Deep, Multimodal Representation Learning for Pan-Cancer Prognosis Prediction

Cheerla, Anika (School: Monta Vista High School)

Estimating the future course of cancer is invaluable to physicians; however, current clinical methods fail to effectively use the vast amount of multimodal data available. To tackle this problem, I constructed a deep neural network model to predict the survival of patients for 33 different cancer types, using gene expressions, miRNA data, clinical data and histopathology images. I developed an unsupervised encoder to compress these four data modalities into a single feature vector for each patient, handling missing data through a resilient, multimodal dropout method. Encoding methods were tailored to each data type - using Dilated DCNNS (Deep Convolutional Neural Networks) to summarize gigapixel-resolution pathology images and using vanilla feedforward networks to extract deep features from genetic and clinical data. I then used these feature encodings to predict survival data, achieving an impressive 0.754 C-index. This research was the first attempt to build a pan-cancer prognosis model - all previous research focused on cancer-specific datasets. Furthermore, my model handles multiple data modalities, efficiently analyzes huge whole-slide images, and summarizes patient details flexibly into an unsupervised, informative profile. I present a powerful automated tool to accurately determine prognosis, a key step towards personalized treatment for cancer patients.

Awards Won:

Samvid Education Foundation: Geno Award of $1000 honoring the literary work of Tamil novelist, Sujatha.