MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-1291

Using Local Context Information to Improve Automatic Mammographic Mass Detection

Marina Velikova[a], Peter J.F. Lucas[a] and Nico Karssemeijer[b]

[a] *Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands*
[b] *Department of Radiology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands*

Abstract

Despite their promising application, current Computer-Aided Detection (CAD) systems face difficulties, especially in the detection of malignant masses–a major mammographic sign for breast cancer. One of the main problems is the large number of false positives prompted, which is a critical issue in screening programs where the number of normal cases is considerably large. A crucial determinant for this problem is the dependence of the CAD output on the single pixel-based locations initially detected. To refine the initial detection step, in this paper, we propose a novel approach by considering the context information between the neighbouring pixel features and classes for every initially detected suspicious location. Our modelling scheme is based on the Conditional Random Field technique and the mammographic features extracted by image processing techniques. In experimental study, we demonstrated the practical application of the approach and we compared its performance to that of a previously developed CAD system. The results demonstrated the superiority of the context modelling in terms of significantly improved accuracy without increase in computation efforts.

Keywords:

Breast cancer, Computer-assisted decision making, Mammography.

Introduction

Breast cancer is a disease that no woman wants to be diagnosed with, but if it does occur, finding it as early as possible can save woman's life. Worldwide there have been developed screening programs where mammographic exams are performed in asymptomatic women. The enormous workload and the very low incidence rates of breast cancer (3-10 per 1000) in these programs, however, create various challenges for radiologists in the interpretation of mammograms.

The current advances in computer technology and screening digitalisation led to the rapid development of Computer-Aided Detection (CAD) systems. In screening programs these systems might be especially useful as a second reader for improving the mammographic analysis [1]. 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.

Computerized programs are currently employed for the detection and classification of masses and microcalcifications–the two major mammographic signs of breast cancer. Applications have shown that CAD tends to perform better in identifying malignant microcalcifications compared with masses [2]. Masses are more difficult to detect due to the great variability in their physical features and similarity to the breast tissue, especially at early stages of development. Hence, the prompt of the current CAD comprises not only the cancer but also a large number of false positive (FP) locations–undesired result in screening where the reading time is crucial.

In this paper, we focus on improving mass detection on mammograms by reducing the number of FPs while keeping the true detection rates high at the very first step of CAD–detection of suspicious pixel-based locations. In our previous work this selection is mostly based on the information provided by a single pixel [3], [4]. This contradicts, however, the basic decision-making principle of radiologists, accounting for context information in the region of interest. To incorporate this knowledge into a CAD system, we propose a novel approach for mass detection by explicitly modelling the dependencies of the neighbouring pixel features and classes. We use Conditional Random Fields (CRF), introduced in [5], as a powerful probabilistic tool to represent context dependencies using undirected graphs. CRFs model the class distribution given a set of observed features, which makes them especially suitable for sequential labelling and classification tasks such as mass detection. The main advantage of CRFs over generative models such as Markov Random Fields (MRF) is their conditional nature, relaxing the need for making a lot of independence assumptions among the observed features, required by MRFs to allow tractable inference.

A number of previous works considered context information for breast cancer detection. In [6] the authors propose an approach using MRF to segment breast masses achieving 90% sensitivity at 2 FP detections per image. In [7], a random field model is used to detect microcalcifications on mammograms. Alternative directed-graph approach based on Bayesian net-

works has been proposed in [8] to predict the type of breast cancer represented by microcalcifications, reaching prediction performance of a human expert.

In contrast to the previous studies where only the dependencies between the neighbouring labels are considered, in this work we also account for the dependencies between the observed features in the neighbourhood employing the CRF representation. In such a way we can better represent the mass characteristics and tackle the problem of noise in the labelling. Next we present our mammographic CRF model.

Materials and Methods

Mammographic Discriminative Model

In screening mammography, it is known that cancer develops mostly in one of the breasts and at one location. In the initial stage of development, malignant masses are often small and may have subtle characteristics, which hardly distinguish them from the surrounding tissue. This implies that scanning the whole breast for a mass would be a cumbersome task in terms of time and computation efforts. Therefore we propose to consider not the whole image but only those parts of it, which indicate certain cancer characteristics. Hence our method consists of two main steps: (1) Detection of suspicious pixel-based locations (PxL) and (2) Building a CRF model for each of the detected PxLs.

Initial pixel-based mass detection

In this step we use the pixel-based mass detection scheme from the CAD system presented in [3]. For all pixels in the segmented breast area, this algorithm calculates at each location i a set of 5 local mass features: 3 for stellate (star-like) lesions and 2 for focal masses. A 3-layer neural network classifier, supervised by the pixel labels $\in \{-1, 1\}$, combines these features into the so-called mass likelihood l_i – the likelihood that a mass is present at location i. The method is applied at a high sensitivity level to ensure that most masses are found. This implies that a large number of FPs would also be observed because the detection was based on 5 features only.

CRF modelling of the initially detected local masses

To further filter out the FP locations, for every selected pixel q we build a CRF model, which is defined with respect to a system of neighbourhood around q. The CRF modelling we propose in this paper is based on the approach in [9], which we extend to account for the specifics of the mammographic mass detection.

For every image we define a set $Q = \{q \mid q \in B,\ l_q > \theta\}$, where B is the set of all pixels in the breast area and θ is a threshold value. We then use every $q \in Q$ as a central point to construct a region (grid) R_q with a size of $M \times N$ pixels. Since the pixel information is very sensitive to small variations, we create non-overlapping groups of $m \times n$ pixels and we call these pixel groups *sites*. For every site s we compute the feature vector $\boldsymbol{y}_s$ as the mean of the feature values of the pixels in the site. We compute the site label x_s by assigning –1 (cancer) if sufficient cancer information is available in the site (e.g., 40% or more of the pixels are cancerous) and 1 otherwise.

Based on this construction we define a CRF over the labelling $\boldsymbol{x}$ given the observations $\boldsymbol{y}$ as an undirected graph with a set S of vertices corresponding to the random variables x_s, which satisfy the Markov property with respect to the graph stating that $p(x_s \mid \boldsymbol{y}, x_{S\setminus\{s\}}) = p(x_s \mid \boldsymbol{y}, x_{\mathbf{N}s})$ where $S\setminus\{s\}$ is the set of vertices excluding $\{s\}$ and $\mathbf{N}_s$ is the set of neighbours of s. In other words, the random variable x_s is conditionally independent of all other variables given the observations and its neighbours. We represent the neighbourhood (direct dependencies) of s in the graph by the set of edges between s and the 4 nearest neighbour sites. Then the conditional distribution of the labelling $\boldsymbol{x}$ given the observations $\boldsymbol{y}$ is factorised into a product of two potential functions, capturing the compatibility of a certain variable configuration: one local potential associating given site with a certain class ignoring the site neighbours and one interaction potential for the site dependencies. After log-transformation, the conditional distribution is defined as:

$$p(\boldsymbol{x} \mid \boldsymbol{y}) = \frac{1}{Z}\exp\left(\sum_{s\in S}\varphi(x_s, \boldsymbol{y}_s) + \sum_{s\in S}\sum_{r\in \mathrm{N}_s}\psi(x_s, x_r, \boldsymbol{y})\right). \tag{1}$$

where Z is a normalizing constant ensuring that p is a valid probability distribution, φ is the local potential and ψ is the interaction potential.

To account for the feature dependencies between the sites, for each site s we consider 2 scales at which $\boldsymbol{y}_s$ is computed–the original $m \times n$ scale and $2m \times 2n$ scale. This results in the doubled feature vector $\boldsymbol{Y}_s$. Next we define $\boldsymbol{t}_s = h(\boldsymbol{Y}_s)$, where h is a second-degree polynomial expansion, to explicitly model feature interdependencies. Given the computed features, the potentials are then defined by

$$\varphi(x_s, \boldsymbol{y}_s) = \exp(x_s \boldsymbol{w}_s \boldsymbol{t}_s) \tag{2}$$

$$\psi(x_s, x_q, \boldsymbol{y}) = \boldsymbol{\beta}\Phi(x_s x_q \boldsymbol{v}_{sq} \boldsymbol{t}_{sq}), \tag{3}$$

where Φ is the probit function, $\boldsymbol{t}_{sq}$ is the edge feature vector created by concatenation of $\boldsymbol{t}_s$ and $\boldsymbol{t}_q$, and the parameters $\boldsymbol{w}_s$ and $\boldsymbol{v}_{sq}$ are weight vectors. Note that we modify the standard definition of the edge potential as done in [9], by adding a parameter set $\boldsymbol{\beta}$ in order to account for different weights with respect to the label interaction. This set contains four values corresponding to the four label combinations. Since we aim at mass detection, the pairwise cliques for which $x_s = x_q = -1$ are expected to get larger potentials, facilitated by the larger value of $\beta_{-1,-1}$. Schematic presentation of the CRF modelling for mass detection is presented in Figure 1.

Parameter learning and inference

The parameters that need to be optimised in the learning process are $\Omega = \{\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{\beta}\}$. Since in the construction of CRFs the observed features are dependent, the corresponding parameters need to be learned simultaneously. This implies that the

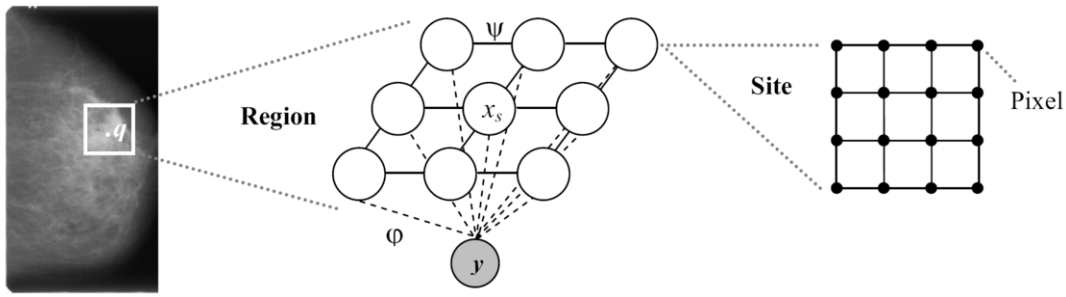

Figure 1- CRF modelling for automatic mass detection

evaluation of Z in Equation (1), which sums over all assignments, becomes intractable for exact inference. Various approximate techniques have been used to overcome this problem: mean-field, loopy belief propagation, stochastic gradient. Here we consider the stochastic gradient descent technique, which has shown fast converging properties. The true gradient is approximated by computing the gradient at an individual training sample. After each step, the new parameter set is adjusted using the approximated gradient based on the update rule of the form $\Omega_d = \Omega_{d-1} - \eta\nabla\Omega_i$, with d ranging over the number of training samples, ∇ being the gradient and η being the learning rate. As prior parameters we assumed that the values of $\boldsymbol{w}$ and $\boldsymbol{v}$ are drawn from the normal distribution with zero mean and variance σ^2.

The goal of inference in our modelling scheme is to find for a new test region the most likely output label sequence $\boldsymbol{x}^*$ over the region sites given the observed site features. We use as inference method belief propagation. To compute the single label probability of the whole region we take the mean over the label sequence on 16×16 window around the centre q. Only the pixels in the close proximity of q are considered as they are most likely to contribute to the correct classification.

Experiments and Results

Design

The proposed approach was evaluated using data collected in the Dutch breast cancer screening program. The dataset consists of mammograms for 164 cases from which 118 were with biopsy-proven cancer and 46 were normal. The total number of images was 220 from which 93 were normal. For every image, which has been digitized at 200 micron (0.2 mm) spatial resolution, the breast area was segmented by the CAD system and a likelihood image was created. The pixel-based locations with a mass likelihood ≥ 0.5 were selected as most suspicious. This resulted in 494 pixel mass locations from which 361 (73%) were normal. Each of these locations was then used as centre to construct a region of size 128×128 pixels (~2.5×2.5cm) for CRF mass detection modelling. Every region was partitioned into 32×32 non-overlapping sites and site features were computed as described above.

Given the classification problem at hand, we trained the CRFs using a random subset of 88 locations with cancer only. Our choice was motivated by the fact that the regions built on normal locations do not provide information about the difference in the labels and features of the sites. The remaining cancerous and all normal regions were used as a test set, accounting for 61% normal test cases. We note that *no* FPs from cancerous cases were used in training and testing, since in the evaluation procedure at a case level they are not included. The results presented next are based on the test data.

We compare the results of the CRF modelling scheme (CAD-CRF) with those obtained from CAD on the likelihood image (CAD-Lik) at two levels: region and case. We evaluated the performance at a region level by the Free-response Receiver Operating Characteristic (FROC) curve whereas at a case level we used the Receiver Operating Characteristic (ROC) [10]. The Area Under the Curve (AUC) is used as a standard evaluation technique. The CRF modelling was done using the freely available CRF toolbox for Matlab [9], where we applied necessary changes to implement the proposed model.

Results

Accuracy

Region level. Figure 2 shows test sample regions and the results on their detection obtained by CAD-CRF. In the test samples the black areas represent cancer whereas the white regions are FP regions detected by CAD-Lik. Clearly a large number of FPs are filtered-out by CAD-CRF, as indicated by the grey regions in the test results on Figure 2. Figure 3 presents the results from both CAD-CRF and CAD-Lik at a region level. We observe improvement in the mass detection rate for the proposed method, starting already at a FP rate of 0.023 per image. At a FP rate of 0.1 per image, for example, CAD-CRF detects 84.6% of the masses whereas for CAD-Lik the detection rate is 71.8%. Furthermore, the context modelling allowed the detection of all masses at a FP rate of 0.3 per image, whereas for the original likelihood image this was achieved at a FP rate of 0.58. The areas under the FROC curves for the FP range [0.01,1] are: AUC(CAD-CRF) = 0.782 and AUC(CAD-Lik) = 0.705, indicating considerable improvement in the detection of the cancer at a local level.

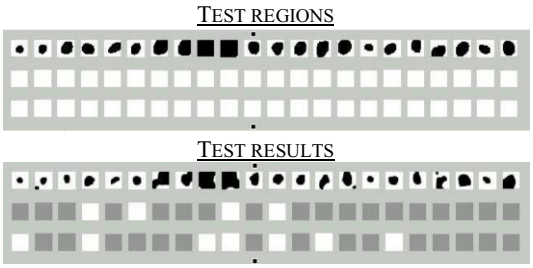

Figure 2- Test regions and corresponding results obtained from CAD-CRF. The black areas represent cancer, the white regions are FPs and the grey regions are true negatives

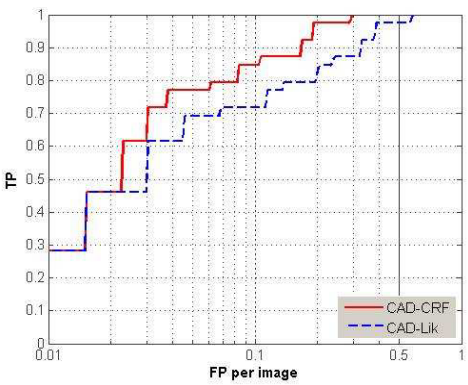

Figure 3- FROC curve for the performance of CAD-CRF and CAD-Lik at a region level

Case level. The probability for a case being cancerous is computed by taking the maximum probability for the regions in the case. Figure 4 presents the ROC curves with the respective AUCs for CAD-CRF and CAD-Lik at a case level. We performed a one-side statistical test for the difference between the AUCs using ROCKIT [11]. The obtained *p*-value is 0.6%, with 95% confidence interval of (0.01, 0.09) indicating the significant improvement in the cancer detection rate achieved by CAD-CRF. Figure 5 also demonstrates the improved performance of CAD-CRF based on filtering out the FP regions detected by CAD-Lik on two normal cases.

Next we examine the effect of the parameters $\boldsymbol{\beta}$ in Equation (3) on the performance of CAD-CRF. After training the model we obtained a vector $\boldsymbol{\beta}$ with values $\beta_{1,1} = 25.03$, $\beta_{1,-1} = 0.5$, $\beta_{-1,1} = 0.5$ and $\beta_{-1,-1} = 50.06$. These results confirm our expectation that the better classification of cancerous regions is facilitated by the large weight for cancerous neighbour pixels. Further evidence supporting this hypothesis is obtained by comparing the results obtained from CAD-CRF where all values of $\boldsymbol{\beta}$ are fixed to 1, as done in [9]. Then the case AUC drops to 0.786 because 17 cancerous test cases are missed.

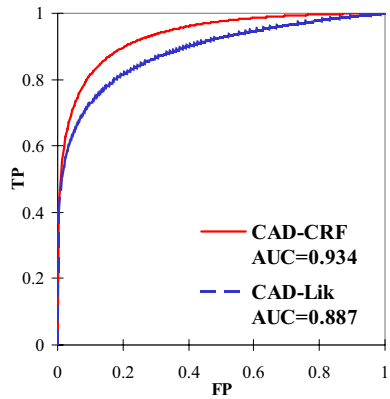

Figure 4- ROC curve for the performance of CAD-CRF and CAD-Lik at a case level

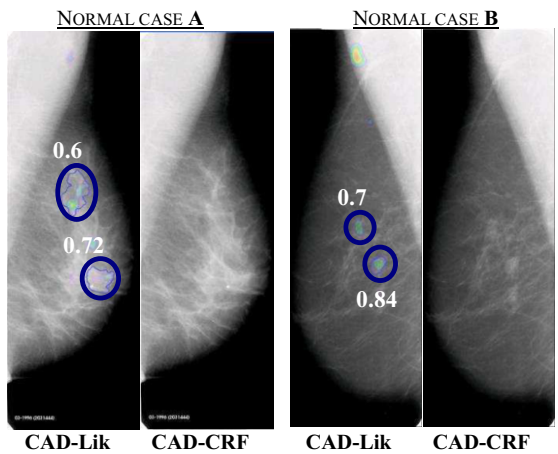

Figure 5- Performance of CAD-CRF and CAD-Lik on two normal cases. The numbers represent the pixel-based likelihoods computed by CAD-Lik for detected FP locations. For the same locations the likelihood computed by CAD-CRF is 0

Computational issues

The learning took 15 iterations in 4.8 minutes on a 2.33GHz Intel Core™2 Duo machine. The learning parameter η was set to 0.0009. The average time of processing a test region was 0.22 ± 0.04 seconds. This implies that for an image with an average number of 3 detected mass locations the testing takes around a second. Thus the improvement in accuracy achieved by using CRF modelling comes at *no* computational costs and can easily be applied in the screening setting.

Discussion

The motivation for the experimental study was to show that extending an existing CAD system by using context information (1) provides significantly better prediction decisions at a region and case level (2) without increasing the computational time. On the one hand, modelling context information leads to significantly better distinction between masses and FPs in comparison to the original CAD system, leading to considerable reduction of FPs while maintaining the high sensitivity rate reached by the CAD system. This improvement is crucial for screening programs where the number of normal cases is much larger than the number of cancerous cases and the sensitivity of human eye decreases with increasing the case disproportion. We also note that the proposed system obtained 100% sensitivity at a FP rate of 0.3 per image, which falls in the radiologists' operating FP range and it is considerably better than the result obtained in [6], for example. Next the current CAD-CRF makes better decisions based on a set of only 5 original pixel-based features. This might also explain the imperfect segmentation achieved by CAD-CRF as shown in Figure 2. However, region segmentation is performed in the subsequent stage of the CAD system once the suspicious pixel-based locations are detected. On the other hand, the extended system remains efficient since the added filtering step does not require additional computational time for providing the prediction decision. This makes the proposed context-based approach also attractive to apply in a screening program where time is an important factor.

Current problem in the original CAD system, which remains in its extended context version, is that a cancerous lesion can be detected by multiple regions. One solution is to perform the testing of the learned classifier for a new region on a larger area consisting of overlapping windows of fixed size. Additional advantage of this approach might be that cancers missed by the CAD-Lik system can eventually be detected.

Conclusion

We presented a novel approach for mass detection based on CRF modelling using local image-based context information. We tested the approach on screening data and compared the results with those obtained from the likelihood image built by a CAD system. The proposed method outperformed the benchmark by reducing significantly the number of FP regions and misclassified normal cases while maintaining high cancer detection rates. The additional filtering step added to the CAD system showed no increase in the computational effort. These results are encouraging for building reliable, fast and accurate CAD systems, which can be employed for assisting radiologists in mammographic screening programs. To make progress in this direction, future work aims to extend the proposed model by including temporal information on the mass development and conducting large-scale experiments with digital screening data.

Acknowledgments

Funded by the Netherlands Organization for Scientific Research under BRICKS/FOCUS grant number 642.066.605.

References

[1] Destounis S. Computer-aided detection and second reading utility and implementation in a high-volume breast clinic. Applied Radiology, 33(9), 2004.

[2] Morton MJ, Whaley DH, Brandt KR, and Amrami KK. Screening mammograms: interpretation with computer-aided detection–prospective evaluation. Radiology, 239:375–383, 2006.

[3] Te Brake GM and Karssemeijer N. Single and multiscale detection of masses in digital mammograms. IEEE Trans. Med. Imag. 18(7):628–639, 1999.

[4] Velikova M, Samulski M, Lucas PJF, and Karssemeijer N. Improved mammographic CAD performance using multi-view information: A Bayesian network framework. Physics in Medicine and Biology, 54:1131–1147, 2009.

[5] Lafferty J, McCallum A, and Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proc. 18th ICML, 2001.

[6] Li HD, Kallergi M, Clarke LP, Jain VK, and Clark RA. Markov random field for tumor detection in digital mammography. IEEE Trans. Med. Imag. 14(3):565–576, 1995.

[7] Karssemeijer N. A stochastic model for automated detection of calcifications in digital mammograms. Proc. 12th IPMI 511:227–238, 1991.

[8] Burnside ES, Rubin DL and Shachter RD. Using a Bayesian network to predict the probability and type of breast cancer represented by microcalcifications on mammography. Proc. 11th MEDINFO 107(1):13–17, 2004.

[9] Kumar S and Hebert M. Discriminative Random Fields: A discriminative framework for contextual interaction in classification. Proc. 9th IEEE ICCV 2:1150–1157, 2003.

[10] Hanley JA and McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology, 143:29–36, 1982.

[11] Metz CE, Wang PL, and Kronman HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. In Information Processing in Medical Imaging. Nijhoff, 432–445, 1984.

Address for correspondence

Marina Velikova, Radboud University Nijmegen, Dept. Model-Based System Development, Institute for Computing and Information Sciences, PO Box 9010, 6500 GL Nijmegen, The Netherlands, email: marina.velikova@gmail.com