# Research proposal Lieke

**Investigating the amount of training data**

In the clinical world, I think the biggest limitation of the usage of deep neural networks is the limitation of the amount of (labelled) data so I think it would be interesting and relevant to look into that.

I think it would be a challenge to try to take an existing model and finetune it so that it has an high accuracy with less training data. We can for instance randomly take parts of the data we have and make 3 to 5 (or more or less) different models with each its own fraction of training data that’s included. We can compare the results of the models and talk about the influence of the amount of data available.

**Unsupervised machine learning**

Labelling histological patches is very time consuming and for several biomedical problems there is not a big dataset with histological labelled images. Therefore, I think it would be interesting to look at unsupervised machine learning and compare this with supervised machine learning. We can take an already existing architecture and edit this to two models (supervised and unsupervised) and compare the results.

One drawback is that convolutional neural networks are supervised thus that we cannot use convolutional layers in our algorithm [1]. All the networks proposed by Myrthe are CNNs, therefore, I think this could be complicated but we can certainly look into this.

[1] <https://www.analyticsvidhya.com/blog/2021/05/introduction-to-supervised-deep-learning-algorithms/>

**Already existing CNNS that are available in python:**

* GoogleNet: <https://gist.github.com/joelouismarino/a2ede9ab3928f999575423b9887abd14>
* VGG-16: <https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c>
* ResNet: <https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/>
* Efficientnet: <https://pypi.org/project/efficientnet/>