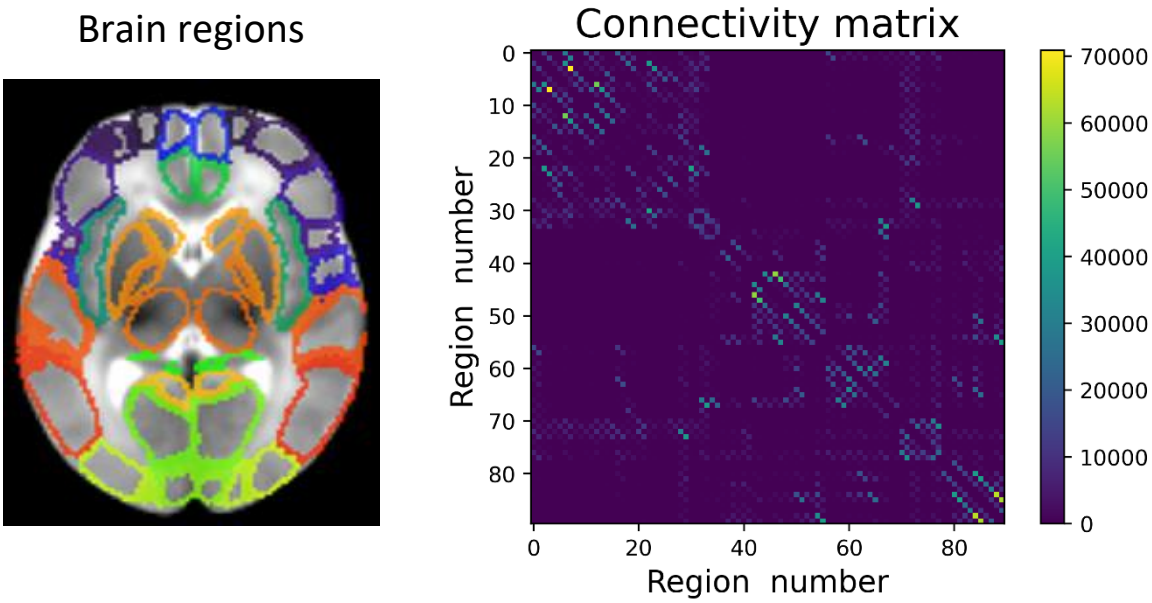
Classification of prematurity using structural connectivity

Mini Project – Part A

In this mini project you are asked to **predict whether the baby is born preterm** from the **structural connectivity matrices**. These matrices contain a measure of connectivity between pairs of brain regions. There is 90 regions and matrices have 90x90 values. They are symmetric and have zeros on the diagonal (the connectivity of a region with itself is not meaningful).



You are given a connectivity matrix for each baby, together with information about their age at scan and indication whether they were born preterm. Your goal is to build an accurate classifier to predict prematurity and identify connections that are most predictive of prematurity.

Structure of the project

Dataset

The data is provided in these files:

- **matrices.p** - the connectivity matrices as a 3D numpy array in pickle format
- **subject_info.csv** – age at scan and prematurity status (1 = premature, 0 = term-born)
- **label_names.csv** – names of the 90 brain regions
- **Mini Project A.ipynb –** a jupyter notebook template for your code. It contains code for loading the data.

Pre-processing

First, you will pre-process and reformat the data from the provided files to prepare them for training.

Analysis of the dataset

Next, you are asked to explore the structure of the data and list potential challenges you might encounter to train a classifier successfully

Evaluation framework

You then need to design your evaluation framework. Think of different sets of data that you will need and performance measures that you will use to tune and compare the models.

Methods

Design a classifier using the techniques from the weeks 2-4. Your goal is a well performing classifier, that addresses challenges you found when exploring the data. Stick to classical techniques, you will be asked to produce an advanced solution in part B of this project (due at the end of the term).

Results

Evaluate performance of your classifier using appropriate measures. Discuss whether you managed to address the challenges that you identified.

Submission

The submission consists of single **Jupyter notebook**, based on the provided template **Mini Project A.ipynb.** The code must be clean and working, otherwise it will not be marked. Results and plots need to be clearly displayed, we will not re-run your code while marking.

Upload your Jupyter notebook on **KEATS**.

Deadline: 1st November 2021, 4pm

Marking Scheme

The marks will be awarded for different sections as follows:

Pre-processing: 7 marks
Analysis of the dataset: 18 marks
Evaluation framework: 15 marks
Classifier design and training: 30 marks
Results: 30 marks

Total: 100 marks. The resulting marks will be converted to **15% of your grade**.