

An Introduction to Methods for Analyzing Large Neuroimaging Datasets

Herve LEMAITRE Co Organizer

IMN-GIN
Bordeaux, FM
France

Robert Whelan, PhD Organizer

Trinity College Dublin
Trinity College Dublin
Dublin, Dublin
Ireland

Sunday, Jun 19: 8:00 AM - 4:30 PM

1150

Educational Course - Full Day (8 hours)

SEC Centre

Room: Hall 1

Timeliness and importance of the topic: Human brain imaging is in a period of profound change. There is growing recognition that sample size must drastically increase to achieve adequate statistical power and reproducibility. Accordingly, several large neuroimaging databases have been established recently. For example, Adolescent Brain and Cognitive Development (ABCD: <https://abcdstudy.org/>) is a 10-year, \$300 million, neuroimaging project that will recruit 10,000 people (currently mid-way through). Importantly, ABCD data will be fully open access, available with minimal restrictions. Similar open-access databases include Alzheimer's Disease neuroimaging initiative (ADNI: <http://adni.loni.usc.edu/>) and Open Access Series of Imaging Studies (OASIS; <https://www.oasis-brains.org/>), sharing platforms (e.g., openneuro.org) or belonging to large consortia (IMAGEN: <https://imagen-europe.com/>). The UK Biobank (<https://www.ukbiobank.ac.uk/>) will collect neuroimaging data from 100,000 people (currently, 50,000 have been scanned), and these data are available to researchers for a very modest access fee. These datasets will open up neuroimaging to a new generation of scientists.

Timeliness and importance of the desired learning outcomes: Despite the recent proliferation of large, easily available, neuroimaging datasets there are few resources that are specifically oriented towards the specialized methods required to analyze them. Rather, extant resources focus on methods appropriate to analysis of a relatively small number of subjects that a single laboratory would obtain. The aim of this course will be to provide researchers with the skills to interrogate these large datasets, and with best practice recommendations for managing and curating data.

Objective

Having completed this course, participants will be able to:

- Access and manage large neuroimaging datasets available to the scientific community
- Work with very large datasets: preprocessing and statistical/algorithmic analyses to examine between-group or individual differences
- Apply best practice methods for ensuring reproducible workflows and for data curation and annotation

Target Audience

The target audience is researchers who have some prior experience working with structural and/or functional MRI (and/or EEG/MEG) data and who are interested in analyzing large neuroimaging datasets. Some level of coding knowledge would be advantageous, although some of the toolboxes can be used entirely through a user interface.

Presentations

Getting Started: downloading and organizing data

A brief introduction to the essential components of how open access data can be accessed and how they are organized (specifically, the BIDS format). This presentation will use the AOMIC-PIOP2 dataset as an example (<https://openneuro.org/datasets/ds002790/versions/2.0.0>).

Presenter

Jivesh Ramduny, Trinity College Dublin
Psychology
Dublin, Dublin
Ireland

Neuroimaging workflow in the cloud

Cloud computing brings scalability and reproducibility to neuroimaging analyses, but leveraging these features cost-effectively requires changes to most neuroimaging workflows. In this session we will describe features of cloud computing that can be leveraged for robust and scalable analyses and how to architect new workflows to cost-effectively take advantage of them.

Presenter

Tara Madhyastha, PhD, Industry Seattle, WA
United States

From dozens to thousands: important lessons when scaling up structural MRI processing using CAT

A typical MRI processing pipeline with 20-50 subjects is not practical/feasible with 200/1000 subjects. This tutorial will start with a theoretical component discussing why an explicit planning phase, the use of provenance tracking, and automatic QC are mandatory for working with large datasets. The use of containers and a scheduler are also desirable. Next, this talk will demonstrate the DataLad-based (www.datalad.org) Computational Anatomy Toolbox (CAT) preprocessing pipeline with an open data example. The presentation will conclude by discussing important aspects of large statistical group models in CAT.

Presenter

Felix Hoffstaedter, Research Centre Jülich
Institute of Neuroscience and Medicine (INM-7)
Jülich, North Rhine-Westphalia
Germany

NeuroImaging PREProcessing tools (NiPreps)

The current neuroimaging workflow has matured into a large chain of processing and analysis steps involving a large number of experts, across imaging modalities and applications. The vision for NiPreps is to provide end-users (i.e., researchers) with applications that allow them to perform quality control smoothly and to prepare their data for modeling and statistical analysis.

Presenter

Oscar Esteban, University Hospital of Lausanne and University of Lausanne Lausanne
Switzerland

Harmonizing data analysis for diverse neuroimaging datasets using HALFpipe

Increasing reproducibility in neuroimaging often means combining datasets for larger sample sizes and/or direct replications of findings. HALFpipe (Harmonized Analysis of Functional MRI pipeline) is a tool that can help in these scenarios, because it can automatically construct processing pipelines to derive task-based activation and functional connectivity from virtually any dataset (see also <https://doi.org/gddf>). In this course, we will demonstrate how to set up HALFpipe for your data using an example. We will also discuss the lessons we learned when processing diverse datasets within the ENIGMA consortium and how these might inform your analysis plan.

Presenter

Lea Waller, Charité Universitätsmedizin Berlin Berlin, Germany
Germany

Machine learning for Neuroimaging in python with the Brain Predictability toolbox

The Brain Predictability toolbox (BPt) represents a unified framework of machine learning (ML) tools designed to work with both tabulated data (in particular brain, psychiatric, behavioral, and physiological variables) and neuroimaging specific derived data (e.g., brain volumes and surfaces). This presentation will present a worked example of how one could use BPt to perform machine learning based analyses on data from the AOMIC: the Amsterdam Open MRI Collection. This example will focus on using fMRI derived “connectome” input brain features in order to predict, in a properly cross-validated manner, a range of different individual phenotypes.

Presenter

Sage Hahn, University of Vermont Burlington, VT
United States

Combining graph theory and machine learning

In this tutorial, I will demonstrate NBS-Predict, a prediction-based extension of the Network-based Statistic (Zalesky et al., 2010). NBS-Predict aims to alleviate the curse of dimensionality, lack of interpretability, and problem of generalizability. By combining the powerful features of machine learning and graph theory in a cross-validation structure, it provides a fast and convenient tool to identify neuroimaging-based biomarkers with high generalizability. Unlike generic machine learning algorithms, results derived from the toolbox are straightforwardly interpretable. NBS-Predict comes with a user-friendly graphical user interface (GUI) developed on MATLAB. Thus, it does not require any programming expertise. The toolbox provides an interactive viewer to visualize the results. The extensive user manual and usage tutorials are included in the toolbox.

Presenter

Emin Serin, MSc, Charite Berlin, Berlin
Germany

Connectome-based predictive modelling (CPM) for large neuroimaging datasets

Connectome-based predictive modelling (CPM) is a data-driven approach which enables the prediction of behavioural and cognitive phenotypes from functional connectivity data. CPM has been applied to successfully predict individual differences in various cognitive and behavioural phenotypes. A previous protocol paper outlined the method (Shen et al., 2017). Here, we present a flexible CPM method that is optimised for large neuroimaging datasets via the use of parallel computing. This approach enables researchers to account for possible site and scanner-related heterogeneity in multi-site neuroimaging datasets by controlling for site and/or scanner type as a covariate or by using leave-site-out cross-validation. Dr Rory Boyle will co-present this session.

Presenter

Yihe Weng, MSc, Trinity College Dublin
School of Psychology
Dublin, Dublin
Ireland

Annotating the timeline of neuroimaging time series data using Hierarchical Event Descriptors (HED)

For historical reasons, accepted standards for annotating a complete and detailed answer to the question, 'What happened during the recording?', have remained minimal. Unfortunately, timeline event annotations in currently archived time series neuroimaging datasets are insufficient for supporting detailed, advanced and/or cross-study ('large data') analysis. The Hierarchical Event Descriptors (HED) system, first proposed in 2013, was accepted into all the BIDS modality standards in 2019. Recently, new (3rd-generation) HED specification, schema, and software tools have been released by our HED Working Group. I will demonstrate working examples of annotating EEG/MEG and fMRI datasets using HED, and will illustrate uses of the resulting HED annotations in data search, extraction, and analysis.

Presenter

Scott Makeig, University of California San Diego San Diego, CA
United States

Establishing a reproducible and sustainable analysis workflow

This tutorial will describe how to use Python notebooks to write reproducible code. We will demonstrate best practice for pipeline/code management and versioning using Git. We will end by recommending approaches for sustainable analyses: optimising work practices and code to reduce resource use. Here, we reference work by the OHBA Sustainability special interest group (<https://neuropipelines.github.io/index>). This session will be co-presented by Mélanie Garcia.

Presenter

Clare Kelly, Trinity College Dublin Dublin, Ireland
Ireland
