CANSSI CRT
Neuroimaging Data Analysis
Team Meeting 2016

CRT Meeting

The meeting will be held at the University of Victoria in the David Turpin building (DTB) room A104, and will take place July13th, 9:15AM - 6:00PM.

Scientific Program

Wednesday July 13^{th}

08:45 - 09:15 **Coffee**

09:15 - 09:30 **Welcome:** Farouk Nathoo and Linglong Kong

Session 1:
Chair: Farouk Nathoo

09:30 - 10:15 Wei Tu, *University of Alberta* p.4
Title: Automatic Brain Hematoma and Edema Segmentation using CT

10:15 - 11:00 (Grace) Cui Guo, *University of Michigan* p.4
Title: Scalar on Image Regression with Application to Multiple Sclerosis MRI Lesion Data

11:00 - 11:30 **Coffee Break**

Session 2
Chair: Linglong Kong

11:30 - 12:15 Yin Song, *University of Victoria* p.4
Title: Bayesian Source Reconstruction with Combined MEG and EEG Data

12:15 - 13:00 Roberto Vega, *University of Alberta* p.5
Title: The Successes and Challenges of Mental Illness Diagnosis using Machine Learning and fMRI Data

13:00 - 14:00 Lunch Break

Session 3
Chair: Farouk Nathoo

14:00 - 14:45 Stephen Rush, *University of Guelph* p.5
Title: Phylogenetic LASSO: Pruning the tree of life

14:45 - 15:30 Emma Smith, *University of Guelph* p.5
Title: Quantifying Health: Measuring the Impact of Fecal Microbiota Transplantation on Quality of Life

15:30 - 16:15 Yang Yu, *University of North Carolina* p.6
Title: iGWAS: Integrated Genome-Wide Association Analysis of Large-Scale Imaging Genetic Data

16:15 - 16:30 Coffee Break

Session 4
Chair: Linglong Kong

16:30 - 17:15 Kelly Sunderland, *Rotman Research Institute Baycrest* p.6
Title: Measuring the Inter-Scanner Variability of fMRI Scans from MRI Scanners in Ontario

17:15 - 18:00 Yao Chen, *Purdue University* p.7
Title: Local Region Sparse Learning for Image-on-Scalar Regression

18:00 **Closing:** Farouk Nathoo and Linglong Kong

Abstracts: Invited Student Talks

- **Speaker**: Wei Tu, University of Alberta

  Title: Automatic Brain Hematoma and Edema Segmentation using CT

  Abstract: An efficient and accurate segmentation of hematoma and edema using Computed Tomography (CT) is critical for patient diagnoses and further treatment. However, due to the substantial overlapping between the edema and surrounding brain tissue, image artifacts and the dynamic change of hematoma and edema, the time and cost required for the labor intensive process of manual localization and segmentation of hematoma and edema has become prohibitive. An automatic segmentation algorithm is presented in this talk. The method has 4 stages: (1) automatic midline detection using the skull, (2) region of interest localization, (3) hematoma and edema segmentation using non-local mean spatial fuzzy clustering, (4) segmentation results refinement based on statistical asymmetry of left and right hemisphere. The obtained results demonstrate the proposed algorithm provides objective and reproducible segmentations that are similar to the manual segmentation results. This is joint work with Dr. Linglong Kong, Dr. Rohana Karunamuni, Dr. Ken Butcher, Lili Zheng and Rebecca McCourt.

- **Speaker**: (Grace) Cui Guo, University of Michigan

  Title: Scalar on Image Regression with Application to Multiple Sclerosis MRI Lesion Data

  Abstract: Multiple sclerosis (MS) is an autoimmune disease that attacks the central nervous system. In particular, the immune system attacks the myelin sheath, which acts as an insulator for signal transmission between neurons, and causes a wide range of disabilities. Magnetic resonance imaging (MRI) plays a central role in the diagnosis and management of MS patients because damage to the myelin is visible on MRI. A research question of interest is whether these MRI images can predict MS subtype.

  Subtype classification is important because disease management and treatment are subtype specific. To answer this question we propose a Bayesian scalar-on-image regression model with scalar outcome (MS subtype) and binary image (presence or absence of lesion at each voxel obtained from MRI) covariates. Parameters of these covariates are spatially varying and are fitted using Gaussian random fields. Scalar covariates such as disease duration are also modeled. Our proposed model is fitted to both simulated data and a real data set consisting of 239 MS patients.

  A Hamiltonian Monte Carlo (HMC) algorithm is proposed to implement full Bayesian statistical inference. HMC can be more statistically efficient than other Markov Chain Monte Carlo methods when covariates are highly correlated. To reduce the computational burden, we code the problem to run in parallel on a graphical processing unit (GPU).

- **Speaker**: Yin Song, University of Victoria

  Title: Bayesian Source Reconstruction with Combined MEG and EEG Data

Abstract: In this talk we begin with an overview of the inverse problem incurred in a medical imaging study when we use MEG/EEG data to measure electromagnetic brain activity over an array of sensors at the scalp and it is of interest to localize the corresponding sources of neural activity within the brain. We review some of the existing approaches to solving this inverse problem and discuss the meso-state-space model (MSM) proposed by Daunizeau and Friston (2007). We then propose a new model that builds on the MSM and incorporates three major extensions: (i) We combine the EEG and MEG data together and formulate a joint model for source reconstruction; (ii) we incorporate the Potts model to represent the spatial dependence in an allocation process that partitions the cortical surface into a small number of meso-sources; (iii) we formulate the meso-state dynamics in a flexible manner so that the model can characterize the functional connectivity between meso-sources. I will discuss strategies for Bayesian computation. This is joint work with Farouk Nathoo.

- **Speaker**: Roberto Vega, University of Alberta

  Title: The successes and challenges of mental illness diagnosis using machine learning and fMRI data

  Abstract: Functional magnetic resonance imaging (fMRI) is an imaging technology that estimates brain activity by measuring changes in blood oxygenation. An interesting question to ask is if it is possible to find brain activity patterns that we could use to diagnose different mental illnesses such as autism, schizophrenia, or depression, among others. Unlike association studies, which aim to create models with high explanatory power, predictive studies create classifiers that make predictions for previously unseen data using a multivariate approach. Machine learning offers tools for this latter purpose. The standard approach involves three steps: 1) preprocessing of the data, 2) extraction of relevant features, and 3) use of a learning algorithm to create a classifier. The high dimensionality of the data, relatively small sample sizes, the presence of time series, and the use of data obtained using different MRI scanners make this a non trivial task. In this talk I will present some of the approaches that the group led by Russ Greiner and Matt Brown at the University of Alberta is taking for tackling these challenges, as well as some encouraging results that have been achieved for classifying healthy controls vs. people with mental illnesses.

- **Speaker**: Stephen Rush, University of Guelph

  Title: Phylogenetic LASSO: Pruning the tree of life

  Abstract: We model the effects of bacterial composition of the human gut microbiome by applying a hierarchical LASSO in parallel to penalize the variables corresponding to high level organismal classification, in essence pruning the tree of life by taxon. Based on Zhu and Zhou's hierarchical LASSO, we employ this in seeking the contributors to recovery in faecal microbiota therapy for recurrent Clostridium difficile infection. This can be extended to coincident '-omics' data, with applications to study of the gut-brain axis.

- **Speaker**: Emma Smith, University of Guelph

Title: Quantifying Health: Measuring the Impact of Fecal Microbiota Transplantation on Quality of Life

Abstract: C. difficile infection (CDI) is infamous for its antibiotic resistance, with 15 to 28% of patients experiencing a recurrence following completion of an antibiotic treatment regimen. Fecal microbiota transplantation (FMT) has become a promising treatment alternative for recurrent CDI, with some studies reporting clinical resolution rates of greater than 90%. It is of interest whether FMT leads to a greater increase in patient quality of life compared to the current standard of care and if so, whether this increase is commensurate of any additional costs of treatment. The RAND 36-item short form survey was employed over the course of a FMT clinical trial in order to obtain a quantitative measure of the impact of this treatment on eight dimensions of a patients health. However, the final data set was negatively impacted by the large number of missing observations; as is often the case with longitudinal and survey data. This presentation will examine the use of multivariate imputation by chained equations (MICE) and imputation via pattern-mixture models, at both the compound (dimension score) and component (individual question) variable levels. A shrinkage estimator proposed by Zhao, Cook, and Wu (2016), which combines estimators resulting from both the compound and component imputations, is introduced and examined in the context of the data. A preliminary comparison of the various estimators is conducted.

- **Speaker**: Yang Yu, University of North Carolina

  Title: iGWAS: integrated genome-wide association analysis of large-scale imaging genetic data

  Abstract: More and more large-scale imaging genetic studies are being conducted to collect a rich set of imaging, genetic, and clinical data to detect putative genes for complexly inherited neuropsychiatric and neurodegenerative disorders. Besides major big-data challenges arising from testing genome-wide associations with signals at millions of locations in the brain from thousands of subjects, current methods also suffer from the drawback that they are unable to incorporate known disease information in the analysis. The aim of this paper is to develop a novel integrated Genome-Wide Association analysis (iGWAS) framework in order to integrate available disease information into the whole-genome analyses of whole-brain data. The iGWAS consists of three components including a heteroscedastic linear model, a global sure independence screening procedure, and a detection procedure based on wild bootstrap methods. Simulation studies show that our iGWAS is more efficient than conventional GWAS approaches in identifying disease-related genetic variants and neuroimaging biomarkers. We have successfully applied iGWAS to a large-scale imaging genetic data analysis of ADNI data. Our iGWAS shall be a valuable statistical toolbox for large-scale disease-oriented neuroimaging genetic analysis.

- **Speaker**: Kelly Sunderland, Rotman Research Institute Baycrest

  Title: Measuring the Inter-Scanner Variability of fMRI Scans from MRI Scanners in Ontario

  Abstract: The Ontario Neurodegenerative Disease Research Initiative (ONDRI) is a

multisite, multimodal, longitudinal study of five neurodegenerative diseases. In order for data to be comparable across so many dimensions, it is important that measurement bias is minimized. This is particularly challenging for the neuroimaging assessment as MRI scanners are known to be different between manufacturers, individual units, and change over time. Here, we focus on the inter-scanner fMRI differences and present a quality control process for quantifying the between-site variability.

fBIRN phantoms are scanned monthly at each of the eight scanning sites for a 12 to 24 month period. Two manufacturers are represented in these eight scanners. Fifteen quality assurance parameters (such as drift, ghosting, and mean intensity) are measured using the fBIRN phantom processing pipeline developed by the Biomedical Informatics Network. Images more than three standard deviations away from previous measurements in any of the parameters indicate an acquisition error or scanner problem and are excluded. Scanner-specific outliers are removed using a robust Mahalanobis distance metric on the first two principal components of an initial principal component analysis (PCA). Therefore, the dataset used for analysis has 230 observations over eight sites, each with 13 to 28 measurements, with the exception of one with 92 measurements, as the phantom was scanned weekly.

A PCA is used to characterize the multivariate structure of the scanners, and 90% of the variance is explained by the first four principal components. Resampling is used to quantify the within-scanner variance and the ways in which it could affect the PCA. Since principal components can change order or be reflected during resampling, a Procrustes rotation is necessary to make each iteration comparable with the PCA of observed data. Finally, the 95% confidence intervals are calculated and the overlap between sites is observed.

In general, scanners have small intra-scanner variability once outliers are removed and cluster well with other scanners of the same manufacturer. However, two scanners deviate from their respective clusters as higher levels of ghosting are identified, exemplifying the differences in scanners, regardless of manufacturer consistencies. Finally, while there are considerable differences in the characteristics of the scanners that are produces by different manufacturers, they are primarily driven by resolution and smoothing differences. The potential to remove these differences through a preprocessing pipeline will be investigated in subsequent analyses. This is joint work with Aditi Chemparathy, Stephen Arnott, Fan Dong, Tom Gee, Christopher Scott, Rob Bartha, Sandra Black, Santanu Chakraborty, Jean Chen, Adrian Crawley, General Leung, Nancy Lobaugh, Michael Noseworthy, Patrick Stroman, Malcolm Binns, Tanya Schmah, Stephen Strother & ONDRI Investigators

- **Speaker**: Yao Chen, Purdue University

  Title: Local Region Sparse Learning for Image-on-Scalar Regression

  Abstract: TBD