

CREATING IMAGING AND DIGITAL LITERATE RESEARCHERS AND CLINICIANS THROUGH THE HEALTH AND BIOSCIENCE IDEAS INTERACTIVE TRAINING PROGRAMME.

APRIL 19TH 2023

UCL EDUCATION CONFERENCE 2023

Theme: Skill development and experiential learning

Dr Maria Tziraki: IDEAS Course leader

Dr David Cash : IDEAS PI and Manager

<https://healthbioscienceideas.github.io>

Health+BioscienceIDEAS

Overview

- The training need that generated the “idea” for Health and Bioscience IDEAS
- Objectives of IDEAS
- Organisation and delivery of workshops in data science and medical imaging
- Development and delivery of an innovative three-week TeamCoders workshop
- Challenges faced and lessons learned from the IDEAS workshops

Imaging data plays a critical role in today's health and bioscience research



- Imaging technologies in bioscience and medicine are becoming cheaper and more accessible
- Researchers have access to many large, publicly available data sets
- Complexity of imaging data and analysis techniques is high, with new software tools and AI create 'entry barriers' to the use of advanced technology.
- Conventional training for clinicians and bioscience does not cover these areas

Industry need researchers with data science skills

Professionals from the Healthcare industry



Real-World Data: (RWD) analysis supports informative decisions in R&D, product strategy and business growth

High demand of skilled personnel in advanced data analytics with a particular interest in clinical data analysis

Cost of RWD and training
\$300 million /year /pharma company

Training format.
Bootcamps for skill development instead of MSc

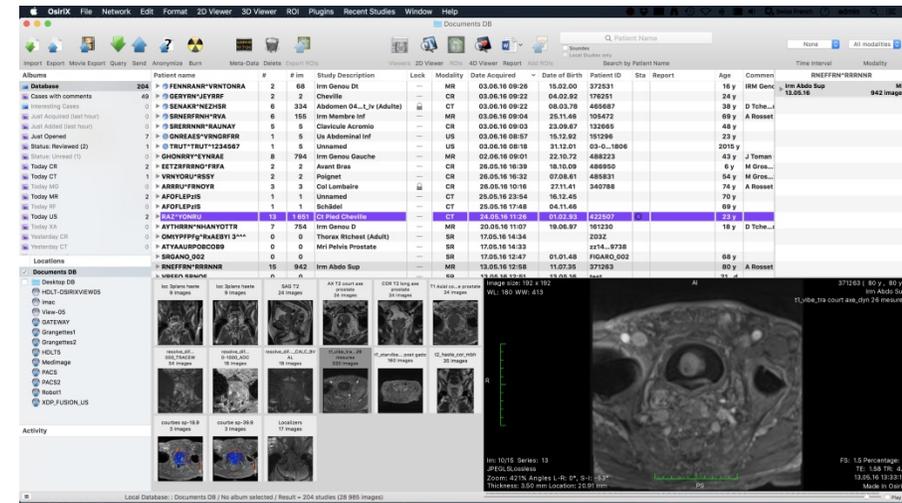
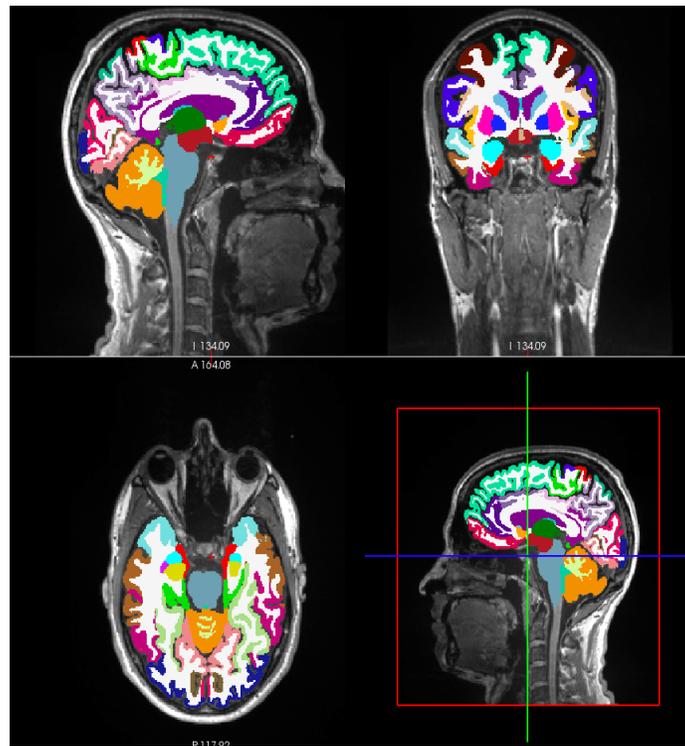
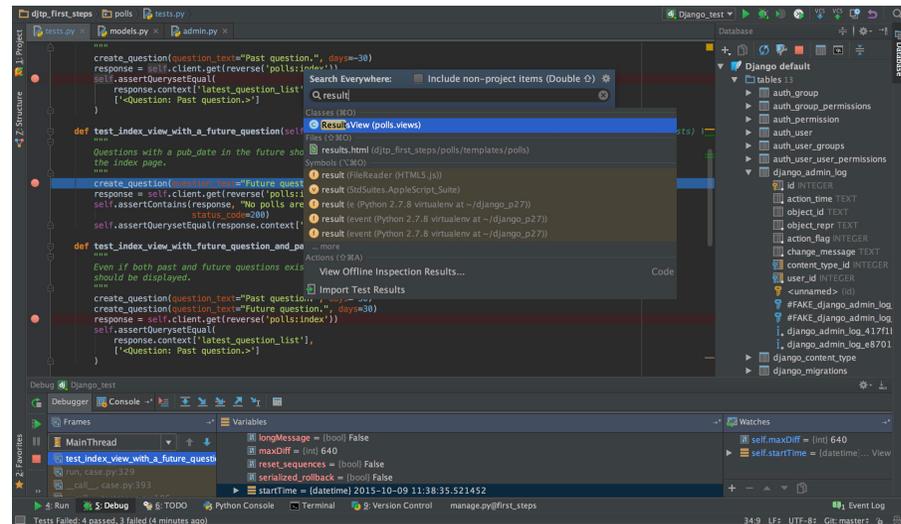
Health+Bioscience IDEAS

Funded by UKRI Innovation Scholars of Data Science Training in Health and Bioscience (DaSH) initiative. Funding period February 2021- September 2023

Software Development

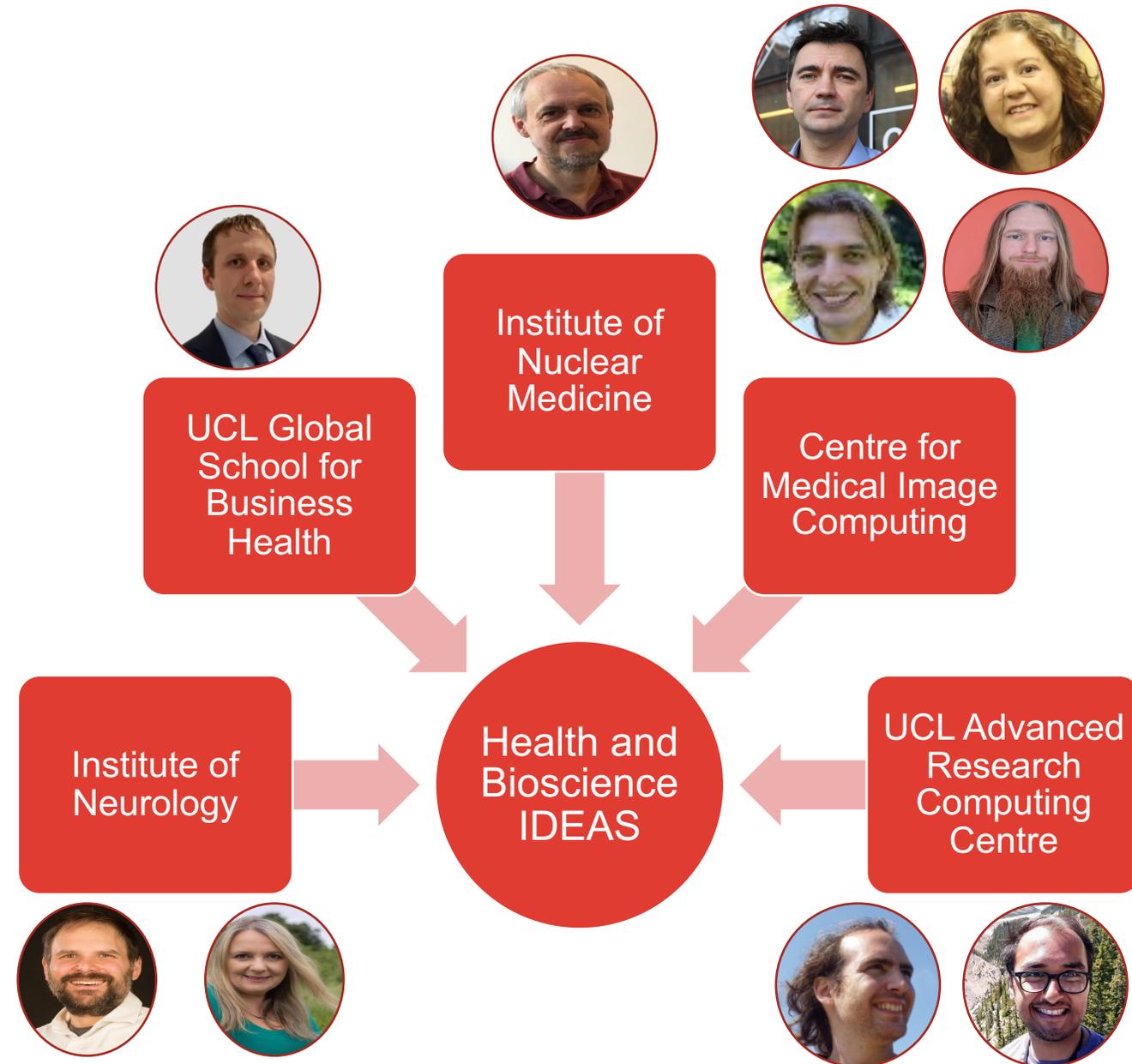
Image Acquisition and Analysis

Data Management



IDEAS Objectives

- To create “digital and imaging literate” health and bioscience researchers, clinicians and healthcare professionals
- To develop and run training activities via, workshops, conference educational sessions, summers schools.
- To provides practical advice and hands-on experience using popular and freely available software.
- To improve reliability of research outcomes and saves significant staff time
- is self-sustaining beyond the lifetime of the grant



Planned offerings for IDEAS courses

Software development

Basic – Software Carpentry

- Introduction to Python
- Introduction to git

Intermediate

- Git: Beyond Basics
- Advanced Python/R
- Sustainable software practices

Advanced

- Performance and scalability
- Making packages
- Automatic code documentation

Image Acquisition and Analysis

Basic

- Intro to medical image Acquisition
- Image formation and manipulation

Intermediate

- Image Registration and Segmentation
- Groupwise analysis

Advanced

- Imaging Genetics
- Machine Learning applications

Data Management

Basic

- Getting started with XNAT
- Medical imaging data management (DICOM, BIDS, information governance)

Intermediate

- Managing XNAT
- Data sharing (FAIR, deidentification)

Advanced

- Automating tasks in XNAT
- Integration with other platforms
- IT planning for imaging studies

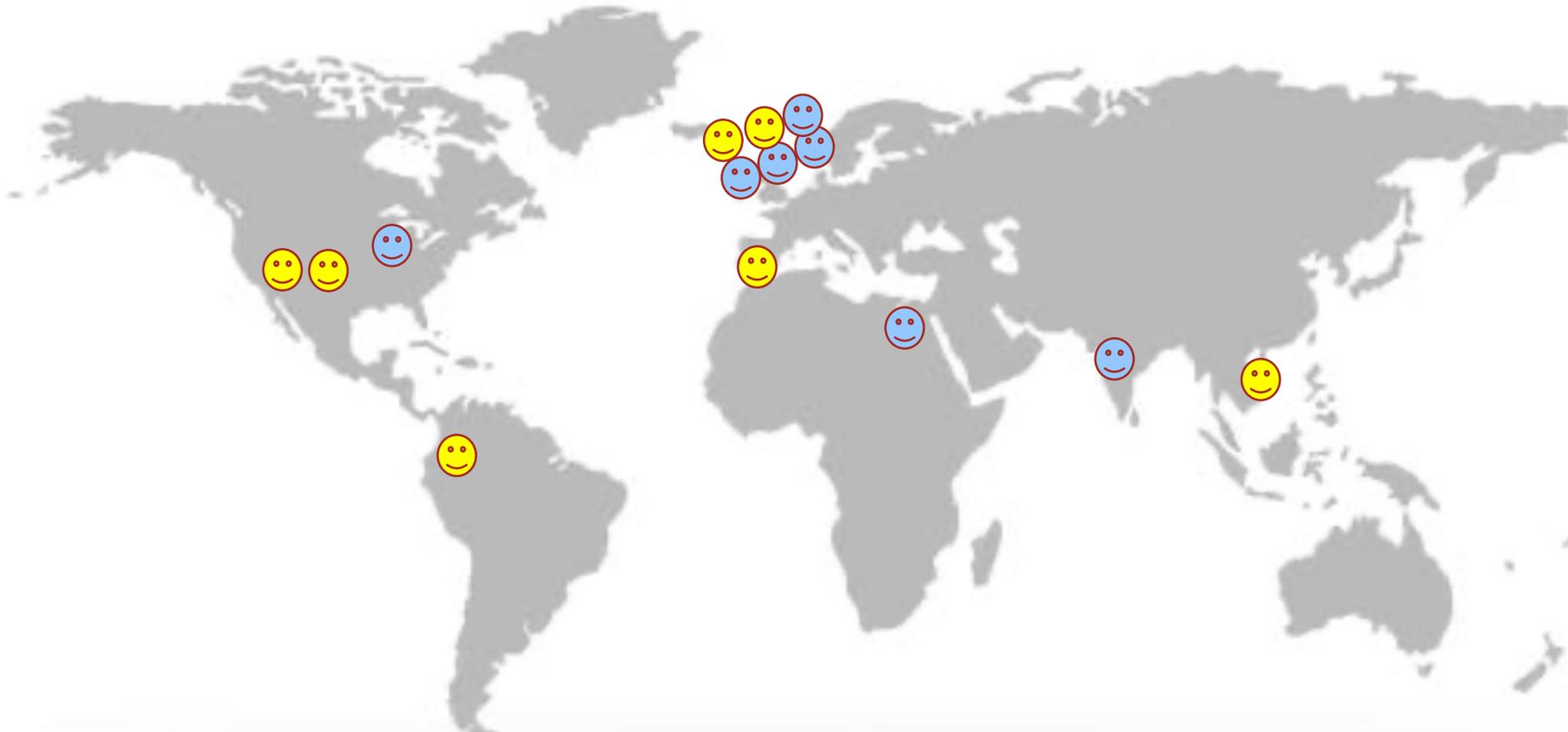
TeamCoders – experiential learning via collaborated coding

Objectives

- Group together individuals from various healthcare disciplines and coding abilities and encourage collaboration via coding practice.
- Follow Experimental learning pattern – Experience- Reflection- Thinking and problem solving – Acting by coding
- Understand the importance of medical imaging in extracting quantitative measures about the given project question.
- Reproduce a seminal paper applying a data-driven disease progression model called the Event Based Model to model the progression of Alzheimer's disease. ([AL Young et al Brain 2014](#))
- Practice standard data cleaning and visualisation techniques to interrogate and investigate the input data before analysis.
- Learn how to implement the Event Based Model on Alzheimer's disease data and use the model to predict future outcomes at different data sets



Team Coders – International and Interdisciplinary participation



Task organization

- Each week (1-3 hours) involved facilitators demonstrating notebooks.
- New skills: command line (Unix Shell), version control with git (software), data cleaning, analysis and visualization with Python. Medical image analysis with Python and the Event Based model Python programming
- Teams of 3-5 individuals worked (programmed) together and solved questions in the exercise notebooks.

Demonstration notebooks

- Led by facilitator
- Data may or may not relate to the project
- Shows both explanation and the code required to complete the exercise

Project (exercise) notebook

- Led by the team
- Data will build up in order to help complete the overall project
- Provides instructions followed by blank code cells for the team to complete with their answers.

Materials were shared online at GitHub repositories

Jupyter Notebooks

Jupyter biomarker_extraction Last Checkpoint: 15/09/2022 (autosaved) Python 3 (ipykernel)

Imaging biomarker extraction walkthrough

This notebook will go over some examples of extracting biomarkers from structural T1 weighted MR images and amyloid PET data. These are some of the most widely used biomarkers that are used in AD research. They are also measures used in the EBM.

Objectives

After completing this notebook, you will be able to:

- Understand how to open and interrogate medical imaging data
- Run basic image processing tasks like image registration and segmentation
- Extract measurements of brain volumes and amyloid burden.

There are many different publicly available, open-source software packages that you can use to obtain these measurements. The specific techniques that we are using in this notebook are meant to demonstrate the key concepts involved in the processing pipelines, but they have not been optimised for this task. The exact decision of what software and pipeline to use for your research will depend on what works best for your data and application.

Preparation

First, we will import the packages that will be used in our notebook. We will be using the following packages:

- [nibabel](#) - a handy package to work with neuroimaging data.
- [Advanced Normalisation Toolbox \(ANTs\)](#) - a popular set of tools that do registration, segmentation and other image processing techniques.
- [Niwidgets](#) - a small package for making interactive viewers in Jupyter notebooks. This is useful for a quick view of an image.
- [fsleyes](#)(OPTIONAL) - a popular, fully featured image visualisation tool. It does have [support for Jupyter notebooks](#). You might find this easier to interactively review your results rather than using the static plots provided here.

```
In [ ]: %capture --no-display
%matplotlib inline
import os
import nibabel as nb
import numpy as np
from niwidgets import NiftiWidget
import pandas as pd
import ants
import matplotlib.pyplot as plt

# If using dark background, then keep the next line in, otherwise comment it out.
plt.style.use('dark_background')
```

We are also going to make a handy function to write out the imaging data so that we can look at it outside of the notebook.

```
In [ ]: def write_image(image_to_write, rootdir, subject, image_type, image_desc, image_space=None):
    if image_space is not None:
        image_name = f"sub-{subject}_space-{image_space}_desc-{image_desc}.nii.gz"
    else:
        image_name = f"sub-{subject}_desc-{image_desc}.nii.gz"
    image_path = os.path.join(rootdir, image_type, image_name)
    ants.image_write(image_to_write, image_path)
```

• Demo Notebook

Jupyter adni_ebm (unsaved changes) Python 3 (ipykernel)

Step 1 - Open the data

First up, load up the spreadsheet that you created from the second project notebook.

```
In [ ]: # Your answer to Step 1
# Below put your code that will load in the data that you created from the ADNI data
```

Step 2 - Initial Processing of data

Our spreadsheets from the first exercise still need a few more tweaks before running the EBM.

Our current MRI features are the raw volumes. However, the volume of most brain structures are highly associated with the head size of the individual. We have a good proxy variable of head size extracted from the image data called *total intracranial volume* or TIV. In order to reduce the effect of head size, Young *et al.* divided all of the brain structures by TIV before putting it into the EBM. So remember to do the same with your features. However, there are two exceptions. The *boundary shift integral* or BSI measures do not measure the size of the brain structure but how much the size has **changed** over one year. These changes have not been shown to be associated with TIV and so they **do not** need to be normalised by TIV before the EBM.

We will discuss this further in future steps, but we need to have a column that provides a special coding of the three values in the baseline diagnosis (DX_bL) column, where:

- Cognitively Normal (CN) = 0
- Alzheimer's Disease (AD) = 1
- Mild Cognitive Impairment (LMCI) = 2

Also, I would suggest to make some Python lists to help hold key sets of column names:

- Key *metadata* that are helpful for processing but not going to be in the EBM, and
- names of *features* that you want to include in the EBM ordering.

```
In [ ]: # Your answer to Step 2
# Below put your code for normalising the MRI volumes by TIV
# and creating helpful lists of metadata variables and feature variables
```

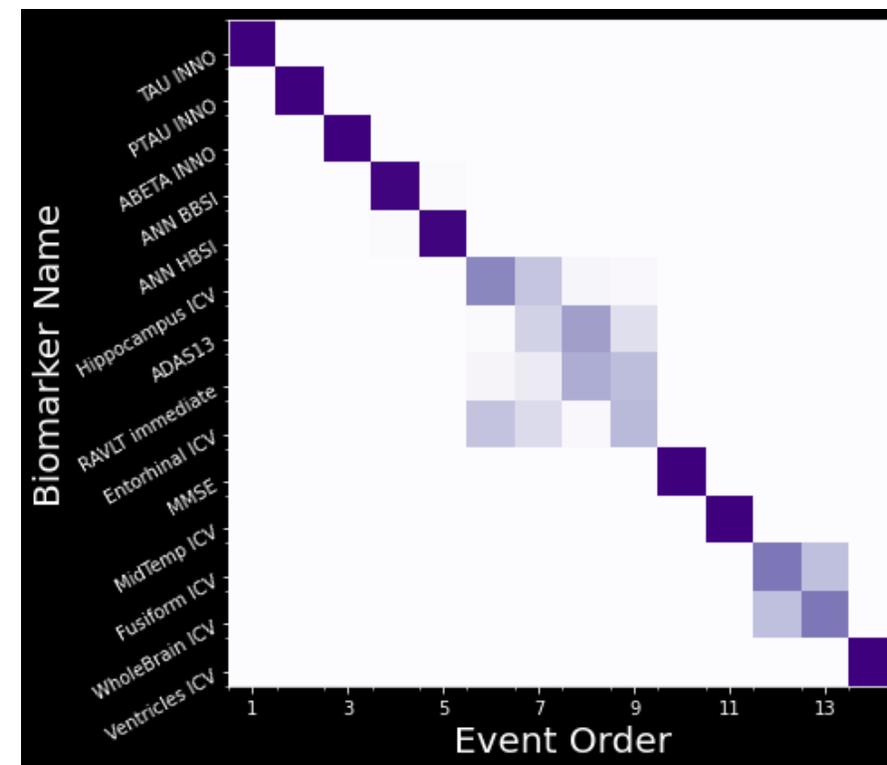
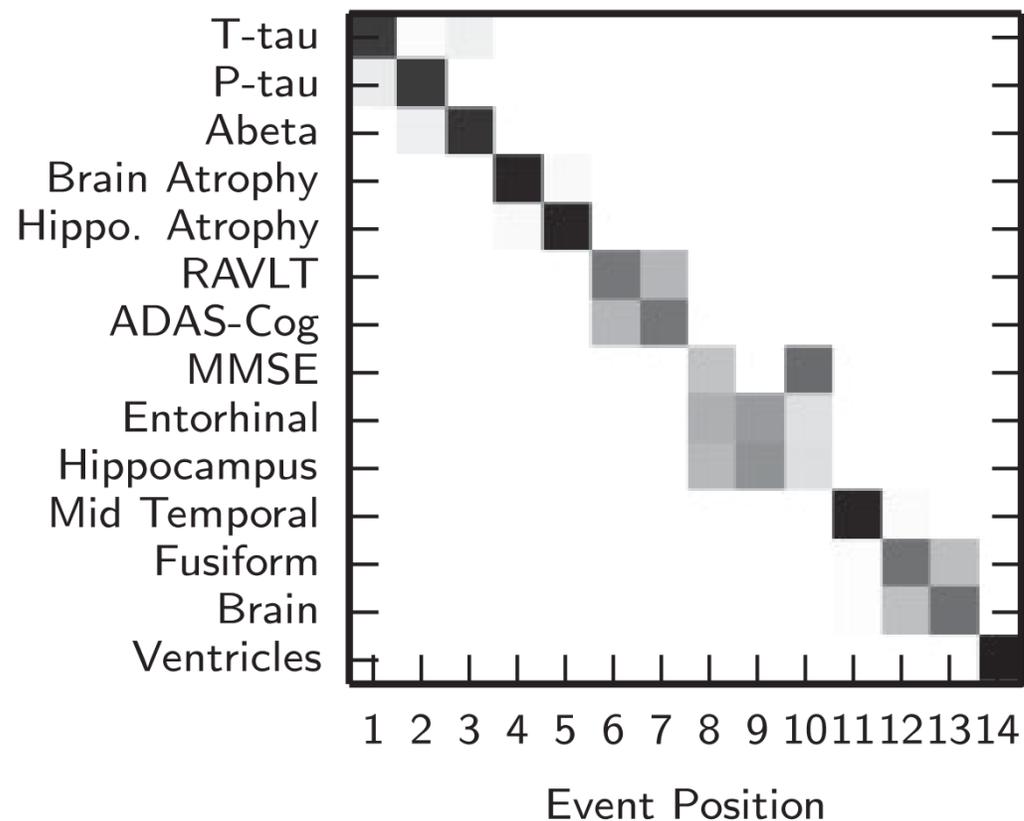
Step 3 - Visualise distributions of features

For each of the features, produce a plot showing the difference between the CN, LMCI, and AD groups.

```
In [ ]: # Your answer to Step 3
# Below put your code for visualising the differences in each features according to disease diagnosis
```

• Exercise notebooks

Example of results recreation



Results of the original paper from A.L Young et all, Brain 2014

Reproduced with KDE-EBM using GMM

Results of the pilot project

By the end of the project, attendees were able to:

- Worked together and practiced coding in groups or individually. Clinicians benefited from the engineers
- Learn from each other practices of data cleaning analysis and visualisation techniques using Python.
- Understood how to work, manipulate and process medical imaging data.
- Learned how to implement the model of the project (the Event Based Model) at the given data.
- Received the material and the knowledge to utilise the Notebooks (algorithms) to analyse their research's data sets.

Team Coders Challenges

- Collaboration issues due to the different time zones of the participants and their allocating time.
- The attendees didn't realize the time commitment requirements and a few of them gave up early.
- The training time in git (version control and collaborating software) was longer than expected. Training in git was a distraction rather than help.

Team Coders Lessons learned

- Reduce no-shows and early withdrawals by setting up more selective application process
- Provide virtual drop-in sessions to solve technical issues with set-ups ahead of the workshop
- Create short videos (basics in Python, git) as recommended pre-course viewing
- Reduce the number of demonstration notebooks but provide the non-essential ones as optional, extended exercises
- Put more weight into time zones when selecting applications and organising teams
- Set fewer tasks with clearer, firmer deadlines.

Conclusions

- IDEAS training programme is designed to cover the needs of healthcare scientists and clinicians to upskill in data and medical image analysis a big data management
- We have delivered two 4 day in-person workshops at UCL and one educational workshop at the Alzheimer Association Conference. Team members organised the MedICSS 2022 summer school and the 2022 XNAT Conference.
- We delivered for the first time TeamCoders a three-week long interactive workshop based in Experiential Learning in which we encouraged collaborative problem-solving via hands-on programming practice.
- We learned from the delivery's format and the challenges we encountered the first time we launched the programme and will apply actions to run future events more smoothly and accurately in time.

Thank you for listening!!

Mary Tziraki : m.tziraki@ucl.ac.uk

IDEAS's website :

<https://healthbioscienceideas.github.io>

Health+BioscienceIDEAS