Causal Probabilistic Modelling for Two-View Mammographic Analysis

Marina Velikova[1], Maurice Samulski[1], Peter J.F. Lucas[2],
and Nico Karssemeijer[1]

[1] Department of Radiology, Radboud University Nijmegen Medical Centre
6525 GA, Nijmegen, The Netherlands
{m.velikova,m.samulski,n.karssemeijer}@rad.umcn.nl
[2] Institute for Computing and Information Sciences, Radboud University Nijmegen
6525 ED Nijmegen, The Netherlands
peterl@cs.ru.nl

Abstract. Mammographic analysis is a difficult task due to the complexity of image interpretation. This results in diagnostic uncertainty, thus provoking the need for assistance by computer decision-making tools. Probabilistic modelling based on Bayesian networks is among the suitable tools, as it allows for the formalization of the uncertainty about parameters, models, and predictions in a statistical manner, yet such that available background knowledge about characteristics of the domain can be taken into account. In this paper, we investigate a specific class of Bayesian networks—causal independence models—for exploring the dependencies between two breast image views. The proposed method is based on a multi-stage scheme incorporating domain knowledge and information obtained from two computer-aided detection systems. The experiments with actual mammographic data demonstrate the potential of the proposed two-view probabilistic system for supporting radiologists in detecting breast cancer, both at a location and a patient level.

1   Introduction

The interpretation of screening mammograms is routine work for radiologists involved in national breast-cancer screening programs. Although routine, it is still a task fraught with difficulty with much space for improvement. The difficulty of reading mammograms is due to a number of factors such as variations in female breast tissue, cancer appearance on a mammogram, and image quality. To facilitate their screening work, radiologists are typically provided with two projections, or *views*, of each breast: mediolateral oblique (MLO), taken under 45° angle and showing part of the pectoral muscles, and craniocaudal (CC), taken head to tail. If cancer is present, then it is expected to be observed in both views. Observing that radiologists still miss too many cancer cases, offering them some form of assistance, for example, through computer-aided detection (CAD) systems is important. However, without exploiting principles that radiologists have

used successfully for decades, it is unlikely that CAD systems will ever outperform highly trained human interpreters. It is only recently that researchers have started to study ways to incorporate such principle into CAD systems.

Bayesian networks are especially promising in bridging this gap between the capabilities of humans and computer-aided interpretation, as they have the virtue of supporting the explicit representation of expert knowledge, handle uncertainty and missing information, and allow combining multiple types of knowledge. One of the principles used by radiologists in analysing mammograms is combining information obtained from different views of the same breast, which provides the basis for the development of a model for two-view mammographic analysis presented in this paper. The aim of the current study was two fold:

- to improve the breast cancer detection rate at a location and a patient level in comparison to a single-view CAD system;
- to get more insight into the mechanisms underlying mammographic analysis, which then can act as a basis for improvement of current CAD systems.

To achieve these goals, we developed a multi-stage system, using (i) a specific class of Bayesian networks, called causal independence models, and (ii) knowledge derived from an analysis of the way radiologists interpret mammograms. The method here builds on our previous research presented in [1,2].

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; see Fig. 2. We call a contoured area on a mammogram a *region*, marked, for example, manually by a human or detected automatically by a CAD system. A region detected by a CAD system is described by a set of continuous (real-valued) *single-view features*, e.g., size, location, contrast. By *link* we denote established correspondence, between two regions in MLO and CC views, respectively. Every link is described by a set of *multi-view features*, such as contrast difference and location difference. The most recent mammographic *exam* for a woman is called *current*, whereas the previous exam(s) are *prior(s)*.

Previous research has already demonstrated the potential of exploring multi-view dependencies to improve the automatic detection of breast cancer on mammograms. The approaches in [3,4] focused on improving the lesion-based results, mostly for prompting purposes, based on the distinction between true and false positive links of regions in MLO and CC views. In other studies multi-view information was used to increase both lesion-based and exam-based performance, i.e., fraction of true positive exams where an exam is true positive if cancer is found in the MLO *or* CC views, in comparison to a single-view CAD system ([5,6]). In contrast to the other research, which mostly explores neural networks or linear discriminant analysis, the probabilistic methodology proposed in the current study has the advantage of providing not only strong predictive power but also explicit modelling of expert knowledge and insight in the results obtained–properties desired especially by medical domain experts.

2   Causal Probabilistic Modelling

2.1   Bayesian Networks

A *Bayesian network* is defined as a pair BN $= (G, P)$, where G is an acyclic directed graph (ADG) $G = (V, E)$ and P is a joint probability distribution of the random variables X. There exists a 1–1 correspondence between the nodes in V and the random variables in X; the (directed) edges, or arcs, $E \subseteq (V \times V)$ correspond 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.,

$$A \perp\!\!\!\perp_G B \mid C \quad \Longrightarrow \quad X_A \perp\!\!\!\perp_P X_B \mid X_C \ ,$$

where A, B and C are sets of nodes of the ADG G and X_A, X_B and X_C are the corresponding sets of random variables, indexed by A, B, and C. 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*), by taking into account the conditional independence information represented by the ADG.

2.2   Causal Independence Models

Causal independence arises when multiple causes (parent nodes) lead to a common effect (child node) through interaction of independent uncertain processes. Causal independence models provide a way to specify interactions among random variables in a compact fashion [7]. A definition of the notion of causal independence is given following the one from [8].

The general structure of a causal-independence model is shown in Fig. 1; it expresses the idea that causes $C_1, \ldots, C_n$ influence a given common effect E through intermediate variables $I_1, \ldots, I_n$; the intermediate variable I_j is considered to be a contribution of the cause variable C_j to the common effect E. The *interaction function* f represents in which way the intermediate effects I_j, and indirectly also the causes C_j, interact. This function f is defined in such way that when a relationship between the I_j's and $E = true$ is satisfied, then it holds that $f(I_1, \ldots, I_n) = true$; otherwise, it holds that $f(I_1, \ldots, I_n) = false$.
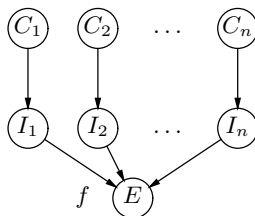

Fig. 1. Causal-independence model

Note that each variable I_j is only dependent on its associated cause C_j and the effect variable E. Furthermore, the graph structure expresses that the effect variable E is conditionally independent of each cause C_j given the associated intermediate variable I_j. It is assumed that absent causes do not contribute to the effect, i.e., $P(I_j = true \mid C_j = false) = 0$. As an example from the breast cancer domain, we can consider the regions detected by a single-view CAD system as cause variables and presence of breast cancer as the effect variable.

2.3   Exact and Threshold Functions

A natural class of interaction functions f are the *symmetric* Boolean functions where the order of the arguments does not matter. There are 2^{n+1} of such functions, with n the number of arguments; typical examples are the logical OR, AND and XOR. In the breast cancer domain, for example, the order of regions does not play a role in determining whether the breast is or is not cancerous.

A useful feature of symmetric Boolean functions is their decomposability in terms of *exact* Boolean functions. The exact function e_k checks whether there are *exactly* k trues among its arguments, i.e., $e_k(I_1, \ldots, I_n) = true$, if $\sum_{j=1}^{n} I_j = k$. In decision making under uncertainty there is a natural tendency to aggregate available uncertain information until a threshold is passed. The *threshold function* τ_k is a symmetric Boolean function that allows us to model this principle; it checks whether there are *at least* k trues among its arguments, i.e., $\tau_k(I_1, \ldots, I_n) = true$, if $\sum_{j=1}^{n} I_j \geq k$. Note that the logical OR function is a threshold function τ_k with $k = 1$ and the AND function is a threshold function τ_k with $k = n$. The conditional probability of the effect variable E given the causes $C_1, \ldots, C_n$ in a noisy threshold model with interaction function τ_k is given by:

$$P_{\tau_k}(e \mid C_1, \ldots, C_n) = \sum_{k \leq l \leq n} \sum_{e_k(I_1, \ldots, I_n)} \prod_{j=1}^{n} P(I_j \mid C_j) \ . \tag{1}$$

From Equation (1), it follows that $P_{\tau_k}(e \mid C_1, \ldots, C_n) \geq P_{\tau_{k+1}}(e \mid C_1, \ldots, C_n)$, for each $k \geq 0$, i.e., with a lower value of threshold k the probability of the effect e is non-decreasing. A more detailed description of exact and threshold functions can be found in [8]. Causal independence models with threshold functions are the basic elements for the model presented in the next section.

3   Causal Modelling for Mammographic Analysis

3.1   Two-View Analysis

The objective of a two-view mammographic analysis is to determine whether or not a breast exhibits cancerous characteristics by establishing correspondences between regions in the two breast views. Fig. 2 depicts the general multi-view detection scheme used in this study.

A lesion (cancerous object) is represented by a circle in a view. Clearly, if a lesion is detected in both views, the breast is cancerous and the patient has breast
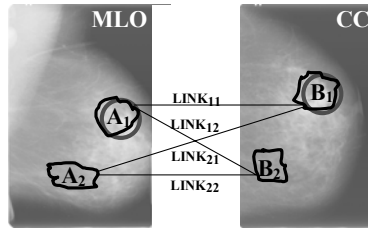
Fig. 2. Schematic representation of mammographic two-view analysis with automatically detected regions. The circle represents a lesion (cancerous object).

cancer. An automatic single-view CAD system attempts to establish whether there are regions that are suspicious for cancer in each view separately. In the figure regions A_1 and B_1 have correctly been detected as lesions, i.e., these are true positive (TP) regions, whereas regions A_2 and B_2 are false positive (FP) regions. Since we deal with projections of the same physical object—the lesion—correspondence between lesions is represented by a *link* (LINK_{ij}) between regions in each view, A_i and B_j. To every link, a value $\text{LINK}_{ij} = \ell_{ij}$ is assigned, where $\ell_{ij} \in \{\text{TPTP}, \text{TPFP}, \text{FPTP}, \text{FPFP}\}$. For every region, breast, exam and patient, a class with values of *true* (presence of cancer) or *false* is assumed to be provided by pathology or a human expert.

3.2   Probabilistic Model

The architecture of our model for two-view mammographic analysis is inspired by the way radiologists analyse images. They do this by distinguishing several levels in the interpretation process. At the lowest (image) level, radiologists look for suspicious regions with cancer characteristics. If suspicious regions are observed on both views of the same breast, then the (individual) suspiciousness of these regions increases implying that a lesion is likely to be present. As a result, the whole breast as well as the exam (patient) is considered suspicious for cancer.

The first steps—identifying suspicious regions and establishing links between them on both views of the same breast—have already been tackled in previous research conducted by our group. Here, we build upon the resulting systems to model the following stages in the mammographic analysis as described above. Fig. 3 presents an overview of the probabilistic model.

We start by modelling the two-view dependencies between the regions in MLO and CC. For each of the four link values ℓ_{ij} we consider the links LINK_{ij} with their respective set of multi-view features **MVFeat** and the correspondence scores $\text{CorrSc}(\text{LINK}_{ij})$, $\text{CorrSc}(\text{LINK}_{ji})$ obtained from the system described in [2]. We have used logistic regression to reliably compute the conditional probability distribution $P(\text{LINK}_{ij} = \ell_{ij} \mid \text{CorrSc}(\text{LINK}_{ij}), \text{CorrSc}(\text{LINK}_{ji}), \textbf{MVFeat})$. Thus, for every link LINK_{ij} we obtain four probabilities corresponding to each link value and every link probability is symmetric.
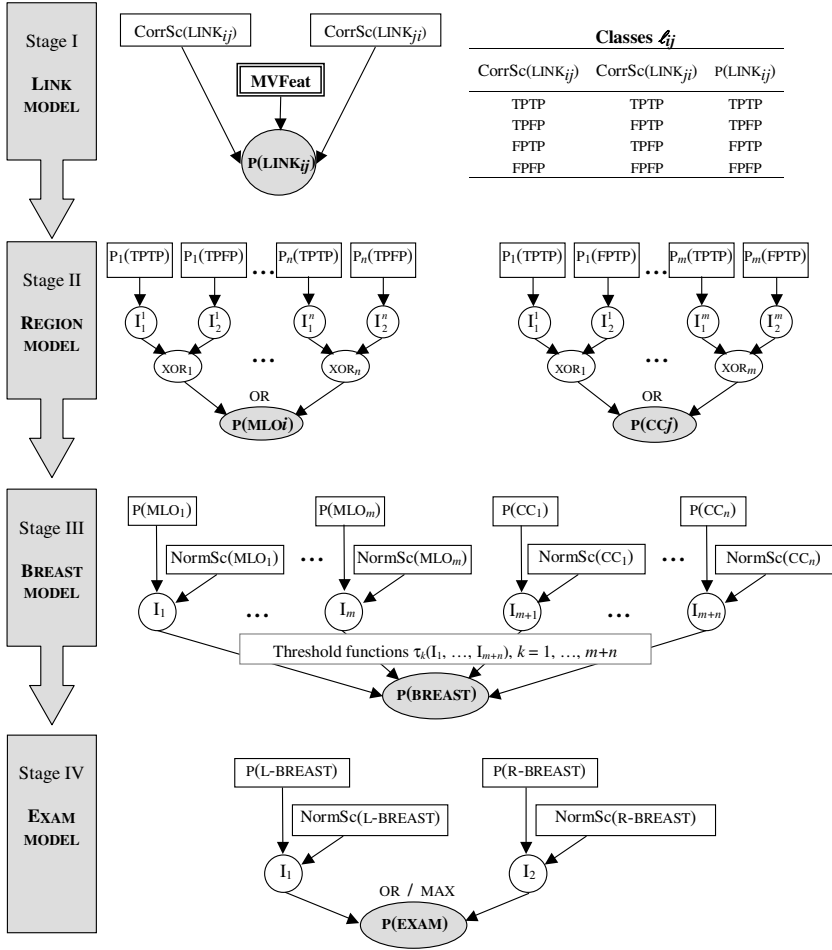
Fig. 3. Multi-stage causal probabilistic model for two-view mammographic analysis

At the second stage, we compute the probability of a region being cancerous given the link information about the regions in the complementary view. This is done by combining link probabilities obtained from the first stage using a causal independence model, where the link probabilities are the cause variables and the region probability is the effect variable. In computing the probability of a region of MLO being cancerous, we combine only the link probabilities for the classes TPTP and TPFP as they correspond to a TP region of MLO. With respect to a region of CC, the link classes considered are TPTP and FPTP. These link probabilities interact through the XOR function, as only one of them can be true. Next the logical OR is used to represent the knowledge that the probability of a region being cancerous is true if at least one of the link probabilities is true.

At the third stage, we focus at the breast level where the region probabilities from the respective MLO and CC views are combined using a causal independence model with a threshold function. In the combining scheme, we also use the suspiciousness measure for the region (NormSc(MLO$_i$), NormSc(CC$_j$)) computed by the single-view CAD system ([2]), which is already a good indicator for discriminating between normal and cancerous regions. By varying the threshold k from 1 to $m + n$ (the maximum number of regions in the breast), we try to get insight into the causal interactions between the regions and the breast. One can expect that models with small threshold values would be able to distinguish well between cancerous and normal breasts whereas models with larger values of k might not be able make the distinction. This expectation follows from the fact that breast cancer in its early stages is mostly unifocal, i.e., located in a single region, and not observed on multiple locations in the breast (view).

At the last stage, we combine the probabilities for the left and right breast and their respective single-view measures for suspiciousness (NormSc(BREAST)) to compute the probability for an exam being cancerous. Two combination functions are used and compared: the logical OR and the MAX function.

Finally, having patients with more than one exam (current and prior(s) available), we compute the probability for the patient having cancer by taking the probability of her current (most recent) exam, which is presumably most informative, i.e. $P(Patient = cancerous) = P(CurrentExam = cancerous)$.

3.3   Data Description

The data set contains 392 (332 current + 60 prior) exams from which 218 (185 current + 33 prior) were cancerous. The exams of one patient were considered as independent. All exams contained both MLO and CC views. All cancerous breasts had one visible lesion in at least one view, which was verified by pathology reports to be cancerous. Lesion contours were marked by a mammogram reader.

For each image (mammogram) we have a number of regions detected as suspicious by the single-view CAD system presented in [2]. This number varies between 1 and 5 per image (2 and 10 per breast). For each region, based on the ground-truth data, we have a class value of $true$ (TP) if the detected region hits a cancerous finding and $false$ (FP) otherwise. Every region from MLO was linked with every region in CC. Every link was described by a set of multi-view features. For every link we assigned one of the four link class values depending on the region class values. We assign binary classes of $true$ (cancerous) and $false$ (normal) for a breast, exam and patient based on the ground-truth information.

3.4   Training, Evaluation and Results

The proposed model has been built, trained and tested using the Matlab-based Bayesian Network Toolbox ([9]). The evaluation of the model is done using ten-fold cross validation with the same data split as the one used in [2]. For every split of the data, the test set is used only for testing and never for training at different stages of the model. Thus, we used an unbiased evaluation procedure.

Table 1. AUCs obtained from `MV-CAD-Causal` and `MV-CAD-kNN` at a *link* level

Link class	Current exams		All exams	
	MV-CAD-Causal	MV-CAD-kNN	MV-CAD-Causal	MV-CAD-kNN
TPTP	0.935	0.918	0.936	0.914
TPFP	0.838	0.660	0.844	0.650
FPTP	0.888	0.809	0.881	0.785
FPFP	0.887	0.829	0.874	0.824

Table 2. AUCs obtained from `MV-CAD-Causal` and `SV-CAD` at a *breast* level

k	Current exams		All exams	
	MV-CAD-Causal	SV-CAD	MV-CAD-Causal	SV-CAD
1	0.919		0.902	
2	0.917	0.904	0.899	0.895
3	0.872		0.859	

Table 3. AUCs obtained from `MV-CAD-Causal` and `SV-CAD` at an *exam* level

k	Current exams			All exams		
	MV-CAD-Causal		SV-CAD	MV-CAD-Causal		SV-CAD
	MAX	OR		MAX	OR	
1	0.903	0.899		0.889	0.879	
2	0.897	0.892	0.877	0.884	0.882	0.865
3	0.897	0.891		0.883	0.877	

At a link level the performance of our multi-view model is compared with the multi-view model based on k-Nearest Neighbour (`MV-CAD-kNN`) presented in [2]. At a region, breast and exam level, the benchmark for comparison is the single-view CAD system (`SV-CAD`). For the latter, the likelihood for a view, breast and exam being cancerous is computed by taking the likelihood of the most suspicious region. The comparison analysis is done using the (Free-response) Receiver Operating Characteristic ((F)ROC) curve and the Area Under the Curve (AUC). We next report the test results for all and current (patient) exams at all levels.

Link level. Table 1 presents the AUCs at a link level obtained from `LinkModel` and `MV-CAD-kNN`. The results show that our link model outperforms `MV-CAD-kNN` for all link values, with a considerable difference for the link types: TPFP, FPTP and FPFP. Except the classification improvement, our method has the advantage of making the links symmetric, i.e., for every two regions there is only one probability per link class.

Region level. To evaluate the performance of both multi- and single-view CAD systems at a region level we use the FROC curve which plots the lesion-detected and localized fraction (y-coordinate) vs. the average number of false positives
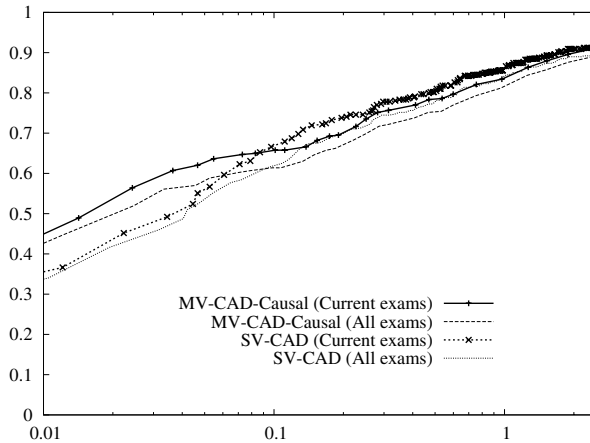
Fig. 4. FROC curves for the performance of `MV-CAD-Causal` and `SV-CAD` based on all and current exams at a *region* level

per image (x-coordinate). Fig. 4 depicts the results for the multi- and single-view CAD system for all and current exams. It is clear that taking two-view information into account helps increasing the cancer detection rate while keeping the number of false positives per image very low. This trend is even clearer for the regions in the current exams. This is a desirable outcome for screening programs where radiologists operate at high specificity (low false positive) rates because of the very low cancer incidence rates.

Breast level. At this level, we compare the performance of `BreastModel` using different threshold functions τ_k, where $k = 1, \ldots, 10$ with respect to `SV-CAD`. Table 2 presents the results for the three best multi-view models for all and current exams. The results confirm our expectation–the correct detection of at least one ($k = 1$) or two cancerous regions ($k = 2$) is sufficient to classify the breast as cancerous. For these thresholds the multi-view CAD system outperforms the single-view CAD system. For threshold functions with $k \geq 4$ the performance of `BreastModel` drops significantly reaching AUCs of 0.506 when $k = 10$ for all and current exams. This result is in line with the domain knowledge and the local nature of the early developing breast cancer.

Exam level. At the screening practice the most important question eventually is whether or not a patient is suspicious for cancer and needs to be referred for further examination. To answer this question, for the current comparison study, we focus on the results obtained from the last stage of our model (`ExamModel`) based on the best three `BreastModel` with $k = 1, 2, 3$; see Table 3. For all and current exams, we observe overall improvement in the breast cancer detection rate achieved by our two-view system. It is interesting to note that the causal independence modelling in `ExamModel` helps improve the performance of the exam model even if the performance of the breast model is less satisfactory as for `BreastModel` with a threshold of $k = 3$. Furthermore, we notice that

using MAX as a combination function for the breast probabilities leads to a better distinction between cancerous and normal exams than using the logical OR. A possible explanation might be that the latter tends to overestimate the probability of normal exams by considering both breasts, whereas the MAX function seems to be more appropriate given that in screening mammography mostly one of the breasts is cancerous.

4   Conclusion

We presented a unified Bayesian network framework for two-view mammographic analysis motivated by the radiologists' practice and the organization of the domain. The foundation of the framework is based on the notion of causal independence where the interaction between the variables is modelled by threshold functions. The definition of link used in this paper better captures the correspondences between the regions in both views, in comparison to our previous approach ([1]), where we used binary links. Through experimental results we showed that for lower thresholds the proposed two-view probabilistic model not only is in line with the domain knowledge but also outperforms a single-view CAD system by increasing the breast cancer detection rate for low false positive rates, both at a location and a patient level.

References

1. Velikova, M., Samulski, M., Lucas, P.J.F., Karssemeijer, N.: Improved mammographic CAD performance using multi-view information: A Bayesian network framework. Physics in Medicine and Biology 54, 1131–1147 (2009)
2. Samulski, M., Karssemeijer, N.: Matching mammographic regions in mediolateral oblique and cranio caudal views: A probabilistic approach. In: Proceedings of SPIE, Medical Imaging, vol. 6915 (2008)
3. Good, W., Zheng, B., Chang, Y., Wang, X., Maitz, G., Gur, D.: Multi-image cad employing features derived from ipsilateral mammographic views. In: Proceedings of SPIE, Medical Imaging, vol. 3661 (1999)
4. van Engeland, S., Karssemeijer, N.: Combining two mammographic projections in a computer aided mass detection method. Medical Physics 34, 898–905 (2007)
5. Paquerault, S., Petrick, N., Chan, H., Sahiner, B., Helvie, M.A.: Improvement of computerized mass detection on mammograms: Fusion of two-view information. Medical Physics 29, 238–247 (2002)
6. Qian, W., Song, D., Lei, M., Sankar, R., Eikman, E.: Computer-aided mass detection based on ipsilateral multiview mammograms. Acad. Rad. 14, 530–538 (2007)
7. Heckerman, D., Breese, J.S.: Causal independence for probability assessment and inference using Bayesian networks. IEEE Trans. on SMC–A 26, 826–831 (1996)
8. Visscher, S., Lucas, P.J.F., Schurink, C.A.M., Bonten, M.J.M.: Modelling treatment effects in a clinical Bayesian network using Boolean threshold functions. Artificial Intelligence in Medicine (2008)
9. Murphy, K.: Bayesian Network Toolbox (BNT) (2007),
   http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html