Division of Statistics and Machine Learning Department
of Computer and Information Science Linköping
University

# Master thesis project ”Biologically informed graph neural networks for drug effect prediction”

*Background*

More than 40% of patients do not respond to standard treatments in common diseases. Accordingly, methods from personalized medicine aim to identify and propose the best treatments for a patient with given characteristics.

Many advanced methods have been developed for prediction of sensitivity of samples to drugs. Some of these methods, like RefDNN [1], are based on deep learning to provide accurate predictions. However, these models are based on so called bulk data measuring an average expression over many cells, and they do not take into account the knowledge of biological signaling that occurs between cells of different type.
Our research group works with so called Multicellular Disease Models (MCDM) which compute graphs representing biological signals between cells of different type, and these models require data per cell type as an input. The idea of this project is first to transform the bulk data into data per cell type by deconvolution methods, such as [2], then compute one MCDM per patient, and finally find an association of MCDMs to the drug response by means of graph neural networks [3].

*Data*

- GDSC database contains drug sensitivities for various drug and sample combinations [4].

*Research questions*

- Which data deconvolution method is most appropriate to transform bulk data into data per cell type in order to compute MCDMs?
- Which graph neural network method (GNN) methodology can be used for predicting drug sensitivities from the MCDM graphs obtained per patient?
- Is accuracy and certainty of the obtained graph neural network predictions high enough for it to be used in a clinical practice?
- How much do predictions obtained through the MCDM-based GNN models differ from predictions of some published ML model that only uses gene expressions but not the biological information?

*Prerequisites*

- Good knowledge of Machine learning and Statistics
- Good programming skills

*Research Team*

- Oleg Sysoev, STIMA, Linköping University
- Mikael Benson, Karolinska Institutet, Stockholm

*Contact and application*

- Oleg Sysoev, oleg.sysoev@liu.se

*References*

[1] Choi, J., Park, S., & Ahn, J. (2020). RefDNN: a reference drug based neural network for more accurate prediction of anticancer drug resistance. *Scientific reports*, *10*(1), 1-11.
[2] Charytonowicz, Daniel, Rachel Brody, and Robert Sebra. "Interpretable and context-free deconvolution of multi-scale whole transcriptomic data with UniCell deconvolve." *Nature Communications* 14.1 (2023): 1350.
[3] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI open*, *1*, 57-81.
[4] Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., ... & Garnett, M. J. (2012). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, *41*(D1), D955-D961.