

## Title: Using neural networks to select what data to label

**Background:** Neural networks can solve incredible difficult problems but in order to attain good results a lot of labelled data is needed. However, in many fields, such as medicine, it can be very expensive to acquire labelled data. Therefore the possibility of training models with little to none labelled data is a field of interest.

**Motivation:** Labelled data can be very expensive to attain, as it requires professionals and experts to look at data points to verify what these are. Active Learning tries to combat this by iteratively selecting which data points offer the best opportunity to improve the model, if these were labelled and added to the training data.

**Problem:** In this project we will improve the accuracy of a Deep learning model by implementing Active Learning. We will do this by training our model on unlabelled data and then use uncertainty sampling in multiple iterations to improve the accuracy. Furthermore we will try to beat the baseline of the model by using one of the following implementation methods:

- Use Bayesian neural networks to improve uncertainty estimates
- Use a generative model to find insightful data points
- Use entropy sampling to measure the impurity or randomness of the output

### Milestones:

11-11-24: Submit synoosis and read references

18-11-24: Make the training code run and train baseline model

25-11-24: Implement one method and compare results to baseline

02-12-24: Improve model and start on report

09-12-24: Gather results and present them on a poster

10-12-24: Poster presentation of results

17-12-24: Finish report

21-12-24: Report hand in

### References:

Ren, Pengzhen, et al. "A survey of deep active learning." *ACM computing surveys (CSUR)* 54.9 (2021): 1-40.