SMI 2023 Program

Monday 5/22

830 - 10am	Tutorial: Introduction to the tidyverse and Quarto *Location: Pinnacle Ballroom*	
10 - 1015am	Break	
1015 - 1145am	Tutorial: An Introduction to the ANTsX ecosystem through R *Location: Pinnacle Ballroom*	
1145am - 110pm	Lunch	
110 - 240pm	Student Paper Competition Winners *Location: Pinnacle Ballroom*	
240 - 250pm	Break	
250 - 420pm	New approaches to analyzing neuroimaging data *Location: Pathways*	Collaborative Case Study: Novel application to neurological and neuropsychiatric diseases using different imaging techniques *Location: Pinnacle Ballroom*
420 - 6pm	Poster Presentations / Mixer *Location: Pinnacle Foyer and Pinnacle Ballroom*	

Tuesday 5/23

830 - 10am	New Statistical Methods to Improve the Spatial-omics Analysis Pipeline *Location: Pathways*	Recent Developments on Brain Imaging Analysis *Location: Pinnacle Ballroom*
10 - 1010am	Break	
1010 - 1110am	Keynote - Kelvin Lim *Location: Pinnacle Ballroom*	
1110am - 1240pm	Lunch	
1240 - 210pm	Bayesian Methods *Location: Pathways*	Collaborative Case Study: Statistical Methods and Findings from Large Consortia Studies *Location: Pinnacle Ballroom*
210 - 220pm	Break	
220 - 350pm	Modern Statistical Methodology on Spatial, Neuroimaging, and Shape Data Analysis *Location: Pathways*	Recent advancements in statistical methods for brain connectome analysis *Location: Pinnacle Ballroom*
350 - 4pm	Break	

4 - 530pm	Recent Advances in Neuroimaging Statistics for Investigating Human Brain Function *Location: Pathways*	Efficient Modeling of Multi-Region High-Dimensional Molecular Data *Location: Pinnacle Ballroom*

Wednesday 5/24

830 - 10am	Recent Advances in Spatial Analysis of Single-cell Imaging *Location: Pathways*	Statistical Methods for Brain Connectomes *Location: Pinnacle Ballroom*
10 - 1010am	Break	
1010 - 1110am	Keynote - Mingyao Li *Location: Pinnacle Ballroom*	
1110am - 1240pm	Lunch	
1240 - 140pm	Founder's Talk - Ranjan Maitra *Location: Pinnacle Ballroom*	
140 - 150pm	Break	
150 - 320pm	Statistical Methods for Analyzing Multiview and Multi-session Imaging Data *Location: Pathways*	Collaborative Case Study: Event-related potential brain-computer interface data present challenges and opportunities for novel statistical methods *Location: Pinnacle Ballroom*
320 - 330pm	Break	
330 - 5pm	Advances in Statistical Methods for Transmission Electron Microscopy *Location: Pathways*	Penalized Regression and Functional Data Analysis *Location: Pinnacle Ballroom*
5 - 510pm	Closing Remarks *Location: Pinnacle Ballroom*	

Tutorial Information

Introduction to the tidyverse and Quarto

David Schneck, Masonic Institute for the Developing Brain

Description: Tidyverse is a collection of open-source R packages designed with a similar philosophy and structure that aim to make data import, tidying, manipulation, and visualization straightforward and easily-reproducible. Tidyverse contains some of the most well-known and useful R packages for any application of data science, and is a must-have for any aspiring or veteran statistician or data scientist. The following packages will be explored in this workshop with time designated for hands-on coding examples:

- dplyr: Data manipulation and pipes
- ggplot2: data visualization and graphic creation
- tidyr: tools to create clean and tidy working data
- tibble: a re-imagining of the typical dataframe structure
- stringr: tools for working with strings
- Other packages that will be touched upon: readr, purrr, forcats

The second portion of this session will focus on the new technical publishing system: Quarto. Quarto builds upon the utility of Rmarkdown but adds several new features for utility, interactability, readability, and dissemination of analyses. Though Quarto can be used with other languages, we will focus on creating and exploring documents in R. Quarto is a cutting edge resource that will surely help in creating beautiful publications and other documents. A zip file containing materials to be used during the tutorial can be found in this Google Drive Directory.

Preliminary steps:

1. If you have yet to do so, download the appropriate version of R for your system at https://cran.r-project.org/.
2. After you have R downloaded onto your computer, follow this link to download Rstudio IDE: https://posit.co/products/open-source/rstudio/
3. If you have issues with either of the above steps, please follow this step-by-step tutorial to complete steps 1 and 2: https://rstudio-education.github.io/hopr/starting.html.
4. Lastly, please download Quarto software for your appropriate system using the following link: https://quarto.org/docs/get-started/.

An Introduction to the ANTsX ecosystem through R

Nick Tustison, University of Virginia
Brian Avants, University of Virginia

Description: The Advanced Normalizations Tools ecosystem, known as ANTsX, consists of multiple open-source software libraries which house top-performing algorithms used worldwide by scientific and research communities for processing and analyzing biological and medical imaging data. More recent enhancements include statistical, visualization, and deep learning capabilities through interfacing with both the R statistical project (ANTsR) and TensorFlow/Keras libraries (ANTsRNet). In this hands-on tutorial, we will showcase much of the available general and application-specific functionality in these R-based libraries which will permit the participant to create their own processing and analysis pipelines.

Github site for the tutorial: https://gist.github.com/ntustison/12a656a5fc2f6f9c4494c88dc09c5621

Preliminary steps:

1. ANTsR should be installed. The easiest route would be to use devtools, i.e.,

   devtools::install_github(“ANTsX/ANTsR”)

   which should also install the dependencies including ITKR and ANTsRCore.
2. Install ANTsRNet the same way:

   devtools::install_github(“ANTsX/ANTsRNet”)

After installation, they can test to see if installation is correct by running a self-contained brain extraction routine: https://gist.github.com/ntustison/12a656a5fc2f6f9c4494c88dc09c5621#brain-extraction, which will test both ANTsR and ANTsRNet installation. (Skip over the Python analog.) If there are any issues, please reach out, preferably through the GitHub repo site.

Keynote and Founder's Talk Abstracts

Causal discovery analysis for measuring brain connectivity

Kelvin O. Lim, M.D.
Drs. T.J. and Ella M. Arneson Land Grant Chair in Human Behavior
Professor and Director of Adult Mental Health Research
University of Minnesota
Chair: Zhengwu Zhang, UNC Chapel Hill

There are roughly 100 billion neurons in the human brain, which are estimated to form approximately 100 trillion connections. Many brain disorders (e.g. schizophrenia, depression, addiction, dementia) are attributed to the disruption of these connections. Understandably, there is tremendous interest in using neuroimaging to map and quantify connections in the living brain in health and disease. Research initiatives such as the Human Connectome Projects, Adolescent Brain Cognitive Development Study, Healthy Brain And Child Development Study and the UK Biobank have used resting fMRI to measure brain connectivity from tens of thousands of participants. The most common connectivity metric used is the Pearson correlation coefficient. Correlations unfortunately do not provide as much information as we often want. In particular, they do not reveal whether the correlation is due to one region causing the other, or whether they cause each other, or whether some other brain region is causing them both. This additional information is critical for understanding brain functioning and for developing interventions that address psychopathology

In this talk, I will first give a brief overview of methods that have been proposed to model causal relationships in brain connectivity and their limitations. I will then describe how we have applied causal discovery analysis to search for causal models from resting fMRI time courses to estimate brain connectivity. Finally, I will present results from clinical neuroimaging studies where causal models are used to examine brain connectivity.

Integrating spatial transcriptomics with histology to infer super-resolution tissue architecture

Mingyao Li, PhD
Professor of Biostatistics
Department of Biostatistics, Epidemiology and Informatics
University of Pennsylvania Perelman School of Medicine
Chair: Eric Lock, University of Minnesota

The rapid development of spatial transcriptomics (ST) technologies has made it possible to measure gene expression within the original tissue contexts. The applications of ST have enabled researchers to characterize spatial gene expression patterns, study cell-cell communications, and resolve the spatiotemporal order of cellular development, which have transformed our understanding of the functional organization of tissues. Previous studies have shown that gene expression patterns are correlated with histological features, suggesting that gene expression can be predicted from histology images. However, these existing methods do not fully utilize the rich cellular information provided by high-resolution histology images. In this talk, I will present methods that we recently developed that aim to integrate gene expression with histology to computationally reconstruct ST data that cover the entire transcriptome with near-single-cell resolution. Through comprehensive analysis of diverse datasets generated from both diseased and normal tissues, we show that our super-resolution gene prediction is accurate and useful for different applications in tissue architecture inference.

Fourier-structured tensor-variate distributions for use in high-resolution imaging applications

Ranjan Maitra, PhD
Professor of Statistics
Department of Statistics
Iowa State University
Chair: Michele Guindani, University of California, Los Angeles

Data in the form of arrays (or tensors) are ubiquitous in imaging and other contexts, and are usually analyzed using methodologies that impose simplified structures on the tensor-variate structure of their mean or variance. We introduce the Fourier tensor-variate (FTV) family of

distributions, with covariance matrices whose eigenvectors are specified by the real discrete Fourier transform (RDFT). An attractive feature of this covariance specification is its ability to capture nonstationarity while maintaining periodicity. Further, a random tensor with the correspondingly-named Fourier covariance structure is element-wise independent after applying an inverse RDFT. Therefore, traditional univariate distributions can be extended to their FTV counterpart, with inference on the induced FTV family mirroring that of their univariate counterparts, while enjoying the computational benefits of using the Fourier transform. Indeed, the estimation of the high-dimensional tensor covariance is delegated to the estimation of its eigenvalues, naturally allowing for principal component analysis (PCA) to summarize variability. Our methods are evaluated in simulations involving bitmap images, and illustrated on applications involving digital imaging, precision agriculture and medical imaging. This work is joint with Carlos Llosa-Vite of Sandia National Laboratories.

Student Paper Competition Winners

Chair: Mark Fiecas, University of Minnesota

Presenter: Jialu Ran, Emory University
Title: Nonparametric motion adjustment in studies of functional connectivity alterations in autism spectrum disorder children
Abstract: Autism spectrum disorder (ASD) is a common neurodevelopmental condition associated with difficulties with social interactions, communication and other behaviors. To study the characteristics of ASD, investigators often use functional connectivity (FC) derived from resting-state functional magnetic resonance imaging. However, participants' head motion during the scanning session can induce motion artifacts. Many studies remove scans with excessive motion, but children that move more tend to have more severe symptoms. Scan exclusion can lead to drastic reductions in sample size and introduce selection bias. In this study, we propose a framework to decompose neural and motion-induced sources of FC group differences between autistic children and typically developing children, without excluding high-motion participants. In particular, we adjust for motion via causal mediation with stochastic interventions, where motion and other covariates are flexibly modeled using an ensemble of machine learning methods. We establish the theoretical large-sample efficiency and multiple robustness of our proposed estimators and validate the statistical properties of the estimators using simulation studies. The framework is applied to estimate the difference in functional connectivity between autistic children and typically developing children. Our analyses indicate that some long-range connections between a seed region in the default mode network and frontal-parietal regions exhibit hyperconnectivity in ASD. Naıvely including high-motion children appears to cause spurious connectivity differences. Naıvely excluding high motion children removed group differences.

Presenter: Zhiling Gu, Iowa State University
Title: TSSS: A Novel Triangulated Spherical Spline Smoothing for Surface-based Imaging
Abstract: Surface-based imaging has been widely observed and analyzed in practice, especially in atmospheric and planetary science. Examples include oceanic near-surface atmospheric data, aerosol optical depth data, and the cosmic microwave background radiation field. In this talk, we introduce a novel nonparametric method to efficiently discover the underlying signals on surface-based complex domains. Our approach involves a penalized spline estimator defined on a triangulation of surface patches, which ensures both signal matching and smoothness. We investigate the asymptotic behavior of the proposed estimator and show that it converges to the desired solution. We conduct simulation experiments and data applications on cortical surface functional magnetic resonance imaging data and oceanic near-surface atmospheric data, demonstrating the superior performance of our method.

Presenter: Yuchen Xu, Cornell University
Title: Dynamic Atomic Column Detection in Transmission Electron Microscopy Videos via Ridge Estimation
Abstract: Ridge detection is a classical tool to extract curvilinear features in image processing. As such, it has great promise in applications to material science problems; specifically, for trend filtering relatively stable atom-shaped objects in image sequences, such as Transmission Electron Microscopy (TEM) videos. Standard analysis of TEM videos is limited to frame-by-frame object recognition. We instead harness temporal correlation across frames through simultaneous analysis of long image sequences, specified as a spatio-temporal image tensor. We define

new ridge detection algorithms to non-parametrically estimate explicit trajectories of atomic-level object locations as a continuous function of time. Our approach is specially tailored to handle temporal analysis of objects that seemingly stochastically disappear and subsequently reappear throughout a sequence. We demonstrate that the proposed method is highly effective and efficient in simulation scenarios, and delivers notable performance improvements in TEM experiments compared to other material science benchmarks.

Presenter: Yizi Zhang, Columbia University
Title: Motion-Invariant Variational Auto-Encoding of Brain Structural Connectomes
Abstract: Mapping of human brain structural connectomes via diffusion MRI offers a unique opportunity to understand brain structural connectivity and relate it to various human traits, such as cognition. However, motion artifacts from head movement during image acquisition can impact the connectome reconstructions, rendering the subsequent inference results unreliable. We aim to develop a generative model to learn low-dimensional representations of structural connectomes that are invariant to motion artifacts, so that we can link brain networks and human traits more accurately, and generate motion-adjusted connectomes. We applied the proposed model to data from the Adolescent Brain Cognitive Development (ABCD) study and the Human Connectome Project (HCP) to investigate how our motion-invariant connectomes facilitate understanding of the brain network and its relationship with cognition. Empirical results demonstrate that the proposed motion-invariant variational auto-encoder (inv-VAE) outperforms its competitors on various aspects. In particular, motion-adjusted structural connectomes are more strongly associated with a wide array of cognition-related traits than other approaches without motion adjustment.

Invited Sessions Abstracts

New approaches to analyzing neuroimaging data

Organizer: Benjamin Risk, Emory University
Chair: Kristin Linn, University of Pennsylvania

Presenter: Meimei Liu, Virginia Tech
Title: Graph Generative Models with Application in Brain Imaging Datasets
Abstract: There has been a huge interest in studying human brain connectomes inferred from different imaging modalities and exploring their relationships with human traits, such as cognition. Brain connectomes are usually represented as networks, with nodes corresponding to different regions of interest (ROIs) and edges to connection strengths between ROIs. Due to networks' high-dimensionality and non-Euclidean nature, it is challenging to depict their population distribution and relate them to human traits. Building on recent advances in deep learning, this work focuses on two tasks in learning brain graphs. (1) Population distribution learning. We develop a nonlinear latent factor model to characterize the population distribution of brain graphs and infer their relationships to human traits. (2) Interpretable transfer learning for graph inference. We aim to extract and transfer the knowledge learned from large-scale studies conducted under different sources to assist the inference in the target small-scaled cases. We applied the developed approaches to two large-scale brain imaging datasets, the Adolescent Brain Cognitive Development (ABCD) study and the Human Connectome Project (HCP) for adults, to study the structural brain connectome and its relationship with cognition.

Presenter: Jun Young Park, University of Toronto
Title: Spatially-enhanced clusterwise inference for testing and localizing intermodal correspondence
Abstract: With the increasing availability of neuroimaging data from multiple modalities—each providing a different lens through which to study brain structure or function—new techniques for comparing, integrating, and interpreting information within and across modalities have emerged. Recent developments include hypothesis tests of associations between neuroimaging modalities, which can be used to determine the statistical significance of intermodal associations either throughout the entire brain or within anatomical subregions or functional networks. While these methods provide a crucial foundation for inference on intermodal relationships, they cannot be used to answer questions about where in the brain these

associations are most pronounced. We introduce a new method, called CLEAN-R, that can be used both to test intermodal correspondence throughout the brain and also to localize this correspondence. Our method involves first adjusting for the underlying spatial autocorrelation structure within each modality before aggregating information within small clusters to construct a map of enhanced test statistics. Using structural and functional magnetic resonance imaging data from a subsample of children and adolescents from the Philadelphia Neurodevelopmental Cohort, we conduct simulations and data analyses where we illustrate the high statistical power and nominal type I error levels of our method. By constructing an interpretable map of group-level correspondence using spatially-enhanced test statistics, our method offers insights beyond those provided by earlier methods.

Presenter: Kaizhou Lei, Florida State University
Title: When mediation analysis faces subgroup heterogeneity
Abstract: Mediation analysis is an essential tool in the imaging genetics study forAlzheimer's disease (AD). The goal is to identify the causal mechanism or pathway that links genetic exposures and neurological outcomes through some neuroimaging mediators. Although various mediation analysis approaches have been proposed to discover the underlying causal pathway in AD, there are several challenges, such as the subgroup heterogeneities in terms of (i) brain connectomes and (ii) causal mechanisms. To address these issues, we propose a novel mediation analysis tool that can simultaneously detect individual brain connectomes and subgroup causal pathways. Specifically, a two-layer structure equation model, including a mixture of conditional Gaussian graphical models, is developed to establish the heterogeneous mediation pathways. A penalized EM algorithm is proposed to estimate both average direct effect and indirect effect. Both simulation studies and a real example analysis using the diffusion tensor imaging data from the ADNI study are conducted to assess the finite sample performance of our method.

Presenter: Zhengwu Zhang, University of North Carolina
Title: Alignment of Continuous Structural Connectivity
Abstract: TBD

Collaborative Case Study: Novel application to neurological and neuropsychiatric diseases using different imaging techniques

Organizer: Dana L. Tudorascu, University of Pittsburgh
Chair: Daniel Rowe, Marquette University

Presenter: Elizabeth Sweeney, University of Pennsylvania
Title: Quantitative Susceptibility Maps in Multiple Sclerosis Lesions
Abstract: Multiple sclerosis (MS) is an inflammatory disease of the central nervous system characterized by lesions in the brain and spinal cord. Magnetic resonance images (MRI) are sensitive to these lesions. A particular type of lesion, called a chronic active lesion, is characterized by a hyperintense rim of iron-enriched, activated microglia and macrophages, and has been linked to greater tissue damage. An MRI technique called quantitative susceptibility mapping (QSM) provides efficient in vivo quantification of susceptibility changes related to iron deposition and identifies these chronic active lesions, called QSM rim positive (rim+) lesions. QSM rim+ MS lesions and their longitudinal behavior have the potential to serve as a biomarker of chronic inflammation and to be utilized to monitor disease progression and evaluate disease-modifying therapies in MS. In this talk, I will discuss the challenges of estimating treatment effects using the longitudinal behavior of QSM rim+ lesions. I will compare two disease-modifying treatments, Tecfidera® and Copaxone®, using linear mixed effects regression models with inverse probability of censor weighting. One of the major limitations of this model is that the inflammatory stage or age of the lesion is unknown, causing misregistration of the lesion-level data. I will also introduce methodology to estimate the age of MS lesions using cross sectional MRI information.

Presenter: Ceren Tozlu, Cornell University
Title: Identification of the sequence of structural disconnections due to multiple sclerosis lesions using an event-based modeling approach
Abstract: Objective: Prediction of disease progression is challenging in multiple sclerosis (MS) as the sequence of lesion development and retention of inflammation within a subset of chronic lesions is heterogeneous among patients. We investigated the sequence of lesion-related regional structural disconnectivity across the spectrum of disability and cognitive impairment in MS.
Methods: In a full cohort of 482 patients, the Expanded Disability Status Scale was used to classify patients into

(i)no or mild vs (ii)moderate or severe disability groups. In 363 out of 482 patients, Quantitative Susceptibility Mapping was used to identify paramagnetic rim lesions (PRL), which are maintained by a rim of iron-laden innate immune cells. In 171 out of 482 patients, Brief International Cognitive Assessment was used to identify subjects with cognitive impairment. Network Modification Tool was used to estimate the regional structural disconnectivity due to MS lesions. Discriminative event-based modeling was applied to investigate the sequence of regional structural disconnectivity due to all representative lesions across the spectrum of disability and cognitive impairment.
Results: Structural disconnection in the ventral attention and subcortical networks was an early biomarker of moderate or severe disability. The earliest biomarkers of disability progression were structural disconnections due to PRL in the motor-related regions. Subcortical structural disconnection was an early biomarker of cognitive impairment.
Discussion: MS lesion-related structural disconnections in the subcortex is an early biomarker for both disability and cognitive impairment in MS. PRL-related structural disconnection in the motor cortex may identify the patients at risk for moderate or severe disability in MS.

Presenter: Seonjoo Lee, Columbia University
Title: Longitudinal psychiatric comorbidity patterns and its neural underpinnings in children and adolescent.
Abstract: A prominent concern in child psychiatry is the high frequency of co-occurring psychiatric disorders in youth, given that comorbidity increases the risk for functional impairment. However, studying comorbidity is often low-powered due to the possible diagnosis combinations. Another approach is based on factor analysis. However, the existing methods do not directly model or account for comorbidity. Alternatively, we propose a dyadic cluster analysis by defining the co-occurrence of two diagnoses as a link, forming the diagnostic comorbidity network. A simulation study showed that dyadic cluster analysis identified heterogeneous comorbidity patterns better than current factor analysis-based cluster analysis. We applied the dyadic cluster analysis to the ABCD baseline data stratified by sex. We identified 4 clusters with different comorbidity patterns for both sexes, and those four clusters showed different brain morphometry in girls.

Presenter: Dana Tudorascu, University of Pittsburgh
Title: Linear Effect of Inter-Scanner Variability: Insights from Paired Cross-Scanner T1-weighted Images in elderly subjects
Abstract: The collection of structural MRI data across sites increases statistical power and enables the generalization of research outcomes; however, due to the variety of imaging acquisition, inter-scanner variability hinders the direct comparability of multi-scanner MRI data. Thus, many harmonization methods have been proposed to reduce inter-scanner variability in the image domain. Although proposed methods, especially incorporating deep learning techniques, have achieved promising performance, interpretability and understanding of inter-scanner variability were still limited. In this study we investigate a small sample of eighteen cognitively normal participants, each scanned on 4 different 3T scanners including GE, Philips, Siemens-Prisma and Siemens-Trio during a short period of time (at most few weeks apart). We applied and extended a statistical harmonization method, ComBat, to the image domain and investigated the linear effect of inter-scanner variability and image quality metrics. Furthermore, we attempted to harmonize cross-scanner images by removing the estimated site effect. Besides estimating parametric maps of site effect, we calculated image quality metrics using MRIQC and similarity index to investigate the manifestation of scanner-related variation. Voxel-based morphometry using CAT12 was used to estimate cortical volumetric measures to compare the difference before and after the harmonization.

New Statistical Methods to Improve the Spatial-omics Analysis Pipeline

Organizer: Simon Vandekar, Vanderbilt University
Chair: Julia Wrobel, Colorado School of Public Health

Presenter: Gregory Hunt, William & Mary
Title: Leveraging design to remove spatial artifacts in high-throughput imaging assays
Abstract: High-throughput profiling of tissues using automated imaging techniques is an increasingly powerful approach used broadly across the life sciences. These approaches allow the simultaneous extraction of thousands of image features from each of thousands of images thereby building up a quantitative profile of the tissue under different conditions. These profiles may then be analyzed with statistical approaches. Like other high-throughput techniques, unwanted technical and biological effects can often obscure the desired biological signal of interest. This talk discusses how a prospective design of experiments can enable powerful normalization of these types of data. In particular, the approach we develop leverages multiple levels of replication to enable the discovery and removal of unwanted effects in the data. This is demonstrated for the microenvironment

microarray (MEMA) platform in a study of cellular microenvironmental perturbation. In this application the proposed approach removes strong, unwanted, spatial artifacts present in the data and thereby enhances biological signal. More generally, the approach shows strong promise for a wide selection of similar biological assays.

Presenter: Mansooreh Ahmadian, University of Colorado Denver-Anschutz Medical Campus
Title: A Platform-Independent Framework for Phenotyping of Multiplex Tissue Imaging Data
Abstract: Multiplex imaging is a powerful tool to analyze the structural and functional states of cells in their morphological and pathological contexts. However, hypothesis testing with multiplex imaging data is a challenging task due to the extent and complexity of the information obtained. Various computational pipelines have been developed and validated to extract knowledge from specific imaging platforms. A common problem with customized pipelines is their reduced applicability across different imaging platforms: Every multiplex imaging technique exhibits platform-specific characteristics in terms of signal- to-noise ratio and acquisition artifacts that need to be accounted for to yield reliable and reproducible results. We propose a pixel classifier-based image preprocessing step that aims to minimize platform-dependency for all multiplex image analysis pipelines. Signal detection and noise reduction as well as artifact removal can be posed as a pixel classification problem in which all pixels in multiplex images can be assigned to two general classes of either I) signal of interest or II) artifacts and noise. The resulting feature representation maps contain pixel-accurate representations of the input data but exhibit significantly increased signal-to-noise ratios with normalized pixel values as output data. We demonstrate the validity of our proposed image preprocessing approach by comparing the results of two well-accepted and widely-used image analysis pipelines.

Presenter: Zheng Li, University of Michigan
Title: Multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies
Abstract: Spatial transcriptomics have enabled gene expression profiling on tissues with spatial localization information, characterizing the transcriptomic landscape of many tissues. Here, we present a statistical method, BASS, for effective spatial transcriptomic analysis that examines the hierarchical organization of tissues at two distinct scales. Specifically, at the single-cell scale, our method performs cell type clustering and clusters cells into cell types. At the tissue regional scale, our method segments the tissue section into distinct spatial domains in a de novo fashion. Importantly, our method performs both analyses in a coherent fashion through a Bayesian hierarchical modeling framework, allowing for seamless integration of gene expression information with spatial information to improve the analyses at both scales. Moreover, our method allows for integrative analysis of spatial transcriptomic data measured on multiple tissue sections in the same anatomic region, allowing us to borrow critical biological information across tissue sections to further enhance analytic performance.

Presenter: Lukas Weber, Johns Hopkins Bloomberg School of Public Health
Title: nnSVG: scalable identification of spatially variable genes using nearest-neighbor Gaussian processes
Abstract: Spatially-resolved transcriptomics enables the measurement of transcriptome-wide gene expression at spatial resolution, thus providing comprehensive information on the spatial organization of cell populations and spatial expression patterns of genes within complex tissues. Feature selection to identify biologically informative genes represents a key analysis step during data-driven analysis workflows applied to these datasets. Here, we present nnSVG, a new, scalable approach to identify spatially variable genes in spatially-resolved transcriptomics data, based on nearest-neighbor Gaussian processes. Our method provides high sensitivity by fitting gene-specific estimates of length scale parameters within the models, can identify genes that vary in expression continuously across the entire tissue or within a priori defined spatial domains, and scales linearly with the number of spatial locations. We demonstrate that our method outperforms existing methods using experimental data from several technological platforms and simulations. An accessible software implementation is available within the R/Bioconductor framework at https://bioconductor.org/packages/nnSVG.

Recent Developments on Brain Imaging Analysis

Organizer: Lin Zhang, University of Minnesota
Chair: Eric Rawls, University of Minnesota

Presenter: Kristin Linn, University of Pennsylvania
Title: Algorithmic Fairness of Models for Predicting Alzheimer's Disease Progression
Abstract: Alzheimer's disease (AD) disproportionately affects marginalized older adults. Machine learning (ML) techniques have the potential to improve early detection of AD. However, ML models may suffer from biases and

perpetuate existing disparities. In this work, we audited the fairness of three ML models for predicting progression from normal cognition to mild cognitive impairment (MCI) and from MCI to AD. We assessed three common fairness metrics—equal opportunity, equalized odds, and demographic parity—measured across subgroups defined by gender, ethnicity, and race. Although the three models demonstrated high accuracy in aggregate, all three models failed to satisfy fairness metrics for subgroups defined by ethnicity and race. The models generally satisfied metrics of fairness for gender. I will discuss potential implications of our findings and place them in context with recently published literature on algorithmic fairness in imaging applications.

Presenter: Zhaoxia Yu, University of California Irvine
Title: Novel Spatial Statistical Methods for Spatial Transcriptomics
Abstract: With recent advances in technology, scientists have made remarkable progress in gene expression analysis, transitioning from tools that provided average gene expression in tissues to devices that measure the expressions of all genes in a single cell. Spatial transcriptomics methods have pushed gene sequencing even further by not only recording gene expression profiles at subcellular resolution but also precisely locating each cell in a tissue. This type of data is critical in understanding human diseases, as many diseases start from individual cells and spread spatially. Beyond disease, spatial transcriptomics data can also lead to the creation of unbiased atlases, such as the brain atlas, which maps and comprehends the spatial organization and distribution of cells in the cortex.However, the high-resolution spatial transcriptomics data generated by these new technologies presents substantial analytical and computational challenges, requiring the development of suitable statistical methods to process the raw data and extract the biological signal while effectively utilizing the high-resolution spatial information. In this talk, I will present a multivariate factor models with factors that vary spatially and by groups, such as developmental stages, aging, experimental conditions, or disease status, while flexibly modeling the simultaneous dependence in space and by group and accommodating the large dimensionality of the data.

Presenter: Veera Baladandayuthapani, University of Michigan
Title: Tumor Radiogenomics in Gliomas with Bayesian Layered Variable Selection
Abstract: We propose a statistical framework to integrate radiological magnetic resonance imaging (MRI) and genomic data to identify the underlying radiogenomic associations in lower grade gliomas (LGG). We devise a novel imaging phenotype by dividing the tumor region into concentric spherical layers that mimics the tumor evolution process. MRI data within each layer is represented by voxel--intensity-based probability density functions which capture the complete information about tumor heterogeneity. Under a Riemannian-geometric framework these densities are mapped to a vector of principal component scores which act as imaging phenotypes. Subsequently, we build Bayesian variable selection models for each layer with the imaging phenotypes as the response and the genomic markers as predictors. Our novel hierarchical prior formulation incorporates the interior-to-exterior structure of the layers, and the correlation between the genomic markers. We employ a computationally-efficient Expectation--Maximization-based strategy for estimation. Simulation studies demonstrate the superior performance of our approach compared to other approaches. With a focus on the cancer driver genes in LGG, we discuss some biologically relevant findings. Genes implicated with survival and oncogenesis are identified as being associated with the spherical layers, which could potentially serve as early-stage diagnostic markers for disease monitoring, prior to routine invasive approaches.

Presenter: Ruben Sanchez-Romero, Rutgers University
Title: Causally-informed activity flow generative models predict empirical task-evoked activations from brain network interactions
Abstract: Brain activity flow models (Cole et al., 2016) estimate the movement of task-evoked activity over brain connections to help explain the generation of task-related functionality. In particular, for a target brain region we model the generation of task-related activations as a function of brain connections and activations from its relevant source regions. To increase the interpretability and intervention potential of these models, we are interested in minimizing the risk of false positives when estimating brain connections via functional/effective connectivity (FC). We first evaluate the impact of FC estimation accuracy on activity flow model predictions by testing a continuum of FC methods ordered according to the amount of statistical conditional independence information and increasing use of statistical causal principles. We test correlation, multiple regression, combinedFC (Sanchez-Romero et al., 2021) and the PC causal discovery algorithm (Spirtes et al., 2000). We first compare these methods in simulations and empirical fMRI data. Next, to illustrate how activity flow models can be used to study specific cognitive functionality we apply an activity flow model to a dorsolateral prefrontal cortex region, demonstrating distributed network mechanisms in

default mode and frontoparietal areas contributing to its selective activation during an n-back working memory task. Together, these results show that FC methods that leverage statistical conditional independence information and causal principles, can be used to construct directed activity flow models with increasing interpretability and intervention potential, thus enabling us to better identify the network mechanisms underlying cognitive computations in the human brain.

Bayesian Methods

Organizer: Mark Fiecas, University of Minnesota
Chair: Caitlin Ward, University of Minnesota

Presenter: Dustin Pluta, Rice University
Title: Improved statistical power of trial-level event-related potentials with Bayesian random-shift Gaussian processes
Abstract: Studies of cognitive processes via electroencephalogram recordings often analyze group-level event-related potentials (ERPs) averaged over multiple subjects and trials. This averaging procedure can obscure scientifically relevant variability across subjects and trials, but has been necessary due to the difficulties posed by inference of trial-level ERPs. We present the Bayesian Random Phase-Amplitude Gaussian Process (RPAGP) model for inference of trial-level amplitude, latency, and ERP waveforms. We apply RPAGP to data from a study of ERP responses to emotionally arousing images to analyze differences in trial-level characteristics across experimental conditions. The model estimates of trial-specific signals are shown to greatly improve statistical power in detecting significant group differences compared to existing methods. Further potential extensions and applications of the RPAGP framework for improving existing neuroimaging analysis pipelines will also be discussed.

Presenter: Rajarshi Guhaniyogi, Texas A&M
Title: Bayesian Multi-Modal Data Integration
Abstract: This talk focuses on a multi-modal imaging data application where structural/anatomical information from grey matter (GM) and brain connectivity information in the form of a brain connectome network from functional magnetic resonance imaging (fMRI) are available for a number of subjects with different degrees of a neurodegenerative disorder (ND). The clinical/scientific goal in this study becomes the identification of brain regions of interest significantly related to the speech rate measure to gain insight into an ND pathway. Viewing the brain connectome network and GM images as objects, we develop a flexible joint object response regression framework of network and GM images on the ND measure. A novel joint prior formulation is proposed on network and structural image coefficients in order to exploit network information of the brain connectome, while leveraging the topological linkages among connectome network and anatomical information from GM to draw inference on brain regions significantly related to the ND. The principled Bayesian framework allows precise characterization of the uncertainty in ascertaining a region being actively related to ND. We will discuss strategies to draw scalable Bayesian inference in these models.

Presenter: Michele Guindani, University of California, Los Angeles
Title: A Bayesian Time-Varying Psychophysiological Interaction (PPI) Model
Abstract: Psychophysiological interaction (PPI) models have been largely employed to study task-modulated seed-based brain connectivity in fMRI studies. However, popular implementations of the PPI framework assume that the partial correlation between the seed region and the ROIs is static in the absence of a stimulus, whereas current developments in neuroimaging suggest that functional connectivity is by nature dynamic. In this talk, I present a Bayesian modeling framework that extends the generalized PPI model and estimates task-modulated time-varying background functional connectivity from an fMRI experiment. In order to model the dynamics of the background regression coefficients, the framework employs a time-varying scale-mixture shrinkage prior that enforces sparsity of the non-zero coefficients. The approach can be parallelized to identify functional connectivity patterns for varying choices of the seed region. Then, the significant partial correlations across runs are selected by using a non-marginal decision-theory-based multicomparison framework, which also leads to reduced spurious non-zero PPI effects. The performance of the model is illustrated in a simulation analysis and in an application to data from a serial reaction time experiment.

Presenter: John Kornak, University of California, San Francisco
Title: Bayesian modeling for longitudinal trajectories of dementia brain changes
Abstract: The 2010 hypothetical ‘Jack’ model attempts to describe the timeline with which different biomarkers change in AD and has sparked much discussion and subsequent research. Understanding this temporal ordering in AD and other forms of dementia has major implications for prediction and clinical trial design. I will present some Bayesian modeling work that aims to

estimate the temporal path of brain biomarker changes (imaging and otherwise) in frontotemporal dementia and determine potential differences in trajectories across genetic subtypes. This is collaborative work with UCSF Memory and Aging Center along with the Berry Consultants group.

Collaborative Case Study: Statistical Methods and Findings from Large Consortia Studies

Organizer: Simon Vandekar, VUMC
Chair: Jun Young Park, University of Toronto

Presenter: Aaron Alexander-Bloch, Children's Hospital of Pennsylvania
Title: Towards lifespan “brain charts” of normal brain growth, aging, and neuropsychiatric disease: recent developments, outstanding hurdles, and future clinical opportunities
Abstract: From the clinical perspective, the ultimate goal for brain charts is to provide reference norms of phenotypes derived from neuroimaging data, analogous to pediatric growth charts for height and weight. In the case of anatomical MRI, brain charts enable brain morphological phenotypes derived from individual MRI scans, such as cortical thickness or gray matter volume, to be quantitatively benchmarked to population norms. The result is highly interpretable metrics that place an individual on a continuous dimension of brain deviation. The ambitious goal of widely adopted brain charts that add value to research and clinical studies has not yet been achievable because of difficulty in amassing multi-site data, image processing challenges (especially early in life), and accurately modeling normative brain development. Our recent work using generalized additive models for location, shape and scale (GAMLSS) based on LBCC data has the goal to overcome these challenges and provide a roadmap for the utility of brain charts in future neuroimaging studies of the human brain. The resulting “brain charts” identified previously unreported neurodevelopmental milestones and provided a standardized and interpretable measure of deviation from typical development, which revealed new patterns of neuroanatomical differences across neurological and psychiatric disorders. Brain charts for the human lifespan are an essential first step towards robust, standardized quantification of deviation from age-related trends for multiple commonly-used MRI measures, providing an anchorpoint for neuroimaging research and quantitative standards for clinical studies.

Presenter: Russell T. Shinohara, University of Pennsylvania
Title: Statistical harmonization methods for the analysis of multi-center consortia
Abstract: Magnetic resonance imaging (MRI) studies often involve large multi-center designs. To address this, an increasingly commonly used approach is ComBat, which was first proposed in the genomics literature. Our group has found these methods to be helpful but sometimes insufficient in complex study designs, especially when joint or predictive modeling is employed. To address this, we have developed a broader framework for multi-scanner harmonization that allows for flexible, non-linear, and multivariate modeling in a variety of imaging science settings.

Presenter: Kaidi Kang, Vanderbilt University
Title: Study Features Impacting Replicability of Brain-wide Association Studies
Abstract: Brain-wide association studies (BWAS) have recently been criticized for having low replicability. Several recent studies have suggested that large sample sizes are required to improve replicability of BWAS because the effect sizes (ESs) are much smaller than was expected from biased ES estimates in smaller studies. However, design considerations, such as longitudinal studies, can also be used to improve the replicability of BWAS. We analyze and meta-analyze 57 neuroimaging datasets from the Lifespan Brain Chart Consortium (LBCC) to demonstrate the influence of covariate sampling distribution and longitudinal/cross-sectional study designs on ESs. The meta-analysis results demonstrate that ESs depend on study design and study population features. ESs are smaller in studies with lower covariate sampling variance, and that ESs for age are dependent on the age of the population being studied, indicating time periods when there is increased sensitivity to detect age-related differences. Longitudinal studies have systematically larger ESs. Since this may compromise the comparability of ESs between cross-sectional and longitudinal studies in meta-analyses, we propose an estimate of the cross-sectional (CS) RESI for longitudinal datasets that improves the comparability. Critically, this also allows us to quantify the benefit of a longitudinal study design using a single longitudinal dataset. Our work demonstrates that large sample sizes may not be the only solution to enhancing the replicability of BWAS. The ESs in BWAS also depend on other study features such as study design and covariate sampling distributions and careful study design can be used to improve ESs.

Modern Statistical Methodology on Spatial, Neuroimaging, and Shape Data Analysis

Organizer: Shan Yu, University of Virginia
Chair: Rajarshi Guhaniyogi, Texas A&M

Presenter: Caitlin Ward, University of Minnesota
Title: Capturing Spatio-temporal Behavioral Change in Bayesian Epidemic Models
Abstract: For many infectious disease outbreaks, the at-risk population changes their behavior in response to outbreak severity, changing the transmission dynamics to change in real-time. In addition, both transmission and the response may vary across regions in a population. This talk will present a spatially stratified behavioral change epidemic model formulation where spatio-temporal transmission is captured by varying levels of “alarm” across the population. Region-specific alarm is specified as functions of recently observed incidence and/or mortality, and the incorporation of spatial heterogeneity in the model allows us to measure the relative importance of local vs. global epidemic trajectory in informing alarm. The model is set in a data-augmented Bayesian framework as epidemic data are often only partially observed and we can use prior information to help with parameter identifiability. The benefit and utility of the proposed approach is illustrated through an application to county-level COVID-19 data from the United States.

Presenter: Yaotian Wang, University of Pittsburgh
Title: High-Dimensional Directed Network Analysis of Human Brains
Abstract: The human brain is a high-dimensional directed network system consisting of many regions as network nodes that exert influence on each other. The directed influence from one region to another is called directed connectivity and corresponds to one directed edge in the directed brain network. To understand how brain regions interact with each other and form different brain network patterns when performing different functions, we develop statistical modeling approaches to reveal high-dimensional directed brain networks using brain data. In this talk, I will present two models. (1) The first model is for studying normal and abnormal directed brain networks of patients with epilepsy using their intracranial electroencephalography (EEG) data. Epilepsy is a directed network disorder, as epileptic activity spreads from a seizure onset zone (SOZ) to many other regions after seizure onset. Intracranial EEG data are multivariate time series recordings of many brain regions. With our proposed model, we revealed the evolution of brain networks from normal to abnormal states and uncovered unique directed connectivity properties of the SOZ during seizure development. (2) The second model characterizes whole-brain directed networks of the population of healthy subjects based on functional magnetic resonance imaging (fMRI) data. We also propose a computationally efficient algorithm to address the challenge of analyzing thousands of subjects’ fMRI data. Using our new model and algorithm, we analyzed the resting-state fMRI data of around one thousand subjects from the Human Connectome Project (HCP). We revealed both population-mean and subject-specific whole-brain directed networks.

Presenter: Yuexuan Wu, University of Washington
Title: Longitudinal elastic shape analysis of brain subcortical surfaces
Abstract: Over the past 30 years, magnetic resonance imaging (MRI) has become a ubiquitous tool for accurately visualizing the development of the brain’s subcortical structures. However, the quantification of complex subcortical structures is still in its infancy due to challenges in shape extraction, representation, and modeling. We develop a simple and efficient framework of longitudinal elastic shape analysis (LESA) for subcortical structure surfaces. Integrating ideas from elastic shape analysis of static surfaces and statistical modeling of sparse longitudinal data, LESA provides a set of tools for systematically quantifying changes in longitudinal surface shapes from raw structural MRI data. LESA can efficiently represent complex subcortical structures using a small number of basis functions and can accurately predict the spatiotemporal shape changes of the surfaces. Besides, by applying LESA to analyze three longitudinal neuroimaging data sets, we showcase its wide applications in estimating continuous shape trajectories, building life-span growth patterns, and comparing shape differences among different groups.

Presenter: Shan Yu, University of Virginia
Title: Nonparametric Regression for 3D Point Cloud Learning
Abstract: Over the past two decades, we have seen an exponentially increased amount of point clouds collected with irregular shapes in various areas. Motivated by the importance of solid modeling for point clouds, we develop a novel and efficient smoothing tool based on multivariate splines over the tetrahedral partitions to extract the underlying signal and build up a 3D solid model from the point cloud. The proposed smoothing method can denoise or deblur the point cloud effectively and provide a

multi-resolution reconstruction of the actual signal. In addition, it can handle sparse and irregularly distributed point clouds and recover the underlying trajectory. The proposed smoothing and interpolation method also provides a natural way of numerosity data reduction. Furthermore, we establish the theoretical guarantees of the proposed method. Specifically, we derive the convergence rate and asymptotic normality of the proposed estimator and illustrate that the convergence rate achieves the optimal nonparametric convergence rate. Through extensive simulation studies and a real data example, we demonstrate the superiority of the proposed method over traditional smoothing methods in terms of estimation accuracy and efficiency of data reduction.

Recent advancements in statistical methods for brain connectome analysis

Organizer: Zhengwu Zhang, UNC Chapel Hill
Chair: Yi Zhao, Indiana University School of Medicine

Presenter: Ying Guo, Emory University
Title: Statistical learning for multimodal brain connectomes using neuroimaging
Abstract: Recent advancements in multimodal neuroimaging such as functional MRI (fMRI) and diffusion MRI (dMRI) offer unprecedented opportunities to combine the strengths of modalities in investigating brain architecture from both functional and structural perspectives. Many existing studies typically examine these modalities in separate analyses. In recent years, multimodal methods are emerging to facilitate joint analyses. We present statistical methods for exploring multimodal brain connectomes to investigate brain organizations in both white matter structural connection and intrinsic functional connection. The proposed methods aim to address the following key questions in such analysis: how to reliably infer underlying structural and functional connectome states from noisy imaging-based connectivity measurements from dMRI and fMRI, what is the relationship between structural and functional connectomes and how the interplay between them depends on individuals' demographic and clinical groups, how brain structural and functional connectomes change with neurodevelopment, aging or brain diseases. The methodological motivation of the models and their estimation methods will be discussed. We will present new insights derived from the proposed methods into the neurodevelopmental of brain structural and functional connectomes during adolescence.

Presenter: Benjamin Risk, Emory University
Title: Counteracting selection bias in functional connectivity studies of autism
Abstract: In resting-state functional magnetic resonance imaging studies, it is common for more than 50% of data to be removed due to participant motion. Motion tends to be higher in children with developmental disorders, such as autism spectrum disorder. This creates the potential for selection bias. In this study, we modify doubly robust estimators of the average treatment effect to address this problem. We propose a permutation test of the difference between two groups for improved type one error rates in finite samples. We compare functional connectivity in autistic children to children without autism.

Presenter: Yize Zhao, Yale University
Title: Bayesian pathway analysis over brain network mediators for survival data
Abstract: Technological advancements in noninvasive imaging facilitate the construction of whole brain inter-connected networks, known as brain connectivity. Existing approaches to analyze brain connectivity frequently disaggregate the entire network into a vector of unique edges or summary measures, leading to a substantial loss of information. Motivated by the need to explore the effect mechanism among genetic exposure, brain connectivity and time to disease onset, we propose an integrative Bayesian framework to model the effect pathway between each of these components while quantifying the mediating role of brain networks. To accommodate the biological architectures of brain connectivity constructed along white matter fiber tracts, we develop a structural modeling framework which includes a symmetric matrix-variate accelerated failure time model and a symmetric matrix response regression to characterize the effect paths. We further impose within-graph sparsity and between-graph shrinkage to identify informative network configurations and eliminate the interference of noisy components. Extensive simulations confirm the superiority of our method compared with existing alternatives. By applying the proposed method to the landmark Alzheimer's Disease Neuroimaging Initiative study, we obtain neurobiologically plausible insights that may inform future intervention strategies.

Presenter: Arkaprava Roy, University of Florida
Title: Nonparametric Group Variable Selection with Multivariate Response for Connectome-Based Modeling of Cognitive Scores
Abstract: In this work, the aim is to identify the brain regions having a significant effect on cognitive functioning. The cognitive profiles are measured in terms of seven cognitive age-adjusted test scores from the NIH

toolbox of cognitive battery. The structural connectomes are represented by adjacency matrices. Most existing works consider the upper or lower triangular section of these adjacency matrices as predictors. An alternative characterization of the connectivity properties is available in terms of the nodal attributes. Although any single nodal attribute may not be adequate to represent a complex brain network, a collection of nodal centralities together can encode different patterns of connections in the brain network. In this article, we consider nine different attributes for each brain region as our predictors. We propose Gaussian RBF-nets with a novel group sparsity inducing prior to modeling the unknown mean functions. We show that the proposed method performs overwhelmingly better than all its competitors. Applying our proposed method to a Human Connectome Project (HCP) dataset, we identify the important brain regions and nodal attributes for cognitive functioning, as well as identify interesting low-dimensional dependency structures among the cognition related test scores.

Efficient Modeling of Multi-Region High-Dimensional Molecular Data

Organizer: Saonli Basu, University of Minnesota
Chair: Zhaoxia Yu, University of California Irvine

Presenter: Eric Lock, University of Minnesota
Title: Classification from multi-region molecular data
Abstract: To capture multiple facets of a complex biological system, molecular "omics" data are increasingly measured across multiple regions or tissue types on the same individuals. For example, magnetic resonance spectroscopy (MRS) data measures the abundance of several metabolites across multiple neurological regions. Such data can be represented as a multi-way array (i.e., as tensor) with dimensions corresponding to the individuals, molecular features, and regions. We describe methods to classify individuals from such multi-way data, using low-rank structure and sparsity to achieve efficient and interpretable results. These methods extend popular single-way classification methods for high-dimensional data, such as support vector machines (SVM), distance weighted discrimination (DWD), and Bayesian regularized regression for a binary outcome. The methods are successfully applied to the classification of neurodegenerative disorders and iron deficiency in early-life, providing new scientific insights.

Presenter: Xi Jiang, Southern Methodist University
Title: iIMPACT: Integrating Image and Molecular Profiles for Spatial Transcriptomics Analysis
Abstract: The breakthrough in spatially resolved transcriptomics (SRT) has enabled comprehensive molecular characterization while preserving spatial information. Meanwhile, histology image analysis powered by artificial intelligence enables the histological characterization of single cells. However, existing state-of-art clustering methods for spatial transcriptomics data analysis mainly rely on molecular information, without fully leveraging the biological features from histology image, which causes compromised algorithm accuracy. To address this problem, we developed iIMPACT, a multi-stage statistical method integrating the image and molecular profiles to analyze SRT data. It utilizes an interpretable Bayesian finite mixture model for analyzing the cellular spatial organization and a regression model for domain-specific spatially variable gene analysis. Applying our method to publicly available SRT datasets, we found that iIMPACT outperforms state-of-art clustering methods in correctly determining spatial domains when comparing with ground truth biological knowledge. The subsequent domain-specific spatially variable gene analysis accurately identified known functional genes failed to be detected by published state-of-art methods. IIMPACT represents a highly accurate and interpretable clustering approach to reveal cellular spatial organization and functional gene landscape from spatial transcriptomics data.

Presenter: Souvik Seal, Medical University of South Carolina Hollings Cancer Center
Title: SMASH: Scalable Method for Analyzing Spatial Heterogeneity of genes in spatial transcriptomics data
Abstract: In high-throughput spatial transcriptomics (ST) studies, it is of great interest to identify the genes whose level of expression in a tissue covaries with the spatial location of cells/spots. Such genes, also known as spatially variable genes (SVGs), can be crucial to the biological understanding of both structural and functional characteristics of complex tissues. Existing methods for detecting SVGs either suffer from huge computational demand or significantly lack statistical power. We propose a non-parametric method termed SMASH that achieves a balance between the above two problems. We compare SMASH with other existing methods in varying simulation scenarios demonstrating its superior statistical power and robustness. We apply the method to four ST datasets from different platforms revealing interesting biological insights.

Presenter: Weihua Guan, University of Minnesota
Title: Digital Spatial Profiling of Kidney Allograft Biopsies with Chronic Allograft Dysfunction: Analysis and Challenges

Abstract: Digital Spatial Profiling (DSP) is a state-of-the-art method for multiplexed spatial profiling of formalin-fixed, paraffin-embedded (FFPE) tissue samples. DSP combines compartmentalized tissue imaging, with photo-cleavable antibodies, and RNA probes for whole transcriptome analysis. Through DSP imaging, regions of interest (ROIs) can be selected for different compartments of kidney biopsies such as glomeruli or tubules, for RNA sequencing. In this preliminary study, we analyzed three kidney transplant biopsy samples, 32 ROIs total, measured with DSP and RNA sequencing by Nanostring GeoMX platform. Using traditional cell type convolution approaches, we found different abundance for mast cell resting, macrophages M1, and regulatory T cells between the graft loss status (FDR < 0.10). Principal component analysis was carried out for ROI clustering. Differential expression (DE) analysis was used to identify individual transcripts associated with graft loss, stratified by compartment types. These preliminary results suggest that DSP of FFPE tissue can be useful for the identification and quantification of specific cell types and molecular transcripts of interest. At the end of the talk, analytical challenges will be discussed on development of statistical methods for DSP data analysis.

Recent Advances in Neuroimaging Statistics for Investigating Human Brain Function

Organizer: Joshua Lukemire, Emory University
Chair: Aidan Neher, University of Minnesota

Presenter: Panpan Zhang, Vanderbilt University
Title: A Bayesian Model for Link Prediction in Functional Brain Networks
Abstract: Link prediction is an important research theme in network analysis. Missing links are common in complex networks due to various reasons. Inference based on incomplete networks may lead to statistical bias. In this paper, we propose a Bayesian model for recovering missing links by simultaneously accounting for observed network structure as well as auxiliary nodal information. The proposed model is compared with competing methods through extensive simulations and applied to neuroimaging data in Alzheimer's disease.

Presenter: Dayu Sun, Emory University
Title: Sparse Partial Logistic Tensor Regression with Application to Neuroimaging Data
Abstract: Tensor data, often characterized as multi-dimensional arrays, have been increasingly prevalent in biomedical studies, for example, in neuroimaging applications. Analyses of tensor data are subject to complications including the high-dimensionality and intrinsic structures within tensor data. The regression of a continuous response on tensor predictors has been well investigated in recent years, but there is relatively limited literature on the regression of a binary outcome on tensor predictors. The existing decomposition-based methods may suffer from a heavy computational burden and the lack of uniqueness for tensor decompositions. In this work, we propose a Sparse Partial Logistic Tensor Regression method for modeling binary outcomes in terms of both tensor and vector/scalar predictors. We utilize novel mode-wise penalized manifold optimization techniques to achieve dimension reduction and sparsity in tensor coefficient estimation that may improve the prediction performance. Extensive simulation studies illustrate the proposed method demonstrates satisfactory performance under various scenarios and outperforms existing methods. We apply our proposed method to investigate the association between the diagnosis of posttraumatic stress disorder (PTSD) and brain connectivity matrices derived from functional magnetic resonance imaging (fMRI) data from a mental health study.

Presenter: Joshua Lukemire, Emory University
Title: A General Framework for Repeated Measures Sparse Bayesian Independent Component Analysis with Applications to Multi-center and Longitudinal Imaging Studies
Abstract: We introduce a general framework of repeated measures Sparse Bayesian independent component analysis (RM-SparseBayes ICA). This general method provides a rigorous and much needed tool for investigating brain networks and their differences in imaging studies with complex study designs including longitudinal and/or multi-center studies. Our approach incorporates effects corresponding to data collection sites, allowing us to include information such as scanner type and field strength. Additionally our approach uses subject-specific effects to accommodate within-subject repeated measures such as those from longitudinal studies. Through simulations, we show that the proposed method has considerably improved performance as compared to other potential approaches. We provide applications to a resting-state fMRI study.

Presenter: Hyunnam Ryu, Columbia University
Title: Persistent Homology-based Functional Connectivity and its Association with Cognitive Ability
Abstract: Brain-segregation attributes in resting-state functional networks have been widely investigated to understand cognition and cognitive aging using various

approaches (e.g., average connectivity within/between networks and brain system segregation). While these approaches have assumed that resting-state functional networks operate in a modular structure, a complementary perspective assumes that a core-periphery or rich club structure accounts for brain functions where the hubs are tightly interconnected to each other to allow for integrated processing. In this paper, we apply a novel method, persistent homology (PH), to develop an alternative to standard functional connectivity by quantifying the pattern of information during the integrated processing. We also investigate whether PH-based functional connectivity explains cognitive performance and compare the amount of variability in explaining cognitive performance for three sets of independent variables: (1) PH-based functional connectivity, (2) graph theory-based measures, and (3) brain system segregation. Resting-state functional connectivity data were extracted from 279 healthy participants, and cognitive ability scores were generated in four domains (fluid reasoning, episodic memory, vocabulary, and processing speed). The results first highlight the pattern of brain-information flow over whole brain regions (i.e., integrated processing) accounts for more variance of cognitive abilities than other methods. The results also show that fluid reasoning and vocabulary performance significantly decrease as the strength of the additional information flow on functional connectivity with the shortest path increases. While PH has been applied to functional connectivity analysis in recent studies, our results demonstrate potential utility of PH-based functional connectivity in understanding cognitive function.

Recent Advances in Spatial Analysis of Single-cell Imaging

Organizer: Jiangmei Xiong, Vanderbilt University
Chair: Christian Coffman, University of Minnesota

Presenter: Brooke L. Fridley, Moffitt Cancer Center
Title: Spatial Segregation of Immune Cells in the Tumor Immune Microenvironment
Abstract: Spatial segregation is a concept that has been used in many studies of social inequalities and ecology. Spatial segregation is defined as the degree of spatial separation of two or more populations in a region of interest. In the context of single-cell spatial protein data, spatial segregation can be conducted from the viewpoint of immune cell co-localization or segregation in the tumor immune microenvironment (TIME). One measure of spatial segregation and co-localization was developed by Dixon based on a contingency table approach, whereby spatial segregation of two populations is determined by looking at the number of times that a cell and its neighbor are from the same population is different than expected by chance. In this presentation, we will describe a modification of this statistic for the setting in which there is a low immune cell abundance in the TIME. We will also illustrate the use of this statistic to assess the spatial segregation of T cell subsets in tumors from a large single-cell protein imaging study of Black women with high-grade serous ovarian cancer (N =93 subjects with 260 intra-tumoral regions of interest (ROIs)). The goal of the analysis is to determine segregation or co-localization of these immune cells in the TIME and how this level of segregation is related to overall survival. Additionally, we will compare the results from the modified Dixon's statistics to another commonly used statistic for co-clustering, bivariate Ripley's K.

Presenter: Siyuan Ma, Vanderbilt University Medical Center
Title: A Flexible Generalized Linear Mixed Effects Model for Testing Cell-Cell Colocalization in Spatial Immunofluorescent Data
Abstract: Existing methods often rely on permutation-based statistics to test for enrichment of cell-cell colocalization events within spatial immunofluorescent images. This type of approach is time-consuming, and importantly, does not generalize well for comparisons between images/conditions. We show that by making essentially the same set of assumptions, cell-cell interaction events can be modeled with a spatial regression generalized linear mixed effect model, thus allowing flexible inclusion and testing of image/condition effects. We exemplify the utility of such a model in with an application in protein immunofluorescent imaging of inflammatory bowel disease tissues.

Presenter: Julia Wrobel, Colorado School of Public Health
Title: Analysis of Immune Cell Spatial Clustering using Functional Data Models
Abstract: The tumor microenvironment (TME), which characterizes the tumor and its surroundings, plays a critical role in understanding cancer development and progression. Recent advances in imaging techniques enable researchers to study spatial structure of the TME at a single-cell level. Investigating spatial patterns and interactions of cell subtypes within the TME provides useful insights into how cells with different biological purposes behave, which may consequentially impact a subject's clinical outcomes. We utilize a class of well-known spatial summary statistics, the K-function and its variants, to explore inter-cell dependence as a function

of distances between cells. Using techniques from functional data analysis, we introduce an approach to model the association between these summary spatial functions and subject-level outcomes, while controlling for other clinical scalar predictors such as age and disease stage. In particular, we leverage the additive functional Cox regression model (AFCM) to study the nonlinear impact of spatial interaction between tumor and stromal cells on overall survival in patients with non-small cell lung cancer, using multiplex immunohistochemistry (mIHC) data. The applicability of our approach is further validated using a publicly available multiplexed ion beam imaging (MIBI) triple-negative breast cancer dataset.

Presenter: Misung Yi, Thomas Jefferson University
Title: Quantification of Spatial Interaction between Cancer and Immune Cells in Tumor Microenvironment
Abstract: Advanced analysis systems for pathology allow capturing spatial coordinates of all cells in immunohistochemistry images of tumor microenvironment (TIME), but there are no established methods for objective quantification of spatial interaction between cancer and immune cells. We consider novel and previously proposed metrics of immune cells density within the tumor stroma and metrics of spatial interactions between cancer and immune cells based on distributions of the nearest neighbor distances. The spatial localization of CD8+ T cells, CD163+ tumor associated macrophages (TAMs) and cancer cells were used to generate distributions of the nearest neighbor distances and other metrics of interaction between cancer and CD8+ cells or cancer and CD163+ cells in breast cancer tissue. The metrics of spatial interaction between cancer and immune cells were considered as predictors of progression-free survival (PFS) of breast cancer patients. Shorter progression-free survival was associated with high density of CD163+ TAMs, shorter median cancer-to-CD163+ nearest neighbor distance, high number of directly adjacent CD163+ TAMs (within juxtacrine proximity <12 micrometers to cancer cells), high number of communicating CD163+ TAMs (within paracrine communication distance <250 micrometers to cancer cells), and low number of communicating CD8+ T cells to cancer cells after multivariable adjustment for clinical and pathological risk factors and correction for optimistic bias due to dichotomization. Part of the results to be presented are published in Maisel, Brenton A., et al. "Spatial metrics of interaction between CD163-positive macrophages and cancer cells and progression-free survival in chemo-treated breast cancer." Cancers 14.2 (2022): 308.

Statistical Methods for Brain Connectomes

Organizer: Selena Wang, Yale University
Chair: Yize Zhao, Yale University

Presenter: Selena Wang, Yale University
Title: Establishing group-level brain structural connectivity incorporating anatomical knowledge under latent space modeling
Abstract: Brain structural connectivity, capturing the white matter fiber tracts among brain regions inferred by diffusion MRI (dMRI), provides a unique characterization of brain anatomical organization. One fundamental question to address with structural connectivity is how to properly summarize and perform statistical inference for a group-level connectivity architecture, for instance, under different sex groups, or disease cohorts. Existing analyses commonly summarized group-level brain connectivity by a simple entry-wise sample mean or median across individual brain connectivity matrices. However, such a heuristic approach fully ignores the associations among structural connections and the topological properties of brain networks. In this project, we propose a latent space-based generative network model to estimate group-level brain connectivity. Within our modeling framework, we simultaneously incorporate the anatomical information of brain regions as the attributes of nodes to enhance the plausibility of our estimation and improve biological interpretation. We name our method the attributes-informed brain connectivity (ABC) model, which compared with existing group-level connectivity estimations, (1) offers an interpretable latent space representation of the group-level connectivity, (2) incorporates the anatomical knowledge of nodes and tests its co-varying relationship with connectivity and (3) quantifies the uncertainty and evaluates the likelihood of the estimated group-level effects against chance. We evaluate the performance of our model through extensive simulations. By applying the ABC model to study brain structural connectivity stratified by sex among Alzheimer's Disease (AD) subjects and healthy controls incorporating the volume, thickness and area attributes of brain regions, our method shows superior predictive power on out-of-sample structural connectivity and identifies meaningful sex-specific network neuromarkers for AD.

Presenter: Yi Zhao, Indiana University School of Medicine
Title: Covariance-on-Covariance Regression

Abstract: A Covariance-on-Covariance regression model is introduced in this manuscript. It is assumed that there exists (at least) a pair of linear projections on outcome covariance matrices and predictor covariance matrices such that a log-linear model links the variances in the projection spaces, as well as additional covariates of interest. An ordinary least square type of estimator is proposed to simultaneously identify the projections and estimate model coefficients. Under regularity conditions, the proposed estimator is asymptotically consistent. The superior performance of the proposed approach over existing methods are demonstrated via simulation studies. Applying to data collected in the Human Connectome Project Aging study, the proposed approach identifies three pairs of brain networks, where functional connectivity within the resting-state network predicts functional connectivity within the corresponding task-state network. The three networks correspond to a global signal network, a task-related network, and a task-unrelated network. The findings are consistent with existing knowledge about brain function.

Presenter: Rongjie Liu, Florida State University
Title: Learning based Principal Parcellation Analysis for Brain Connectomes
Abstract: Our understanding of the brain's structure and its relation to human traits is influenced by how we represent the brain's connections. The standard approach is to divide the brain into regions of interest and use an adjacency matrix to measure connectivity between pairs of these regions. However, this approach is limited by the arbitrary choice of regions. To overcome this limitation, we propose a new method that uses tractography and autoencoder to cluster fibers and create a data-adaptive parcellation. This parcellation is designed to better explain variations among individuals and predict human traits. The method is called learning-based principal parcellation analysis and it represents individual brain connections as compositional vectors based on a system of fiber bundles that captures population-level connectivity.

Presenter: Xiao Xu, Indiana University
Title: Novel Penalized Regression Method to Study Alcohol Drinking and Brain Functional Connectivity Link
Abstract: The intricate associations between brain functional connectivity (FC) and clinical outcomes are not easily estimable. Traditional statistical and machine learning approaches do not satisfactorily address the dependence of the outcomes on the intercorrelated connectivity patterns. Specifically, entries in the FC matrix are interrelated and can jointly and/or synergistically affect the outcomes. Here, we utilize a novel penalized regression approach named SpINNEr (Sparsity Inducing Nuclear Norm Estimator) to facilitate identification of brain FC patterns predicting outcomes in a study of alcohol use disorder (AUD) risk. This is accomplished by imposing a nuclear norm penalty to ensure a low-rank structure while an L1-norm encourages entry-wise sparsity to suppress small regression coefficients. The stability of the associations between the FC matrices and empirically derived drinking composite is ensured with bootstrap replications with a threshold of 70% repeatability. We compare our method with the conventional connectome-based predictive modeling (CPM). Unlike the SpINNEr method, the conventional connectome-based predictive modeling (CPM) approach generated many findings that were not replicated in bootstrap repetitions. Thus, SpINNEr is a plausible alternative to CPM that is generalizable to a wide range of clinical applications that assess relationships between high-dimensional predictors and scalar outcomes.

Statistical Methods for Analyzing Multiview and Multi-session Imaging Data

Organizer: Suprateek Kundu, UT MD Anderson Cancer Center
Chair: Thierry Chekouo, University of Minnesota

Presenter: Suprateek Kundu, UT MD Anderson Cancer Center
Title: Bayesian Tensor Approaches for Integrative Imaging Analysis
Abstract: We propose two novel Bayesian tensor approaches for integrative imaging analysis. The first approach develops a spatially aware tensor-based harmonization approach for T1w-MRI data called Tensor-ComBat (TC) that results much higher preservation of biological signals by accounting for the spatial configurations of voxels/ROIs instead of treating them as independent as in state of the art ComBat methods. We develop an efficient Bayesian MCMC implementation of the proposed method. We apply the approach to over 2000 longitudinal ADNI T1w-MRI data spanning 21 sites, and illustrate conclusive advantages in terms of preserving biological signals and removing batch effects over ComBat methods applied to both voxel-level data as well as ROI level features. Our research illustrates that tensor-based ComBat image harmonization can potentially replace the routinely used ComBat data harmonization without significant trade-off in computational costs. The second part of the talk focuses on a tensor-based approach for mapping the genetic

architecture of longitudinal cortical thickness changes in AD patients. We develop a novel approach for joint learning of multiple related tensor on vector regressions for this problem. Using imaging data from multiple visits bolsters the signals of the important genetic signatures and results in greater power for feature selection and improved prediction accuracy at the level of image voxels. An efficient MCMC implementation is proposed and extensive numerical studies presented to showcase the advantages in terms of operating characteristics, and the method is applied to longitudinal ADNI-1 dataset.

Presenter: Vince Calhoun, Georgia State University
Title: Carving Data at its Joint-ness: Multimodal Data-Driven Fusion of NeuroImaging Data
Abstract: Prior studies have shown the advantages of leveraging multimodal data, such as brain structure and brain function, for the purposes of brain biomarker development or prediction. However it can be challenging to combine or fuse high dimensional neuroimaging data. In this talk I will review some of the approaches we have developed to try to benefit from the joint information between modalities, including in cases where the data have mis-match dimensionality, such as with functional MRI which has a time dimension where structural MRI does not. I will provide a few examples of some data fusion approaches we have developed and also review some of the models that we have developed. There is still much work to be done, but it is clear that we can improve our understanding of brain health and disorder by "carving data at its joint-ness" by leveraging the joint information only available by fusing multiple neuroimaging modalities.

Presenter: Xin Ma, Florida State University
Title: Combining Multi-session Imaging Data for Better Prediction and Feature Detection
Abstract: Recent imaging studies offer access to longitudinal imaging data in single modality as well as data from multiple imaging modalities. These multi-session imaging data are usually inter-correlated. Treating them separately could result in loss of power in prediction. Imaging data are also known to contain high-dimensional features which require penalized or constrained treatments in the estimation. In this process, pooling information from all sessions would be vital in detecting important imaging features. In this talk, I will introduce latest methodology developments in analyzing the combined multi-session imaging data. The first project considers imaging data as functional objects from multiple inter-related sources in a regression framework with clinical phenotype outcome. We propose to represent the functional objects with wavelet bases and use a grouped penalty to control the sparsity level across the data sources. We also consider the situation when the imaging data are measured with errors and propose a correction procedure. Compared to penalized methods estimating data sources separately, our proposed method offers superior performance in both prediction and feature selection with noiseless as well as noisy imaging data. The second project investigates combining functional magnetic resonance imaging (fMRI) data from multiple task and rest sessions using the large-scale Adolescent Brain Cognitive Development study with built-in feature selection mechanism in a deep learning framework that is based on state-of-the-at Transformers approach. Our analysis highlights the importance of accounting for the dynamic variability in fMRI time series and discovers shared and differential brain regions in predicting various intelligence metrics. We conclusively illustrate the predictive advantages resulting from the combination of temporally-varying fMRI features across multiple experiments, over prediction based on single modality, as well as prediction using static FC features that do not account for temporal variations over time.

Collaborative Case Study: Event-related potential brain-computer interface data present challenges and opportunities for novel statistical methods

Organizer: Tianwen Ma, Emory University
Chair: Dustin Pluta, Rice University

Presenter: Jane E Huggins, University of Michigan
Title: Event-related potential brain-computer interface design and analysis needs
Abstract: Brain-computer interfaces (BCIs) are intended to provide a method for technology access that does not require movement. Instead, brain activity is directly interpreted to identify the BCI user's desired action. The event-related potential (ERP) BCI design, commonly called a P300 speller, presents an on-screen keyboard of stimuli from which the user can select their desired key. The brain activity after each stimulus is analyzed to identify ERPs related to recognition of a target stimulus. Analysis of the brain activity in response to multiple stimuli enables identification of the user's desired key. Even after decades of BCI studies, BCI performance remains slow and the accuracy of performance varies. Improved methods are needed for the basic task of identifying the target stimulus as quickly as possible. Additional usability challenges include accurately

interpreting data from people with physical impairments and identifying time periods in which the user is not actively trying to use the BCI. Improved methods are also needed to minimize the amount of labeled data needed to calibrate the BCI for a specific user. Creation of robust real-time analysis methods for practical BCI use as a daily communication tool presents many challenges and opportunities for new statistical analyses.

Presenter: Lexin Li, University of California Berkeley
Title: Sequential Top Arm Identification with Application to Brain-Computer Interface
Abstract: A brain-computer interface (BCI) is a system that detects brain activity patterns and operates the external devices to assist people with disabilities in communication. P300 event-related potential (ERP)-based speller is designed to identify a sequence of tokens that disabled people want to type in through EEG stimulus presentation paradigm. However, traditional non-adaptive paradigms treat each token selection independently, leading to a lengthy training process. To improve the sample efficiency, we formulate the problem as solving a sequence of top arm identification tasks in multi-armed bandits. Powered by the recent success of pre-trained large language models (LLMs), the knowledge learned from previous tasks could naturally form a strong informative prior for the next task. We propose a sequential top-2 Thompson sampling algorithm that utilizes those prior information in a coherent way. Under a fixed-confidence setting, we derive the sample complexity bound that indicates the role of prior information. We evaluate our algorithm using both synthetic data and a real P300 ERP-based BCI speller simulator.

Presenter: Tianwen Ma, Emory University
Title: Bayesian Keep-or-Merge Training Framework for Data Integration in ERP-based Brain-Computer Interface
Abstract: An event-related potential (ERP)-based BCI speller helps disabled people with normal communications. Existing methods constructed binary classifiers to detect target ERP responses. Current training strategy uses data from participants themselves only with lengthy training time, causing attention shifts and mental fatigue. To resolve this issue, we propose a Bayesian Keep-or-Match (BKM) method for data integration. BKM specifies the joint distribution of stimulus-specific EEG signals among new and source participants via a Bayesian hierarchical mixture model. We refer to the baseline cluster as the one for the new participant. For inference, we apply a keep-or-merge strategy such that if source and new participants are similar, they share the same set of model parameters, otherwise, they keep their own sets of model parameters. The similarity is determined by a binary selection indicator vector. The parameter set for source participants can be pre-computed to save time. For prediction, we predict on the testing data using the baseline cluster. We demonstrate the advantages of BKM using extensive simulation studies and show the real data analysis from UMDBI Lab.

Penalized Regression and Functional Data Analysis

Organizer: Mark Fiecas, University of Minnesota
Chair: Zhiling Gu, Iowa State University

Presenter: Jeffrey Morris, University of Pennsylvania
Title: Quantile Functional Regression for Distributional regression of biomedical imaging data
Abstract: In many areas of science, technological advances have led to devices that produce an enormous number of measurements per subject, including biomedical imaging data. Frequently, researchers deal with these data by extracting summary statistics from these data (e.g. mean or variance) and then modeling those, but this approach can miss key insights when the summaries do not capture all of the relevant information in the raw data. One of the key challenges in modern statistics is to devise methods that can extract information from these big data while avoiding reductionist assumptions. In this talk, we will discuss methods for modeling the entire distribution of the measurements observed for each subject and relating properties of the distribution to covariates, with possible smooth nonlinear covariate and longitudinally varying affects. We apply this method to two biomedical imaging applications: one computing how the distribution of pixel intensities within a glioblastoma region relate to various biological and clinical factors, and the second using quantitative susceptibility mapping measuring inflammatory processes in brain imaging from multiple sclerosis patients. This general approach has many important applications, including many biomedical imaging applications, as well as wearable device data from accelerometers, blood pressure, and blood sugar monitors, as well as other types of high frequency data streams.

Presenter: Todd Ogden, Columbia University
Title: Nonparametric functional data modeling of pharmacokinetic processes with applications in dynamic PET imaging
Abstract: Modeling a pharmacokinetic process typically involves solving a system of linear differential equations and estimating the parameters upon which the functions depend. In order for this approach to be valid, it is

necessary that a number of fairly strong assumptions hold, assumptions involving various aspects of the kinetic behavior of the substance being studied. In many situations, such models are understood to be simplifications of the "true" kinetic process. While in some circumstances such a simplified model may be a useful (and close) approximation to the truth, in some cases, important aspects of the kinetic behavior cannot be represented. We present a nonparametric approach, based on principles of functional data analysis, to modeling of pharmacokinetic data. We illustrate its use through application to data from a dynamic PET imaging study of the human brain.

Presenter: Gang Chen, NIH
Title: BOLD response is more than just magnitude --improving detection sensitivity through capturing hemodynamic profiles
Abstract: Typical FMRI analyses assume a canonical hemodynamic response function (HRF) with a focus on the overshoot peak height, while other morphological aspects are largely ignored. Thus, in most reported analyses, the overall effect is reduced from a one-dimensional curve to a single scalar. Here, we adopt a data-driven approach to HRF estimation at the whole-brain voxel level, without assuming a profile at the individual level. Then, we estimate the BOLD response in its entirety with a roughness penalty at the population level to improve predictive accuracy, inferential efficiency, and cross-study reproducibility. Through a fast event-related FMRI dataset, we demonstrate the extent of under-fitting and information loss that occurs when adopting the canonical approach. We also address the following questions:
How much does the HRF shape vary across regions, conditions, and clinical groups?
Does an agnostic approach improve sensitivity to detect an effect compared to an assumed HRF?
Can examining HRF shape help validate the presence of an effect complementing statistical evidence?
Could the HRF shape provide evidence for whole-brain BOLD response during a simple task?

Advances in Statistical Methods for Transmission Electron Microscopy

Organizer: David S. Matteson, Cornell University
Chair: David Schneck, Masonic Institute for the Developing Brain

Presenter: David S. Matteson, Cornell University
Title: Exploring Blob Detection to Determine Atomic Column Positions and Intensities in Time-Resolved TEM Images with Ultra-Low Signal-to-Noise
Abstract: Spatially resolved in situ transmission electron microscopy (TEM) is a suitable technique to record information about the atom-scale dynamics with millisecond temporal resolution from materials. However, characterizing dynamics or fluxional behavior requires processing short time exposure images which usually have severely degraded signal-to-noise ratios. The poor signal-to-noise associated with high temporal resolution makes it challenging to determine the position and intensity of atomic columns in materials undergoing structural dynamics. To address this challenge, we propose a noise-robust, processing approach. In particular, a blob detection algorithm has been tailored to deal with noisy TEM image series from nanoparticle systems. In the presence of high noise content, our blob detection approach is demonstrated to outperform the results of other algorithms, enabling the determination of atomic column position and its intensity with a higher degree of precision.

Presenter: Andrew M. Thomas, Cornell University
Title: The DetecTDA Algorithm: Feature Detection and Hypothesis Testing for Extremely Noisy Images
Abstract: In this talk, we propose the detecTDA algorithm for feature detection and hypothesis testing in images with ultra-low signal-to-noise ratio using persistent homology. Our main application is in the identification of atomic columns and other features in transmission electron microscopy (TEM). Cubical persistent homology is used to identify local minima and their size in subregions in the frames of nanoparticle videos, which are hypothesized to correspond to relevant atomic features. We compare the performance of our algorithm to other employed methods for the detection of columns and their intensity. Monte Carlo goodness-of-fit testing using real-valued summaries of persistence diagrams derived from simulated images (generated from pixels residing in the vacuum region of an image) is developed and employed to identify whether the proposed atomic features generated by our algorithm in observed images are distinct from noise. A guarantee on the false discovery rate for multiple Monte Carlo testing of identical hypotheses is also established.

Poster Abstracts

Title: A Two-Stage Approach for Segmenting Spatial Point Patterns Applied to Tumor Immunology
Presenter: Alvin Sheng, North Carolina State University
Co-authors: Brian Reich and Ana-Maria Staicu
Abstract: In tumor immunology, clinical regimes corresponding to different stages of disease or responses to treatment may manifest as different spatial arrangements of tumor and immune cells. Spatial point pattern (SPP) statistics can be used to partition tissue images according to these regimes. To this end, we propose a two-stage approach: first, local intensities and pair correlation functions (PCF) are estimated from the SPP of cells within each image. PCFs are reduced via functional principal components analysis. Second, the estimates are clustered in a Bayesian mixture model with spatially dependent cluster labels. The clusters correspond to regimes of interest that are present across subjects; the cluster labels segment the images according to those regimes. Through Markov Chain Monte Carlo (MCMC) sampling, we jointly estimate and quantify uncertainty in the subregions corresponding to each cluster, the spatial characteristics of each cluster, and the spatial dependence parameter. The number of clusters is found through cross-validation. Simulations demonstrate the performance of the method, and it is applied to a set of multiplex immunofluorescence images of pancreatic tissue.

Title: Quantification of crowd size with mobile phones and social media Big data as a solution to Covid19 Pandemic and Longevity Risk Management.
Presenter: Leonard Mushunje, Columbia University
Abstract: The global pandemic of Covid19 is considered an ultimate result of the human to human close interactions. In most cases, it is difficult to maintain the suggested social distance measures and this is triggering the spread and incidence of the disease. As such, there is a need to insert some beautiful ways and measures to mitigate and reduce such devastating issues. This study aims to provide data based way of measuring and quantifying the number of people in any area, mostly in developing circles. Such information helps in several ways: In the health sector, it enhances better policy making and swift provision of emergency evacuations-during the pandemic. In insurance business, it fosters effective management of Longevity risks. Insurers can adjust their capital reserves for life insurance products and longevity linked products (annuities) by age and gender factors. Using the O.R. Tambo airport data, we derived a number of correlation results. Firstly, we found a strong correlation between the number of people in restricted places and the captured mobile phone based social activities. Secondly, we found a strong positive correlation of number of attendees, social median and Internet activities, and Covid19 spikes. Lastly, we found that insurance firms in South Africa were at risk of increasing longevity risk levels due to unanticipated death of policyholders. This was indicated by sharp increases in demand for claim reserves during the pandemic. The approach used and the results generated poses that big data mining and extraction from various social media platforms offers valuable and insightful patterns which helps in planning and measurements in line with the current global pandemic crisis and optimal reserving.

Title: MultiComBat: ComBat harmonization of multiple batch variables
Presenter: Hannah Horng, University of Pennsylvania
Co-authors: Despina Kontos, Russell T. Shinohara
Abstract: ComBat is a promising harmonization method for biomedical imaging data acquired on multiple scanners, but it cannot harmonize for multiple batch effects. In this work, we generalize the ComBat model to incorporate multiple batch effects and modify the estimation algorithm to estimate the corresponding corrections to transform the data to have the same mean and variance as samples with a specified reference batch combination. We evaluate the MultiComBat approach by conducting a simulation study of batch effects for multiple batch variables in radiomic features and compare the harmonization performance of MultiComBat with standard ComBat harmonization by batch variable combination. Standard ComBat demonstrated a lower percentage of features with significant differences in distribution detected using the Anderson-Darling test than MultiComBat, likely due to over-harmonization without respect for the multiple batch variable structure. However, MultiComBat was more effective in batch group means to the specified reference, indicating that MultiComBat obtains more generalizable correction estimates. Our findings suggest that MultiComBat can be used to harmonize by multiple batch effects and demonstrates greater accuracy in comparison to standard ComBat.

Title: Novel Scalar-on-matrix Regression for Unbalanced Feature Design Matrices
Presenter: Jeremy Rubin, University of Pennsylvania, Department of Biostatistics, Epidemiology, and Informatics
Co-authors: Jarcy Zee
Abstract: Imaging features from whole slide images (WSIs) of kidney biopsies can be used to comprehensively characterize histologic objects and develop new biomarkers of kidney disease outcomes. Each individual's imaging features can be represented by a matrix whose entries are a common set of features (columns) that are measured for each histologic object (rows) from that individual's WSI. However, since each individual can have a different number of histologic objects in their WSI, the imaging feature matrices have different numbers of rows across individuals. We propose the CLUstering Structured laSSO (CLUSSO), a novel scalar-on-matrix regression technique to predict scalar clinical outcomes from imaging feature matrices that are unbalanced across individuals. CLUSSO clusters histologic objects into subgroups and averages features within subgroups before using scalar-on-matrix regression with the structured lasso. Simulation study results indicate that CLUSSO can accurately identify the features which truly affect the outcome when there is a small number of imaging features. Further, the bias in coefficient estimates from CLUSSO is near optimal. In contrast, a naïve method which applies the standard lasso to the features averaged across all histologic objects for each individual exhibits increasing false positive rates for increasing sample size as well as larger biases in the coefficient estimates relative to CLUSSO. CLUSSO was applied to predict kidney disease function from imaging features of tubules from kidney biopsy WSIs and compared to the naïve method for illustration.

Title: An MCMC Approach to Bayesian Image Analysis in Fourier Space
Presenter: Konstantinos Bakas, King Abdullah University of Science and Technology (KAUST)
Co-authors: John Kornak, Hernando Ombao
Abstract: Bayesian methods are commonly applied to solve image analysis problems such as noise-reduction, feature enhancement and object detection. A primary limitation of these approaches is the computational complexity due to the interdependence of neighboring pixels which limits the ability to perform full posterior sampling through Markov chain Monte Carlo (MCMC). To alleviate this problem, we develop a new posterior sampling method that is based on modeling the prior and likelihood in the space of the Fourier transform of the image. One advantage of Fourier-based methods is that many spatially correlated processes in image space can be represented via independent processes over Fourier space. A recent approach known as Bayesian Image Analysis in Fourier Space (or BIFS), has introduced parameter functions to describe prior expectations about image properties in Fourier space. To date BIFS has relied on Maximum a Posteriori (MAP) estimation for generating posterior estimates; providing just a single point estimate. The work presented here develops a posterior sampling approach for BIFS that can explore the full posterior distribution while continuing to take advantage of the independence modeling over Fourier space. As a result computational efficiency is improved over that for conventional Bayesian image analysis and mixing concerns that commonly have to be dealt with in high dimensional Markov chain Monte Carlo sampling problems are avoided. Implementation results and details are provided using simulated data.

Title: Application of Closed-form Gamma Mixture Model in mxIF Cell Gating
Presenter: Jiangmei Xiong, Vanderbilt University
Co-authors: Eliot T McKinley, Joseph T Roland, Robert Coffey, Martha J Shrubsole, Ken S.Lau, Simon Vandekar
Abstract: Multiplexed immunofluorescence (mIF) imaging is a sub-cellular resolution technology where cell expression levels are captured with multichannel images of stained tissue samples that identify up to 30 proteins. Marker gating is the process of identifying cell phenotypes based on cell marker expression values. Traditional manual gating is slow, and results are not reproducible, while recent software for phenotyping assumes a log-normal model, which is not appropriate for non-modal cell population densities. We introduce a closed-form gamma mixture model to perform cell gating, which removes the subjectivity introduced by traditional manual gating procedures and can easily incorporate biological information while speeding up the estimation process. In this work, we derived the closed-form gamma mixture model, and illustrate the analysis pipeline to perform automatic gating of segmented cell data. As the model is applied to a large number of slides and marker channels, we introduce diagnostic plots that aid in fine-tuning the results. We also evaluate our method by comparing with silver standard and another recently developed method for phenotyping mIF data.

Title: An Analytical Framework for Quantifying Cell-Cell Interactions from Multiplex Imaging Data
Presenter: Maria Masotti, University of Michigan Biostatistics
Co-authors: Nathaniel Osher, Joel Eliason, Arvind Rao, Veera Baladandayuthapani
Abstract: The tumor microenvironment (TME) is a complex ecosystem containing tumor cells, other surrounding cells, blood vessels, and extracellular matrix. Recent advances in multiplexed imaging

technologies allow researchers to map over 50 cellular markers in the TME at the single cell level while preserving their spatial locations. Evidence is mounting that cellular interactions in the TME can promote or inhibit tumor development and contribute to drug resistance. Current statistical approaches to quantify cell-cell interactions do not readily scale to the outputs of new imaging technologies which can distinguish over 20 unique cell types in one image. We propose a scalable analytical framework and accompanying R package, DM.ME, to quantify, visualize, and model cell-cell interactions in the TME.

Title: A Highly-Accelerated Image Reconstruction Method by Simultaneous Encoded Slices with CAIPI in FMRI
Presenter: Ke Xu, Marquette University
Co-authors: Daniel B. Rowe
Abstract: A topic of fMRI studies is to accelerate the number of images per unit of time to create each volume. Techniques such as SENSE and GRAPPA measure fewer data in an image slice but are able to reconstruct an image. Controlled Aliasing in Parallel Imaging (CAIPI) shifts is a technique where the field-of-view is shifted for decreasing the influence of the geometry factor. The simultaneous multi-slice (SMS) techniques provide an alternative reconstruction method that multiple slices are acquired and aliased concurrently. A novel SMS technique called A CAIPI Approach of Multi-Coil Separation of Parallel Encoded Complex-Valued Slices will be presented. It combines the image shift method of CAIPI and the CAIPI with view angle tilting (CAIPIVAT) technique, along with the Hadamard phase-encoding aliasing technique which allows different combinations of the aliased slices and increases the size of the aliasing matrix by applying a unique Hadamard multi-band pulse sequence. The bootstrap sampling method and the artificial aliasing of calibration images are included to reduce correlation. The least square estimation function is applied to separate the reconstructed images.

Title: A Full Bayesian Approach to GRAPPA Reduces Noise in fMRI Image Reconstruction
Presenter: Chase J. Sakitis, Department of Mathematical and Statistical Sciences, Marquette University
Co-authors: Daniel B. Rowe
Abstract: Functional Magnetic Resonance Imaging (fMRI) is a medical imaging technique used to noninvasively observe the human brain in action. The machine measures arrays of complex-valued spatial frequencies called k-space which are then reconstructed into images using an inverse Fourier transform (IFT). In fMRI, capturing brain activation during a physical task is dependent on how quickly volume brain images are acquired. The acquisition of full volume k-space arrays can take a considerable amount of scan time and is the limiting factor in the fMRI process. GeneRalized Autocalibrating Partial Parallel Acquisition (GRAPPA) is a parallel imaging technique that allows under sampling, or subsampling, of k-space arrays but can still produce full images. To fill in the missing values of the subsampled k-space, GRAPPA uses localized weights estimated from full pre-scan calibration k-space arrays before the fMRI experiment. However, GRAPPA uses the same estimated weights for every time point in the series of subsampled k-space arrays and discards valuable information from the pre-scan calibration k-space arrays. Here, we propose a Bayesian approach to GRAPPA where prior distributions for the unobserved spatial frequencies, localized weights, and uncertainty are quantified from the a priori calibration k-space information. Unobserved spatial frequencies, localized weights, and k-space noise variance are estimated jointly from the maximum a posteriori estimate via the Iterated Conditional Modes algorithm or individually estimated from the marginal posterior means via Markov chain Monte Carlo using Gibbs sampling. Our Bayesian GRAPPA (BGRAPPA) technique incorporates valuable prior image information to estimate the unobserved k-space values from the posterior distribution. BGRAPPA and GRAPPA are applied to simulated fMRI data with preliminary results yielding less noisy reconstructed images, improving image quality.

Title: Component Analysis Feature Selection for Classification of High-dimensional Neuroimaging Data
Presenter: Matthew A. Snodgress, University of Minnesota, Twin Cities
Co-authors: Nathaniel Helwig, Ph.D.
Abstract: Developing a useful classification rule for high-dimensional neuroimaging data requires selecting features (i.e., classification inputs) that are both (a) informative about group differences and (b) interpretable with respect to underlying brain processes, which is an open methodological problem. Component analysis (CA) methods have been shown to be useful for decomposing high-dimensional neuroimaging data into interpretable components that reflect differences in brain processes across individual persons. However, there has been less work investigating how such methods can be used to extract features for classifying high-dimensional neuroimaging data contained within three-way (or higher) arrays (i.e., tensors). In this paper, we explore the utility of CA models—specifically, principal component analysis (PCA); independent component analysis (ICA); and Parallel Factor Analysis (PFA)—when the goal is to develop a binary classification rule. We propose a cross-validation procedure that tunes the hyperparameters of CA models and binary

classifiers, which results in a flexible approach for classification of high-dimensional neuroimaging data. Using event-related potential data, we demonstrate that CA models can produce features that are useful inputs for standard binary classification methods, such as penalized logistic regression (PLR). Our results reveal that the proposed feature selection method can produce classification rules that (1) perform as well or better than other methods, and (2) are more interpretable with respect to latent brain networks than typical machine learning rules. In particular, our findings suggest that CA feature selection, in combination with penalized logistic regression, can provide clear insights regarding differences in ERP waveforms between alcoholics and controls while also providing good classification accuracy.

Keywords: Neuroimaging, Classification, ICA, Parallel Factor Analysis, LASSO

Title: Integrating segmentation uncertainty into statistical analyses of brain volumes
Presenter: Christina Chen, University of Pennsylvania
Co-authors: Russell T. Shinohara
Abstract: Multi-atlas image segmentation is a widely used approach in imaging studies that involve, for example, estimating the volume of a region of interest (ROI). However, current practices typically treat these images equally without incorporating the fact that the registration quality might vary among subjects. We propose a method that estimates the variance of the ROI volume estimate for each subject due to the multi-atlas segmentation procedure and thus provides a way of reweighting these estimates to increase efficiency in downstream estimation problems.

Title: Deep Causal Feature Extraction and Inference with Neuroimaging Genetic Data
Presenter: Yuchen Yao, School of Statistics at University of Minnesota
Co-authors: Dipnil Charkraborty, Lin Zhang, Xiaotong Shen, Wei Pan
Abstract: Alzheimer's disease (AD) is one of leading causes of death while the etiology of AD is not fully understood. In previous work, pre-defined features of some regions of interest (ROIs) were used in the study of the etiology of AD but they might not be causal. Inspired by the huge success of deep learning (DL), we propose a new DL based instrumental variable (IV) regression, Deep Feature Extraction via Instrumental Variable Regression (DeepFEIVR), extracting causal features for an outcome (e.g. AD status) from high dimensional imaging data (e.g. MRI scans), then using separate and large-scale GWAS summary statistics to test for associations between the extracted features and the outcome. In simulations, we show DeepFEIVR can correctly detect causal relationships, and the results from DeepFEIVR using summary statistics are nearly the same as those obtained from individual-level data. In real data analysis, we apply DeepFEIVR to the Alzheimer's Disease Neuroimaging Initiative (ADNI) data to extract causal features from brain MRI scans to predict the AD status, then use the International Genomics of Alzheimer's Project (IGAP) AD GWAS summary data to test for associations between the extracted features and AD. Finally, to facilitate interpretation, we investigate the relationships between the extracted features and brain regions.

Title: Compensating for selection bias in functional connectivity study of autism
Presenter: Liangkang Wang, Emory University
Co-authors: Benjamin Risk
Abstract: The exclusion of high-motion participants in functional Magnetic Resonance Imaging (fMRI) studies is a common practice to reduce motion-related artifacts. However, this exclusion can introduce biases by altering the distribution of clinically relevant variables, leading to a non-representative study sample. This paper aims to introduce a framework that employs the Average Inverse Probability Weighted Estimator (AIPWE) method to address these biases by treating excluded scans as missing data.

Using simulated datasets, we tested the AIPWE method on single-region and multi-region scenarios with varying block correlations to evaluate its effectiveness in addressing selection bias. Our results demonstrate that the AIPWE method effectively mitigates the impact of the selection bias in these simulations, providing more accurate estimates of functional connectivity.

We applied the AIPWE method to real-world data from 396 children aged 8-13 (144 with autism spectrum disorder and 252 typically developing) from the Autism Brain Imaging Data Exchange (ABIDE) datasets. Our findings reveal that autistic children are more likely to be excluded compared to typically developing children. To address data loss and resulting biases, we adapted the AIPWE method in conjunction with an ensemble of machine learning algorithms.

The proposed approach identified more edges with differing functional connectivity between autistic and typically developing children compared to the standard approach, highlighting the potential of our framework to improve the study of heterogeneous populations where motion is prevalent. Overall, this study underscores the importance of considering missing data, where even in the absence of selection bias, AIPWE can lead to improved statistical power in functional connectivity analyses.

Title: The Effects of Early Alcohol Consumption on Mental Health Outcomes: Results from the Adolescent Brain and Cognitive Development (ABCD) Study
Presenter: Ana Ferariu, Drexel University
Co-authors: Hansoo Chang, Alexei Taylor, Fengqing Zhang
Abstract: Early alcohol exposure has been shown to be associated with an increased risk of alcohol consumption and substance use disorder later in life. Alcohol use affects psychopathology and personality traits. In this ongoing study, we model the change in alcohol sipping trajectory over time using data from the Adolescent Brain and Cognitive Development (ABCD) study, the largest longitudinal study of brain development and youth health in the United States. We then examine the effects of early alcohol use patterns on personality and mental health outcomes, such as impulsivity, behavioral inhibitory system (BIS), behavioral activation system (BAS) and mania. In addition, we analyze the task fMRI data from the stop signal paradigm to capture the neural correlates of response inhibition. Results show that alcohol sipping patterns and activation of inferior frontal gyrus are differentially associated with several aspects of mental health outcomes prospectively. Findings from our study hold the potential to provide important insights into the complex and multifaceted relationship between alcohol use, personality, and mental health.

Title: Linking Biological Age and Brain Age to Child Health in the Adolescent Brain Cognitive Development (ABCD) Study
Presenter: Hansoo Chang, Drexel University
Co-authors: Alexei Taylor, Ana Ferariu, Fengqing Zoe Zhang
Abstract: Biological age and brain age estimated with either biological markers or neuroimaging measures have recently emerged as surrogate aging biomarkers shown to be predictive of various health outcomes. These surrogate aging biomarkers are not disease-specific and capture health at the whole person level. Most of the existing studies examined these surrogate biomarkers in middle-aged and older adults. More research is needed to evaluate the utility of biological and brain age in children with longitudinal data. Using the Adolescent Brain Cognitive Development (ABCD) study, we utilized blood-based sample and brain imaging data to calculate biological and brain age with the Klemera-Doubal method in a cohort of young children across the United States. In addition, we examined the relationship between biological and brain age with a multitude of physical, mental, and social health outcomes in children. Our results show that biological age and brain age are predictive of different child health outcomes and provide complementary information.

Title: Covariance-Assisted Sparse Multivariate Additive Regression (Co-MAdRe)
Presenter: Neel M. Desai, University of Pennsylvania
Co-authors: Veera Baladandayuthapani, Russell T. Shinohara, Jeffrey S. Morris
Abstract: This article develops a robust, computationally efficient solution for the simultaneous selection and estimation of multiple sparse additive models with correlated errors. Our method (Co-MAdRe) simultaneously selects between null, linear, and smooth non-linear effects for each predictor while incorporating jointly estimated sparse residual structure among responses for potential gains both in selection accuracy and in statistical efficiency in a manner analogous to the principles of seemingly unrelated regressions (SUR). Our method is constructed in a computationally efficient manner that allows the selection and estimation of linear and non-linear covariates to be conducted in parallel across responses. Compared to single-response approaches that marginally select linear and non-linear covariate effects, we demonstrate with extensive designed simulations that our approach leads to gains in both statistical efficiency and estimation accuracy, particularly in settings where signal is moderate compared to the level of noise. We apply our approach to protein-mRNA expression from 8 known breast cancer pathways obtained from The Cancer Proteome Atlas (TCPA), characterizing both mRNA-protein associations and protein-protein subnetworks for each pathway.

Title: Group distributional ICA for multi-subject DTI data
Presenter: Guangming Yang, Emory University
Co-authors: Ben Wu, Jian Kang, Ying Guo
Abstract: Diffusion tensor imaging (DTI) is a frequently used imaging modality to investigate white matter fiber connections of human brain. DTI provides a unique structural perspective for studying human brain organization which accounts for its increasing popularity. Analysis of images derived from this modality involves common goals such as dimension reduction, denoising, and structure network extraction. However, there has been very limited work for conducting blind source separation on multi-subject DTI data. Due to the special characteristics of the 3D diffusion tensor measured in DTI, existing methods such as standard independent component analysis (ICA) can't be directly applied. Motivated by a new DICA framework (Wu et al., 2021)^{1}, the proposed group DICA aims to fill this gap by performing a fundamentally new blind source separation approach that separates the parameters in the distribution function of the observed imaging data as a mixture of independent source signals. To

the best of our knowledge, the proposed group DICA is the first ICA method that directly decompose 3D diffusion tensors in multi-subject DTI studies. When applying the method to multi-subject DTI data, the extracted structural networks were found to correspond to several major white matter fiber bundles. We also evaluate the performance of the method as compared with the existing DTI source separation method through simulation studies and reproducibility evaluations.

Title: Unveiling spatial architecture in multiplexed imaging data with a spatial topic model
Presenter: Xiyu Peng, Memorial Sloan Kettering Cancer Center
Co-authors: James Smithy, Ronglai Shen, Katherine Panageas
Abstract: Multiplexed imaging technologies allow for in-depth examination of biological tissues at the single-cell level while retaining their spatial information. However, the complex spatial patterns of tissue organization remain poorly understood and have yet to be systematically studied, due to a lack of statistical and computational methods. We propose a novel spatial topic model to investigate tissue architecture in multiplexed imaging data. This model considers both cell phenotype and spatial information, facilitating the identification of complex spatial tissue structures. In the model, spatial information is incorporated into the design of documents, which represent densely overlapped regions in images. By establishing a flexible cell-to-document relationship, the model clusters co-occurring and spatially adjacent cells into the same topic. The model is implemented using a collapsed Gibbs sampling algorithm for inference. Our proposed methodology was successfully demonstrated by analyzing multiplexed imaging data from tumor tissue samples, accurately identifying and quantifying spatial tissue structures without human intervention.

Title: Sparse Independent Component Analysis with an Application to Cortical Surface fMRI Data
Presenter: Zihang Wang, Emory University
Co-authors: Benjamin B. Risk, Irina Gaynanova
Abstract: Independent component analysis (ICA) is widely used to estimate resting-state networks (spatial components) and their time courses in neuroimaging studies. However, existing ICA algorithms are hard to interpret due to estimated components being non-sparse. Previous approaches to sparse ICA replace the non-smooth objective function with smooth approximations, resulting in components that do not achieve exact zeros. We propose a novel Sparse ICA method that enables sparse estimation of independent source components by solving a non-smooth optimization problem via the relax-and-split framework. The proposed Sparse ICA method is computationally efficient and balances statistical independence and sparsity simultaneously. Our simulations demonstrate improvements over existing approaches in terms of the estimation accuracy of both source signals and signal time courses. Computationally, our method is very fast, being on par with the popular Fast ICA approach. A single initialization takes less than three seconds to estimate thirty ICs on fMRI cortical surface data with 59,412 vertices. We apply our method to resting-state fMRI in school-aged children, which reveals differences in functional connectivity patterns between autistic and typically developing children.

Title: Neural Orientation Distribution Fields
Presenter: Will Consagra, Harvard Medical School
Co-authors: Yogesh Rathi
Abstract: Inferring brain structure in-vivo requires accurate estimation of the orientation distribution function (ODF), which encodes key local tissue properties. However, accurately estimating ODFs and quantifying uncertainty from diffusion MRI (dMRI) data is a complex inverse problem due to several factors, such as significant noise, high-dimensional parameter spaces, and sparse measurements. In this work, we model the dMRI signals as discrete, noisy observations of a latent continuous function-valued random field. We expand this field using a random series representation parameterized by a neural field (NF), facilitating the estimation of a deep data-driven basis system that learns the complex spatial correlation structures in the data, thereby improving the statistical efficiency for sparse sample and noisy regimes. Given explicit priors on the smoothness of the field, we derive an analytic form for the approximate predictive posterior distribution which can be used to quantify the uncertainty in the ODF estimate at any spatial location. We demonstrate our method on a simulated 3D phantom and apply it to a real diffusion dataset, showing improvements over existing approaches.

Title: Cross-sectional and longitudinal investigations of neural correlates of suicidal ideation in childhood and early adolescence
Presenter: Timothy J. Wanger, University of Minnesota, School of Medicine, Department of Psychiatry and Behavioral Sciences
Co-authors: Andrea Wiglesworth, Bonnie Klimes-Dougan, Mark Fiecas, Bryon Mueller, Monica Luciana, Kathryn Cullen
Abstract: Background: Suicide is a prominent, and growing issue in young individuals in the United States. Identifying neural profiles of individuals who report suicidal ideation (SI) is a critical step towards understanding the development of suicidal thoughts and behaviors. To achieve this, longitudinal studies such as the Adolescent Brain Cognitive Development study (ABCD) can be leveraged to

identify SI predictors, assess the reliability of observed SI correlates, and chart developmental trajectories of individuals reporting SI.
Methods: Our cohort consisted of 6340 youths from the ABCD 3.0 Release that had full datasets at both baseline (when they were 9-10 years old) and at the two-year follow-up. Following up on our previous study which identified neural correlates of SI during the baseline visit (Wiglesworth 2021), here we employed identical methods (including the same brain metrics of resting state functional connectivity and brain activation patterns during an emotional n-back task within brain regions encompassing the default mode and salience networks) to determine whether these correlation patterns were also evident at the second-year follow-up. Additionally, we used linear discriminant analysis to test whether SI at the two-year follow-up visit could be predicted by neural indicators at the baseline visit.
Results: The group of adolescents reporting SI at the two-year follow-up had minimal overlap with the group had reported SI at baseline. Furthermore, examination of self-reported SI revealed inconsistencies (e.g., some adolescents who had reported SI at baseline reported no past history of SI at follow-up) which may complicate analyses. Cross-sectional findings did not replicate correlations observed in the baseline visit– though we observed new associations between SI and decreased resting state functional connectivity in the default mode network (B= -0.118, p = 0.046), and increased dorsomedial prefrontal cortex and medial temporal gyrus connectivity (B = 0.632, p = 0.014). The linear discriminant analysis model to predict whether adolescents at the two-year follow-up had SI in the present, past, or never had limited sensitivity (0.28, 0.30, 0.92) and positive predictive value (0.13, 0.05, 0.98) respectively.
Conclusions: While findings do not suggest consistent patterns of neural correlates of SI across the baseline and 2-year ABCD assessments, they do provide some evidence that dysregulation of neural networks (default mode, salience) is associated with SI across time points. Possible factors to explain the pattern changes may include brain developmental changes occurring over this time period, heterogeneity across individuals in the neural underpinnings of SI, and inconsistencies in SI reporting. With future follow-up visits expected for this cohort, future work can focus on the neural changes occurring in cohorts whose developmental trajectories may include occurrence and reoccurrence of SI.

Title: Analysis of Neuroimaging Experiments in R
Presenter: Joerg Polzehl, WIAS
Co-authors: Karsten Tabelow
Abstract: R provides a wide range of tools for modeling and data analysis for neuroimaging experiments.

In our book 'Magnetic Resonance Brain Imaging - Modeling and Data Analysis Using R' we present examples for the analysis of
four types of experiments:
- functional MRI,
- diffusion weighted imaging,
- multi parameter mapping and
- inversion recovery MRI.
The examples are based on packages available from Neuroconductor or/and CRAN and use publicly available data sets, e.g. from the openNeuro repository.

For the second edition the content of the book has been extended. References and resources have been updated. The R code used to produce the book has been reorganized for better readability. All code and data are available for complete reproducibility of the analyses.

Title: Mitigating inter-scanner biases in high-dimensional neuroimaging data via spatial Gaussian process
Presenter: Rongqian Zhang, Department of Statistical Sciences, University of Toronto
Co-authors: Linxi Chen, Lindsay D. Oliver, Aristotle N. Voineskos and Jun Young Park
Abstract: In neuroimaging studies, combining data collected from multiple study sites is becoming common to increase the reproducibility of scientific discoveries. However, it has been shown that unwanted variations are associated with using different scanners in data acquisitions, termed “inter-scanner biases.” Despite existing methods for correcting scanner effects (e.g. ComBat), these methods mostly focus on removing scanner-specific means and variances. They are limited to addressing high-dimensional data that reveals a strong spatial autocorrelation, such as cortical thickness. To address these challenges, we develop a novel multivariate normalization method called Spatial Autocorrelation Normalization via Gaussian process (SAN-GP) that models and reduces inter-scanner biases for vertex-level cortical thickness data. We also consider integrating our method with the existing method (i.e., ComBat) to achieve better harmonization. We evaluate and compare our method with existing methods in removing scanner effects and retaining spatial auto-correlation by using extensive simulation studies.

Title: Threshold-Free Identification of Group-Difference Networks in Neural Connectivity Data
Presenter: Julia M. Fisher, BIO5 Institute, Statistics Consulting Laboratory, University of Arizona, Tucson, AZ
Co-authors: Antonio M. Rubio, Aidan Dolby, Chidi Patrick Ugonna, Edward J. Bedrick, Nan-kuei Chen

Abstract: A common question in cognitive neuroscience is how populations (e.g., people with a particular disease, age-matched healthy controls, etc.) differ in their neural connectivity. These neural connections can be represented using a network where nodes represent specific brain areas or regions of interest (ROIs) and edges reflect the strength of connection between ROIs. However, even with a coarse-grained parcellation of the human brain, the number of ROI-to-ROI connections is large enough that the expected number of false positive results without any familywise error correction is prohibitively high. The Network-Based Statistics (NBS) approach to analyzing neural connectivity achieves weak control of family wise error by thresholding a connectivity matrix of linear modeling group contrasts and doing inference on supra-threshold connected clusters (difference networks) of ROIs and edges. The resulting difference networks (and subsequent scientific interpretation) can be highly dependent on the choice of threshold, with low thresholds tending to produce larger clusters than higher thresholds.

We propose a modification of NBS that removes the dependence of difference networks on the choice of threshold by using the connection modeling t statistics as weights in the Louvain community (i.e., cluster) detection algorithm. The Louvain algorithm seeks to partition a graph into tightly connected modules with sparser connections between them via the maximization of a modularity index. Since the algorithm is not deterministic, we run it 100 times and construct a coclassification matrix (i.e., a matrix of the number of times each pair of ROIs were placed in the same module). The Louvain algorithm is run once more on the coclassification matrix to produce a more robust partition of the original connectivity matrix than any single run of the Louvain algorithm would likely produce. The modularity index of this final partition is compared to a null distribution of modularity indices. The null distribution is constructed by repeating the above process on a set of randomization data sets, where for each data set the subject group labels have been randomly permuted. As the modularity index is sensitive to the number of modules in a given partition, inference is conditional on the number of modules in the final partition of the true data set.

Initial simulation results indicate that the false positive rate is well maintained for the conditional modularity inference; in a set of 1000 simulated data sets with 15 subjects per group, a 64-ROI connectome, and no group differences, 4.2% of data sets had statistically significant modularity when the significance level alpha was set to 5%; for alpha = 0.01, it was 0.6%; for alpha = 0.1, it was 10%. In a simulation where a single fully-connected network in one group (alpha = 0.8) was split into four disconnected sub-networks in a second group with different within-sub-network correlations (rho = 0.2, 0.4, 0.6, 0.75), the proposed community detection approach resulted in statistically significant group difference modularity in 67.3% of 1000 data sets (with rho = 0.05). Comparisons of the final Louvain partition to the true partition were made with the Rand index, a measure of similarity between two partitions that ranges from 0 to 1 with 1 indicating perfect agreement. The average Rand index across the 1000 data sets with the above simulated group difference was 0.8095 (95% CI: 0.8094 – 0.8095). In contrast, NBS partitions at three quantile-based thresholds (95th , 99th, and 99.9th percentiles) in line with the thresholds used in the original NBS paper revealed average Rand indices of 0.3145 (95% CI: 0.3137 – 0.3153), 0.5757 (95% CI: 0.5720 - 0.5794), and 0.7723 (95% CI: 0.7703 – 0.7744) for the same data. Paired t tests of the Rand indices in the Louvain versus original NBS partitions were all statistically significant (all $p < 0.05$) indicating improved community detection via the Louvain approach. Additional simulation results and application of the method to a data set of patients with Parkinson's Disease versus healthy controls will be shown. Furthermore, the proposed approach will be combined with threshold-free NBS (an NBS extension that allows familywise error-corrected inference on individual connections) to give simultaneous threshold-free difference network identification and individual connection differences.

Title: Multivariate Residualization in Medical Imaging Analysis
Presenter: Kevin Donovan, University of Pennsylvania
Co-authors: Nicholas Tustison, Kristin Linn, Russell Shinohara
Abstract: Nuisance variables in medical imaging research are common, complicating association and prediction studies based on image data. Medical image data are typically high dimensional, often consisting of many highly correlated features. As a result, computationally efficient and robust methods to address nuisance variables are difficult to implement. By-region univariate residualization is commonly used to remove the influence of nuisance variables, as are various extensions. However, these methods neglect multivariate properties and may fail to fully remove influence related to the joint distribution of these regions. Some methods, such as functional regression and others, do consider multivariate properties when controlling for nuisance variables. However, the utility of these methods is limited for data with many image regions due to computational and model complexity. We develop a multivariate residualization method to estimate the association between the image and nuisance variable using a machine learning algorithm and then compute the orthogonal projection of each

subject's image data onto this space. We illustrate this method's performance in a set of simulation studies and apply it to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

Title: Leveraging the Adjusted Rand Index for Feature Selection: An Empirical Comparison of Graph Theory Metrics for Resting-State Functional Connectivity Subtyping Using HYDRA
Presenter: Leyla R Brucar, MSc., Graduate Program in Cognitive Science. Department of Psychiatry, University of Minnesota
Co-authors: Eric Rawls, PhD; Anna Zilverstand, PhD
Abstract: The inefficacy of standard treatments for psychiatric disorders is linked to the 'heterogeneity problem': different causal mechanisms underlie psychiatric disorders for different subsets of individuals. Brain-based subtyping may be an effective tool to reduce neurobiological heterogeneity evident in psychiatric disorders, but the methodological variation in these studies has yielded irreproducible results. Using Alcohol Use Disorder (AUD) as an example, we identified an optimal metric for resting-state functional connectivity (RSFC)-based subtyping using graph theory (GT) and a semi-supervised clustering algorithm, HYDRA. We leveraged the Adjusted Rand Index (ARI), a quantitative measure of cluster performance, to compare the reproducibility of GT metrics for RSFC subtyping.

We included 668 participants from the Human Connectome Project with complete phenotypic and resting-state data (AUD Subjects: 2+ DSM-IV-TR symptoms; n=147; 38% female). Each participant's weighted RSFC matrix was computed from a parcellated functional time series, and used to calculate five common graph metrics: nodal global efficiency, local efficiency, betweenness centrality, participation coefficient and strength. Each of the five GT measures were used as input to HYDRA assessing cluster solutions from k = 2:10, using 10-fold cross validation based on the ARI. This analysis was repeated 100 times, resulting in 100 ARI observations for each cluster solution (k = 2:10) for each metric. A two-way repeated measures ANOVA was implemented, using GT metric and cluster k as within-subject factors, to compare stability and performance between the metrics.

All GT metrics demonstrated their best cluster performance for the cluster solution k = 2, with a significantly higher ARI score as compared to all other solutions. The participation coefficient had the highest overall ARI score for this cluster solution (k = 2), significantly greater than all other metrics ($p < 0.001$). Local efficiency displayed the second highest ARI, followed by equally performing global efficiency and strength, and then betweenness centrality.

We leveraged the ARI, an empirical measure of cluster performance, to identify the most reliable GT metric for RSFC subtyping in AUD using HYDRA. Future work will use the identified GT metric, participation coefficient, to assess its ability to detect underlying neurobiological heterogeneity in AUD. If distinct and reliable RSFC subtypes can be identified, this will pave the way for the development of more targeted, neurobiologically-grounded, personalized interventions.

Title: Penalized model-based clustering of fMRI data
Presenter: Andrew S. DiLernia, Grand Valley State University
Co-authors: Karina Quevedo, Jazmin Camchong, Kelvin Lim, Wei Pan, Lin Zhang
Abstract: Effectively describing functional connectivity (FC) of the brain is important for improving understanding of neurological and mental disorders. We present a penalized model-based clustering method, the random covariance clustering model (RCCM), to use fMRI data to simultaneously cluster subjects and provide interpretable estimates of both subject- and cluster-level FC networks. We demonstrate the utility of the RCCM by analyzing a multi-subject resting-state fMRI data set collected on healthy controls and participants diagnosed with varying degrees of schizophrenia, finding evidence to support the disconnection hypothesis among those diagnosed with schizophrenia.

Title: A fast and powerful spatial-extent inference for testing variance components in reliability and heritability studies
Presenter: Ruyi Pan, University of Toronto
Co-authors: Erin W. Dickie, Colin Hawco, Nancy Reid, Aristotle N. Voineskos, Jun Young Park
Abstract: Clusterwise inference is a popular approach in neuroimaging to increase sensitivity, but most existing methods are currently restricted to the General Linear Model (GLM) for testing mean parameters. Statistical methods for variance components testing, which are critical in neuroimaging studies that involve estimation of narrow-sense heritability or test-retest reliability, are seriously underdeveloped due to methodological and computational challenges, which would potentially lead to low power. To fill this gap, we propose a fast and powerful test for variance components called CLEAN-V. CLEAN-V models global spatial dependence structure of imaging data and computes a locally powerful variance component test statistic by data-adaptively pooling neighborhood information. Correction for multiple comparisons is achieved by permutations to control family-wise error

rate (FWER) accurately. Through analysis of task-fMRI data from the Human Connectome Project across five tasks and comprehensive data-driven simulations, we show that CLEAN-V outperforms existing methods in detecting test-retest reliability and narrow-sense heritability with significantly improved power, with the detected areas aligning with activation maps. The computational efficiency of CLEAN-V also speaks of its practical utility, and it is available as an R package.

Title: Bayesian inference on Brain-Computer Interface using the GLASS Model
Presenter: Bangyao Zhao, University of Michigan
Co-authors: Jane Huggins, Jian Kang
Abstract: The brain-computer interface (BCI) enables individuals with severe physical impairments to communicate with the world. BCIs offer computational neuroscience opportunities and challenges in converting real-time brain activities to computer commands and are typically framed as a classification problem. This article focuses on the event-related potential (ERP) BCI design, known as the P300 BCI, where the primary challenge is classifying target/non-target stimuli. We develop a novel Gaussian latent group model with sparse time-varying effects (GLASS) for making Bayesian inferences on the P300 BCI. GLASS adopts a multinomial regression framework that directly addresses the dataset imbalance in BCI applications. The prior specifications facilitate i) feature selection and noise reduction using soft-thresholding, ii) smoothing of the time-varying effects using global shrinkage, and iii) clustering of latent groups to alleviate high spatial correlations of EEG data. We develop an efficient gradient- based variational inference (GBVI) algorithm for posterior computation and provide a user-friendly python module available at https://github.com/BangyaoZhao/GLASS. Our application study identifies important EEG channels (PO8, Oz, PO7, Pz, C3) that align with existing literature. GLASS further reveals a group effect from channels in the parieto-occipital region (PO8, Oz, PO7), which is validated in cross-participant analysis.

Title: Heritability of subcortical volumes in adolescents
Presenter: Christian Coffman, University of Minnesota
Co-authors: Eric Feczko, Saonli Basu
Abstract: Understanding heritability of neuroanatomical structures in adolescents is important to characterize the developmental trajectory of the human brain. Subcortical volumes are highly heritable in adults and teenage years, but likely differ in adolescence. The Adolescent Brain Cognitive Development study (ABCD) provides data to characterize heritability of subcortical structures in adolescence. Site effects, often correlated with the genetic effect could bias estimates. Here is a proposed random-effect model- based method of moments approach to estimate heritability on regional brain volumes. Theoretical as well as simulations results demonstrate that our proposed approach provides an unbiased estimation when site effects are correlated with genetic effects, such as when ethnicities are unequally distributed across sites. We compare several heritability estimation approaches on regional brain volumes while adjusting for site specific effects. The most heritable regions were cerebellum cortex (≈ 60%) and the hippocampus (≈ 50%). Our proposed approach is computationally efficient and provides a significant improvement over existing approaches.

Title: A robust multivariate, non-parametric outlier identification method for scrubbing in fMRI
Presenter: Fatma Parlak, Indiana University
Co-authors: Damon D. Pham, Amanda F. Mejia
Abstract: Functional magnetic resonance imaging (fMRI) data contain high levels of noise and artifacts due to head motion, scanner instabilities, and other sources. To avoid contamination of downstream analysis, fMRI-based studies must identify and remove these sources of noise prior to statistical analysis. One common approach is through “scrubbing” or “censoring” of fMRI volumes that are thought to contain high levels of noise. However, existing scrubbing techniques are based on subject head motion measures or ad-hoc measures of signal change. Here, we consider scrubbing through the lens of outlier detection, where volumes containing artifacts are thought of as multidimensional outliers. Robust multivariate outlier detection methods have been proposed using robust distances, which are related to the Mahalanobis distance. These robust distances have a known distribution when the data are i.i.d. Gaussian, and that distribution can be used to determine an appropriate threshold for outliers. However, in the fMRI context, we observe clear violations of the assumptions of Gaussianity and independence. Here, we develop a robust multivariate outlier detection method that is applicable to non-Gaussian data. The objective is to obtain threshold values to flag outlying volumes based on their robust distances. We propose two threshold candidates that embark with the same two steps, but the choice of which depends on a researcher’s purpose (i.e., greater data retention vs. greater sensitivity). The two main steps of our procedure are: (1) dimension reduction and selection to identify primarily artifactual latent directions in the data, (2) robust univariate outlier imputation to remove the influence of outliers from the distribution , and (3) estimating a threshold for outliers based on the upper quantile of the outlier-free distribution of

RD, we propose two threshold choices. The first threshold is an upper quantile (e.g. 99th) of the empirical distribution of robust distances obtained from the imputed data. The second threshold is an estimation of the upper quantile of the robust distances distribution employed by a nonparametrc bootstrap to account for uncertainty in the empirical quantile. We compare our proposed approach with existing approaches for scrubbing in fMRI, including motion scrubbing, of data-driven scrubbing and existing multivariate outlier detection method based on restrictive parametric assumptions.

Acknowledgments

Organizing Committee

Mark Fiecas (Chair), University of Minnesota
Suprateek Kundu, UT MD Anderson Cancer Center
Elizabeth Sweeney, University of Pennsylvania
Simon Vandekar, Vanderbilt University
Caitlin Ward, University of Minnesota
Lin Zhang, University of Minnesota
Tingting Zhang, University of Pittsburgh

Student Paper Competition

Andreas Anastasiou, University of Cyprus
Andrew Brown, Clemson University
Joshua Cape, University of Wisconsin-Madison
Ivor Cribben (Chair), University of Alberta
Ani Eloyan, Brown University
Timothy Johnson, University of Michigan
Hakmook Kang, Vanderbilt University
Jun Young Park, University of Toronto
Elizabeth Sweeney, University of Pennsylvania
Yuping Zhang, University of Connecticut
Jing Zeng, University of Science and Technology of China
Xin Zhang, Florida State University

Staff

Bernadette Gloeb
Sarah Harris
Kevin Loecke
Beth Olson
Mike Sawyer
Megan Schlick

Student Volunteers

Jessica Butts
Souradipto Ghosh Dastidar
Jonathan Kim
Zexi Rao
Kailash Rohit Velusamy
Zilin Wang