B-SCREEN: Bayesian Decision Support in Medical Screening

Perry Groot[1], Peter Lucas[1], Nivea de Carvalho Ferreira[1]
Nico Karssemeijer[2], Maurice Samulski[2], Marina Velikova[2]

[1]Institute for Computing and Information Sciences
Radboud University Nijmegen
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands
{perry,peterl}@cs.ru.nl

[2]Department of Radiology, UMC St. Radboud
Postbus 9101, 6500 HB Nijmegen, The Netherlands
{n.karssemeijer,m.samulski}@rad.umcn.nl

Radboud University Nijmegen

UMC St Radboud

NWO Netherlands Organisation for Scientific Research

Background

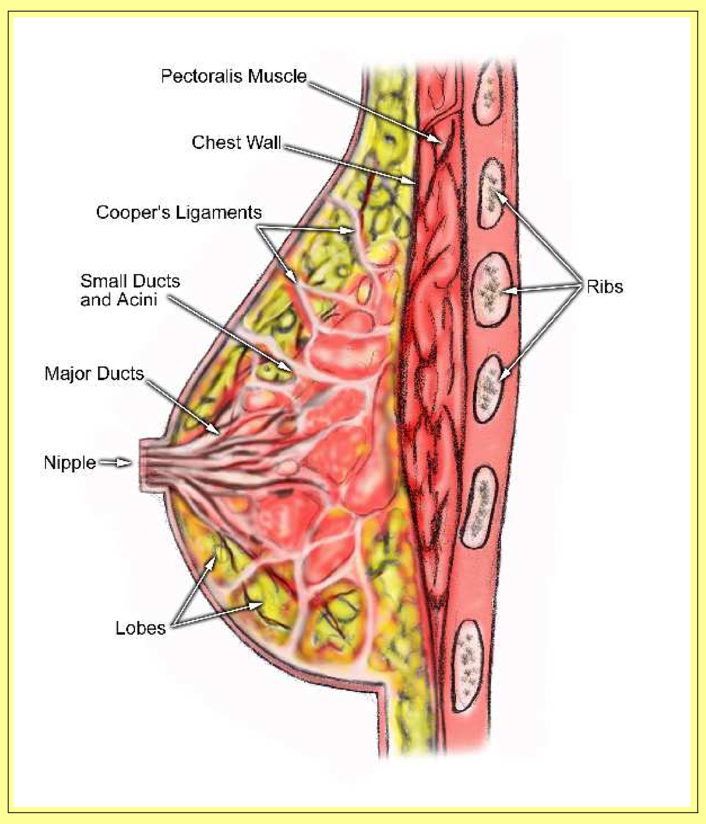
The availability of huge data sets may lead to important changes in health care by mining those data for the development of decision support sytems. Probabilistic grapical models are considered appropriate tools for data mining. Soon, the digitization of the Dutch breast cancer screening will start giving a unique opportunity to mine these data for the development of decision support systems in Radiology.

Aims

The further development of CAD technology using Bayesian Networks to address the problem of interpretation failures by radiologists.

- Develop novel classification methods such that:
  - Medical knowledge can be incorporated
  - Classifiers are faithful w.r.t. the data
- Develop new image representation techniques
- Extend mammographic breast cancer data sets
- Determine experimentally how to use CAD as a decision aid in a practical context

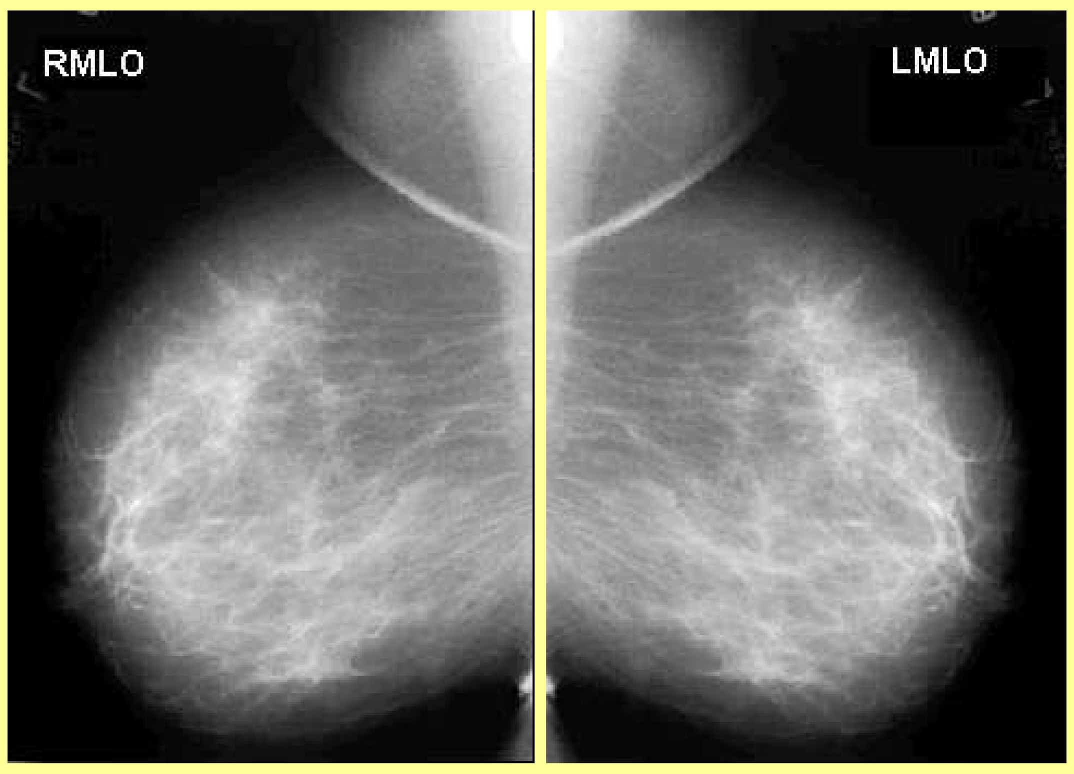
The Breast



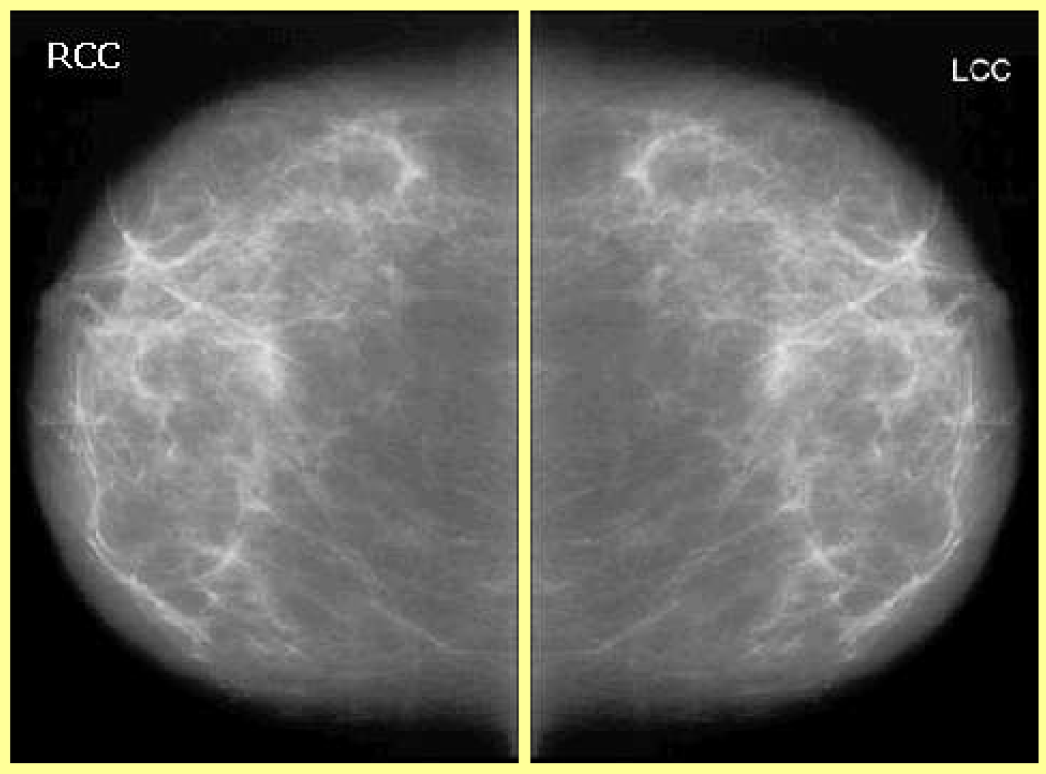
The female breast is mainly made up of lobules (milk-producing glands), ducts (milk passages that connect the lobules and the nipple), fatty and connective tissue surrounding the ducts and lobules, blood and lymphatic vessels.

Breast cancer is the result of uncontrolled growth of breast cells. Most breast cancers originate in the cells of the ducts or of the lobules. The early stage of ductal cancer is referred to in-situ, i.e., the cancer remains refined to the ducts. When the cancer invades the surrounding tissue (and possibly has spread to other organs), it is referred to as invasive.

Mammography



MLO or medial lateral oblique view

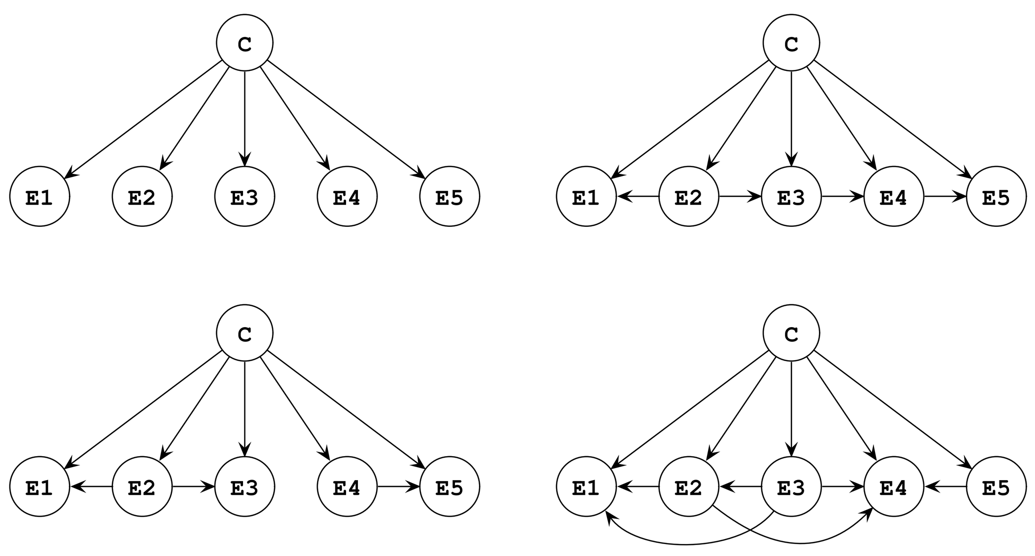


CC or cranio-caudal view

Mammography is based on the difference in absorption of X-rays between various tissue components. The projection of the breast can be made from different angles. The medial lateral oblique (MLO, side view taken at an angle), and the cranio caudal (CC, top to bottom view).

Methodology

Learning Bayesian networks from data involves learning the *network topology* and learning the *parameters*. As the possible network structures are exponentially large, we restrict them to Bayesian classifiers. Variables will represent continuous images features, which need to be reduced using for example feature selection and principal component analysis. Most effort will be put though in using expert knowledge to train a series of simple classifiers on separate subsets (representing different signs, views, or distinct aspects like lesion boundary and effect of a lesion on its surroundings). The output of these simple classifiers can then be incorporated as input to another Bayesian network.

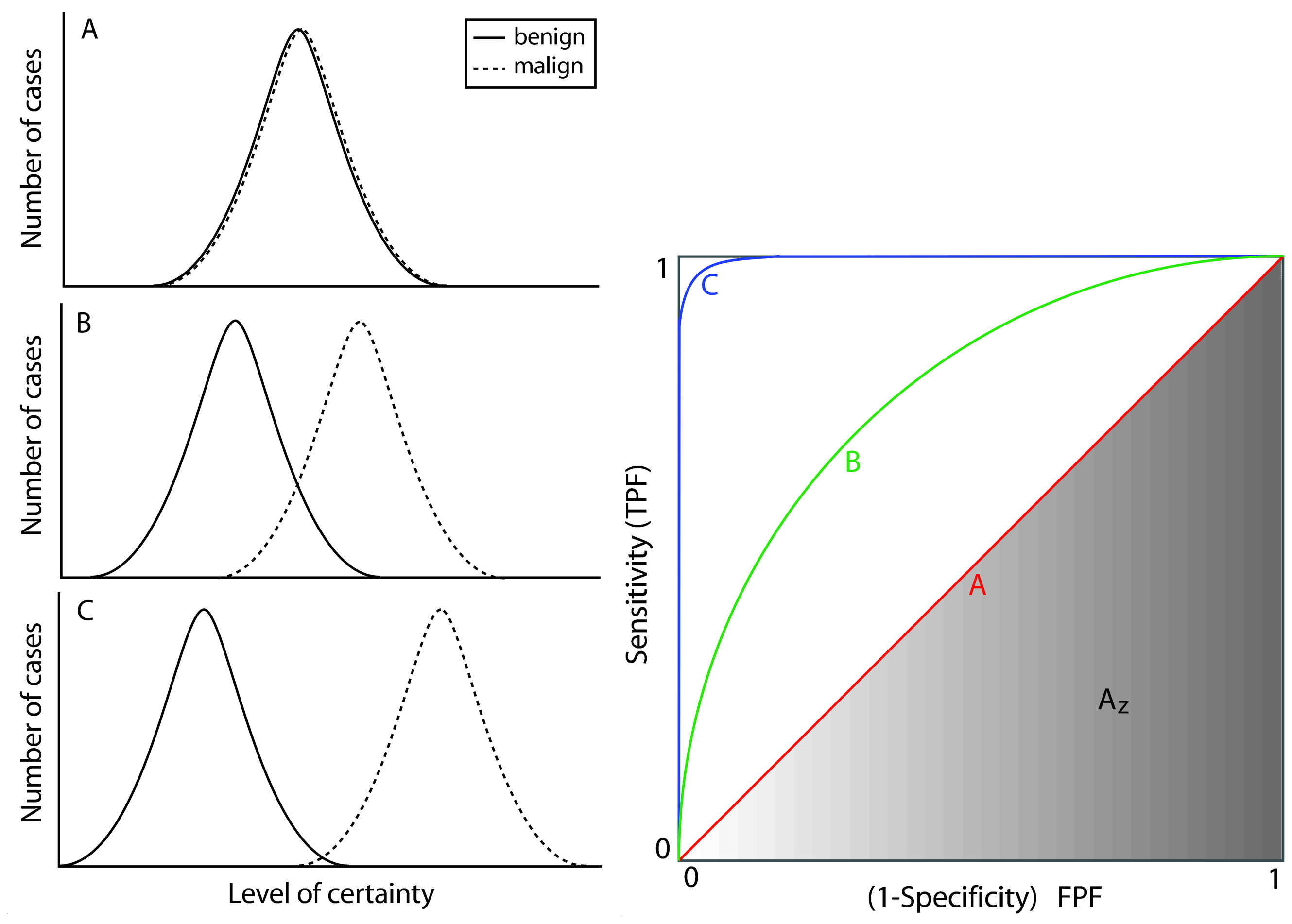


Bayesian Network classifiers based on different models for representing dependencies between evidence variables E_i for a class C. The Naive Bayesian classifier (NB, top left), Tree-augmented Naive Bayesian classifier (TAN, top right), Forrest-augmented Naive Bayesian classifier (FAN, bottom left), and Bayesian Network augmented Naive Bayesian classifier (BAN, bottom right).

In the right picture, on the left the certainty of a tumour being benign relative to malign is given. On the right the corresponding ROC curves are shown. The area under the ROC curve, the value A_z, is used for measuring the performance. The better we are to distinguish malign from benign cases, the higher A_z will be for the corresponding ROC curve.

Classifier training will be started using existing algorithms and N-fold cross-validation. A preliminary investigation already has shown adequate support for continuous variables (possibly after data normalisation techniques). Preliminary results shows that a Naive Bayesian classifier can give comparable results with a Support Vector Machine, which currently gives the best classifier performance.

[1] M. Samulski, N. Karssemeijer, P. Lucas, and P. Groot. Classification of mammographic masses using support vector machines and Bayesian networks. In *SPIE Medical Imaging*, 2007.



The B-Screen project is part of the

The Nijmegen Health Care Computing Initiative

funded by NWO, hefboom program, Pr. Nr. 641.000.403
Perry Groot, Marko van Eekelen, Peter Lucas