**Afbeelding met Lettertype, tekst, Graphics, grafische vormgeving

Automatisch gegenereerde beschrijving**Reading assignment – Group 5

The reviewed paper was by N. Wu et al., (2020). Deep Neural Networks Improve Radiologists’ Performance in Breast Cancer Screening. IEEE Transactions on Medical Imaging, vol. 39, no. 4, pp. 1184-1194, doi:10.1109/TMI.2019.2945514.

The paper introduces a neural network that is built to classify breast cancer on the basis of mammography images. It evaluates different techniques for implementing such a neural network to reach the best Area Under the Curve (AUC). This paper aims to develop a model that can be used by a radiologist to assist in predicting breast cancer on image data from a screening mammography. The idea is that with this model, breast cancer will be easier to detect at screening, making follow-up investigations like a needle biopsy more informed and only necessary for those who actually need it.

Various machine learning methods are used to acquire the best predictive capacities. The basis is a convolutional neural network (CNN), where multiple structures of the network are tested using different ways of connecting the different screening mammography views. Average pooling is used to combine the information in the right and left breast images before being sent to two fully connected neural layers. The softmax function is used to get probability densities for 8 classes, labelled as; right/left side and benign/malignant which can be positive/negative. These are averaged across the relevant images (right for right etc.) and on the basis of this, a classification is made. The ResNet-based 22-layer networks are used as the layers. The model was optimised using Adam optimization where the predictions are optimised per view (image) by minimising the binary cross-entropy. The model weights were regularized using L2-regularization. Early-stop is used when the AUC does not improve after 20 epochs where every epoch a new selection of the training data is shown to the model.

Apart from training a neural network on the full images, patches of the mammogram are generated where clinicians have produced labels on a pixel level predicting the presence or absence of malignant and benign findings in a given patch. Two heatmaps are generated which highlight areas with higher probabilities of finding benign and malignant elements for each pixel. These heatmaps are used as an input in the breast-level neural network, which allows the main classifier to benefit from the pixel-level information while not needing to repeat the computation of these pixel probabilities each time.

Additionally, transfer learning is used. The weights used in the main classifier are initialised using the weights generated in another model (BI-RADS model) This leads to a better initialisation of the network and thus enhances the learning rate, which is important as only a relatively small number of biopsied training data is available. Of course, the ensemble technique for the model is also used during model training to improve the predictive capabilities.

With regard to the evaluation metric, the AUC and the precision-recall AUC (PRAUC) are used to evaluate the model.

Many machine learning methods are thus combined to provide a robust model with strong predictive capabilities. The transfer learning of the weights on a more noisy BI-RADS classification task is definitely a strong point that adds to the strength of the model. The use of pixel-level mammograms to generate more input (heatmaps) for the CNN is another addition that shows the advances this paper brings to the table. The AUC is used as an evaluation criterion, which is very commonly used for these types of classification tasks. Adding the PRAUC, a good summary of the classifier’s performance is given.

As with most complex neural networks, there are many hyperparameters to be set/tuned. The choices for these can have significant impact on the outcome of the model. Another thing to note is that because there are so few images with a biopsy attached, the training process was altered. Every training epoch all the images where a biopsy was done are shown and an equal number randomly selected from the remaining training set. So the model sees some images more often than others. This could affect model performance on a validation set. Due to the fact that the model was shown all exams with biopsies in the training set, the model can be made more robust by getting more exams with biopsies, since the model does not get to see this data each training set.

This model could be further improved with the use of Generative Adversarial Networks (GANs). GANs can generate realistic mammography images, serving as an effective data augmentation technique. By enlarging the dataset, the model’s ability to generalise to various imaging conditions can be improved. This can also be done by varying pathological conditions, so the model gets provided with a more diverse set of training examples. Even a step further would be the use of a conditional GAN (cGAN), where the class label information is incorporated. Encouraging the generator to produce not only realistic mammography images but also correspond to the specified class. And the discriminator learns to distinguish between real and fake samples while considering the associated class information.