Dissertation Defense
Doctor of Philosophy in Intelligent Systems Program

“INTERPRETABLE DEEP LEARNING FOR ADVANCING PRECISION ONCOLOGY”
by Shuangxia Ren

Date: July 24, 2024
Time: 3:00 p.m. – 5:00 p.m.
Place: Room 6106 Sennott Square, 210 S. Bouquet Street, Pittsburgh, PA 15260
https://pitt.co1.qualtrics.com/jfe/form/SV_a3RiU792GiD3OJw

Committee:

• Xinghua Lu, Professor, Intelligent Systems Program and Department of Biomedical Informatics, University of Pittsburgh
• Gregory Cooper, Professor, Intelligent systems program and Department of Biomedical Informatics, University of Pittsburgh
• Harry Hochheiser, Associate Professor, Intelligent systems program and Department of Biomedical Informatics, University of Pittsburgh
• Lujia Chen, Assistant Professor, Department of Biomedical Informatics, University of Pittsburgh

Abstract:
The goal of precision oncology is to provide each patient with the most appropriate cancer treatment. This approach involves understanding the impact of genomics on individual cancer cases and utilizing this understanding to tailor cancer therapies targeting the genetic mutations that drive the malignancy. Achieving this level of personalization in cancer treatment can be realized by integrating artificial intelligence (AI) to predict how a patient’s cancer cells will react to anticancer medications based on their genomic data.

This thesis introduces three novel deep learning methods for drug sensitivity prediction: 1) We employed graph regularized matrix factorization to disassemble the drug response matrix into vectors representing cell lines and drugs. Subsequently, we developed mapping functions to transform observed molecular characteristics into the representation vectors for cell lines and drugs. 2) A model was created to learn a mapping function from Somatic Genetic Alterations (SGAs) to gene expression. The hidden representations between these observed datasets served as the features of cell line to predict drug response. 3) An extended version of the Variational Autoencoder (VAE) model was utilized to condense gene expression and SGAs data into a succinct, informative, low-dimensional representation of cell lines. Following the acquisition of these learned representations, they were employed instead of raw omics data for drug sensitivity prediction.

While all three models leverage representation learning for cell lines and subsequently apply the representation for drug sensitivity prediction, notable differences distinguish them. The first model excels in prediction performance, accurately forecasting drug responses for unseen drugs and cell lines but lacks interpretability. Conversely, the second model, though interpretable, falls short in performance due to its reliance on SGAs for predicting drug responses. To achieve a balance between predictive efficacy and interpretability, we developed the third model which takes gene expression and SGAs as input, utilizing a Deep Generative Model (DGM) and self-attention mechanism to enhance interpretability. This approach contributes to the broader evolution of cancer medicine by aligning predictive accuracy with interpretive insights.