

Improved mammographic CAD performance using multi-view information: a Bayesian network framework

Marina Velikova¹, Maurice Samulski¹, Peter J F Lucas²
and Nico Karssemeijer¹

¹ Department of Radiology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

² Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands

E-mail: m.velikova@rad.umcn.nl

Received 14 October 2008, in final form 4 December 2008

Published 27 January 2009

Online at stacks.iop.org/PMB/54/1131

Abstract

Mammographic reading by radiologists requires the comparison of at least two breast projections (views) for the detection and the diagnosis of breast abnormalities. Despite their reported potential to support radiologists, most mammographic computer-aided detection (CAD) systems have a major limitation: as opposed to the radiologist's practice, computerized systems analyze each view independently. To tackle this problem, in this paper, we propose a Bayesian network framework for multi-view mammographic analysis, with main focus on breast cancer detection at a patient level. We use causal-independence models and context modeling over the whole breast represented as links between the regions detected by a single-view CAD system in the two breast projections. The proposed approach is implemented and tested with screening mammograms for 1063 cases of whom 385 had breast cancer. The single-view CAD system is used as a benchmark method for comparison. The results show that our multi-view modeling leads to significantly better performance in discriminating between normal and cancerous patients. We also demonstrate the potential of our multi-view system for selecting the most suspicious cases.

1. Introduction

Breast cancer is the most common form of cancer among women worldwide and its early detection does improve the chances of successful treatment and recovery (Breast cancer and screening 2008). Therefore, many countries have introduced breast cancer screening programs

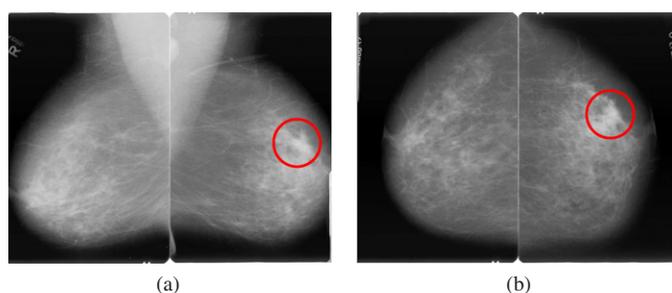


Figure 1. (a) MLO and (b) CC views of a right and left breast of a patient. The circle depicts a (cancerous) lesion in the left breast.

with periodic mammographic examinations in asymptomatic women. In contrast to the clinical situation, in the screening setting the detected lesions are usually small and due to the breast compression they are sometimes difficult to observe in both views. In other words, while the correct detection and location of a cancerous region is important, in breast cancer screening the crucial decision based on the mammographic exam is whether or not it is likely that a woman has breast cancer, and if the answer is positive, is referred to the clinic for further examination.

A screening mammographic examination usually consists of four images, corresponding to each breast scanned in two views—mediolateral oblique (MLO) view and craniocaudal (CC) view (see figure 1). The MLO projection is taken under 45° angle and shows part of the pectoral muscle. The CC projection is a top-down view of the breast. In reading mammograms, radiologists judge whether or not a lesion is present by comparing both views and breasts. The general rule is that a lesion is to be observed in both views.

To guarantee high detection rates, independent double reading by two radiologists is a widely used standard in breast cancer screening. Due to its complexity and the variability in human performance, however, mammographic reading and decision making appear to be difficult tasks. Radiologists are usually confronted with two main problems in the mammographic analysis: (i) perceptual oversight where an abnormality is present, but is missed and (ii) interpretation failure where an abnormality is seen but its significance is misinterpreted. There are two main types of abnormalities: microcalcifications and masses. In this work, we deal with the second, more frequently occurring type. There is strong evidence that for masses misinterpretation is a more common cause of missing cancers in screening than perceptual oversight.

In an attempt to support radiologists in overcoming these problems, a large number of mammographic computer-aided detection (CAD) systems have been developed and tested in the past 20 years. Essentially, the working principle of current CAD systems comprises a multi-stage process based on identification of regions of interest using image processing and pattern recognition techniques, extraction of a feature vector for each of these regions and classification of the regions as cancerous (abnormal) based on supervised learning techniques such as neural networks.

Despite the reported evidence about their potential benefit, most CAD methods suffer from certain limitations due to the uncertainty inherent in the domain. For example, misclassification can arise between a region of interest and its extracted feature vector or lack of separability between regions of interest that have similar features. One reason for these problems is that, opposite to the radiologist's practice, most computerized systems are based on a single-view

principle where each view and the regions within a view are analyzed independently. Hence, the multi-view and multi-region dependences in the breast are ignored and the breast cancer detection can be obscured. As a result, such systems perform worse than the human experts, which limits their practical application and usability.

To tackle these problems statistical modeling of the domain can be applied to the automatic detection process. In this paper, we propose a Bayesian network framework for exploiting multi-view dependences for the analysis of screening mammograms. Given the goal of screening programs, we focus primarily on the breast cancer detection at a patient level, rather than on the location of the cancer in the mammogram. The main idea of our methodology lies in combining the information available as detected regions from a single-view CAD system in MLO and CC to obtain a single likelihood measure for a patient being cancerous. In comparison to previous methods, we can outline a number of advantages of our probabilistic framework:

- Handling noise and missing information: specifying and learning the network parameters in a probabilistic manner allows uncertain information to be incorporated based on the values of all the non-missing variables.
- Incorporating domain knowledge: unlike black-box approaches such as neural networks, our framework captures explicitly view dependences through the Bayesian network structure and the definition of the conditional probability tables.
- Using context information over the whole breast: breast classification is done on the basis of simultaneous consideration of the regions automatically detected in each breast view and their links to the other view of the same breast.

We adopt the following terminology from the breast cancer domain throughout this paper. By *lesion* we refer to a physical cancerous object detected in a patient (the circle in figure 1). We call a contoured area on a mammogram a *region* (for example, marked manually by a human or detected automatically by a CAD system). A region can be true positive (TP), i.e., correct detection of the lesion (cancer) or false positive (FP). A region detected by a CAD system is described by a number of continuous (real-valued) *features* (e.g., size, location, contrast). By *link* we denote matching (established correspondence) between two regions in MLO and CC views, respectively. The term *case* refers to a patient who has undergone a mammographic exam. The most recent case for a patient is called *current* whereas the previous case(s) are *prior(s)*.

The remainder of the paper is organized as follows. In the following section we briefly review previous research in multi-view breast cancer detection. In section 3 we describe the general problem of multi-view detection, introduce basic definitions related to Bayesian networks and then we present a general Bayesian network framework for multi-view detection. The proposed approach is evaluated on an application of breast cancer detection using actual screening data. The evaluation procedure and the results are presented in section 4. Conclusions and directions for extension of our model are given in section 5.

2. Previous research

A number of previous works deal with the problem of automatic multi-view breast cancer detection on mammograms. Good *et al* (1999) have proposed a probabilistic method for true matching of lesions detected in both views, based on Bayesian network and multi-view features. The results from experiments demonstrate that their method can significantly distinguish between true and false positive links of regions. Van Engeland *et al* describe another linking method in van Engeland *et al* (2006) based on linear discriminant analysis

(LDA) classifier and a set of view-link features to compute a correspondence score for every possible region combination. For every region in the original view the region in the other view with the highest correspondence score is selected as the corresponding candidate region. The proposed approach demonstrates an ability to discriminate between true and false links. van Engeland and Karssemeijer (2007) extend this matching approach by building a cascaded multiple-classifier system for reclassifying the region level of suspiciousness of an initially detected region based on the linked candidate region in the other view. Experiments have shown that the lesion-based detection performance of the two-view detection system is significantly better than that of the single-view detection method.

Paquerault *et al* (2002) also consider established correspondence between suspected regions in both views to improve lesion detection. LDA is used to classify each object pair as true or false. By combining the resulting correspondence score with its one-view detection score the lesion detection improves and the number of false positives reduces. In this study, the authors also report improvement in the case-based performance (fraction of TP cases where a case is TP, if cancer is found in MLO *or* CC view) based on multi-view information, especially for cases where the lesion has been detected in both views.

In two recent studies, Sun *et al* (2001), Qian *et al* (2007) also demonstrate the superior performance of a multi-view CAD system over its single-view counterpart. The approach consists of multiple steps starting with advanced single-view image processing for region segmentation, followed by multi-view feature extraction and final classification of the detected regions of interest based on neural networks with Kalman filtering. Using iterative processing between the single- and multi-view stages, the authors show a reduction at the false positive rates of masses per image as well as an increase at the case-based detection rate.

However, in all these works the main focus is on improving the localized detection of breast cancer, mostly for prompting purposes, rather than the detection at a case level. Therefore, the likelihood for cancer in a case is often determined by the region with the maximum likelihood. In contrast, in the current study we aim at building a CAD system that discriminates well between normal and cancerous cases—the ultimate goal of breast cancer screening programs—by considering all available information (in terms of regions) in a case. In the following section, we describe such a system based on a probabilistic methodology and we demonstrate its practical potential on a case study.

3. Bayesian multi-view detection

3.1. Problem description

In multi-view medical imaging, two-dimensional (2D) projections of the organ(s) of interest (e.g. breast) are acquired from two or more viewing angles. The objective of the multi-view detection then is to determine whether or not the object has certain characteristics (e.g., being cancerous) by establishing correspondences between the 2D image characteristics of regions (subparts) in multiple object views (projections). Figure 2 depicts the general multi-view detection scheme.

We have a physical object referring to an organ (displayed as a gray cloud), which is projected in two views, *View-A* and *View-B*. Suppose we have a cancerous physical subpart of the object represented by the ovals in both projections; hence, the whole object is cancerous. In both views an automatic single-view system detects potential cancerous regions described by a number of real-valued extracted features. In the figure regions A_1 and B_1 are correct detection of the cancerous physical subpart, i.e., these are TP regions whereas A_2 and B_2 are FP regions. Since we deal with projections of the same physical object we introduce *links*

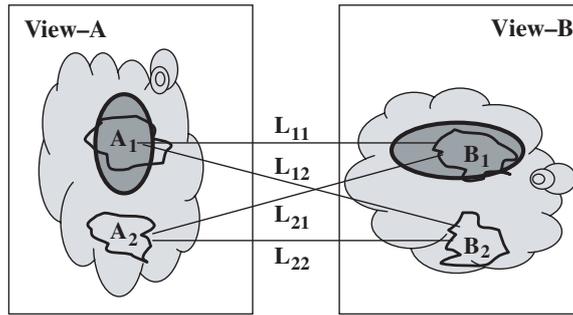


Figure 2. Schematic representation of multi-view analysis of a physical object with automatically detected regions.

(L_{ij}) between the detected regions in both views, A_i and B_j . Every link has a class (label) $L_{ij} = \ell_{ij}$ defined as follows

$$\ell_{ij} = \begin{cases} true & \text{if } A_i \text{ or } B_j \text{ are TP,} \\ false & \text{otherwise.} \end{cases} \quad (1)$$

This definition allows us to maintain information about the presence of cancer even if there is no cancer detection in one of the views. A binary class C with values of *true* (presence of cancer) and *false* for region, view or the whole object (organ) is assumed to be provided by pathology or a human expert.

In any case, multiple views corresponding to the same cancerous part contain correlated characteristics whereas views corresponding to normal parts tend to be less correlated. For example, in mammography an artifactual density might appear in one view due to the superposition of normal tissue whereas it disappears in the other view. To account for the interaction between the breast projections, in this paper we develop a Bayesian network framework for mammographic analysis. The power of Bayesian networks lies in their ability to (i) explicitly and efficiently encode causal dependences in a domain and (ii) model and reason about uncertainty in a probabilistic fashion. This makes them a suitable modeling tool for the multi-view detection problem. The following section gives some general background about Bayesian networks.

3.2. Bayesian networks

Consider a finite set \mathbf{U} of random variables, where each variable U in \mathbf{U} takes on values from a finite domain $\text{dom}(U)$. Let P be a joint probability distribution of \mathbf{U} and let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be disjoint subsets of \mathbf{U} . We say that \mathbf{X} and \mathbf{Y} are conditionally independent given \mathbf{Z} , denoted by $\mathbf{X} \perp\!\!\!\perp_P \mathbf{Y} \mid \mathbf{Z}$, if for all $\mathbf{x} \in \text{dom}(\mathbf{X}), \mathbf{y} \in \text{dom}(\mathbf{Y}), \mathbf{z} \in \text{dom}(\mathbf{Z})$, the following holds:

$$P(\mathbf{x} \mid \mathbf{y}, \mathbf{z}) = P(\mathbf{x} \mid \mathbf{z}), \text{ whenever } P(\mathbf{y}, \mathbf{z}) > 0.$$

In short, we have $P(\mathbf{X} \mid \mathbf{Y}, \mathbf{Z}) = P(\mathbf{X} \mid \mathbf{Z})$.

A Bayesian network is defined as a pair $\text{BN} = (G, P)$ where G is an acyclic directed graph (ADG) $G = (V, E)$ with a set of nodes V corresponding to the random variables in \mathbf{U} and a set of edges (arcs) $E \subseteq (V \times V)$ corresponding to direct causal relationships between the variables. We say that G is an *I-map* of P if any independence represented in G , denoted by $A \perp\!\!\!\perp_G B \mid C$ with $A, B, C \subseteq V$ mutually disjoint sets of nodes, is satisfied by P , i.e.,

$$A \perp\!\!\!\perp_G B \mid C \implies \mathbf{X}_A \perp\!\!\!\perp_P \mathbf{X}_B \mid \mathbf{X}_C,$$

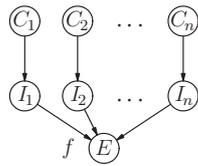


Figure 3. Causal-independence model.

where A , B and C are sets of nodes of the ADG G and \mathbf{X}_A , \mathbf{X}_B and \mathbf{X}_C are the corresponding sets of random variables. The acyclic directed graphical part of a Bayesian network G is by definition an I-map of the associated joint probability distribution P . A Bayesian network BN offers a compact representation of the joint probability distribution P in terms of local *conditional probability distributions (CPDs)*, or, in the discrete case, in terms of *conditional probability tables (CPTs)*, associated with the individual nodes. The conditional probability distributions are usually more compact than in the general case, as they take into account the conditional-independence information represented by the ADG. For a more detailed recent description of Bayesian networks, the reader is referred to Jensen and Nielsen (2007).

Causal-independence models. It is known that the number of probabilities in a CPT for a certain variable grows exponentially in the number of parents in the ADG. Therefore it is often infeasible to define the complete CPT for variables with many parents. One way to specify interactions among statistical variables in a compact fashion is offered by the notion of *causal independence* (Heckerman and Breese 1996). Causal independence arises in cases where multiple causes (parent nodes) lead to a common effect (child node). Here we present the formal definition of the notion of causal independence as given in Lucas (2005).

The general structure of a causal-independence model is shown in figure 3; it expresses the idea that causes C_1, \dots, C_n influence a given common effect E through intermediate variables I_1, \dots, I_n ; the intermediate variable I_k is considered to be a contribution of the cause variable C_k to the common effect E . The *interaction function* f represents in which way the intermediate effects I_k , and indirectly also the causes C_k , interact. This function f is defined in such way that when a relationship between the I_k 's and $E = true$ is satisfied, then it holds that $f(I_1, \dots, I_n) = true$; otherwise, it holds that $f(I_1, \dots, I_n) = false$. Note that each variable I_k is only dependent on its associated cause C_k and the effect variable E . Furthermore, the graph structure expresses that the effect variable E is conditionally independent of each cause C_k given the associated intermediate variable I_k .

An important subclass of causal-independence models is obtained if the deterministic function f is defined in terms of separate binary functions g_k ; it is then called a *decomposable* causal-independence model (Heckerman and Breese 1996). Usually, all functions $g_k(I_k, I_{k+1})$ are identical for each k . Typical examples of decomposable causal-independence models are the noisy-OR (Diez 1993) models, where the function g represents a logical OR. These models express that the presence of any of the causes C_k with absolute certainty will cause the effect $E = true$. A simple example of a noisy-OR model is given in the Appendix.

In our modeling framework, presented in the following section, we apply such a decomposable causal-independence model with the logical OR. Our choice is motivated by two major features of the representation of the noisy-OR models. First, from the definition of the noisy-OR model it follows that the higher the number of causes influencing the effect the higher the probability that the effect occurs. This rule is definitely applicable in the domain

of breast cancer detection where the more the evidence (e.g., in terms of detected regions) is added the higher the probability for cancer. Another important feature from a computational point of view is that the representation of the noisy-OR models has a linear complexity with respect to the number of causes.

3.3. Model description

Our modeling scheme is based on two Bayesian networks with a hand-constructed (fixed) structure to explicitly represent the multi-view dependences in the detection problem. Consider again the detection scheme presented in figure 2. The regions A_i and B_j are generally conditionally independent given the case. However, they become dependent once we have evidence that they are the projections of the same lesion in two views. In the context of Bayesian networks, this region dependence can be modeled by (i) three nodes: two for the regions and one for the link and (ii) the so-called *v-structure* where directed arcs are drawn from the region nodes to the link node: $A_i \longrightarrow L_{ij} \longleftarrow B_j$. Such a representation of the dependence between a relation (link) and its parts (regions) has also been advocated by other researchers in the field of vision perception (Sarkar and Boyer 1993). Note that swapping the arc direction from the link to the regions would imply that the regions are conditionally independent given the existence of a link, which contradicts our intuition.

Furthermore, by definition the link variable is discrete and the regions are represented by a vector of real-valued features (x_1, x_2, \dots, x_n) extracted from an automatic detection system. Therefore we apply logistic regression to compute $P(L_{ij} = \ell_{ij} | A_i, B_j)$:

$$P(L_{ij} = \ell_{ij} | A_i, B_j) = \frac{\exp(\beta_0^{\ell_{ij}} + \beta_1^{\ell_{ij}} x_1 + \dots + \beta_{2n}^{\ell_{ij}} x_{2n})}{1 + \exp(\beta_0^{\ell_{ij}} + \beta_1^{\ell_{ij}} x_1 + \dots + \beta_{2n}^{\ell_{ij}} x_{2n})},$$

where β 's are the model parameters to be estimated and the index $2n$ is the total number of region features from both views. We note that other estimators such as multilayer neural networks can be also used to define $P(L_{ij} = \ell_{ij} | A_i, B_j)$ but we choose logistic regression as it ensures in a straightforward way that the outputs $P(L_{ij} = \ell_{ij} | A_i, B_j)$ are probabilities.

In our multi-view detection problem the object (organ) contains a number of links where every region in one view is connected to all the regions in the other view. Hence, it is intuitive and straightforward to construct a causal structure where all the links are modeled in parallel (see the first top layer in the network depicted in figure 4(a)). Thus, using the context modeling capabilities of Bayesian networks we consider at once all information available about the object.

Next we estimate the probabilities $P(C_{A_i} = \text{true} | \{L_{ij} = \ell_{ij}\}_{j=1}^{N_B})$ and $P(C_{B_j} = \text{true} | \{L_{ij} = \ell_{ij}\}_{i=1}^{N_A})$ where $C_{A_i}(C_{B_j})$ is the class of region $A_i(B_j)$, $N_A(N_B)$ is the total number of regions in *View-A* (*View-B*) and $\{L_{ij} = \ell_{ij}\}_{j(i=1)}^{N_B(N_A)}$ denotes the set of all links containing $A_i(B_j)$. Given our link class definition in (1), we can easily model these conditional dependences through a causal model using the logical OR. We refer to this Bayesian network as *RegNet* (see figure 4(a)).

Recall that our main goal is to optimize classification globally in terms of the whole object (organ). Therefore, we construct a second Bayesian network to combine the computed region probabilities from *RegNet* to obtain the probability of a view being *true*. We use a causal-independence model with the logical OR where the cause nodes C_i are the region probabilities, the intermediate nodes I_i are the region classes and the only leaf node is the view probability. This Bayesian network is depicted in figure 4(b) and we refer to it as *ViewNet*. The whole multi-view model based on *RegNet* and *ViewNet* is called *MV-CAD*.

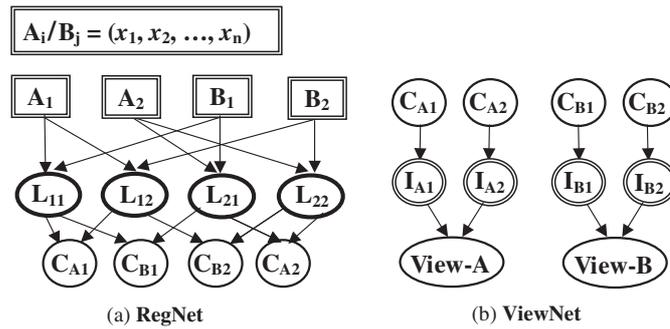


Figure 4. Bayesian network framework for representing the dependences between multiple views of an object.

Finally, we combine the view probabilities obtained from ViewNet into a single probabilistic measure for the object (organ) as a whole by using different schemes. The first scheme MV-CAD-Avg is straightforward—simply averaging both view probabilities. In another more advanced scheme MV-CAD-LR, we take into account the class of the object (*false* or *true*) by using a logistic regression model with the estimated view probabilities as input variables.

4. Application to breast cancer detection

As mentioned in the introduction, multi-view analysis plays a crucial role in the breast cancer detection on mammograms. Here, we describe the application of the proposed Bayesian network framework in this domain.

4.1. Single-view CAD system

The inputs for our multi-view detection scheme are the regions detected by a single-view CAD system (van Engeland *et al* 2006) that consists of the following main steps (see figure 5):

- (i) Segmentation of the mammogram into background area, breast, and for MLO, the pectoral muscle.
- (ii) Initial detection of pixel-based locations of interest. For each location in the breast area a number of features are computed that are related to tumor characteristics such as presence of spicules and focal mass. Based on these features, a neural network (NN) classifier is then employed to compute likelihood for cancer. The locations with a likelihood above certain threshold are selected as locations of interest.
- (iii) Region extraction with dynamic programming using the detected locations as seed points. For each region a number of continuous features are computed based on breast and local area information.
- (iv) Region classification as ‘normal’ and ‘abnormal’ based on the region features. A likelihood for cancer is computed based on supervised learning with a NN and converted into *normality score* (NormSc): the average number of normal regions in a view (image) with the same or higher cancer likelihood. Hence, the lower the normality score the higher the likelihood for cancer.

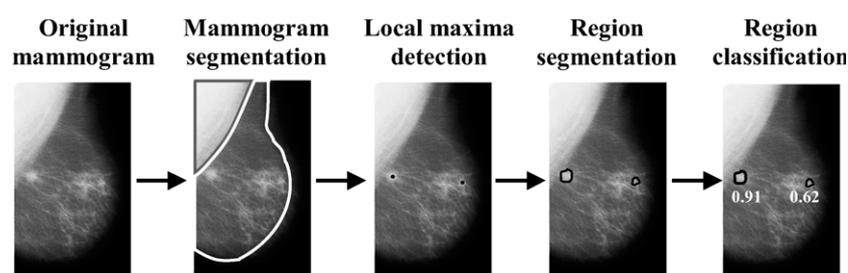


Figure 5. Stages in the single-view CAD system.

4.2. Data description

The data set we use in this study contain 1063 screening exams (cases) from which 385 were cancerous. The data is a mixture of 795 current cases and 268 prior cases; 33 were cancerous priors with the cancer visible in retrospect. We considered the exams of one patient as independent cases. All exams contained both MLO and CC views. The total number of breasts were 2126 from which 388 had cancer. All cancerous breasts had one visible lesion representing a mass, architectural distortion, or asymmetry in at least one view, which was verified by pathology reports to be malignant (cancerous). Lesion contours were marked by, or under supervision of, an experienced screening radiologist.

For each image (mammogram) we have a number of regions detected by the single-view CAD system. Every region is described by 11 real-valued features automatically computed by the system. These features include the neural network's output from the single-view CAD and lesion characteristics such as spiculation, focal mass, size, contrast, linear texture and location coordinates. Since the only certain information we have about the findings is the one related to the cancer, for each region, based on the ground-truth data, we have a class value of *true* ('cancerous') if the detected region hits a cancerous finding and *false* ('normal') otherwise, which may also include hits of benign findings. Every region from MLO view was linked with every region in CC view. For every link we added the binary class values of *false* ('normal') and *true* ('cancerous') following the definition in equation (1). We assign analogous binary classes for view, breast and case based on the ground-truth information.

We construct the data such that every row corresponds to one breast observation represented by all feature vectors for the regions in MLO, followed by the regions in CC. The sequence of regions per view was determined by the level of suspiciousness, starting with the most suspicious one. In this study we conduct experiments with two datasets where we select five and three regions per view with the lowest NormSc. The two datasets are described in table 1. The selection of five regions per view leads to data where in only five out of 385 cancerous cases there is no TP detected region and thus no *true* link is available in these cases for training the networks. On the other hand, we have a large number of MLO and CC regions, which are mostly FP. Therefore for the second data set we choose three regions per view. However, this leads not only to considerably less FP regions in MLO and CC but also to a higher number of missed TP regions—in total 13 out of 385 cases.

4.3. Training and evaluation

To train and evaluate the proposed multi-view CAD system, we used two-fold cross validation: the dataset is randomly split into two subsets with approximately equal number of observations

Table 1. Description of the two datasets used in the current study.

Parameter	Dataset-1 (Data5reg)	Dataset-2 (Data3reg)
Number of regions per view	5	3
Total number of regions (MLO/CC)	10478/10343	6358/6328
Number of cancerous cases without <i>true</i> links	5	13

and proportion of cancerous cases. The data for a whole case belonged to only one of the folds. Each fold is used as a training set and as a test set. At every level (region, view, breast and case) the same data folds were used. Although we use the results from the single-view CAD system, we want to emphasize that the random split for the multi-view CAD system is done independently—the single-view CAD system was trained and tested with ten-fold cross validation on a much larger dataset including regions from cases without CC views.

Bayesian network training. Both RegNet and ViewNet have been built, trained and tested by using the Bayesian Network Toolbox in Matlab (Murphy). The learning has been done using the EM algorithm, which is typically used to approximate a probability function given incomplete samples (in our networks the OR-nodes are not observed) (Dempster *et al* 1977).

Breast data training. As we discussed in the description of our model, we apply two combining schemes—averaging and logistic regression—to compute the probability for a breast being cancerous given the respective view probabilities. For the logistic regression, the input contains view information represented by the probabilities for MLO and CC obtained from ViewNet and the minimum NormScs for each view, which are also indicators for view suspiciousness.

Case classification. We compute the likelihood of a case being cancerous based on the computed right and left breast probabilities. The first simplest approach is to take the maximum out of both probabilities. Furthermore, for the MV-CAD-LR model, which accounts for the breast classes, we presume that further improvement can be achieved by using the case class. Therefore we perform logistic regression using two inputs: the maximum out of both breast probabilities and the single-view measure for suspiciousness. Thus from the multi-view CAD system, we obtain in total three measures for a case being cancerous: MV-CAD-Avg-max, MV-CAD-LR-max and MV-CAD-LR-LR.

The performance of our multi-view model is compared with that of the single-view CAD system (SV-CAD). For the latter, the likelihood for a view, breast and case being cancerous is computed by taking the likelihood (NormSc) of the most suspicious region. The comparison analysis is done using the receiver operating characteristic (ROC) curve (Hanley and McNeil 1982) and the area under the curve (AUC), a standard performance measure in the radiologists' practice. The significance of the differences obtained in the AUC measures is tested using the ROCKIT software for fully paired data: for each patient we have a pair of test results corresponding to MV-CAD and SV-CAD Metz *et al* (1984).

4.4. Experiments and results

4.4.1. Individual view, breast and case classification. Based on the results from ViewNet, figures 6(a) and (b) present the classification outcome with the respective AUC measure per MLO and CC view for Data5reg and Data3reg. First we observe that for both MV-CAD

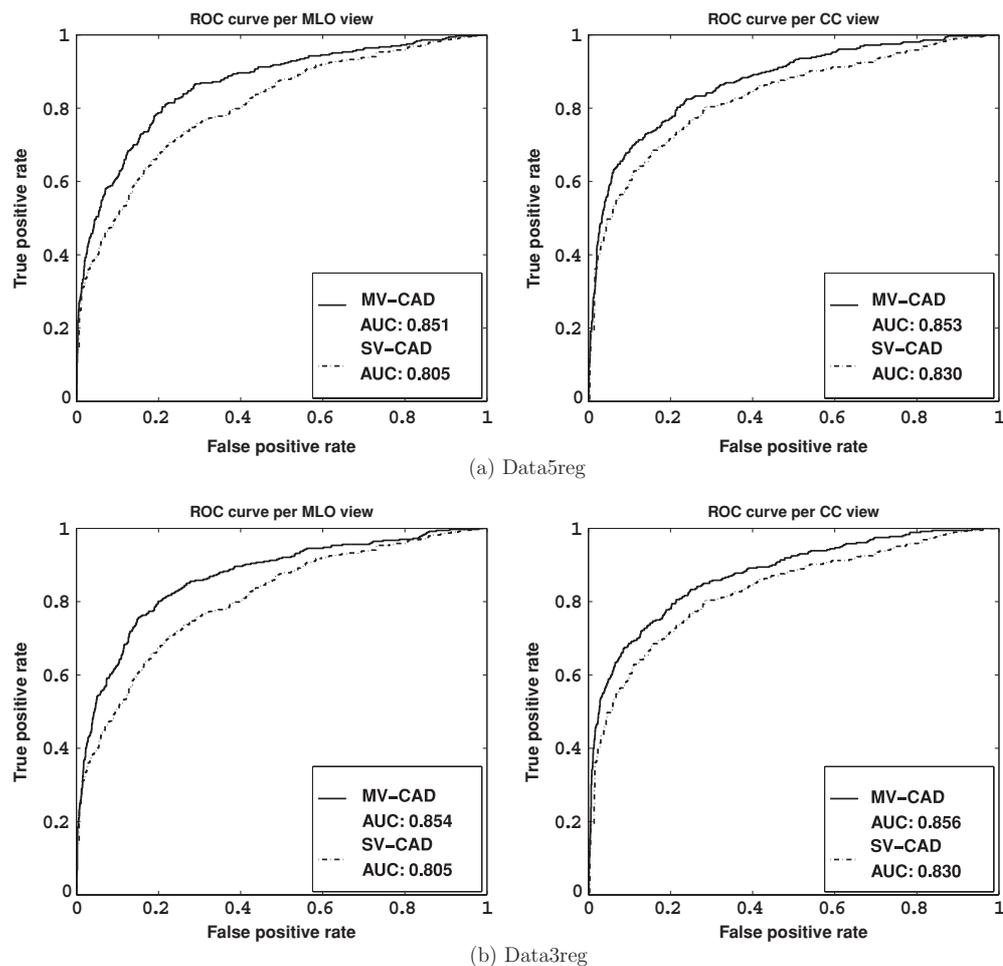


Figure 6. ROC analysis per MLO and CC view.

and SV-CAD the performance for CC view in terms of AUC is better than that for MLO view. This can be explained by the fact that the classification of CC views is generally easier than that of MLO views due to the breast positioning. At the same time our multi-view system improves considerably upon the single-view CAD system in better distinguishing cancerous from normal MLO views whereas for CC views this improvement is less. Another interesting result is that the largest improvement, especially for MLO view, is observed in the lower scale of the false positive rate (<0.5).

To check the significance of the difference between the AUC measures we test the hypothesis that the AUC measures are equal against the one-sided alternative hypothesis that the multi-view system yields higher AUC for MLO and CC views. Table 2 summarizes the statistical test results by providing the corresponding p -values and 95% confidence intervals of the difference between the AUC measures. The results clearly indicate an overall improvement in the discrimination between cancerous and normal views for both MLO and CC projections. Such an improvement is expected as the classification of each view in our multi-view system takes into account region information not only from the view itself but also from the regions in the other view.

Table 2. AUC (std. error) obtained from the single- and multi-view system per MLO and CC with the respective one-sided p -values and 95% confidence intervals for the difference.

View	Method	Data5reg	p -value	Data3reg	p -value
MLO	SV-CAD	0.805 (0.013)	–	0.805 (0.013)	–
	MV-CAD	0.851 (0.011)	0.000 (0.028, 0.063)	0.854 (0.011)	0.000 (0.031, 0.067)
CC	SV-CAD	0.830 (0.012)	–	0.830 (0.012)	–
	MV-CAD	0.853 (0.011)	0.004 (0.006, 0.039)	0.856 (0.011)	0.001 (0.009, 0.044)

Table 3. AUC (std. error) obtained from the single- and multi-view system at a *breast* level with the respective one-sided p -values and 95% confidence intervals for the difference.

Method	Breast			
	Data5reg	p -value	Data3reg	p -value
SV-CAD	0.849 (0.012)	–	0.849 (0.012)	–
MV-CAD-Avg	0.862 (0.011)	0.047 (–0.002, 0.029)	0.860 (0.011)	0.094 (–0.005, 0.026)
MV-CAD-LR	0.868 (0.011)	0.010 (0.003, 0.034)	0.865 (0.011)	0.024 (0.000, 0.031)

While the view results are very promising from a radiologists' point of view it is more important to look at the breast and case level performance. Tables 3 and 4 present the respective AUC (standard error) obtained from MV-CAD and SV-CAD systems as well as the one-sided p -values and 95% confidence intervals obtained from the tests on the differences between our multi-view model and the single-view system. Although the simple averaging method MV-CAD-Avg (MV-CAD-Avg-max) tends to show better distinction between normal and cancerous breasts (cases) with respect to the SV-CAD, the difference in the AUC measures is statistically insignificant. However, taking into account the breast classes and performing new training as done in the more advanced MV-CAD-LR leads to a significant improvement in the classification outcome. The best performance for both datasets at a case level is achieved for MV-CAD-LR-LR, confirming our expectation that further improvement can be achieved by training using the case class. Furthermore, we note that for both datasets, MV-CAD-LR-LR yields the same AUCs but with slightly different p -values. To explain this difference we plot the ROC curves; see figure 7. We see that for Data5reg improvement in the breast cancer detection is observed over the whole range of false positive rates whereas for Data3reg it is achieved for false positive rates <0.6 .

4.4.2. Use of CAD for prescreening of cases. The results so far presented demonstrate the superior performance of the multi-view system in comparison to its single-view counterpart in terms of individual view, breast and case classification. Here we demonstrate another potential application of the multi-view CAD system to support mammographic decision making, namely automated prescreening of cases. The objective is to group cases into two basic categories: 'suspicious' and 'normal' in order to handle these by a different reading protocol.

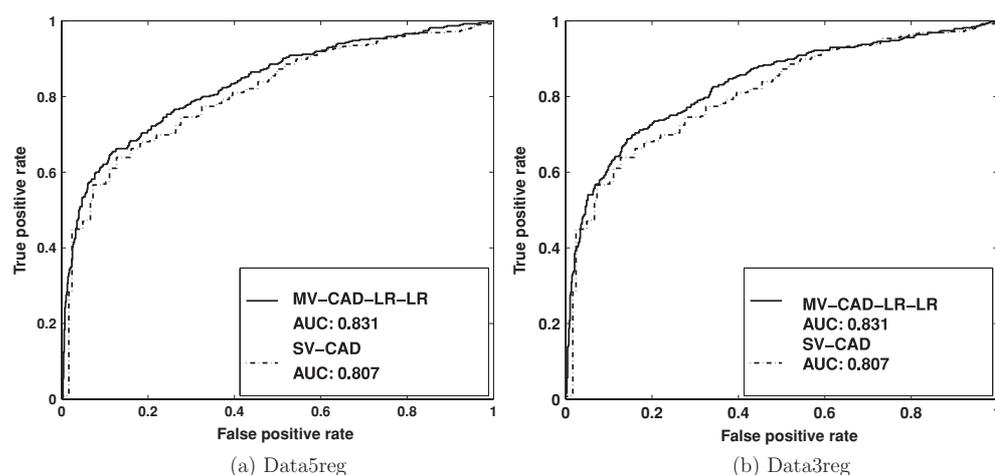


Figure 7. ROC analysis per case.

Table 4. AUC (std. error) obtained from the single- and multi-view system at a case level with the respective one-sided p -values and 95% confidence intervals for the difference.

Method	Case			
	Data5reg	p -value	Data3reg	p -value
SV-CAD	0.807 (0.014)	–	0.807 (0.014)	–
MV-CAD-Avg-max	0.825 (0.014)	0.040 (–0.002, 0.040)	0.819 (0.014)	0.104 (–0.008, 0.035)
MV-CAD-LR-max	0.830 (0.013)	0.014 (0.003, 0.045)	0.828 (0.014)	0.037 (–0.002, 0.041)
MV-CAD-LR-LR	0.831 (0.013)	0.007 (0.005, 0.043)	0.831 (0.013)	0.008 (0.004, 0.043)

One can argue that cases selected as ‘suspicious’ would benefit most from receiving more attention from radiologists. If resources in a screening program only allow for single-reading programs, for example, one might consider a modest extension of the program by double reading only the most suspicious cases. On the other hand, if double reading is practiced and resources are limited, one might consider to use single reading for a subset of cases selected by a CAD system as highly normal, leading to a considerable reduction in the workload. Alternatively, if both radiologists in a double reading setting do not find an abnormality in a case that is judged highly suspicious by a CAD system, one could present such a case to a third reader performing arbitration, similar to the procedure that is often followed if both readers disagree.

The problem of case prescreening has already been addressed in the literature introducing the concept of using specially trained, non-physician personnel for mammographic prescreening. With the increasing demand for mammography, required training times and shortage of manpower, however, it may be more beneficial to use CAD systems as a prescreening tool. To our knowledge only a few studies discussed so far the application of CAD systems for prescreening of cases (Kalman *et al* 1997, Astley *et al* 2002). The current

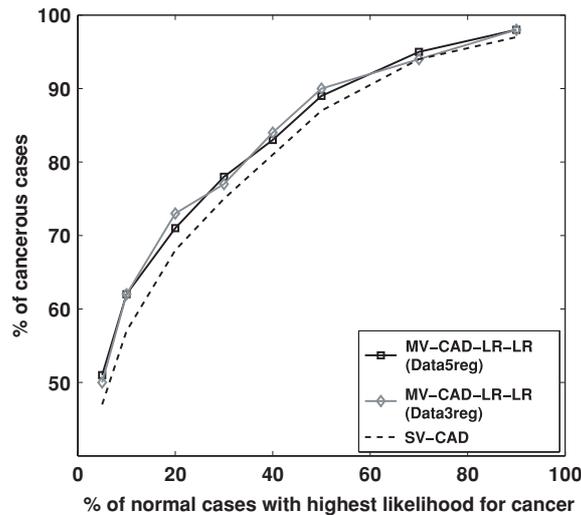


Figure 8. Percentages of cancerous cases within subsets of most suspicious normal cases for the single- and multi-view systems.

work contributes to the fund of knowledge in this area by considering the use of the multi-view and the single-view CAD system for the selection of most and least suspicious cases.

The aim of prescreening is optimizing the detection rate while reducing the workload. Because in the breast cancer screening programs the number of normal cases is far larger than that of cancerous cases, the workload is determined by the number of normals to be read. Therefore, for the prescreening task we consider the percentage of detected cancers as a function of a percentage of normal cases with highest likelihood for cancer. Figure 8 depicts the results for MV-CAD-LR-LR and SV-CAD.

The results demonstrate that using multi-view information leads to overall increase in the number of detected cancers when a subset of normal cases with highest likelihood for cancer is selected. For example, if 10% of the normals are selected then MV-CAD-LR-LR on Data5reg and Data3reg detects 62% and 62% of the cancers against 57% of SV-CAD. This trend is especially observed for the lower range (<20%) of selected normal cases.

When considering the other prescreening task—selection of the least suspicious cases—we would like to minimize the number of cancers missed when a subset of highly normal cases is chosen. In this respect, looking at the upper range of the percentage of selected normal cases in figure 8 (these are the least suspicious cases), we see that the multi-view leads only to a slight reduction in the number of misclassified cancerous cases in comparison to the single-view CAD system. We note that the result that both CAD systems do not detect all the cancers at smaller subsets of normal cases could be explained by the fact that 9% of the cancerous cases included in our study were priors, i.e. cancers that were not detected by the radiologists at the screening stage.

5. Conclusions and future research

In this paper we proposed a Bayesian network framework for multi-view mammographic analysis. We showed that the incorporation of expert knowledge in a probabilistic manner

led to a higher detection rate of breast cancer compared to a single-view CAD system. This improvement was achieved at a view, breast and case level and it is due to a number of factors. First, we built upon a single-view CAD system that demonstrates a good performance at local breast cancer detection. Second, following the radiologists' practice, we considered multi-view dependences between MLO and CC views to obtain a single measure for the view, breast and case being cancerous. This was done by: (i) defining links between the regions detected by the single-view CAD system in MLO and CC views (ii) building a probabilistic causal model where all detected regions with their feature vectors and the established region links are considered simultaneously, and (iii) using the logical OR to compute the region and view probability for cancer. Our multi-view scheme also benefits from its Bayesian nature allowing to handle noisy or incomplete information such as the lack of detected or visible lesions in one of the views.

Except the improvement in the individual case-based performance, in the current study we also demonstrated the potential of the multi-view CAD system for prescreening purposes. In contrast to the traditional prompting CAD systems, in this work we considered the problem of breast cancer detection in screening mammography at a case level. From this perspective, the proposed CAD system could be used to select the most suspicious cases or to group them for batch reading, as a set of difficult cases. In this way, the selected cases would get more attention from radiologists, for example, by providing additional reading. This could help increase the breast cancer detection rate.

Furthermore, our experiments show that the proposed Bayesian network framework is relatively stable with respect to the number of selected regions per mammogram detected by the single-view CAD. In the current study, we used two versions of the same set of patient cases: one with five regions and the other with three regions per mammogram. The results indicate that the performance of the models built on both datasets is comparable on individual view, breast and case classification as well as on the selection of most suspicious cases.

Although we demonstrate that the proposed framework has the potential to assist screening radiologists to improve the evaluation of breast cancer cases, we consider a number of directions for extension. First, the current model is based on features that are independently computed per region. However, it is natural to include multi-view features such as the distance to the nipple or correlation features. In such a way, we can explicitly represent multi-view dependences not only in a qualitative way (through the Bayesian network's structure) but also in a quantitative way (through the input information). This can help improve the system's detection performance. Another possible extension is based on the model structure. Following our Bayesian network framework with using logistic regression and OR-function at a link and view level, we can also apply similar combining schemes at a breast and case level. Thus we can allow for better handling of missing or noisy information in the estimation of the breast/case likelihood for cancer. A third interesting extension of the proposed CAD system is the incorporation of temporal information. In the screening practice, the decision whether a patient has cancer depends not only on the breast multi-view comparison but also on the comparison of current mammograms with previous mammograms of the same patient. The appearance of a new or developing lesion is a strong indication for suspiciousness. Therefore, by integrating multi-view with temporal information in our Bayesian network framework, we can better represent and more accurately model the decision-making process in screening mammography.

Finally, we note that the straightforward nature of the proposed Bayesian network framework allows its relatively easy application to any domain where the goal is computerized multi-view (object) detection.

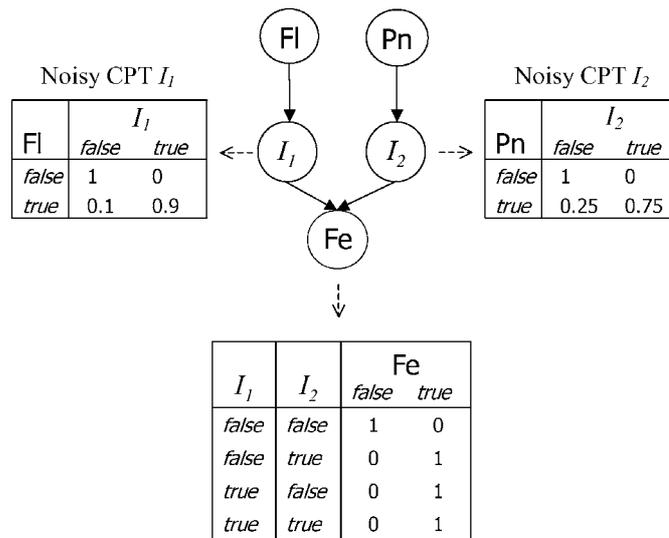


Figure A1. Example of a noisy-OR model.

Acknowledgments

Funded by the Netherlands Organisation for Scientific Research under BRICKS/FOCUS grant number 642.066.605. We would like to thank the anonymous reviewers for their constructive comments that helped improve this paper.

Appendix

Figure A1 depicts an example of a causal-independence model with two cause variables *Flu* (Fl) and *Pneumonia* (Pn) and one effect variable *Fever* (Fe). Probability distributions $P(I_1|Fl)$ and $P(I_2|Pn)$ represent a noise. The interaction function $f(I_1, I_2)$ for the effect *Fever* is the logical OR.

Then the probability of having fever given the states of Fl and Pn is computed as follows:

$$\begin{aligned}
 P(\text{Fe} = \text{true} | \text{Fl}, \text{Pn}) &= \sum_{f(I_1, I_2) = \text{true}} P(\text{Fe} = \text{true} | I_1, I_2) P(I_1 | \text{Fl}) P(I_2 | \text{Pn}) \\
 &= P(I_1 = \text{true} | \text{Fl}) P(I_2 = \text{true} | \text{Pn}) \\
 &\quad + P(I_1 = \text{true} | \text{Fl}) P(I_2 = \text{false} | \text{Pn}) \\
 &\quad + P(I_1 = \text{false} | \text{Fl}) P(I_2 = \text{true} | \text{Pn}).
 \end{aligned}$$

For example, given the evidence of Fl = *true* and Pn = *true* then we obtain

$$P(\text{Fe} = \text{true} | \text{Fl}, \text{Pn}) = 0.9 \cdot 0.75 + 0.9 \cdot 0.25 + 0.1 \cdot 0.75 = 0.975,$$

indicating the combined influence of both causes on the probability of having fever.

References

- Astley S, Mistry T, Boggis C and Hiller V 2002 Should we use humans or a machine to pre-screen mammograms? *Proc. of 6th Int. Workshop in Digital Mammography, Bremen, Germany* (Berlin: Springer)

- Breast cancer and screening 2008 *Technical Report* World Health Organization <http://www.emro.who.int/ncd/publications/breastcancerscreening.pdf>
- Dempster A, Laird N and Rubin D 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc., Ser. B* **39** 1–38
- Diez F 1993 Parameter adjustment in Bayes networks: the generalized noisy or-gate *Proc. of the 9th Conf. on UAI* (San Francisco, CA: Morgan Kaufmann)
- Good W, Zheng B, Chang Y, Wang X, Maitz G and Gur D 1999 Multi-image cad employing features derived from ipsilateral mammographic views *Proc. SPIE, Med. Imaging* **3661** 474–85
- Hanley J A and McNeil B J 1982 The meaning and use of the area under a receiver operating characteristic (ROC) curve *Radiology* **143** 29–36
- Heckerman D and Breese J S 1996 Causal independence for probability assessment and inference using Bayesian networks *IEEE Trans. Syst. Man Cybern. A* **26** 826–31
- Jensen F V and Nielsen T D 2007 *Bayesian Networks and Decision Graphs* (Berlin: Springer)
- Kalman B L, Reinus W R and Kwasny S C 1997 Prescreening entire mammograms for masses with artificial neural networks: preliminary results *Academic Radiol.* **4** 405–14
- Lucas P J F 2005 Bayesian network modelling through qualitative pattern *Artif. Intell.* **163** 233–63
- Metz C E, Wang P L and Kronman H B 1984 A new approach for testing the significance of differences between ROC curves measured from correlated data *Inform. Process. Med. Imaging* (The Hague: Nijhoff)
- Murphy K 2007 Bayesian Network Toolbox (BNT) <http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html>
- Paquerault S, Petrick N, Chan H, Sahiner B and Helvie M A 2002 Improvement of computerized mass detection on mammograms: fusion of two-view information *Med. Phys.* **29** 238–47
- Qian W, Song D, Lei M, Sankar R and Eikman E 2007 Computer-aided mass detection based on ipsilateral multiview mammograms *Academic Radiol.* **14** 530–8
- Sarkar S and Boyer K L 1993 Integration, inference, and management of spatial information using Bayesian networks: perceptual organization *IEEE Trans. Pattern Anal. Mach. Intell.* **15** 256–74
- Sun X, Qian W, Song D and Clark R A 2001 Ipsilateral multi-view cad system for mass detection in digital mammography *Proc. of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (IEEE Computer Society)*
- van Engeland S and Karssemeijer N 2007 Combining two mammographic projections in a computer aided mass detection method *Med. Phys.* **34** 898–905
- van Engeland S, Timp S and Karssemeijer N 2006 Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views *Med. Phys.* **33** 3203–12