UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Mammography-based Breast Cancer Risk Prediction

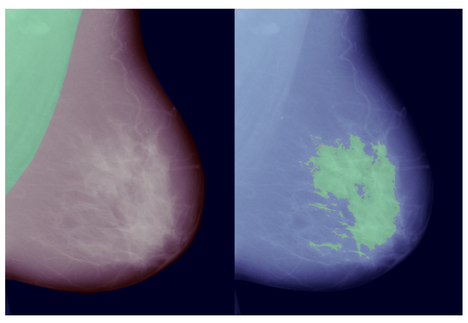
Andreas D. Lauritzen

Mammographies contain non-trivial information related to the lifetime risk of breast cancer including mammographic density and texture patterns. Deep learning methods enable precise and quick quantization of such risk factors and can improve breast cancer risk predictions. Due to the easy collection of large screening cohorts, hierarchical features of mammography can be learned in an unsupervised fashion using autoencoders.
Generative adversarial networks have been shown to improve medical image tasks such as data augmentation, image segmentation in several image medical modalities. Here, a strategy to incorporate generative models into breast cancer risk prediction is outlined.

Introduction

Breast cancer is the most commonly diagnosed cancer among women in Denmark with around 4,700 new diagnoses each year.
Correct diagnosis and early detection of cancerous tissue are critical for treatment and mortality rates. Consequently, Women in Denmark between the age of 50 and 69 are offered breast cancer screening which entails biyearly examinations to detect early signs of breast cancer using a low-dose X-ray image called mammography.

Mammographies contain biomarkers and crucial important information about lifetime risk of developing breast cancer such as fibro-glandular (dense) tissue volume and spatial distribution of parenchymal tissue also referred to as texture. Such information is visually non-trivial and difficult to characterize by human readers, but recent developments in machine learning allow algorithms to learn such features automatically.



(a) (b)

Fig 1. Example of mammography segmentation masks annotated by radiologist. Depicted are (a) breast tissue/pectoral muscle segmentation and (b) dense tissue segmentation.

The Percentage mammographic density (PMD) is defined as the percentage of breast tissue area that is classified as dense tissue, not including the pectoral muscle area.

Examples of texture patterns could be co-occurrence features, coarseness, structural features, variations in gray level and so on. These features were previously manually designed. Present-day texture risk scoring approaches, usually based on deep learning, solve this task by learning features of the given domain that contribute optimally to the best segregation of cancer cases and non-cancer cases.

Thus primary tasks when estimating risk values entail

i) segmenting pectoral, breast, and background
ii) segmenting dense tissue
iii) texture-based risk score prediction

Personalized risk profiles are created by combining PMD, texture risk score, age, hereditary and lifestyle risk factors into a single risk model.

Methods

Breast/Pectoral Segmentation

A 5-layer CNN is used to process mammographic patches classifying the center pixel as either breast tissue, pectoral muscle or background when trained.

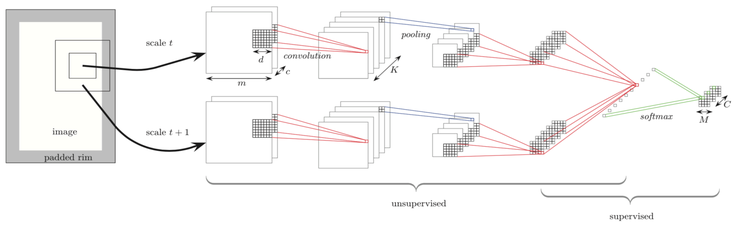


Fig 2. 5-layer CNN. First three layers are trained unsupervised while the last layers are trained separately. This final stage of finetuning is task-specific (segmentation of risk prediction). Input to the network is multi-scale patches samples from a Gaussian scale-space.

The unsupervised part of the network learns features of the image domain independent of the specific task at hand and is trained layer-wise using autoencoders that encode the input data to a sparse overcomplete representation and decodes it again. The loss is computed as the sum of the mean squared error and two sparsity terms.

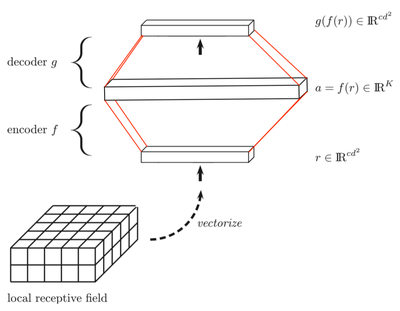


Fig 3. Reconstruction of local receptive field in an overcomplete autoencoder. To prevent overfitting population sparsity and lifetime sparsity are applied, limiting the number of active units per example and number of examples for which a specific unit is active.

The last two layers are trained in a separate supervised stage using the breast/pectoral masks as labels. During this stage, the previous layers are locked. An entire image is processed by a sliding window, predicting each pixel individually.

Dense Tissue Segmentation

For dense tissue segmentation, a modified U-net type network is trained on patches extracted from mammographies using weighted cross-entropy loss.

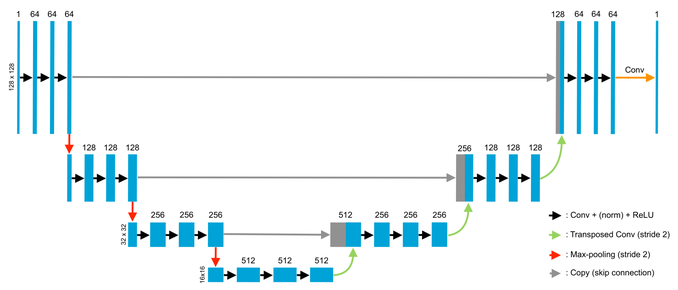


Fig 4. U-Net used for dense tissue segmentation. The U-net preserves high-frequency information supported by skip-connections while learning an under complete representation of the input images enforced by max-pooling operations.

During prediction of a whole image, overlapping patches are extracted and predicted separately and merged with adjacent patches to form a complete segmentation mask.

Alternatively, the previously mentioned 5-layer CNN can successfully be trained for dense tissue segmentation similarly as breast/pectoral segmentation. However, the U-Net is substantially faster when processing higher resolution images.

Texture Risk Scoring

For texture risk scoring, the pre-trained 5-layer CNN is used. The weights of the final two layers are trained to separate between patches from breasts without cancer and patches from with cancer using cross-entropy as loss. During training, only contralateral mammograms were used to remove signs of cancerous tissue or tumors.

Ultimately, the network predicts the probability that a patch comes from a breast with cancer-prone mammographic texture/structure.
For a whole image 500 patches are extracted and the final probability is computed as the mean of the 500 prediction probabilities.

Data

To train and evaluate segmentation and scoring performance two datasets are used. Both have been collected from the Dutch breast cancer screening program between 2003 and 2012.

Segmentation dataset

493 mammograms (MLO and CC view) of healthy women annotated by Danish radiologist using a polygon tool and Cumulus-thresholding for dense tissue segmentation.

Texture risk scoring dataset

1576 mammograms (MLO view only) with 394 cancers, and 1182 healthy controls matched on age and acquisition date.

Results

To train and evaluate segmentation and texture risk scoring performance two datasets are used. Both have been collected from the Dutch breast cancer screening program between 2003 and 2012.

Segmentation Tasks

Breast/pectoral class	DICE
Background	0.99 ± 0.002
Pectoral muscle	0.94 ± 0.069
Breast tissue	0.99 ± 0.008

Dense class	DICE
Non-dense	0.63 ± 0.190
Dense tissue	0.95 ± 0.080

Dataset	AUC
Pearson correlation between PMD estimated with CNN and radiologist	0.93 (0.92 – 0.94)
PMD cancer cases	0.19 ± 0.11
PMD controls	0.15 ± 0.11
AUC using PMD as biomarker for cancer	0.59 (0.56 – 0.62)

Texture Risk Scoring and Validation in cohort study

Texture risk scoring performance is measured in by the area under the ROC-curve (AUC) and validated using 5-fold cross validation.
Furthermore, the texture scoring have been validated in a separate Dutch screening cohort study using 51400 mammograms.

Dataset	AUC
Texture risk scoring dataset (5-fold CV)	0.61 (0.57-0.66)
Screening cohort	0.61 (0.57–0.64)

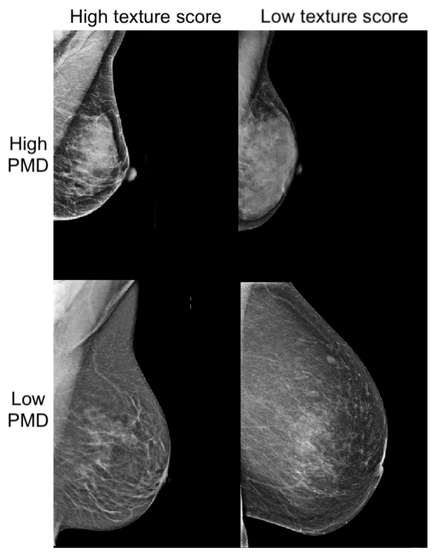


Fig 5. High and low texture and high and low density mammogram combinations.

Top row:
High PMD (14%) and high texture (0.58). High PMD (19%) and low texture (0.47) scores

Bottom row:
Low PMD (5%) and high texture (0.51) and low PMD (4%) and low texture scores (0.44)

Future work

Improving segmentation performance with cGAN

A conditional generative adversarial network (cGAN) consists of two sub-networks. The objective of the generator network is to learn a segmentation mask by approximating the ground-truth mask while producing a segmentation mask that is indistinguishable from the ground-truth mas as seen by the discriminator network. The objective of the discriminator is to learn the classification of a mask being a generated mask or real mask (sampled from ground truth domain).

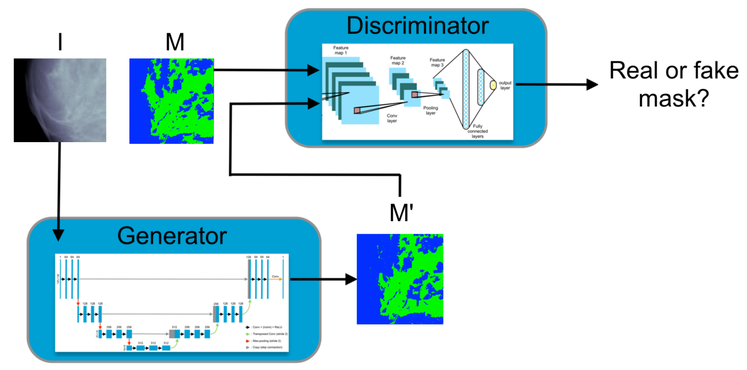


Fig 6. Forward pass through a conditional generative adversarial network. I is a mammographic patch M is the ground-truth segmentation mask and M´ is the generated segmentation mask.

Augmenting Cancer Cases Using GAN

In cohort breast screening datasets controls are often oversampled comparing to cancer cases. Furthermore, gathering cancer labels for large datasets is often difficult. Augmenting cancer cases might be feasible in terms of producing better risk scores.

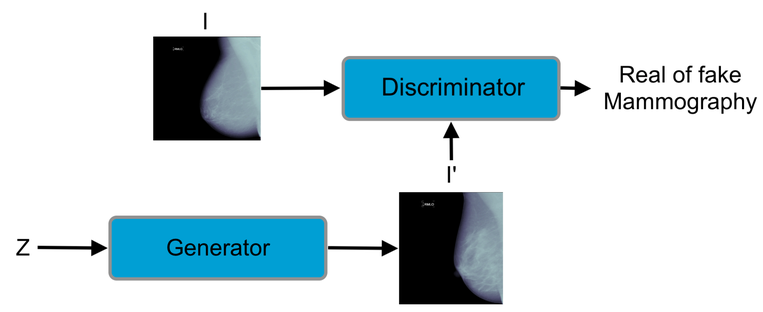


Fig 6. Forward pass through a generative adversarial network. I is a mammography, I´ is synthetic mammography, and Z is a random variable vector.

Conclusion

Here is presented an unsupervised feature learning method for breast/pectoral segmentation and automatic texture scoring and using a supervised approach to segment dense tissues in mammographies.

The results suggest that breast/pectoral muscle segmentation yield overall high DICE similarity, that is satisfactory for calculating PMD.

The results achieved in dense tissue scoring highly correlates with radiologist annotated segmentation masks, yet the DICE similarity suggests that the segmentation performance could be improved significantly.

Texture risk scores estimated with a 5-layer CNN was shown to be related to breast cancer risk. This automatic scoring was validated on a large screening cohort with the same results.

Finally, several approaches of how to improve breast cancer risk modeling using generative models have been proposed including using a cGAN for improving segmentation and a GAN for augmenting dataset with more cancer cases. These propositions will be subject for experimentation in the near future.

References

1. J. O. P. Wanders et al. The combined effect of mammographic texture and density on breast cancer risk: a cohort study. *Breast Cancer Research*. **2018,** *20,* 36.

2. M. Kallenberg et al. Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring. IEEE Transactions on Medical Imaging. **2016,** *35,* 1322-1331.

3. K. Petersen et al. Breast Tissue Segmentation and Mammographic Risk Scoring Using Deep Learning. International Workshop on Digital Mammography. **2014,** 8539, 88-94.