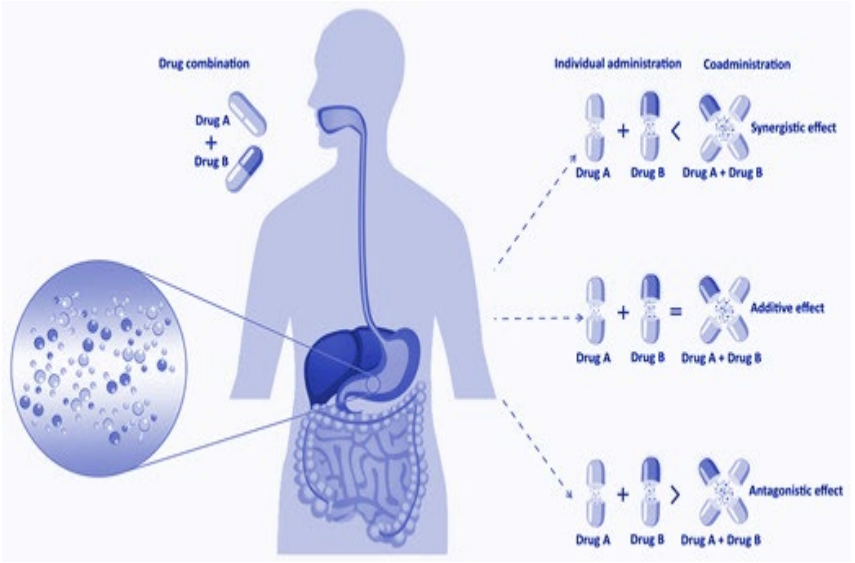
# Cancer Drug Synergy Prediction using Graph Convolutional Neural Networks





**Group 2**

Nitanshu Joshi
Dhanusha Duraiyan
Supriya Bidanta
Farzaneh Behzadnia
Paveethran Swaminathan

# Introduction

Drug A → Specific response for A on cell line

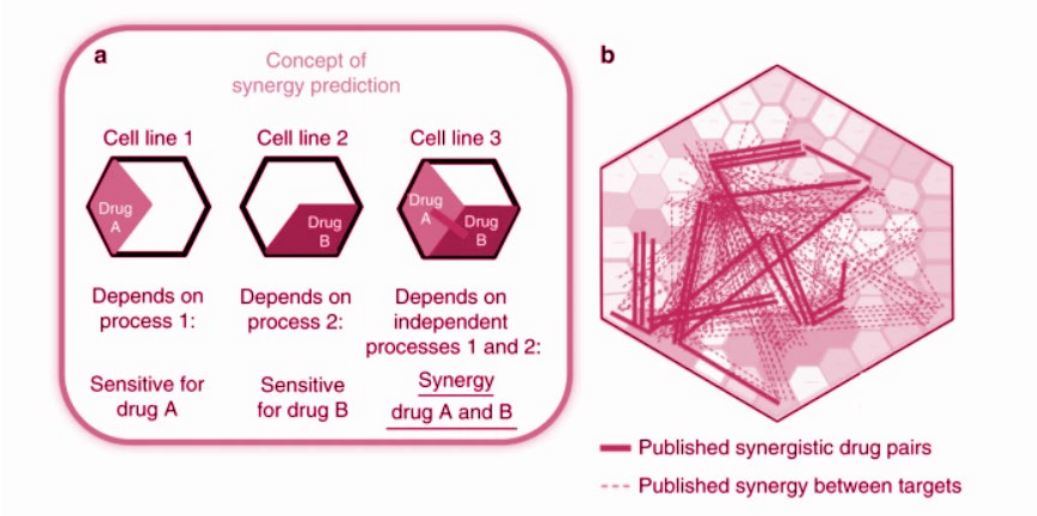
Drug B → Specific response for B on cell line

Synergy of A and B →

Combined response of

A and B on cell line.

*Synergy could be additive or inhibitive.*



# Methods to predict synergy-scores

- Naive method: based on clinical experiences - trial and error method

- High-throughput Screening (HTS)

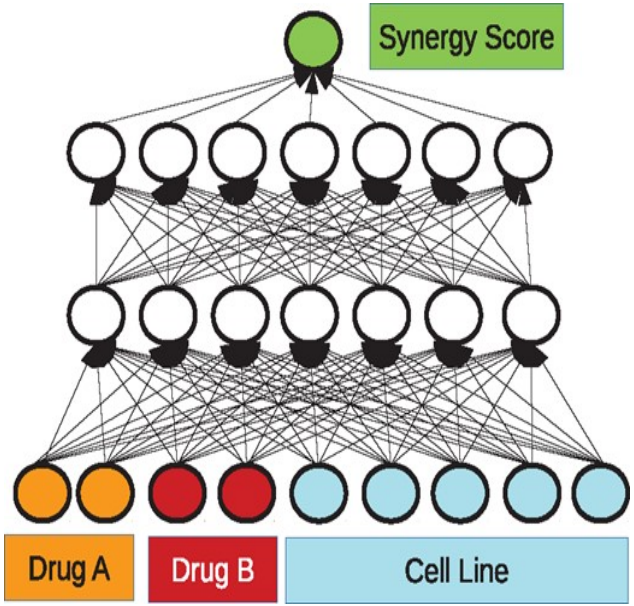
- Machine learning approach

# Primary Research Paper

Finding synergistic drug pairs for a particular cancer type is important for improving efficacy of anticancer treatment.

Each sample:

two compounds + a cell line




[Kristina Preuer, Richard P I Lewis, Sepp Hochreiter, Andreas Bender, Krishna C Bulusu, Günter Klambauer, DeepSynergy: predicting anti-cancer drug synergy with Deep Learning, *Bioinformatics*, Volume 34, Issue 9, 01 May 2018, Pages 1538–1546, https://doi.org/10.1093/bioinformatics/btx806]

# **Data:** Main Drug Data with Synergy Scores

Drug details: http://www.bioinf.jku.at/software/DeepSynergy/

| | Unnamed: 0 | drug_a_name | drug_b_name | cell_line | synergy | fold | drug_a_structure | drug_b_structure |
|---|---|---|---|---|---|---|---|---|
| 0 | 5-FU_ABT-888_A2058 | 5-FU | ABT-888 | A2058 | 7.693530 | 2 | O=c1[nH]cc(F)c(=O)[nH]1 | CC1(c2nc3c(C(N)=O)cccc3[nH]2)CCCN1 |
| 1 | 5-FU_ABT-888_A2780 | 5-FU | ABT-888 | A2780 | 7.778053 | 2 | O=c1[nH]cc(F)c(=O)[nH]1 | CC1(c2nc3c(C(N)=O)cccc3[nH]2)CCCN1 |
| 2 | 5-FU_ABT-888_A375 | 5-FU | ABT-888 | A375 | -1.198505 | 2 | O=c1[nH]cc(F)c(=O)[nH]1 | CC1(c2nc3c(C(N)=O)cccc3[nH]2)CCCN1 |
| 3 | 5-FU_ABT-888_A427 | 5-FU | ABT-888 | A427 | 2.595684 | 2 | O=c1[nH]cc(F)c(=O)[nH]1 | CC1(c2nc3c(C(N)=O)cccc3[nH]2)CCCN1 |
| 4 | 5-FU_ABT-888_CAOV3 | 5-FU | ABT-888 | CAOV3 | -5.139971 | 2 | O=c1[nH]cc(F)c(=O)[nH]1 | CC1(c2nc3c(C(N)=O)cccc3[nH]2)CCCN1 |

| | drug_a_structure | drug_b_structure | cell_line | synergy |
|---|---|---|---|---|
| 0 | Cc1nc(Nc2ncc(C(=O)Nc3c(C)cccc3Cl)s2)cc(N2CCN(C... | Cn1c(=O)n(-c2ccc(C(C)(C)C#N)cc2)c2c3cc(-c4cnc5... | ZR751 | 21.548383 |
| 1 | CS(=O)(=O)CCNCc1ccc(-c2ccc3ncnc(Nc4ccc(OCc5ccc... | COC1CC2CCC(C)C(O)(O2)C(=O)C(=O)N2CCCCC2C(=O)OC... | VCAP | 14.617611 |
| 2 | COc1cccc2c1C(=O)c1c(O)c3c(c(O)c1C2=O)CC(O)(C(=... | NC(=O)c1cccc2cn(-c3ccc(C4CCCNC4)cc3)nc12 | PA1 | -13.653373 |
| 3 | COC12C(COC(N)=O)C3=C(C(=O)C(C)=C(N)C3=O)N1CC1NC12 | CCN(CC)CCNC(=O)c1c(C)[nH]c(C=C2C(=O)Nc3ccc(F)c... | RPMI7951 | -5.641747 |
| 4 | CS(=O)(=O)CCNCc1ccc(-c2ccc3ncnc(Nc4ccc(OCc5ccc... | O=C(O)C1(Cc2cccc(Nc3nccs3)n2)CCC(Oc2cccc(Cl)c2... | LOVO | 28.431355 |

# **Data:** Cancer Cell Line

Cell line: https://sites.broadinstitute.org/ccle/

| | # | #.1 | CellLine | CHL1 | HMCB | HS852T | HS695T | A101D | HS294T | SNU466 | ... | HEL9217 | HEL | UT7 | SET2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | # | # | Tissue | skin | skin | skin | skin | skin | skin | central nervous system | ... | haematopoietic and lymphoid tissue | haematopoietic and lymphoid tissue | haematopoietic and lymphoid tissue | haematopoietic and lymphoid tissue |
| 1 | GeneSym | NaN | GeneID/NA | na | na | na | na | na | na | na | ... | na | na | na | na |
| 2 | LBH | na | 81606 | -0.0 | -0.0 | 0.0 | -0.0 | -0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | GLI2 | na | 2736 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | -0.0 | 0.0 | -0.0 | -0.0 |
| 4 | PAPPA | na | 5069 | 0.0 | -0.0 | 0.0 | -0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 |

5 rows × 1040 columns

# **Data:** Protein-Protein Interactions

PPI: https://snap.stanford.edu/biodata/datasets/10000/10000-PP-Pathways.html

| | Protein1 | Protein2 |
|---|---|---|
| **0** | 1394 | 2778 |
| **1** | 122704 | 54460 |
| **2** | 2597 | 2911 |
| 3 | 4790 | 79155 |
| **4** | 109 | 27115 |
| **5** | 324 | 10982 |
| **6** | 26268 | 6500 |
| **7** | 3609 | 3954 |
| **8** | 152485 | 57504 |
| **9** | 1537 | 55967 |
| **10** | 8452 | 7278 |

# Performance Metrics

## Mean Squared Error (MSE):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

$\mathrm{MSE}$ = mean squared error

$n$ = number of data points

$Y_i$ = observed values

$\hat{Y}_i$ = predicted values

## R2 Score:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \mu)^2}.$$

$$R = sign\left(R^2\right)\sqrt{|R^2|},$$

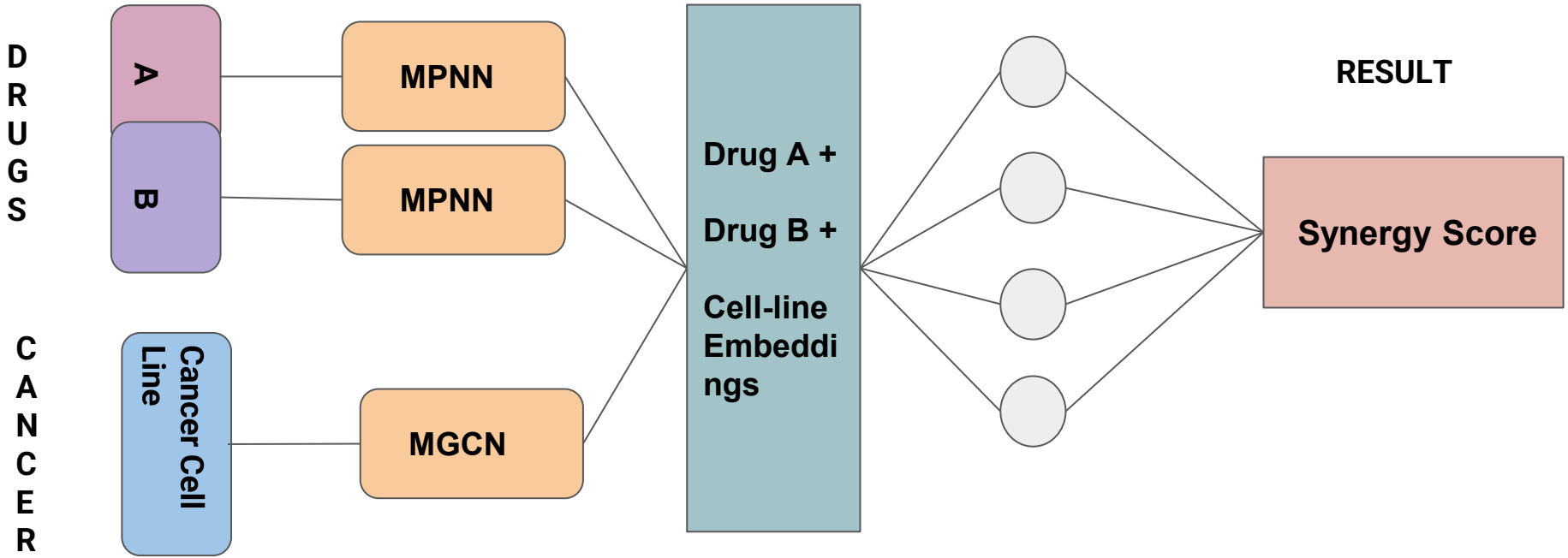
(Image Source: https://emilia-orellana44.medium.com/not-nice-square-error-2d18c248391c)

# Model Architecture

1) Graph Neural Network - Message Passing Neural Network
2) Molecular Graph Convolutional Neural Network

And a feed forward neural network that takes input from these two embeddings and predicts a synergy score.

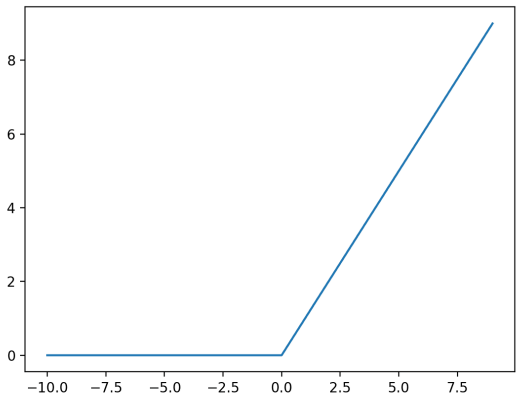
# Flow of Modelling



# Model: Activation Functions

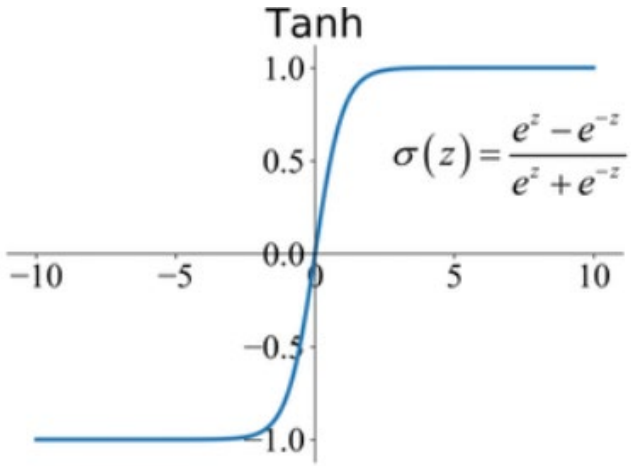
ReLU: (rectified linear activation unit)

f(x) = max(0, x)



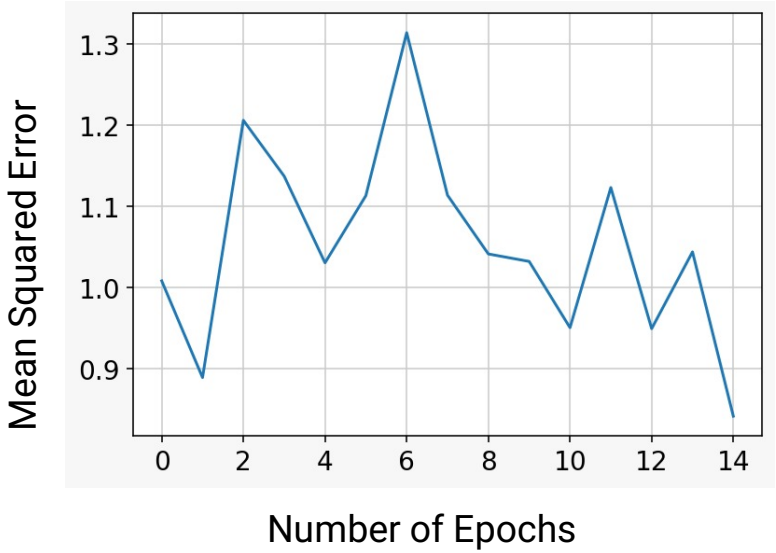

https://machinelearningmastery.com/wp-content/uploads/2018/10/Line-Plot-of-Rectified-Linear-Activation-for-Negative-and-Positive-Inputs.png

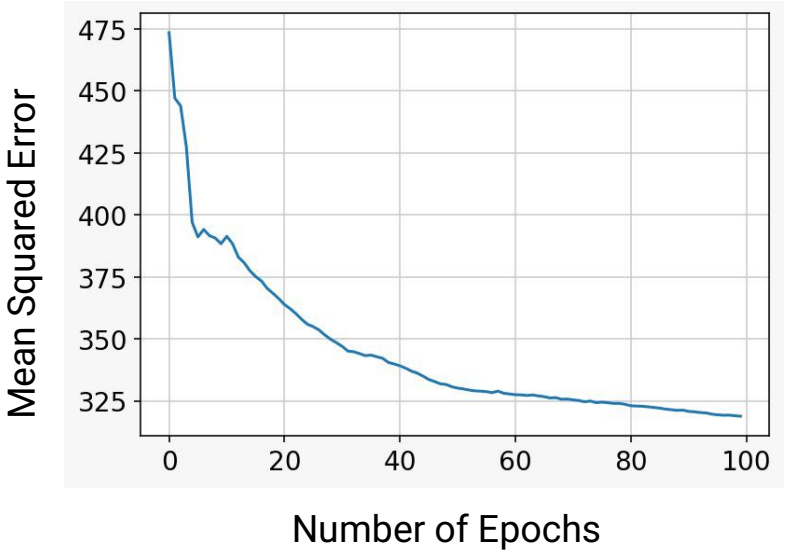
TanH



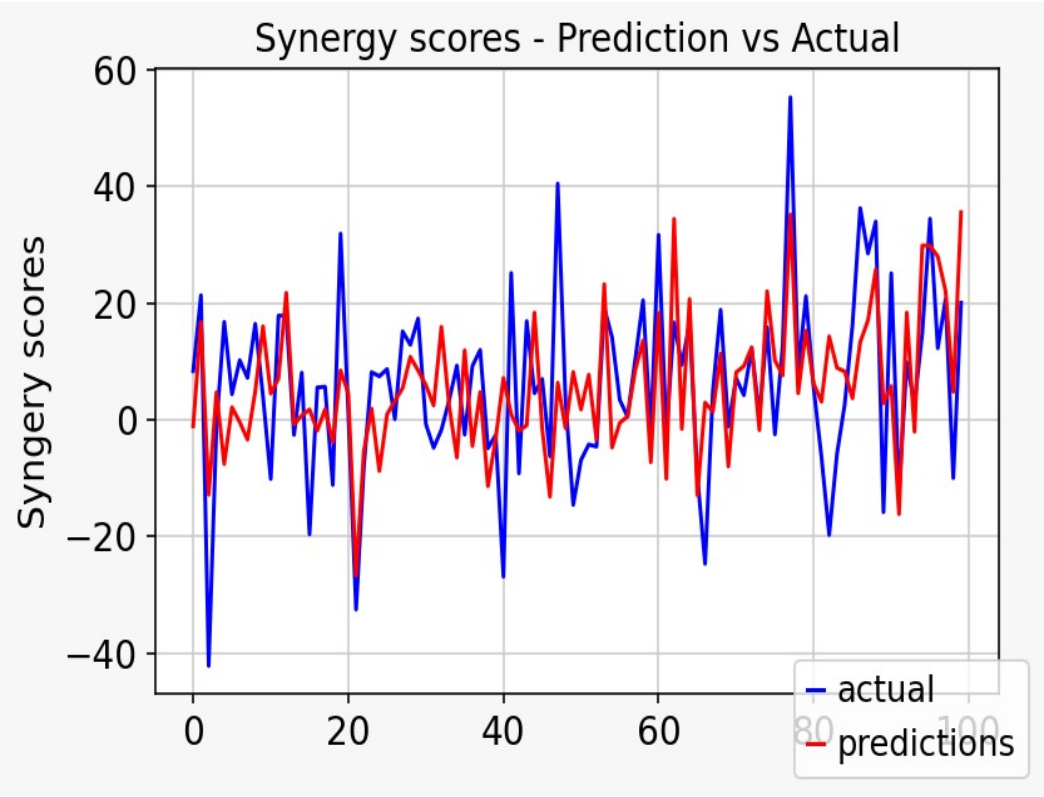

https://www.researchgate.net/figure/Commonly-used-activation-functions-a-Sigmoid-b-Tanh-c-ReLU-and-d-LReLU_fig3_335845675

# Results: Training Curve



# Results: Predictions




**Training (MSE) = 318.786**

**Testing (MSE) = 160.579**

**Testing R-Squared Value = 0.354**

# Conclusion:

1. We have developed a novel graph based deep learning method to predict synergy scores.

1. Since it requires more number of epochs, this method showed difficulty in generalizing on novel drugs-cell line combination when trained on lower parameters. Also, need to look at other activation functions and also decay weights in some epochs.

1. We've demonstrated that this method can also perform decent enough on external unseen test dataset

1. Overall, with more improvement, this could be a valuable tool for selecting drug-combinations for cancer cell-lines.

# Limitations

- Need large number of epochs (at-least 500) to get a good convergence.
- Due to the complexity of the model, it requires too much computational time for training.
- Since the dataset has only a limited number of different drugs and cell lines all methods show difficulties to generalize well across novel drugs and cell lines.
- Evaluating performance of ML models on individual cancer-cell types.
- Since the models are extremely complex, their interpretability is difficult.

# References:


- https://doi.org/10.1093/bib/bbac075
- https://doi.org/10.1158/1535-7163.MCT-15-0843
- https://doi.org/10.3389/fchem.2019.00509
- https://doi.org/10.1093/bioinformatics/btx806
- https://www.nature.com/articles/s41598-021-90923-y
- https://github.com/KristinaPreuer/DeepSynergy
- https://keras.io/examples/graph/mpnn-molecular-graphs/
- https://github.com/tkipf/pygcn
- http://snap.stanford.edu/deepnetbio-ismb/ipynb/Graph+Convolutional+Prediction+of+Protein+Interactions+in+Yeast.html
- https://jhoonline.biomedcentral.com/articles/10.1186/s13045-020-00850-0