IMAGING GENOMICS

HENG HUANG

Department of Electrical and Computer Engineering, University of Pittsburgh
Benedum Hall 3700 O'Hara Street, Pittsburgh, PA 15261
E-mail: heng.huang@pitt.edu

LI SHEN

Center for Neuroimaging, Department of Radiology and Imaging Sciences
Center for Computational Biology and Bioinformatics
Indiana University School of Medicine
355 West 16th Street Suite 4100, Indianapolis, IN 46202
E-mail: shenli@iu.edu

PAUL M. THOMPSON

Imaging Genetics Center, Stevens Institute for Neuroimaging & Informatics
4676 Admiralty Way, Marina del Rey, CA 90232
E-mail: pthomp@usc.edu

KUN HUANG

Department of Medicine, Indiana University School of Medicine
340 W 10th St #6200, Indianapolis, IN 46202
E-mail: kunhuang@iu.edu

JUNZHOU HUANG

Department of Computer Science and Engineering, University of Texas at Arlington
500 UTA Blvd., Arlington, Texas, 76019
E-mail: jzhuang@exchange.uta.edu

LIN YANG

Department of Biomedical Engineering, University of Florida
1275 Center Drive, Biomedical Sciences Building JG-56, Gainesville, FL, 32611
E-mail: lin.yang@bme.ufl.edu

1. Introduction

As an emerging research field, the goal of Imaging Genomics is to study the integrative high-throughput imaging (such as histopathological images in cancer research, MRI and PET images in brain study) and omics (such as SNP, DNA sequence, RNA expression, methylation, epigenetic markers, proteomics, and metabolomics) data, which created new opportunities for exploring relationships between genotypes, phenotypes, and clinical outcomes using quantitative methods. Imaging Genomics research holds great promise for precision medicine to better understand diseases, from genetic and genomic determinants to the complex interplay of phenotypic traits. The unprecedented scale and complexity of the Imaging Genomics data have presented critical computational bottlenecks requiring new biomedical data science tools. The

technological advance in this field is urgently needed and has the potential to significantly contribute to multiple national health priority areas such as the BRAIN Initiative,[1] the Precision Medicine Initiative,[2] and the BIGDATA Initiative.[3]

The objective of this Imaging Genomics Session at PSB 2018 is to encourage discussion on fundamental concepts, novel methods and innovative applications. We hope that this session will become a forum for researchers to exchange ideas, data, and software, in order to speed up the development of innovative technologies for hypothesis testing and data-driven discovery in Imaging Genomics.

2. Overview of Contributions

Our session includes seven accepted papers covering a variety of the subjects in the imaging genomics field. The papers address the imageing genomics research questions from genotype-phenotype association studies of complex brain disorder and cancerous diseases to survival analysis and disease characterization. The computational methods range from the convolutional neural network methods to the low-rank based multi-task learning model. The large-scale imaging and omics data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, The Cancer Genome Atlas (TCGA), the Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) Consortium, and The Cancer Imaging Archive (TCIA) have been analyzed in the accepted papers.

Miller *et al.* analyzed the rare variants from whole-genome sequencing from the ADNI cohort and identified several genes as significantly associated with the imaging phenotype using only synonymous variants that affected codon frequency. Their study showed that the codon bias may play a role in Alzheimer's Disease and that it can be used to improve detection power in rare variant association analysis. Huang *et al.* developed a deep survival learning model to predict patients' survival outcomes by integrating the multi-dimensional TCGA data. In order to take the advantage of both histopathological image information and molecular profiles from imaging-omics data, an integrative pipeline based on deep learning model was created. The unsupervised training and supervised fine tuning processes are combined to conduct survival prediction using a limited number of patient samples. Chidester *et al.* introduced a Discriminative Bag-of-Nuclear-Words (DBoNW) method to predict genomic markers using imaging features, which addresses the challenge of summarizing histopathological images by representing nuclei with learned discriminative nuclear codewords. A reliable patch-based nuclear segmentation scheme using convolutional neural networks was also developed to extract the nuclear features. Huo *et al.* proposed a new multi-task learning model to analyze the associations between single nucleotide polymorphisms (SNPs) and quantitative traits (imaging measures). The low-rank structure in the new multi-task learning model is beneficial to uncover the correlation between genetic variations and imaging phenotypes, such that the candidate genes or loci which is relevant to the biological etiology of the disease can be identified. Adhikari *et al.* developed a multi-site resting state functional MRI (rsfMRI) analysis pipeline to allow research groups around the world to process rsfMRI scans in a harmonized way, to extract consistent and quantitative measurements of connectivity and to perform coordinated statistical tests. The ENIGMA-rsfMRI analysis pipeline was used to ver-

ify and replicate the assertion that there is moderately strong genetic influence on the resting state signal. Han and Kamdar used a bi-directional convolutional recurrent neural network model to predict the methylation state of the MGMT regulatory regions in Glioblastoma Multiforme (GBM) patients via their brain MRI scans collected from TCIA combined with methylation data from TCGA. Finally, Srivastava *et al.* studied the coherent "trans-omics" features that characterize varied clinical cohorts across multiple sources of TCGA data for more descriptive and robust disease characterization. The results showed that the histology images outperformed molecular features while predicting cancer stages, transcriptomics held superior discriminatory power for ER-Status and PAM50 subtypes, and there exist a few cases where all data modalities exhibited comparable performance.

References

1. M. McCarthy. US to launch major brain research initiative. *BMJ*, 346:f2156, 2013.
2. F. S. Collins and H. Varmus. A new initiative on precision medicine. *N Engl J Med*, 372(9):793–5, 2015.
3. L. Ohno-Machado. NIH's Big Data to Knowledge initiative and the advancement of biomedical informatics. *J Am Med Inform Assoc*, 21(2):193, 2014.