Workshop Title: Data Science and AI Applications in Cancer Research

Organizers:

W. Jim Zheng, Ph.D., Professor
School of Biomedical Informatics
University of Texas Health Science Center at Houston
7000 Fannin St | Suite 600 | Houston, Texas, 77030
Phone: (713) 500-3641 | Email: wenjin.j.zheng@uth.tmc.edu

Contact person: W. Jim Zheng (wenjin.j.zheng@uth.tmc.edu)

Organizer's background, prior experience of organizing similar workshops

W. Jim Zheng, Ph.D., MS, joined UTHealth School of Biomedical Informatics in 2013 as an associate professor and associate director of the Center for Computational Biomedicine. Dr. Zheng's research interests focus on integrating, modeling, visualizing, and mining eukaryotic genome information for translational medicine. Dr. Zheng established and directs the Data Science and Informatics Core for Cancer Research (DSICCR). DSICCR is a cutting-edge data science resource with faculty expertise that includes high-throughput genomic data analysis, systems biology, electronic health record mining, and clinical data warehouse development. DSICCR developed a cutting-edge computing infrastructure tailored for data science and AI research by working with fellow faculty, the IT team, and vendors. In addition, DSICCR helped publish over 110 papers and supported obtaining over $58 million in extramural grant awards. This proposed workshop will build on the success of DSICCR to demonstrate how data science and AI can advance cancer research.

Dr. Zheng has been serving on the programming committee of several international conferences, including BIBM and ICIBM, and organized the ISMB Special Session on Computational Methods for Elucidating Nuclear Structure and Dynamics as an organizing committee member in 2012.

*** Topic:** Abstract, theme, the rationale for the workshop, issues to be addressed, goals of the workshop, relevance to ICHI, previous workshops that this proposed workshop is building on (if applicable).

Cancer as a leading cause of death has been studied at all levels, ranging from molecules to cells to individuals and populations. While specific informatics approaches have extensively analyzed data generated at each level, the analysis faces significant challenges, including the enormous volume and heterogeneity (variety) of the data. Furthermore, new technologies applied to cancer research can generate data at extremely high speed. These high volume, high velocity, and high variety data ("big data") pose a significant challenge to conventional data analysis methodologies. However, these challenges can be met by data science and AI that develops novel approaches to analyze big data, providing integrative solutions to analyze cancer data at all levels. This workshop aims to provide informative demonstrations on how data science and AI can be applied to advance cancer research by providing deeper insight into disease mechanisms

and facilitating translation to clinical practice. This goal is perfectly aligned with the scope of ICHI concerning the application of data science and informatics principles to address problems and support research in health, medicine, and life science.

*** Audience:** Target audience, communities that the workshop addresses, anticipated number of attendees.

The targeted audience of this workshop would be: 1) cancer researchers seeking cutting-edge data science and AI methodologies to advance their cancer research; and 2) data science and AI researchers interested in finding new opportunities in cancer research to develop and apply their methodologies. We anticipate 25-100 attendees, given the popularity of the topics and the large local communities in cancer, data science, and AI research at the Texas Medical Center.

*** Publicity:** Measures foreseen to announce the workshop and attract a sufficient number of participants; envisaged follow-up activities after the workshop (if applicable).

The organizer is a member of several professional societies, including AMIA, ACM, ISCB, AACR, and ASHG, and can advertise the workshop through the societies. In addition, the Texas Medical Center has several mailing lists reaching out to a wide range of researchers, including the Gulf Coast Consortium mailing list with seven member institutions. We will advertise our workshop through these mailing lists and the mailing list of individual institutions at the Texas Medical Center.

*** Workshop Structure:** Workshop length (half-day/3 hours or one-day/6 hours), tentative workshop program, and activities are foreseen to foster interaction and communication.

We plan to have a half-day workshop, with one keynote speaker and four workshop presentations on the following data science and AI topics: 1) EHR data mining; 2) Imaging analysis; 3) Multi-omics data analysis; and 4) Cancer Drugs.

*** Reviewing:** Planned reviewing process, program committee (if applicable).

The submitted presentations will be reviewed by 16 participating faculty of DSICCR, as well as DSICCR staff. In addition, we may invite the participating faculty and the collaborators of DSICCR to present their works for the workshop.

*** Call for Papers:** Tentative call for papers with workshop deadlines.

We plan to send out a Call for Papers in January 2023. Then, after review, the decision and invitation will be sent out in March/April 2023.