**Paper selection group 14**

Towards radiologist-level cancer risk assessment in CT lung screening using deep learning. Trajanovski S, Mavroeidis D, Swisher CL, Gebre BG, Veeling BS, Wiemker R et al. *Computerized Medical Imaging and Graphics.* 2021 Mar; 90: 1010883.

Gives a summary of the used (Machine Learning) methodology and evaluation metrics (300 words)

The machine learning framework used in this work consisted of two steps. First two nodule detectors localized all the nodules in the scans. These two were based on hierarchical support vector machines and on semantic segmentation by a deep neural network. Around each nodule a cube was segmented and of each cube three different two-dimensional projections were extracted. This data augmentation was performed to reduce overfitting of the model to the nodule images.

Then a neural network based on ResNet estimated the cancer risk based on the ten largest nodules. ResNet (residual neural network) is a neural network that uses skip connections in order to avoid the vanishing gradient problem. This neural network was deep and wide, since it had a high number of both layers and inputs. Next to nodule localization, additional features such as nodule size and confidence score of the nodule detector were added as inputs to the deep neural network.

In this research four different datasets were combined which gave a total of 8598 scans. Of these 4807 scans were used as training set, and the remaining 3791 scans were used as test set. Data overlaps between the training and the test sets were removed from the test sets. For all the cancer cases in the datasets the diagnosis was confirmed with a test such as a biopsy, but such a ground truth was not available for the non-cancer cases. The diagnosis of a patient was used to label all the nodule images of this patient, so the data was weakly labelled.

The framework was evaluated in three different ways. Its robustness was evaluated by testing the model performance on different datasets. This model performance was reported by the AUC (area under the curve). The performance of the model was tested against the performance of radiologists and against the performance of currently existing models. These results were visualized in a ROC (receiver operating characteristic) curve.