brain interactions
  - Over-fitting issue: need careful validation

**Goal:**
(1) **Imaging profile** that covariate with phenotype (clinical severity)
(2) Individual **computer aided diagnosis**
(3) Individual predictor of **response to treatment**
Data: Iconic / Structural

Iconic (image based) methods

Registration & Pre-processing:

- aMRI
- Segmentation
- FA map
- DTI Qball
- GLM
- fMRI
- PET
- Global scaling

Structural methods

Automatic structures identification

- aMRI
- Sulci/Gyri identification
- Fibre identification

(work in progress)

#features

- aMRI: $\sim 10^{5}, 10^{6}$
- DTI Qball: $\sim 10^{5}$
- GLM: $\sim 10^{5}$
- Global scaling: $\sim 10^{5}$

#features

- $\sim 10^{3}, 10^{4}$
Structural Based Morphometry: summary

- Grey/White segmentation
- Sulci extraction
- Automatic sulci identification
- Machine learning

Central sulcus
Inferior pre-central sulcus
Posterior and anterior parts of the right inferior temporal sulcus
Anatomical MRI: compare sulci measurements

Superimposed sulci of 70 subjects

Compare the same anatomical structures (sulci) across all subject without registration (just linear normalization)
Classification based on sulci: predict gender

Dimension reduction
1/T-test
2/SFS objective function: cross-validation error
3/number of support vectors

Correct prediction rate: 85%

[Duchesnay et al. IEEE-TMI 2007]
Guess the gender from sulci

Correct recognition rate: 85%.
Guess left/right hemisphere from sulci

Use only “size” descriptors: sulcus length, depth, surface

Correct recognition rate : 96%.
ASD subject prediction based on PET-scans

rCBF PET scans of 45 low-functioning ASD and 13 low-functioning non-ASD

Left-out sample cross-validation

Cerebral Blood Flow
PET scan

Learn mapping

Left out
samples

Apply mapping

Assess the generalisation power of the learning algorithm on independent data (toward reproducibility)
**Linear classifier**

**Prediction** rule of linear discriminant classifier (**combine** features):

\[
\text{Link function} (\begin{array}{c}
\text{weight 0} \\
\times \\
\text{feat. 1}
\end{array} + \begin{array}{c}
\text{weight 1} \\
\times \\
\text{feat. 2}
\end{array} + \cdots + \begin{array}{c}
\text{weight P} \\
\times \\
\text{feat. P}
\end{array}) = \text{predicted target}
\]

**Learn:** How to learn \( \mathbf{w} \) the weight vector such:

\[
\text{Link function} (\begin{array}{c}
\mathbf{X} \\
\times \\
\mathbf{w}
\end{array}) = \mathbf{Y}
\]

- Train data (images)
  - \( p \) number of features \( \sim 10^5 \)
  - \( n \) number of samples \( \sim 100 \)

- Predicted target
- True target
Linear classifier: Parametric method

Estimate model parameters:
- **Means** $\mu_1, \mu_2$
- "Dispersion" (Var./Covar. matrix) (Within covariance matrix)

Discriminant projection:
$$\Sigma^{-1} (\mu_1 - \mu_2)$$

Probabilistic generative (LDA)
$$p(y_i = g | X_i, \theta) = \frac{\pi_g \mathcal{N}(x_i | \mu_g, \Sigma)}{\sum_{g \in \{1,2\}} \pi_g \mathcal{N}(x_i | \mu_g, \Sigma)}$$
Linear classifier: discriminative approaches

Find \( \mathbf{w} \) that minimize a prediction error on training data:

\[
\|\mathbf{w}\|_2^L + C \sum_i \xi(\mathbf{w}; \mathbf{x}_i, y_i)
\]

Penalization Loss (Error) function between prediction And true label

**L2 penalization: SVM**
- \( L = 2 \)
- Hinge loss: \( \xi(\mathbf{w}; \mathbf{x}_i, y_i) = \max(1 - y_i \mathbf{x}_i' \mathbf{w}, 0) \)

**L1 penalization: Lasso Logistic Regression**
- \( L = 1 \)
- Logistic loss: \( \xi(\mathbf{w}; \mathbf{x}_i, y_i) = \log(1 + e^{-y_i \mathbf{x}_i' \mathbf{w}}) \)

→ Minimisation of misclassification: favour most numerous class
→ Poor specificity
1) **Samples re-weighting** simple for both SVM and Lasso Logistic Regression
→ Good sensitivity (detection of the most numerous class)
→ **Poor specificity** (detection of the least numerous class)

2) **Sub-sampling** of the most numerous class: can afford to drop some of the few 45 samples of ASD group

3) **Two separate one class learning**
   ~ **generative methods** ie.: learn the conditionals $p(x_i|y_i)$.
   The predictive = conditional * explicit priors
   → Linear Discriminant Analysis (**LDA**)  

Pb.: Overfitting $P(P+5)/2$ estimated parameters (intraclass variance)

→ **Dimension reduction**
Dimension reduction: 1/2 regional features

Goal driven regional feature extraction:
Univariate statistics (GLM)

Thresholding

Average signal within clusters
**Dimension reduction: 1/2 methods**

**Dimension reduction**
- Look for low dimensional data representation

**Unsupervised (data driven)** → **Feature extraction**
- Maximum image variability

**Supervised (goal driven)** → **Feature selection**
- Maximum image/target covariance

**Linear** (Max var.)
- PCA
- ICA

**Non linear** (Manifold learning)
- Isomap
- LLE
- Kernel PCA

**Univariate**
- Filters GLM
  - "Voxel based analysis"
  - "Genome Wide Assoc. Studies"

**Multivariate**
- Wrapper
  * RFE
  * SFFS
- Embedded
  * L1

Feature selection = Feature subset ranking + model selection

Feature subset ranking produce sets of features \( (F_k) \) of increasing size \( k \)
- Filter and RFE: nested sets are nested
- Lasso, SFFS: eventually non-nested sets
Model selection

Select $F_k$ that maximizes some criteria
Here Choose feature subset $F_k$ made of $k$ (regional) features

1) CV → Computational issue: 3 levels of nested sampling loop

2) Penalized likelihood

$$\ln p(y|X^k, F_k) \sim \ln p(y|X^k, \theta^k, F_k) - a \frac{1}{2} k \ln N$$

Evidence \hspace{1cm} Log likelihood \hspace{1cm} Penalisation

Many fixed penalty criteria BIC, AIC, etc.

Under penalization (ignore feature selection)

Data driven calibration of the penalty [Birgé 07]

Add a free parameter “$a$”

Calibrated with random permutation
Comparison methodology

Feature extraction

Dimension reduction

Model selection

Classification

Voxels
Regions

no
GLM
RFE
Lasso

CV
Penalized Likelihood

Lin SVM (L2)
Lasso Log. Reg (L1)
LDA
Results (Validation)

**ROC analysis**

Regional features + Multivariate feature selection + Generative

**Leave-One Out Cross validation**

Accuracy 87%**, Sensi. 91%**, Speci. 77%*

Significance calibrated with permutation (*** p<0.001, * p<0.05)
Discriminative pattern

hyper-perfusion in postcentral area (in ASD)

hypo-perfusion in Superior Temporal Sulcus (in ASD)

The combination of the signal in the two regions enable a clear separation of ASD vs controls

- **Good stability**: same pattern is selected across all re-sampling
- **Shared** pattern that discriminates all ASD from controls
- Multiple etiologies of ASD + numerous neuroimaging findings suggests that several others brain patterns may exist across the autistic spectrum
- Next step: look for the more specific multiple patterns associated with the multiple etiologies
Imaging-genetics

Neuroimaging (PET) → Final phenotype Clinic/behavioural

Application to autism

Genetic (DNA array) → Neuroimaging (fMRI)

Application: asymmetries in language processing

E. Duchesnay, I²BM/NeuroSpin
MLNI Workshop, Marseille, 2011
Functional MRI data of the experimental dataset

1) Inter-subject normalization

2) Intra-subject analysis

For each subject, activation maps:

Extraction of contrast maps for:
- a **reading** task
- a **speech comprehension** task
3) **Choice of brain regions of interest**
- According to the data: **maxima of activation**
- According to the literature: **involved in dyslexia and language networks**

4) **Computation of 34 lateralization indexes**

\[
\hat{\beta}_s^{\text{index}} = \frac{|\hat{\beta}_s^{\text{left}} - \hat{\beta}_s^{\text{right}}|}{\sqrt{(\hat{\beta}_s^{\text{left}})^2 + (\hat{\beta}_s^{\text{right}})^2}}
\]

Q=34 imaging phenotypes

N=94 subjects
Genetic data of the experimental dataset

**DNA microarray (Illumina)**

1,054,068 **Single Nucleotide Polymorphisms** (SNPs)

SNPs: Most variable nucleotides across the genome

For each SNP 3 possible values: AA, AB, BB

**Pre-processing:**

- **Filtering**: (1) Minor Allele Frequency (MAF) at least 10%  
  (2) call rate at least 95%  
  (3) Hardy-Weinberg test not significant at 0.005

- **Coding**: for each SNP, number of minor alleles {0,1,2}

- **Missing SNP** data were imputed with their corresponding median

N=94 subjects  

p=622, 534 SNPs

[Pinel et al., 2007]
Simulated dataset

**Genetic data**
Simulate realistic genetic data

gs algo. [Li & Chen 08]

Hapmap CEU Panel

\( n = 500 \) unrelated subjects

\( p = 85,772 \) SNPs (chromosome 1)

**Imaging data**
Sample from multivariate ()

Parameters \( \mu_1 \mu_2 \Sigma \) estimated

on experimental data

\( n = 500, q = 34 \)

**Genetic effect (additive model)**
- Randomly select 10 SNPs with MAF=0.2
- Two causal patterns each involves 5 SNPs → 4 ROIs
- For each causal patterns \( i \) in 1...2
  * Average the 5 SNPs \( \bar{x}_i \)
  * For each ROI \( j \) in 1...4:
    \[ y_{ij}^* = y_{ij} + \beta_{ij} \bar{x}_i \]
    \( \beta_{ij} \) Controls for explained variance of \( \bar{x}_i \) on \( y_{ij}^* \)
- SNPs in high LD (\( R^2 > 0.8 \)) with true causal are considered causal (56 SPNs)
- Strip of haplotype blocks in the causal SNPs neighbourhood (198 SNPs)
  and move them at the beginning of the dataset
What is currently done in imaging genetics?

**Genome Wide Association Studies (GWAS)**

Massive univariate testing of each SNP versus each imaging phenotype independently *(simple linear regression)*

**Problems:**
- no SNP/phenotype association survived **multiple comparisons**
- **Multivariate nature** of the imaging/genetics link not taken into account
Goal
Study the link between genetic and neuroimaging data by taking into account the interactions between genes and the interactions between brain regions

→ looking for associations between two co-varying networks of genes and brain regions

Problem:
Curse of dimensionality: multivariate methods overfit in high dimensional settings (find associations just by chance)
Multivariate methods based on latent variables

- Canonical Correlation Analysis
- Partial Least Squares

Other two blocks methods in imaging-genetic context
→ Parallel ICA [Calhoun 09]
Partial Least Squares (PLS) Regression

Maximizes the covariance between the two latent variables:
\[
\max \frac{u_h'X_{h-1}Y_{h-1}v_h}{\|u_h\|_2=\|v_h\|_2=1}
\]
→ Solved by an iterative algorithm (NIPALS)
→ Further pairs of components obtained after deflation of X and Y

Canonical Correlation Analysis (CCA)

Maximizes the correlation between the two latent variables:
\[
\max \frac{u_h'X_{h-1}Y_{h-1}v_h}{\|u_h\|_2=\|v_h\|_2=1, \sqrt{u_h'X'Xu_h}\sqrt{v_h'Y'Yv_h}}
\]
Numerical issues: dual formulation of CCA: Kernel CCA (KCCA)
Regularisation

**L2 regularised CCA (rCCA)**
Add diagonal term to intra-block scatter matrices
\[
X'X \rightarrow X'X + \lambda_1 I \quad Y'Y \rightarrow Y'Y + \lambda_2 I
\]

Extreme case of regularisation, scatter matrices → \(I \leftrightarrow rCCA \sim PLS\)

**L2 regularised CCA (rCCA)**
- [Waaijenborg08] Elastic Net: lasso and ridge \(X'X = Y'Y + \lambda I\)
- [Parkhomenko09] Soft-thresholding, \(X'X = \text{diag}(X'X)\)
- [Witten09] \(X'X = Y'Y = I\)

→ such extreme regularization on \(X'X\) makes CCA \(\sim PLS\)

**L1 regularised PLS: sparse PLS (sPLS)**
Add L1 penalisation on the SNPs weights \((u)\)
\[
\min \quad -u'X'Yv + \lambda_1 X \|u\|_1
\]
\(\|u\|_2 = \|v\|_2 = 1\)
- [Chun & Keles07] (regression)
- [LeCao08]
  * Soft thresholding within PLS iterations
Sparse PLS: NIPALS revisited

Bi-convex in \((u)\) and \((v)\), add soft-thresholding within the NIPALS loop

**NIPALS\_soft\_thresholding** \((X, Y)\)

\[X_0 = X, \ Y_0 = Y\]

Iterate over components \((h \text{ in } 0...H)\):

1. Initialize \(u\) and \(v\) using for instance the first pair of singular vectors of the matrix \(X'Y\) and normalize them.

2. Until convergence of \(u\) and \(v\):
   
   (a) For fixed \(v\), find \(\hat{u} = \text{arg min}_{\|u\|_2 = 1} - u'X'Yv + \lambda_1 X \|u\|_1\)
       
       \[\hat{u} = g_{\lambda X} (X'Yv); \quad u = \hat{u}/\|\hat{u}\|_2\]
       
       Where \(g_\lambda(y) = \text{sign}(y)(|y| - \lambda)_+\) is the soft-thresholding function.

   (b) For fixed \(u\), find \(\hat{v} = \text{arg min}_{\|v\|_2 = 1} - u'X'Yv\)
       
       \[\hat{v} = Y'Xu; \quad v = \hat{v}/\|\hat{v}\|_2\]

3. Compute latent variables, loadings and deflation
   
   \[\begin{align*}
   &x^* = X_h u; \quad x_{\text{load}} = X_h x^*/\|x^*\|; \quad X_{h+1} = X_h - x^* x_{\text{load}}' \\
   &y^* = Y_h v; \quad y_{\text{load}} = Y_h y^*/\|y^*\|; \quad Y_{h+1} = Y_h - y^* y_{\text{load}}'
   \end{align*}\]

**return** \((x^*, y^*)\)
Dimension reduction

- We are interested in inter-blocks correlation → CCA
- However CCA overfitt: poor estimation of intra-bloc “variance” (scatter) matrix
- Sparse PLS show promising results on simulated but failed on experiential data
- Add dimension reduction to remove unwilling intra-bloc variance?

→ PCA
→ Feature selection based on filtering
Cross-validation (CV) and significance assessment

For each fold $i$ of the 10-fold CV:

1. Run the studied method on "training" subjects
   - $X^i$ and $Y^i$ (left out subjects at fold $i$)
   - $u^{-i}$ and $v^{-i}$

2. Test on left-out subjects
   - $\text{corr}(X^i u^{-i}, Y^i v^{-i})$

3. $1000$ permutations $\rightarrow$ p-value of the average out-of-sample correlation

Out-of-sample correlation between latent variables averaged over folds
Results on simulated dataset

Comparison of penalisation strategies

SPLS: sparse (L1 regularised) SPLS
rkCCA: L2 regularised kernelized CCA

→ Sparse PLS outperformed other methods
Results on simulated dataset

Comparison of dimension reduction strategies

MULM: Massive Univariate Linear Model
PC = PCA, f = filter
fsPLS = filter + sparse PLS

→ Combined filter + sparse PLS outperformed other methods
Results on experimental dataset

Classical univariate analysis

→ no significant SNP/phenotype associations after correction for multiple comparisons

Multivariate methods

→ Average out-of-sample correlation between latent variables:

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS</td>
<td>-0.09</td>
</tr>
<tr>
<td>Sparse PLS</td>
<td>0.19</td>
</tr>
<tr>
<td>Filtering + Sparse PLS</td>
<td>0.43*</td>
</tr>
</tbody>
</table>

* significant p-value ($p<0.05$), computed with permutation correction for the multiple experiments using maxT

→ Gain in sensitivity compared to univariate analysis
→ Both filtering and sparsity seem necessary
Results: Weights assigned to selected SNPs

→ 50 SNPs selected on all chromosomes
→ Only 14 Genes
→ PLS weights != univ. ranking
→ Some (rare) correlated neighboring SNPs (in linkage disequilibrium) selected together
→ PPP2R2B and RBFOX1 ataxia and a poor coordination of speech and body movement
→ Poor stability: difficult to assess

SPLS weights for 1000s univ. best ranked SNPs
Results: Weights assigned to selected phenotypes

→ 17 selected lateralization phenotypes mainly from the reading task
Two blocks latent methods

Open questions
- Causality
- Structure (gene/ima.)
- Multi-blocks gene.>Ima.>Clinic
...

Classification
Thanks

LNAO & Genim program @ NeuroSpin
- Edith Le floch
- Vincent Frouin
- Bertrand Thirion
- JB Poline
- Alexis Barbot
- Denis Rivière

Unicog @ NeuroSpin
- Philippe Pinel
- Stanislas Dehaene

Supelec
- Arthur Tenenhaus
- Laura Trinchera

St-Anne (AP-HP, Descartes)
- Arnaud Cachia

Necker (INSERM, CEA)
- Monica Zibovicius