Gene Guided Machine Learning
A Path Towards Personalized Breast Cancer Therapy

Kajsa Ledesma Eriksson

Breast cancer is the most prevalent form of cancer and accounts for most deaths caused by cancer in women worldwide. As technology improves, the demand for the most beneficial cancer treatment is increased, where treatment on the patient level is sought after. In today's clinical practice, oncologists can look at which genes are active, or expressed, in the breast cancer tumour to decide which treatment the patient should get. Although, gene expression analysis is very costly and is not available to most breast cancer patients, calling for a more cost-effective and reproducible method. In the following thesis, it is examined if machine learning can be used to get the crucial gene information from actual images of the breast cancer tumour.

For a human eye, it is impossible to say what underlying biological processes that cause a tumour to progress by simply looking at a picture of the tumour. Although, for a computer, this seems to be possible. In my thesis, I have studied whether a machine can learn to compare images of tumours and determine how similar they are on a genetic level.

Let's think of it as two different cakes being ordered from two different bakeries, each having its own special recipe. By looking at the cakes they seem similar, but tasting each one and comparing them to each other, we can get an idea of what ingredients each one is made up of. The same goes for the algorithm. Each tumour has its own recipe; the genes that are active in the tissue sample, and can be compared, not by a human, but by a neural network.

A neural network is a self-learning computing system that, inspired by the neurons in our brains, allows the computer to see structures and information from data; in our case images.

To train the neural network, it was given pairs of images from two different breast cancer tumours as well as information about how active the genes were in the corresponding tumours. By comparing how similar the network predicted the tumour pair to be, and how similar they actually were in the gene expression, the network was guided towards learning to extract information in the images alike the information given by the gene expressions. Compared to the previous analogy, this would be the same as training a taster to be specialized in predicting how similar the recipes of different cakes are. This by letting the taster observe several pairs of cakes and their corresponding recipes.

From the thesis, it was seen that the network could relate the images of breast cancer tumours so that images that were similar, according to the model, actually were expressing the same levels of activity in the genes. This proposes that the gene-guided model could serve as a more cost-effective and reproducible alternative to gene expression analysis.

To put this approach to use, more research needs to be done, but the thesis is proposed as a stepping stone on the path toward a future with more personalized breast cancer therapy.