### **02460 Advanced Machine Learning**

This document describes a project to be carried out during the course 02460. The project will be carried out by the following fantastic group of people:

| **Name** | **DTU Student ID** | **E-mail** |
| --- | --- | --- |
| Sophie Kristine Vilter  Angelos Daglaroglou | s212403  s202915 | s212403@student.dtu.dk  s202915@student.dtu.dk |

**Brief project description (max. 10 lines):**

*The goal of our project is to use a Bayesian Neural Network to classify the body part present in an X-ray image, and to be able to explain how our network made this classification. We will analyze our network’s decision making process by obtaining posterior distributions for the network’s weights, identifying the regions of the X-ray our network considers relevant in making this classification, and evaluating the certainty of the relevant features. A convolutional neural net will be built and translated into a Bayesian Neural Network using Laplace Approximation. The Gradient explanation method will be used to attribute relevance to features of the X-ray images. Uncertainty analysis will be performed on the strength of these relevancies using the Union and Intersection Explanation proposed in the paper “Explaining Bayesian Neural Networks”.*

**Learning objectives: By the end of this project, we want to be able to…**

* Recall how a Convolutional Neural Network functions by building one in Pytorch for image classification with multiple layers and high accuracy (above 75%)
* Describe how the Gradient explanation method mathematically detects relevant X-ray image features compared to other explanation methods
* Apply Bayesian inference on the CNN weights by using the ‘laplace’ package in Python
* Determine which image features our model found most relevant when classifying bone X-rays by using the Gradient explanation method
* Apply the Union and Intersection Explanation to evaluate feature relevance strength by following the method outlined in the paper “Explaining Bayesian Neural Networks”
* Evaluate how incorporating uncertainty analysis to feature revelancey enhances understanding of our BNN’s decision making process
* Argue why an explanation of how a Bayesian Neural Network makes decisions is important, and how this explanation can improve the interpretation a BNN’s results

Works Cited

Bykov, Kirill, et al. "Explaining bayesian neural networks." *arXiv preprint arXiv:2108.10346*

(2021).