NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

School of Computer Science and Engineering
College of Engineering

Deep Learning Approaches to Predict Drug Responses in Cancer

Using A Multi-Omics Approach

Student: Lyu Xintong **Supervisor:** Prof Jagath Chandana Rajapakse

1. Background

Cancers are genetically heterogeneous, and therefore the same anti-cancer drug may have varying degrees of effectiveness on patients due to their different genetic profiles. Oftentimes it is a trial and error process and patients need to try many different anticancers drugs that are ineffective with significant side effects before finding an effective drug for them.

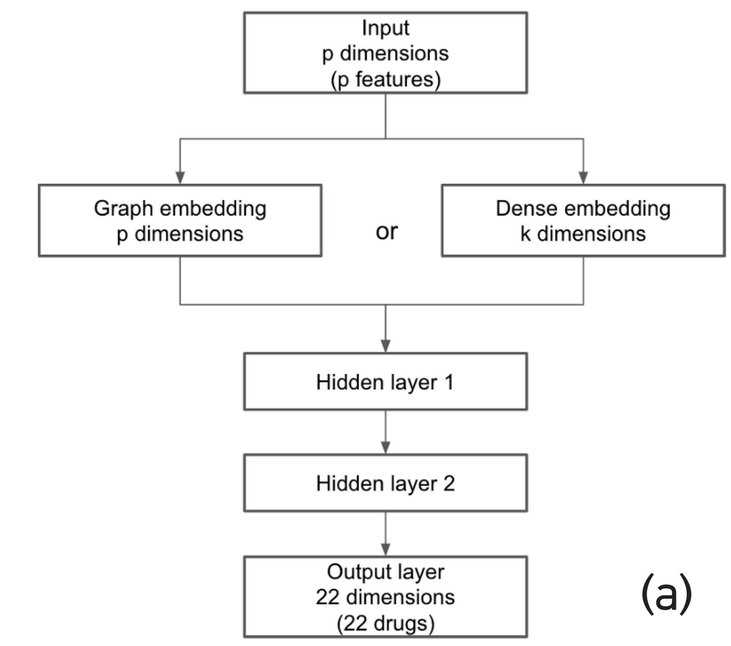
2. Research Motive and Objective

The mechanisms of cancers are also an extremely complex, a multi-omics approach where multiple types of omics data are integrated might provide a more holistic perspective on pharmacogenetics cancer research. The recent development in neuron networks revolutionised bioinformatics and introduced many new techniques to integrate large and complex omics data. The main objective of this project is to use deep learning to predict the response of tumours to different anticancer drugs using a multi-omics approach. So that doctors will be able to take a more customized approach to prescribe anti-cancer drugs that are likely to be more effective.

3. Methodology

5 different omics data are used, including gene expression, gene mutation, gene copy number variation (CNV), metabolomics and reverse phase protein arrays (RPPA). The drug response, measured by the IC50 value is predicted for a list of 22 anticancer drugs. An individual neural network is created for each omics, and a combined neural network is created.

The architecture of individual network is shown in (a). One major challenge was the n<<p issue, which is an inherent problem in genetic data with extremely high dimensionality (p predictors) but is very low in sample size (n samples), and there exist complex interactions and correlations between the predictors. To solve this issue, a graph-embedded network with the gene interaction network as the graph embedding layer is used. The architecture of the combined model is shown in (b). It has an attention layer to combine the output of the 5 individual models.



(a)

(b)

4. Result

The result is shown in table (c) below in terms of MSE of IC50. Another metric that we devised for judging the DNN results is the percentage of times that the predicted most effective drug was in fact the most predicted drug. The chance of randomly choosing a drug that turns out to be the most effective drug is less than 5%. Compared to that, our models performed quite well, scoring 57% for the combined model.

#	Dataset	Model	Train MSE	Test MSE
1	Gene Expression	Graph Embedding DNN	3.0	2.9
2	Gene Mutation	Graph Embedding DNN	1.5	3.6
3	Gene CNV	Graph Embedding DNN	3.3	3.5
4	Metabolomics	Dense Embedding DNN	2.5	3.1
5	RPPA	Dense Embedding DNN	2.3	2.8
6	All omics	Combined DNN with attention	1.8	2.4

(c)