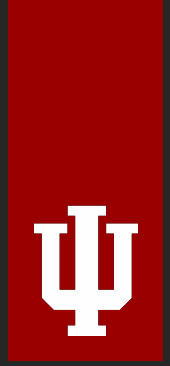


CANCER DRUG RESPONSE PREDICTION

By,
Ruchik D, Sumanth G, Harsha R, Siddhant M

INTRODUCTION

- Drug response prediction is a well-studied problem in which the molecular profile of a given sample is used to predict the effect of a given drug on that sample. Effective solutions to this problem hold the key for precision medicine. In cancer research, genomic data from cell lines are often utilized as features to develop machine learning models predictive of drug response

- Recent advances in artificial intelligence, including machine learning and deep learning, have generated considerable interest in solving classic biomedical problems. Whereas popular applications include disease diagnosis from biomedical images , interpretation of electronic medical records, etc.

- Owing to the significant molecular heterogeneity observed across tumors, there are often many different molecular features and feature combinations that can lead a model to predict particular drug response

- In this study, we aim to use Deep learning models improve Predicting the response of cancer cell to drugs

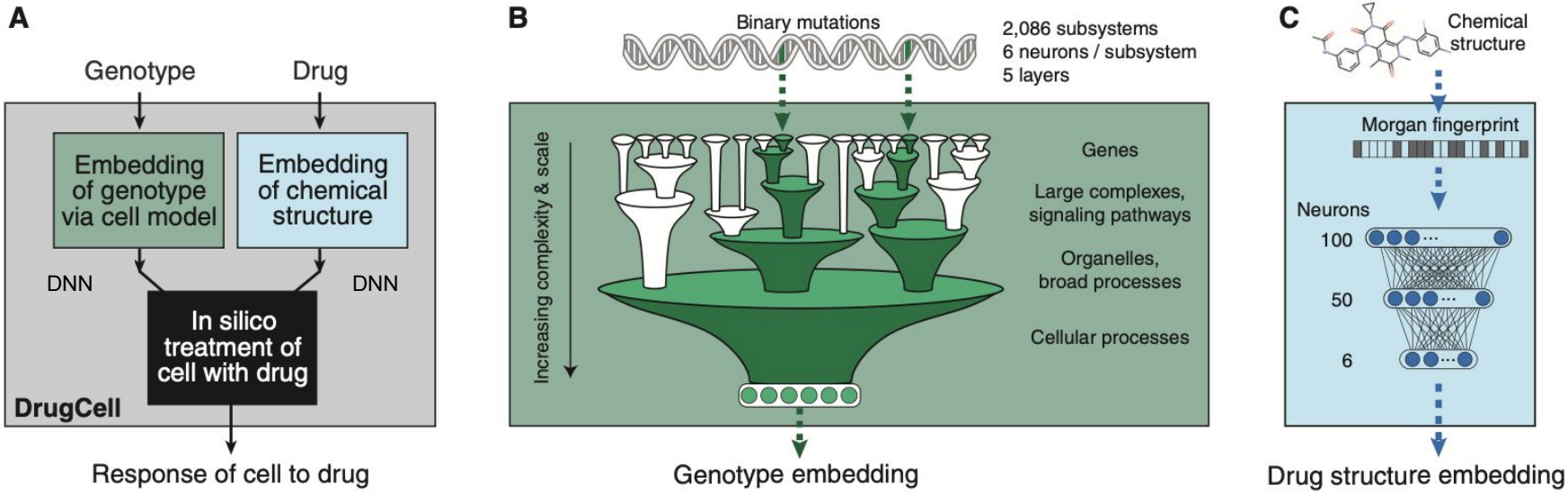
EXPERIMENTAL SETUP

- The datasets used in the project undertaken are obtained from following databases:
  - Cancer Therapeutics Response Portal (CTRP) v2 : It was developed by the Center for the Science of Therapeutics and contains several hundreds of thousands of drug dose-response curves
  - Genomics of Drug Sensitivity in Cancer (GDSC): It is the largest public resource for information on drug sensitivity in cancer cells and molecular markers of drug response.
- The combined dataset consisted of 509,294 cell line drug pairs, covering 684 drugs and 1,235 cell lines.
- The implementation of different models are assessed using the Pearson correlation, which is calculated between predicted and observed values. The higher the correlation, the higher is the prediction accuracies for various drugs. This is further elaborated in the results section.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient

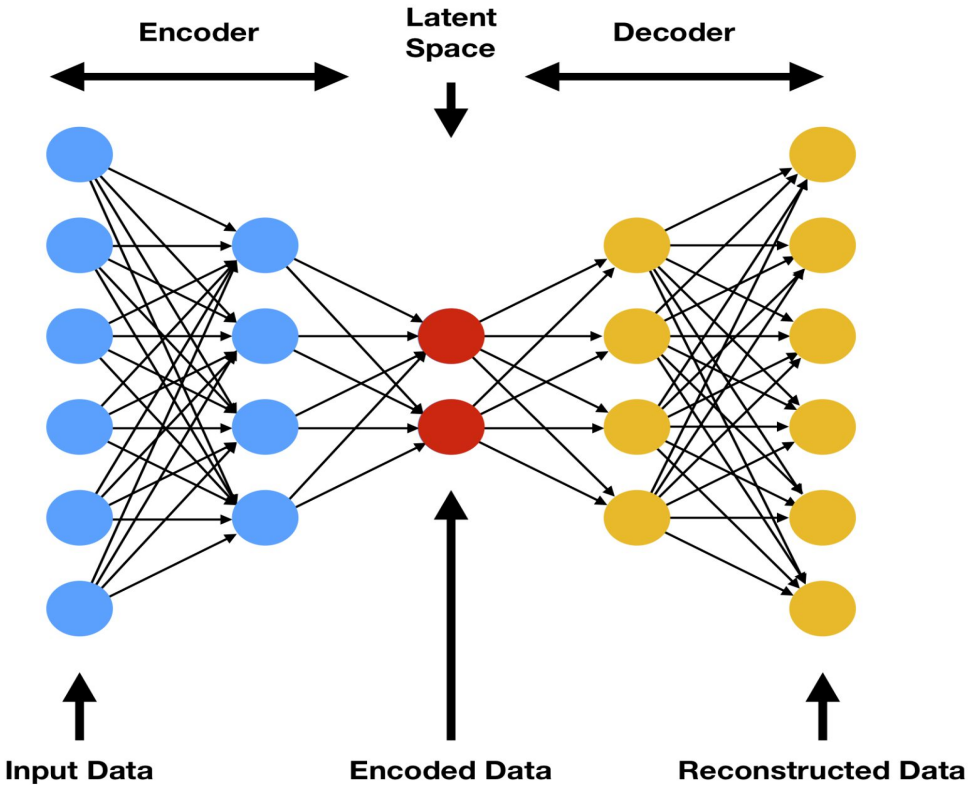
Model Design



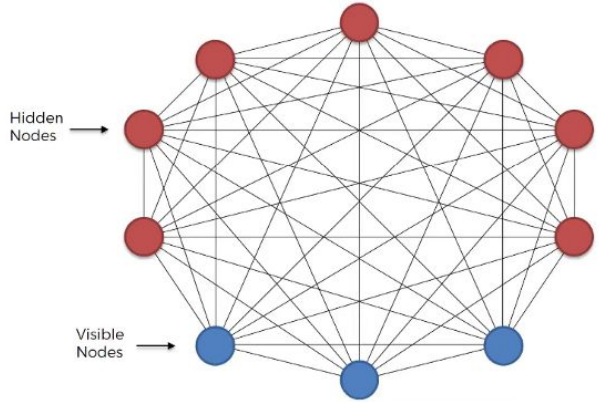
Referred from: https://www.cell.com/cancer-cell/fulltext/S1535-6108(20)30488-8

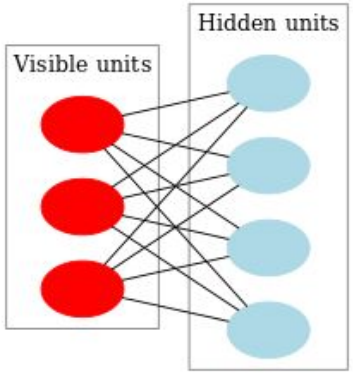
AutoEncoder ARCHITECTURE

- To compress and encode the gene and the drug data we used Autoencoders.
- The network architecture for the autoencoders was a deep neural network with 4 hidden layers.
- We train an autoencoder model on the drug fingerprint and the cell to mutation dataset to encode our train data.



BOLTZMANN ARCHITECTURE





- Boltzmann Machines is an unsupervised DL model in which every node is connected to every other node, even the visible neurons connected to each other and hidden neurons also connected to each other.(Fig 1)

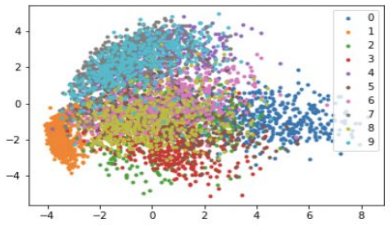
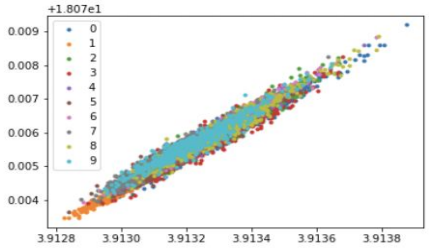
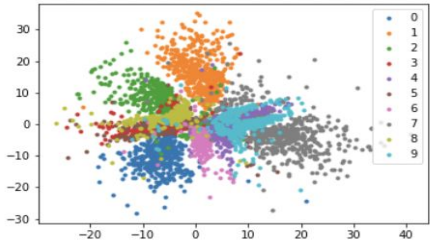
- The training data is fed into the Boltzmann Machine and the weights of the system are adjusted accordingly. Boltzmann machines help us understand abnormalities by learning about the working of the system in normal conditions.

- What makes RBMs different from Boltzmann machines is that visible node isn't connected to each other, and hidden nodes aren't connected with each other. Other than that, RBMs are exactly the same as Boltzmann machines.(Fig 2)

BOLTZMANN ARCHITECTURE

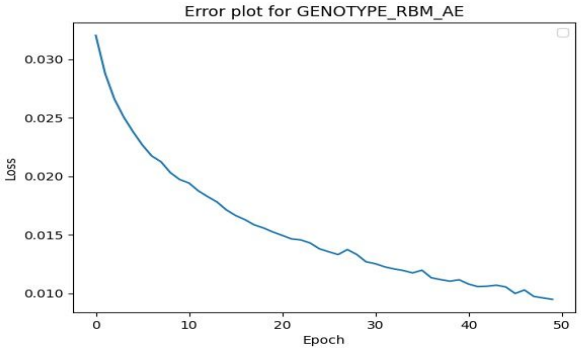
$$E(v,h) = -\sum_{i \in input} b_i v_i - \sum_{j \in features} b_j h_j - \sum_{i,j} v_i h_j w_{i,j}$$

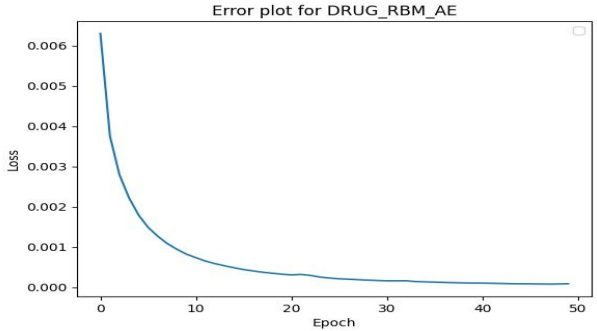
- In RBM, the probability distribution defined by energy of network is matched with the marginal probability of visible units.
- For example :-(input)3008 → 1024 → 512 → 256 → 128 (output)
- Each pair of network layers we will train a RBM and store the weights.
- Initialize the stored weights to the encoder of the Auto-Encoder network.

PCA	DAE Naive Training	DAE Initialized with Pretrained RBMs
		

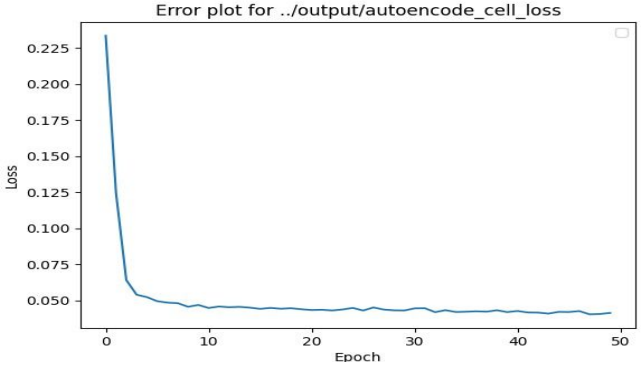
RESULTS

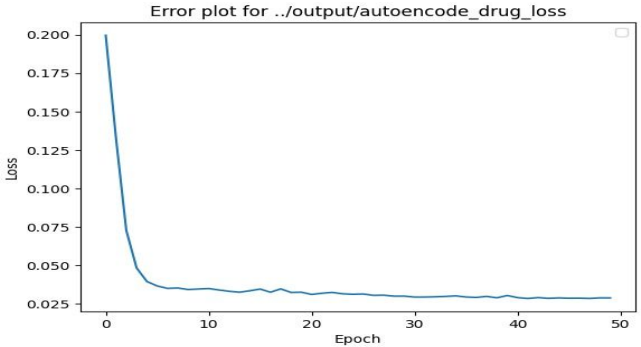
Error Plots for RBM AE





Error Plots for AE with HE initializer





RESULTS

MODELS	AUTOENCODER PEARSON CORRELATION	BOLTZMANN PEARSON CORRELATION
SVR	0.60	0.79
LINEAR REGRESSION	0.45	0.45
RANDOM FOREST	0.72	0.69
MLP	0.45	0.62

RESULTS

Boxplots of the Pearson correlation

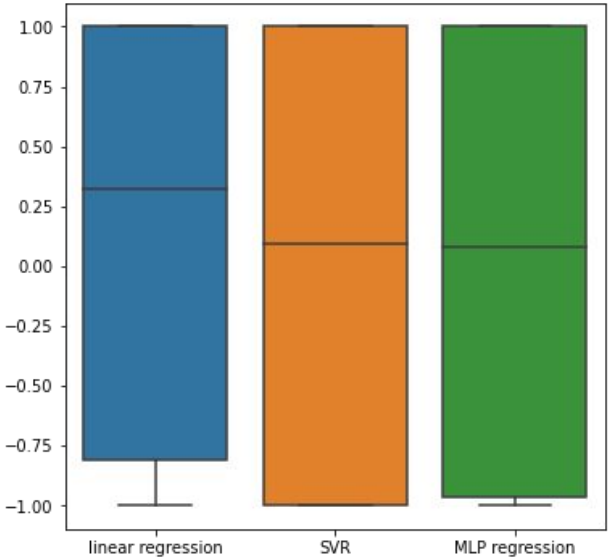


Fig 1 AutoEncoder

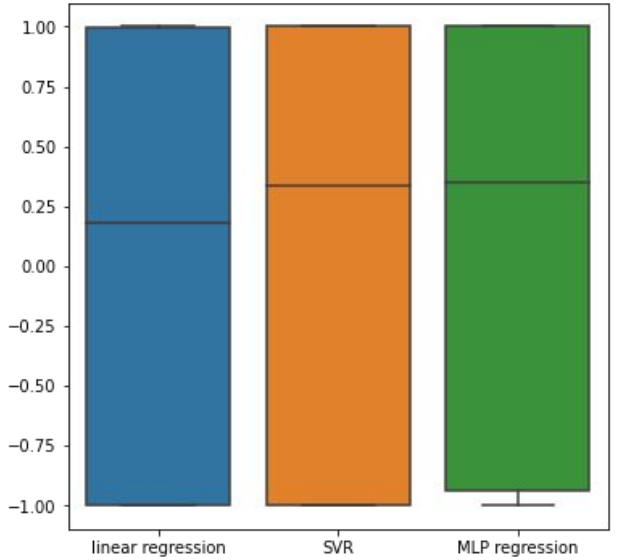


Fig 2 RBM AE

Conclusion

- RBM based AutoEncoder performed better than simple AutoEncoder in drug response prediction.
- DrugCell mainly concentrate on interpretation of model, RBM-AE is lacking of interpretation to Geneotype embedding.
- Defining a hierarchical visible architecture in Restricted Boltzmann Machine AutoEncoder Network.

REFERENCES

- B. M. Kuenzi et al., “Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells”, Vol. 38, Issue 5, pp. 672-684, November 2020, doi: https://doi.org/10.1016/j.ccell.2020.09.014
- G. Adam et al., “Machine learning approaches to drug response prediction: challenges and recent progress”, Vol 4, Issue 19, June 2020, doi: https://doi.org/10.1038/s41698-020-0122-1
- V. Dumoulin et al., “Adversarially Learned Inference”, ICLR, February 2017
- Cancer Therapeutics Response Portal (CTRP) v2, https://pharmacodb.pmgenomics.ca/datasets/2
- Genomics of Drug Sensitivity in Cancer, https://www.cancerrxgene.org/downloads/bulk_download
- Image: https://www.compthree.com/blog/autoencoder/
- Seashore-Ludlow et al., 2015; Yang et al., 2013


Q
A