Experiments of Machine Learning for Neurodiagnosis

by
Alexandre Savio

PhD Advisor:
Prof. Dr. Manuel Graña Romay
at The University of the Basque Country UPV/EHU

Donostia - San Sebastián

November 8th, 2013
The contents of the thesis deals with:

- the process and analysis of different magnetic resonance imaging (MRI) modalities.
- Machine Learning methods to find biomarkers and good discrimination performance of certain neurological disorders and neurodegenerative diseases.
Abstract

- The contents of the thesis deals with:
  - the process and analysis of different magnetic resonance imaging (MRI) modalities.
  - Machine Learning methods to find biomarkers and good discrimination performance of certain neurological disorders and neurodegenerative diseases.
Outline

1 Introduction
2 The data
3 Pre-processing
4 Feature selection
5 Classification
6 Results
7 Conclusions
Motivation

- A rising population of patients with neurological disorders.
- The increasing of demand of primary attention for these patients.
Motivation

- A rising population of patients with neurological disorders.
- The increasing of demand of primary attention for these patients.
Objectives

- Construction of MRI-based automatic diagnostic support tools
- and search for new biomarkers.
Objectives

- Construction of MRI-based automatic diagnostic support tools
- and search for new biomarkers.
Data

- Anatomical MRI (T1)
  - Grey Matter PVE
  - Displacement Fields
- Diffusion Tensor Imaging (DTI)
  - Fractional Anisotropy
  - Mean Diffusivity
- Functional MRI (T2*)
  - Regional Homogeneity
  - Fractional Amplitude of Low Frequency Fluctuations
  - Amplitude of Low Frequency Fluctuations
  - Voxel-Mirrored Homotopic Connectivity

Preprocessing

- Modulated GM
- Deformation Jacobian
- Geodesic Anisotropy
- Displacement Norm

Feature Selection

Supervised
- Pearson Correlation
- Bhattacharyya Distance
- Welch's t-test
- General Linear Model
- Genetic Algorithm

Unsupervised
- Atlas ROIs

Classification

Supervised
- SVM
- Neural Networks
- Random Forest
- LVQ
- ELM
- Ensembles

Cross-Validation
In this thesis we’ve done studies related with:

- Alzheimer’s disease (AD)
- Schizophrenia
- Myotonic Dystrophy Type 1 (MD1)
Alzheimer’s Disease (AD)

- One of the most important causes of disability in the elderly population.
- Progressive neurodegenerative disease showing gradual deterioration in:
  - cognition,
  - function and
  - behavior.
- Experiments with the OASIS and Hospital Santiago databases.
Alzheimer’s Disease (AD)

- One of the **most important** causes of disability in the elderly population.
- Progressive **neurodegenerative disease** showing gradual deterioration in:
  - cognition,
  - function and
  - behavior.
- Experiments with the OASIS and Hospital Santiago databases.
Alzheimer’s Disease (AD)

- One of the most important causes of disability in the elderly population.
- Progressive neurodegenerative disease showing gradual deterioration in:
  - cognition,
  - function and
  - behavior.
- Experiments with the OASIS and Hospital Santiago databases.
Schizophrenia

- Disabling psychiatric disorder characterized by:
  - hallucinations,
  - delusions,
  - disordered thought/speech,
  - disorganized behaviour,
  - emotional withdrawal, and
  - functional decline.

- Diagnosis is made almost exclusively on self-report, observation, and clinical history.

- Experiments with the COBRE and NAMIC databases.
Schizophrenia

- Disabling psychiatric disorder characterized by:
  - hallucinations,
  - delusions,
  - disordered thought/speech,
  - disorganized behaviour,
  - emotional withdrawal, and
  - functional decline.

- Diagnosis is made almost exclusively on self-report, observation, and clinical history.

- Experiments with the COBRE and NAMIC databases.
Schizophrenia

- Disabling psychiatric disorder characterized by:
  - hallucinations,
  - delusions,
  - disordered thought/speech,
  - disorganized behaviour,
  - emotional withdrawal, and
  - functional decline.

- Diagnosis is made almost exclusively on self-report, observation, and clinical history.

- Experiments with the COBRE and NAMIC databases.
Myotonic Dystrophy Type 1

- A neuromuscular disease characterised by the presence of varying degrees of muscle weakness and myotonia, as well as:
  - cataracts,
  - heart conduction problems, and
  - endocrine, ophthalmological and central nervous system (CNS) disorders.

- The most common neuromuscular disease in humans.
- Experiments with the database from the Departamento de Neurología.
Myotonic Dystrophy Type 1

- A neuromuscular disease characterised by the presence of varying degrees of muscle weakness and myotonia, as well as:
  - cataracts,
  - heart conduction problems, and
  - endocrine, ophthalmological and central nervous system (CNS) disorders.
- The most common neuromuscular disease in humans.
- Experiments with the database from the Departamento de Neurología.
Myotonic Dystrophy Type 1

- A **neuromuscular disease** characterised by the presence of varying degrees of muscle weakness and myotonia, as well as:
  - cataracts,
  - heart conduction problems, and
  - endocrine, ophthalmological and **central nervous system (CNS) disorders**.

- **The most common** neuromuscular disease in humans.

- **Experiments with the database from the Departamento de Neurología.**
The data modalities

- Anatomical MRI
- Functional MRI (fMRI)
- Diffusion tensor MRI (DTI)
Anatomical MRI allows **high resolution and contrast** of biological **tissues**.

- Good distinction between **grey matter**, **white matter** and the other tissues of a head.
- Provides a **base modality** for inter-subject **registrations** and anatomical **region** finding.
Anatomical MRI

- Anatomical MRI allows high resolution and contrast of biological tissues.
- Good distinction between grey matter, white matter and the other tissues of a head.
- Provides a base modality for inter-subject registrations and anatomical region finding.
Anatomical MRI allows high resolution and contrast of biological tissues.

Good distinction between grey matter, white matter and the other tissues of a head.

Provides a base modality for inter-subject registrations and anatomical region finding.
Anatomical MRI

Figure: An OASIS patient.
Functional MRI

- fMRI is a technique for measuring signs of brain activity in a time interval.
- It detects changes in blood oxygenation and flow that correlates to some extent with brain activity.

Figure: fMRI activation map overlaid on an anatomical MRI.
Functional MRI

- fMRI is a technique for measuring signs of brain activity in a time interval.
- It detects changes in blood oxygenation and flow that correlates to some extent with brain activity.

Figure: fMRI activation map overlaid on an anatomical MRI.
Resting-state functional MRI

- **Functional brain imaging** of a subject who is **not** performing an explicit task.

- Slow fluctuations in activity allows to find **local or correlated activity** between brain regions and determine resting state networks.
Resting-state functional MRI

- Functional brain imaging of a subject who is not performing an explicit task.
- Slow fluctuations in activity allows to find **local or correlated activity** between brain regions and determine resting state networks.
Diffusion-weighted MRI

- Diffusion weighted MRI provides structural tissue information using the phenomenon of diffusion of water molecules to give contrast to images.
Diffusion tensor MRI

- DTI is a set of DWI estimating diffusion along specific and different orientations.
- Diffusion profiles in different tissue regions allows to study:
  - the integrity of the White Matter (WM) fibres, and
  - structural connectivity between brain regions.
• DTI is a set of DWI estimating diffusion along specific and different orientations.

• **Diffusion profiles** in different tissue regions allows to study:
  
  - the **integrity** of the White Matter (WM) fibres, and
  - **structural connectivity** between brain regions.
Figure: Whole brain fibre tracts estimations
Outline

1. Introduction
2. The data
3. Pre-processing
4. Feature selection
5. Classification
6. Results
7. Conclusions
The raw data needs certain processing for population-based studies, between others:

- Inhomogeneity corrections
- Motion correction
- Slice timing correction
- Eddy currents correction
- Face removal
- Intra-subject registration
- Registration to standard space
- Skull-stripping
- Tissue segmentation

Each modality requires different pre-processing procedures.
Pre-processing

- The raw data needs certain processing for population-based studies, between others:
  - Inhomogeneity corrections
  - Motion correction
  - Slice timing correction
  - Eddy currents correction
  - Face removal
  - Intra-subject registration
  - Registration to standard space
  - Skull-stripping
  - Tissue segmentation

- Each modality requires different pre-processing procedures.
Registration or spatial normalization is a procedure to ensure that each voxel has the same anatomical meaning across:

- all subjects in the study and
- optionally, a standard template for anatomical region identification.
• **Registration** or spatial normalization is a procedure to ensure that each voxel has **the same anatomical meaning** across:
  
  • all **subjects** in the study and
  • optionally, a **standard template** for anatomical region identification.
Registration or spatial normalization is a procedure to ensure that each voxel has the same anatomical meaning across:

- all subjects in the study and
- optionally, a standard template for anatomical region identification.
Figure: Non-linear registration
Voxel-based Morphometry (VBM)

- Allows measurement of structural differences within a group or across groups through voxelwise comparisons of anatomical brain images.
- VBM consists of a long series of pre-processing methods to finally measure voxels significance through a General Linear Model and statistical inference.
Voxel-based Morphometry (VBM)

- Allows measurement of structural differences within a group or across groups through voxelwise comparisons of anatomical brain images.

- VBM consists of a long series of pre-processing methods to finally measure voxels significance through a General Linear Model and statistical inference.
Deformation-based Morphometry

- Analysis based on the deformation fields (DF) obtained from non-linear registration of anatomical images.
- These DF contain information on the estimated anatomical correspondences with the reference image.
- Scalar measures from the DF are used in feature extraction to find differences between groups.
Deformation-based Morphometry

- Analysis based on the deformation fields (DF) obtained from non-linear registration of anatomical images.
- These DF contain information on the estimated anatomical correspondences with the reference image.
- Scalar measures from the DF are used in feature extraction to find differences between groups.
Deformation-based Morphometry

- Analysis based on the deformation fields (DF) obtained from non-linear registration of anatomical images.
- These DF contain information on the estimated anatomical correspondences with the reference image.
- **Scalar measures** from the DF are used in **feature extraction** to find differences between groups.
Deformation-based Morphometry

Figure: Pipeline of the image pre-processing steps.

\[
J = \begin{pmatrix}
\frac{\partial (x - u_x)}{\partial x} & \frac{\partial (x - u_x)}{\partial y} & \frac{\partial (x - u_x)}{\partial z} \\
\frac{\partial (y - u_y)}{\partial x} & \frac{\partial (y - u_y)}{\partial y} & \frac{\partial (y - u_y)}{\partial z} \\
\frac{\partial (z - u_z)}{\partial x} & \frac{\partial (z - u_z)}{\partial y} & \frac{\partial (z - u_z)}{\partial z}
\end{pmatrix}
\]

\[
\sqrt{x^2 + y^2 + z^2}
\]
The following scalar measures are used in the deformation-based experiments:

1. Determinant of the deformation Jacobian matrix (*jacs*)
2. Trace of the deformation Jacobian matrix (*trace*)
3. Norm of the displacement field vectors (*norms*)
4. Modulated grey matter (*modgm*)
5. Geodesic anisotropy (*geodan*)
Deformation-based Morphometry

Figure: (a) Jacobian determinant, (b) modulated GM, (c) displacement norm, (d) Jacobian matrix trace and (e) geodesic anisotropy.
Multivariate imaging modalities

• fMRI and DTI are known as multivariate imaging modalities because for one voxel they have more than one measure taken at different moments during the acquisition.

• They require corrections for:
  • motion,
  • Eddy currents and
  • slice timing.
Diffusion tensor image pre-processing

- DTI
  - Motion and Eddy currents correction
    - DTI corrected
  - Diffusion tensors fitting
    - FA
    - MD
  - Apply transformation
    - FA MNI
    - MD MNI

- Anatomical
  - Skull Stripping
    - Anatomical brain
    - Anatomical brain mask
  - Affine Registration
    - Affine Matrix
  - Non-linear Registration
    - Deformation fields
    - MNI152 template
Typical scalar measures from DTI:

- Fractional anisotropy (FA)
- Mean diffusivity (MD)
Measures from DTI

**Figure**: FA and MD maps of one subject.
The majority of rs-fMRI studies incorporate functional connectivity to analyse their data. Functional connectivity delivers little insight about local properties of spontaneous brain activity in singular regions.
• The majority of rs-fMRI studies incorporate functional connectivity to analyse their data.

• Functional connectivity delivers little insight about local properties of spontaneous brain activity in singular regions.
Local activity measures from rs-fMRI

- May serve as **discriminant features** for individual variations or dysfunction.
- In this thesis we used the following measures:
  - Amplitude of Low Frequency Fluctuations (**ALFF**) and fractional Amplitude of Low Frequency Fluctuations (**fALFF**),
  - Voxel-Mirrored Homotopic Connectivity (**VMHC**) and
  - Regional Homogeneity (**ReHo**).
Local activity measures from rs-fMRI

- May serve as discriminant features for individual variations or dysfunction.
- In this thesis we used the following measures:
  - Amplitude of Low Frequency Fluctuations (ALFF) and fractional Amplitude of Low Frequency Fluctuations (fALFF),
  - Voxel-Mirrored Homotopic Connectivity (VMHC) and
  - Regional Homogeneity (ReHo).
Feature selection

- Provides the **input data** for the classifiers.
- We are following a general procedure that provides:
  - the **anatomical locations** where the data is extracted
  - including, when possible, **relevance rankings** of anatomical locations, based on the classification results.
Feature selection

• Provides the input data for the classifiers.

• We are following a general procedure that provides:
  • the anatomical locations where the data is extracted
  • including, when possible, relevance rankings of anatomical locations, based on the classification results.
VBM-based feature selection

**Figure**: Flow diagram of the **VBM based feature selection** process from the subjects GM segmentation volumes.
Voxel significance measures

- We tested other voxel significance measures:
  -Pearson’s correlation (PC) between voxel data and the subject class labels.
  -Bhattacharyya distance (BD)
  -Welch’s t-test (WT)
  -Spearman’s correlation

- Thresholding is used to create feature selection masks.
Voxel significance measures

- We tested other voxel significance measures:
  - Pearson’s correlation (PC) between voxel data and the subject class labels.
  - Bhattacharyya distance (BD)
  - Welch’s t-test (WT)
  - Spearman’s correlation

- Thresholding is used to create feature selection masks.
Voxel significance measures

Figure: A summary of the pipelines of the feature selection methods. Each path from left to right specifies a feature selection process tested in the computational experiments.
Atlas-based feature selection

**Figure**: Flow diagram of the feature extraction and ensemble of classifiers.
Pre-selection of ROIs

Figure: Slices of the MNI standard template showing the *a priori* maps of affected regions in the **left temporal area (left)**, corresponding to **mild AD** (middle) and to **normal AD** (right) used for ensemble decision.
Outline

1. Introduction
2. The data
3. Pre-processing
4. Feature selection
5. Classification
6. Results
7. Conclusions
Classification

- Classification of the data into its corresponding classes is the final step in the computational pipeline.
- To assess how the classification results generalize to an independent data set we apply cross-validation.
- The most used cross-validation (CV) methods were: stratified 10-fold CV and leave-one-out.
Classification

- Classification of the data into its corresponding classes is the final step in the computational pipeline.
- To assess how the classification results generalize to an independent data set we apply cross-validation.
- The most used cross-validation (CV) methods were: stratified 10-fold CV and leave-one-out.
Classification

- Classification of the data into its corresponding classes is the final step in the computational pipeline.
- To assess how the classification results generalize to an independent data set we apply cross-validation.
- The most used cross-validation (CV) methods were: stratified 10-fold CV and leave-one-out.
Double dipping:

- **training** must be as **independent** as possible of the test data.

- Intuitively: we should exclusively use for test totally unknown data.

- **Supervised feature selection** should be within the cross-validation process.
Double dipping:

- *training* must be as *independent* as possible of the test data.

- Intuitively: we should exclusively *use for test totally unknown data*.

- *Supervised feature selection* should be within the cross-validation process.
Double dipping:

- training must be as independent as possible of the test data.
- Intuitively: we should exclusively use for test totally unknown data.
- Supervised feature selection should be within the cross-validation process.
Figure: Pipeline for circularity free validation of supervised classification with supervised feature selection.
Classification algorithms

- Support Vector Machines (SVM)
- Artificial Neural Networks (ANN)
  - Multi Layer Perceptron trained with Backpropagation (MLP)
  - Radial Basis Function Networks (RBFN)
  - Probabilistic Neural Networks (PNN)
- Learning Vector Quantization (LVQ2)
- Random Forests (RF)
- Extreme Learning Machines (ELM)
Classification algorithms ensembles

- Bootstrapped Dendritic Classifiers (BDC)
- Extreme Learning Machines (ELM)
  - Ensembles of ELM
  - Hybrid Extreme Rotation Forest (HERF)
- Ensemble of independent SVM classifiers
- Adaptive Boosting and Diverse AdaBoost with SVM
Experiment 1

- Alzheimer’s disease
- Deformation-based Morphometry (DBM) features from OASIS: jacs, norms, modgm, trace and geodan
- Pearson’s Correlation (PC), Bhattacharyya’s Distance (BD) and Welch’s t-test (WT) as feature selection
- SVM for classification
Figure: **Average and standard deviation** of Accuracy, Precision, Recall (Sensitivity) and ROC area over the 10-fold cross-tests results of a linear SVM trained to minimize the error rate, features selected using Pearson’s correlation varying the voxel significance threshold.
### Table: Mean (standard deviation) of Accuracy, Precision, Sensitivity, and ROC area of the full 10-fold cross-validation classification results for features selected with the 95% percentile.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>jacs</strong></td>
<td>88.10 (0.00)</td>
<td>50.76 (1.70)</td>
<td>58.33 (3.73)</td>
<td>92.32 (1.53)</td>
</tr>
<tr>
<td><strong>norms</strong></td>
<td>88.10 (2.75)</td>
<td>84.14 (6.12)</td>
<td>58.33 (8.98)</td>
<td>94.79 (2.53)</td>
</tr>
<tr>
<td><strong>modgm</strong></td>
<td>92.07 (1.12)</td>
<td>95.83 (5.89)</td>
<td>86.67 (4.71)</td>
<td>96.67 (0.44)</td>
</tr>
<tr>
<td><strong>trace</strong></td>
<td>89.43 (2.70)</td>
<td>79.55 (5.77)</td>
<td>65.00 (5.00)</td>
<td>94.27 (0.43)</td>
</tr>
<tr>
<td><strong>geodan</strong></td>
<td>88.15 (2.25)</td>
<td>87.74 (3.96)</td>
<td>70.00 (0.00)</td>
<td>93.49 (1.04)</td>
</tr>
</tbody>
</table>
E1: DBM and BD

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>jacs</td>
<td>91.27 (1.22)</td>
<td>86.44 (1.89)</td>
<td>81.67 (7.45)</td>
<td>95.36 (1.54)</td>
</tr>
<tr>
<td>norms</td>
<td>89.89 (1.77)</td>
<td>82.62 (6.94)</td>
<td>79.67 (3.73)</td>
<td>93.12 (1.48)</td>
</tr>
<tr>
<td>modgm</td>
<td>89.74 (1.77)</td>
<td>82.94 (4.14)</td>
<td>78.00 (5.00)</td>
<td>97.37 (0.29)</td>
</tr>
<tr>
<td>trace</td>
<td>89.52 (1.78)</td>
<td>82.43 (4.15)</td>
<td>80.33 (7.45)</td>
<td>95.67 (1.08)</td>
</tr>
<tr>
<td>geodan</td>
<td>92.09 (2.60)</td>
<td>88.09 (5.33)</td>
<td>80.00 (4.00)</td>
<td>95.37 (0.99)</td>
</tr>
</tbody>
</table>

Table: Mean (standard deviation) of Accuracy, Precision, Sensitivity, and ROC area of the full 10-fold cross-validation classification results for features selected with the 95% percentile.
Table: Mean (standard deviation) of Accuracy, Precision, Sensitivity, and ROC area of the full 10-fold cross-validation classification results for features selected with the 95% percentile.
E1: Localization of discriminant voxels

Figure: Discriminant voxel sites selected applying a 95% percentile on the empirical distribution of the Pearson’s correlation (left), Bhattacharyya’s distance (middle) and Welch’s t-test (right) computed on the modulated GM (modgm).
Experiment 2

- Schizophrenia
- Local activity measures from COBRE
- Four different pre-process configurations tested: w/o temporal filtering and w/o global signal regression.
- Pearson’s Correlation and Bhattacharyya’s Distance as feature selection
- SVM as classifier
E2: Performance on ReHo and BD

Figure: Classification performance using TPF-GSR ReHo data and the Bhattacharyya distance varying the voxel significance threshold.
### E2: Accuracies

<table>
<thead>
<tr>
<th></th>
<th>Pearson</th>
<th>Bhattacharyya</th>
<th>Welch t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ALFF</strong></td>
<td>Lin-SVM</td>
<td>0.65 (0.02)</td>
<td>0.61 (0.01)</td>
</tr>
<tr>
<td></td>
<td>RBF SVM</td>
<td>0.67 (0.01)</td>
<td>0.54 (0.01)</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>0.72 (0.01)</td>
<td>0.72 (0.02)</td>
</tr>
<tr>
<td><strong>fALFF</strong></td>
<td>Lin-SVM</td>
<td>0.68 (0.01)</td>
<td>0.63 (0.01)</td>
</tr>
<tr>
<td></td>
<td>RBF SVM</td>
<td>0.64 (0.01)</td>
<td>0.62 (0.01)</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>0.66 (0.02)</td>
<td>0.63 (0.01)</td>
</tr>
<tr>
<td><strong>ReHo</strong></td>
<td>Lin-SVM</td>
<td>0.73 (0.01)</td>
<td>0.76 (0.01)</td>
</tr>
<tr>
<td></td>
<td>RBF SVM</td>
<td>0.76 (0.02)</td>
<td>0.64 (0.02)</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td><strong>0.80 (0.02)</strong></td>
<td>0.73 (0.01)</td>
</tr>
<tr>
<td><strong>VMHC</strong></td>
<td>Lin-SVM</td>
<td>0.75 (0.01)</td>
<td>0.74 (0.01)</td>
</tr>
<tr>
<td></td>
<td>RBF SVM</td>
<td>0.58 (0.01)</td>
<td>0.58 (0.01)</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>0.73 (0.02)</td>
<td>0.76 (0.02)</td>
</tr>
</tbody>
</table>

**Table:** Maximum accuracies and variances across feature selection thresholds for each experiment on the TPF-GSR data.
E2: Localization of discriminant voxels

Figure: Voxels selected from the TPF-GSR ReHo data using Bhattacharyya distance thresholded at 95%. The colormap indicates the voxel selection frequency during cross-validation.
Figure: Voxels selected from the TPF-GSR ReHo data using Pearson correlation thresholded at 90%. The colormap indicates the selection frequency during cross-validation of each voxel.
Methodological conclusions

- We have studied the application of machine learning to **different modalities** of MRI data.
- We have built and tested a **general computational pipeline** that can be applied to the various data modalities contemplating the special case of supervised feature selection.
- We found **good classification results** that encourage the realization of CAD systems for many cases.
- Anatomical **feature localizations** are in good agreement with the findings reported in the literature, validating the approach from the medical point of view.
Operational conclusions

• An extensive exercise in the management of large image data of diverse modalities related to specific diseases.

• We have followed in the measure of our possibilities the policy of sharing our software and data in order to enhance the reproducibility of our studies, and foster new experiments and results on our datasets.
Experiments of Machine Learning for Neurodiagnosis

by
Alexandre Savio

PhD Advisor:
Prof. Dr. Manuel Graña Romay
at The University of the Basque Country UPV/EHU

Donostia - San Sebastián

November 8th, 2013