MetaImaGen Project

Author(s) :
Marco Lorenzi
Boris Gutman
Andre Altman
Barbara Bardoni
Santiago S. Silva R.

Version 1.0
July 30, 2018

Version Control

Version Number	Description / Change	Author	Date
1.0	Initial Version	Santiago-Smith Silva-Rincon	09/05/2018

Contents

Introduction

At this moment, databanks worldwide contain brain images and individual genomes in previously unimaginable numbers. Combined with developments in data science, these massive data provide the potential to fully understand the genetic underpinnings of brain disease. However, as it often happens, society lags behind innovation: different chunks of data, which are stored at different institutions, cannot always be shared directly due to privacy concerns and legal complexities, thus, preventing exploitation of big data in the study of brain disorders.

In this project we aim at pursuing our work on online learning in distributed medical databases. We apply a novel computational paradigm for the meta-analysis of large-scale medical datasets distributed across clinical centers. The rationale of the project is to pool different data parts without sharing individual information, through advanced multivariate data analysis tool based on online and distributed learning. Our proposal extends the state-of-art methodology beyond the simplistic effects of individual genes on individual brain components (mass-univariate analysis), by analytically exploring how combinations of genes and brain areas interact in concert (multivariate online-learning). This promising methodology will be applied for the first time within the large-scale ENIGMA imaging-genetics consortium, providing data of thousands of subjects from several clinical centers distributed around the world. Thanks to the expertise and feedback of our UCA excellence partner in biology and clinic (CNRS IPMC, Fondation CHU- Lenval), the project presents a unique opportunity for the identification and validation of sets of candidate genetics variants underpinning autism and neuropsychiatric disorders.[1]

---

[1]Requeriment document from

Chapter 1

Installation guide

1.1 Requirements

1.1.1 Hardware requirements

These are the minimum hardware requirements

- CPU: 4 cores @ 2 GHz

- RAM: 8 GB

- Storage: 4 GB

1.1.2 Software requirements

Operating system

This software runs **ONLY** on Unix-based OS like[1]:

- Linux: Fedora, CentOS, Debian, Ubuntu, etc.

- macOS: El Capitan 10.11 or higher.

More information can be found here.

---

[1]**IMPORTANT:** *Windows platforms are not still supported.*

Other software

For running the MetaImaGen software your system needs to have:

- Docker version 17 or higher (Installation instructions)

- Docker Compose Version 3.0 or higher (Installation instructions)

1.1.3   Getting Started with MetaImaGen

1. Clone the repository by executing:
   ```
   git clone https://github.com/sssilvar/meta-imagen.git
   ```

2. Run the test program to check if everything is in order:
   ```
   bash meta-imagen/bin/test.sh
   ```

3. Run the analysis of your images by executing:
   ```
   bash meta-imagen/bin/run_metaimagen.sh [data_folder_path][center_id]
   ```

   Where `[data_folder_path]` is the folder containing all the subjects processed by **FreeSurfer** and a file called *groupfile.csv* that contains the subjects to be processed (IDs) and the diagnostic group where they belong to. An example is shown in table 1.1

sid	dx
002_S_0954	MCIc
002_S_1070	AD
...	...

Table 1.1: *groupfile.csv* structure.