Project 1: Incorporating prior knowledge into machine learning algorithms to predict outcome for mental health conditions from brain connectivity data

Magnetic Resonance Imaging (MRI) enables us to image patients' brains, in a safe, non-invasive way, and can also be used to estimate connectivity between macroscopic brain regions (see Fig.1). Over the last 20 years, many studies have used MRI to study the brain's structure and function, showing for example that different regions of the brain are functionally specialised for specific tasks (Yeo et al, 2011), and differentially related to cognition and psychopathology (Xia et al, 2018). Different brain regions are also associated with different gene expression profiles (Hawrylycz et al, 2012).

Recently, there has been substantial interest in using MRI brain connectivity to predict patient outcomes for mental health conditions, for example schizophrenia. Predicting patient outcomes with high accuracy could help ensure that existing treatments are focussed on the patients at highest risk, whilst also elucidating the neural mechanisms that predispose patients to poor outcome, potentially opening the door to new treatments.

However, most machine learning studies to date have taken a primarily data driven approach and incorporating prior domain knowledge could improve both interpretability and accuracy. This is particularly important given that MRI brain scans are expensive to collect and therefore neuroimaging datasets available for machine learning are relatively small compared to more traditional machine learning fields. In this project, we will explore new approaches to integrate prior information about the brain's structure and function into machine learning algorithms. Possible directions include labelling brain regions or connections according to previously identified functions or gene expression profiles, semi-supervised methods in which machine learning algorithms first learn dimensions of the data which have previously been identified as important, or other approaches according to the student's interests.

This project would suit a student with an interest in Neuroimaging data, Network Science and applications of Machine Learning to Science.

The project will contribute to the Accelerate Programme for Scientific Discovery, an interdisciplinary research team that uses the power of machine learning to advance the frontiers of science. Students working on this project will join a community of committed researchers, working to promote and facilitate the use of machine learning techniques across research domains.

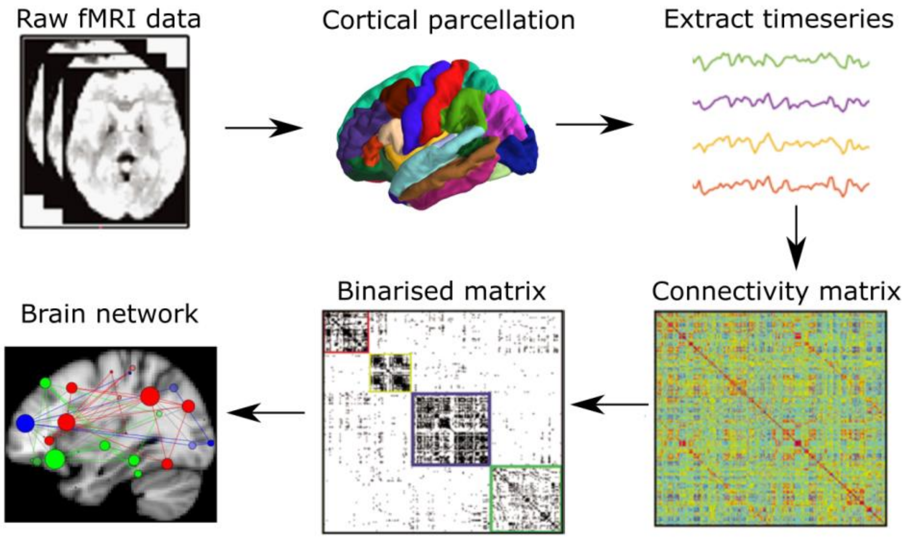

Figure 1: Constructing a functional brain network. We first parcellate the brain into approximately 100 macroscopic brain regions, and extract regional timeseries of brain activity. We then calculate the similarity of these timeseries pairwise between brain regions, to obtain a connectivity matrix. Figure adapted from (Homan et al, 2018).

Project 2: New network approaches to detect incoherent speech for psychotic disorders

Patients with psychotic disorders such as schizophrenia often exhibit subtle disorders of language, sometimes described as incoherent or disorganised speech (Liddle et al, 2002). Recent work suggests that automated NLP markers can detect reduced semantic coherence in speech transcripts from patients (Fig. 2A) and when coupled with machine learning may also be able to predict disease outcomes for individual patients (Corcoran et al 2018, Mota et al 2017), potentially opening the doorway to new diagnostic approaches and treatments. However, the NLP markers explored in the literature to date are relatively coarse-grained, and there is a clear need to develop more advanced NLP features of incoherent speech in psychosis to enable more accurate prediction of patient outcomes.

In this project, we will develop new tools to represent transcribed speech excerpts from patients as structured semantic networks (see Fig. 2B). We will use a range of methods to identify the key entities in the transcription and the relationships between them. We will then engineer features from these networks and begin to assess whether they might have additional predictive power compared to previously identified markers of incoherent speech.

This project would suit a student with an interest in NLP and Network Science.

The project will contribute to the Accelerate Programme for Scientific Discovery, an interdisciplinary research team that uses the power of machine learning to advance the frontiers of science. Students working on this project will join a community of committed researchers, working to promote and facilitate the use of machine learning techniques across research domains.

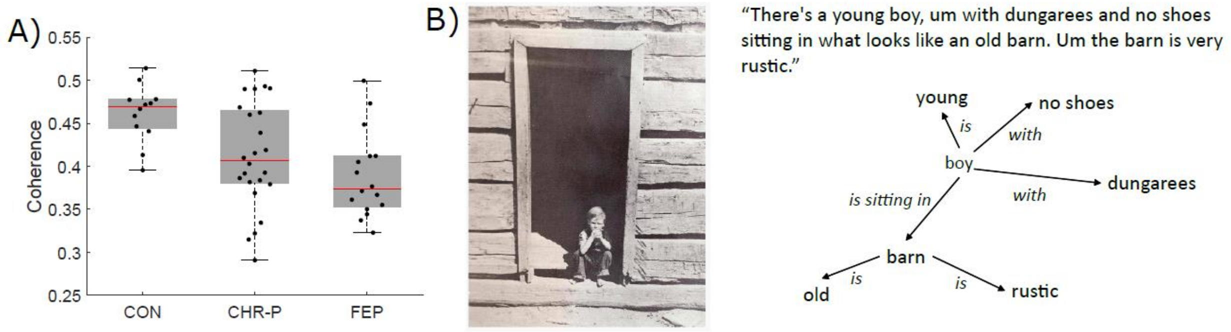

Figure 2: Semantic speech measures. A) Semantic coherence was decreased in first episode psychosis (FEP) patients compared to healthy controls (CON). B) Schematic of how a structured semantic speech network might look, for part of a patient's description of an image from the Thematic Apperception Test (Murray et al, 1943).