A Novel Breast Cancer Recurrence Risk Predictor Using a Deep Learning Multi-omics Data Integration Framework

Rahman, Ariana (School: BASIS Chandler)

Over the last few decades, many breast cancer patients have become disease-free due to the advancements in detection, diagnosis, and treatment. However, cancer can come back for almost 30% of patients who are disease free after initial treatment. Current methods of recurrence risk prediction use clinical or phenotypic characteristics which may not give a complete understanding of the underlying genetic causes of the recurrence. The aim of this study was to utilize multi-omics high-throughput data such as gene expression, copy number variation, mutation, and microRNA, in recurrence rate prediction. For this project, a modified deep learning–based autoencoder structure known as AIME was used to extract accurate integrated embedding. Unlike a typical autoencoder, where the input and output layers are the same, AIME uses two different data types such as gene expression and copy number variation and, at the same time, adjusts for the confounding factors which in this case, was the Estrogen Receptor status. The integrated embedding data from AIME was then applied to a supervised machine learning approach called Random Forest algorithm to classify samples into disease free and tumor regressed categories. As part of the AMIE output, a list of the 25 most significant gene for breast cancer recurrence was identified which could be effectively applied as molecular biomarkers to improve patient diagnostics, develop targeted therapies, and investigate tumor progression and recursion. This approach can be expanded to include other cancer types to predict and understand tumor recursion.

Awards Won:
Third Award of $1,000