

Overview of the domain of breast cancer

Nivea and Marina

Maart, 2007

Abstract

1 Introduction

The aim of this brief report is to present a general overview of the domain of breast cancer. More precisely, we describe the terms and concepts related to domain, and we try to derive some general rules expressing the expert knowledge in the area. In addition, we discuss briefly the CAD systems at UMCN and the ones based on Bayesian networks that will be of interest for the further work on the B-Screen project.

2 The problem of breast cancer

Robra *et. al* present in the [4] an account on the epidemiology and risk factors for breast cancer. As stated there, breast cancer is the most frequent malignant tumor among women, for instance, in the most European countries and North America. Besides, it also stands as the most common cause of death in women aged between 35 and 55 in these countries. Developed countries show higher rates than less developed ones [8].

Analysis of geographical and ethnical discrepancies in the incidence of breast cancer indicate that an association between both social factors (such as access to medical facilities and diagnosis at an advanced stage of the disease) and nonhereditary lifestyle factors with breast cancer.

Factors that have been established as risk factors for the development of breast cancer include age, genetic predisposition, history of breast cancer, history of benign breast disease, exogenous hormones, exposure to radiation, and dietary and other lifestyle factors.

2.1 Risk factors

The greatest risk factor for nonhereditary breast cancer is old age [4]. The incidence increases rapidly with age during the reproductive years and at a

slower rate after menopausal years [8, 7]. Postmenopausal women are at a lower risk than premenopausal women of the same age [8].

The age at which a woman begins her menstrual cycle is inversely related to breast cancer risk. That is, the earlier the age of menarche, the higher the risk of breast cancer. Yet another reproductive factor is the protective relation of childbearing against the disease [8]. As Key *et al.* presents, women who have had one full-term pregnancy have around a 25% reduction in breast cancer risk in average, when compared to nulliparous women. The number of pregnancies affects a woman's risk independently of age at first birth [8, 7]. Besides, it seems not yet clear whether spontaneous or induced abortion during reproductive year is associated (positively or negatively) to the risk of breast cancer.

Hormonal stimulation, through oral contraceptives or estrogen-replacement therapy, increases the breast cancer risk. The increase is about 25% in users of combined oral contraceptives. Research on the effect of hormonal therapy for the menopause indicates an increased risk for women under this therapy, which increases depending on the duration of the use. However, the excess risk reduces after cessation of use, reaching a normal level after (about) 10 years [4, 8].

Nutritional and other life style factors, such as alcohol consumption and moderate levels of physical activity, seem important when analyzed as (small/moderate) risk factors. Dietary determinants of risk (e.g., high-fat diets, consumption of meat, fibres and vegetables) cannot be easily (and accurately) linked to breast cancer, due to errors in measurement during studies [8]. Light to moderate alcohol consumption seems to linearly increase the risk of breast cancer: an intake of (around) 3 units per day (approximately 30g) shows a 30% increase in breast-cancer risk [8]. In postmenopausal women, relative overweight has been strongly associated to breast cancer, while physical activity lower the risk [4]. While in [4] it is said that a relation between physical activity and breast cancer cannot be established for premenopausal women, in [8] we find that (not entirely consistent) data shows somewhat stronger risk decreasing for premenopausal than postmenopausal women. This leads us to believe that more conclusive research is needed in order to make that relation clear. Other factor that might show some influence on breast cancer risk include adult height. Meanwhile, smoking and stressful life do not show any significant associability.

Ionizing radiation can also be listed as having a causal relation to breast cancer. Studies indicate that risk from multiple exposure at relatively low dosage, such as multiple chest fluoroscopies in the treatment of tuberculosis, is similar to the risk of one large dose of similar magnitude (perhaps as seen in the high-dose x-rays treatment for acute mastitis) [7]. It remains unclear the effect of very low doses of radiation, as in medical diagnosis exposure, and exposure after 40 years [8].

Elevation in breast cancer risk occurs in women with history of benign breast cancer disease, being generally associated to the so called proliferative lesions [8, 7]. Mammographically dense breast tissue have an increased risk for breast cancer [4, 8]. Increased risk can be also seen in women with previous breast cancer diagnosis, having a three- to four-fold increase of a second primary cancer in the contralateral breast [4, 7]. The risk for a secondary cancer is specially

high in women with family history for breast cancer.

Hereditary factors exist in about 10% of all breast cancer cases [4]. Women with a first-degree relative who has breast cancer are (about) twice as much in risk of developing the disease, compared to the normal population [8, 7].

3 Mammography

Some of the risk factors presented in the previous section may be associated to preventive measures against the development of breast cancer, some other certainly not. This makes some secondary measures, such as screening mammography, to have a crucial value in the course of prevention or, at least, early detection of suspicious lesions.

That is the reason why there is an effort on introducing breast cancer screening, as well as educational, programs. In the Netherlands, women are encouraged to undertake a mammographic examination every two years after becoming 50-years-old. Identification of an early-stage lesion can be decisive for establishing a curative therapy (in contrast to a palliative treatment necessary when the disease is beyond cure).

Next, we give a brief overview on identification and classification of breast abnormalities on mammographic images¹. Later in this section we describe the main features of circumscribed lesions, calcifications, architectural distortion and asymmetry. The information found in this section is mainly based on the descriptions given in Chapter 8 of [4].

3.1 Terminology

First, in order to avoid ambiguity and misunderstanding, we want to describe the terminology used throughout the text (when referring to certain aspect concerning the domain).

Three main concepts are defined to refer to different levels of data representation: (i) *image (mammogram)*—mammographic image of one breast in certain view; (ii) *exam (screening)*—mammographic screening made at given time, which includes four images (for both breasts and two views); (iii) *case*—a patient under study that has several exams over a period of time.

Mammographic views (or, projections) are the usually the mediolateral oblique (MLO) or craniocaudal (CC) positionings.

In mammograms, a area of interest will be denominated *region*. This region might present some *features* that might suggest (or not) a certain level of suspiciousness for breast cancer development. When analyzing those features (e.g., size, density, location) we might also refer to the region as *lesion* or *abnormality*.

In terms of temporal aspects, mammographic exams are usually referred to as *current*, representing the last, up-to-date result; and *prior*, to refer to the exam previous to the current. Some cases have more than two screenings, in

¹Notice that here we just concentrate on the identification of abnormalities without discussing other issues such as technical quality or analyzability of mammograms.

which case we can talk about a series of 1 to n priors. In this case, n represents the oldest screening.

Finding is one and the same region of interest found in different views of the same exam, and/or different exams as well. That is, having observed a lesion in one view, we talk about correlating it to a (possible) lesion in the other view and also in the previous exam.

3.2 On viewing mammograms

Usually, when visualising and analysing mammograms, it is not the major changes that bring concern, but the subtle ones that are not always clearly or easily detectable. Besides, the identification of one abnormality might result in having other abnormalities being overlooked (either in the same or in the other breast).

In order to identify subtle changes it is of vital importance to have a preceding mammogram of the patient, which we call *prior*, available for comparison. However, it is necessary to bear in mind that minor changes might only be noticeable after long intervals (e.g., when comparing the current exam to the one previous to the prior exam).

Even though the occurrence of a lesion can be expected in any area of the breast, most cancers are located in the upper outer quadrant (and less frequently in the retromamillary area).

Essentially, breast cancer is an unilateral, segmental disease. However, 1–2% are bilateral when detected. A solitary lesion with secondary malignant features such as local retraction of glandular tissue may reveal its true character.

Comparing the two breasts for symmetry is also important for the identification of changes. For instance, contextual information such as the patient's breast density pattern can be taken into account through that comparison. Sometimes, breasts images can present a certain difference in density due to uneven compression during examination. In such cases, it should be made clear that the difference is related to an unequal amount of glandular tissue - features such as architecture and distribution of adipose tissue should remain similar.

For the analysis (or, interpretation) of the abnormalities found in a mammographic image, the clinical context play an important role. Information such as patient's age, undertaking of hormone therapy, previous surgery or radiotherapy, and presence of (other) risk factors are considered.

As Bun states in [4], real mammographic changes located in the breast are classified into one of the following groups:

- Circumscribed densities
- Stellate densities
- Calcifications (micro and macro)
- Distortion of the architecture and asymmetry
- Changes of the skin

- Lymph nodes
- A combination of two or more of the above

Superficial malignant process, benign inflammation or previous surgery, can cause subcutaneous changes and thickening of the skin. This appears in the mammogram as a blurring of the (usually sharp) interface between skin and subcutaneous fat. Even though a subcutaneous carcinoma may cause a retraction of thickened skin, mammographic thickening of the skin does not represent an absolute sign of such a tumor infiltration. On the other hand, skin thickening may be the only evidence of a mammographically occult cancer.

The presence of one or more enlarged, homogeneous, dense lymph nodes might help on the correct identification of a malignant process. However, changes in lymph nodes might have other causes. For instance, dermatological disorders. It might be that mammography alone cannot be sufficient for such a differentiation.

A more detailed review of the other possible mammographic changes are presented below. We describe here the main features used for an (reasonably) accurate identification and classification of those.

3.3 Circumscribed lesions

Circumscribed densities are classified into *nodular* or *stellate lesions*. Eccentrically located lesions might be visible only in one of the mammogram projections.

In terms of the shape, nodular lesions varies from round to oval. A coarse-lobulated aspect generally indicates a benign nature for lesions, while fine lobulations are suspicious for malignancy.

The most relevant criterion for differing between malignant and benign densities is the *margin*. A nodular lesion with entirely or partially ill-defined margin is suspicious for malignancy, specially when the presence of spicules is also observed. Margins of benign lesions are generally smooth and well-defined - a sharp margin, as to speak.

Unrestrained proliferation of cells in an irregular directional pattern make the margin of the lesion ill-defined. With continuous growth, the blurring aspect of the image may increase, making the center of the lesion to become quite dense. Short spicules with broad basis represent infiltration of tumor into adjacent normal tissue.

When comparing an identified lesion with its surrounding tissue, many malignant ones appear as being hyperdense, while some can be isodense. When a lesion has the same density as the surrounding tissue it might remain mammographically invisible. “Of all palpable breast cancers, 5 – 7% are mammographically occult.” As a consequence, it seems common-sense that a palpable lesion that is occult mammographically asks for further diagnostic evaluation.

A lesion with spicules longer than the diameter of its central density is referred to as a stellate lesion. A direct relationship between the diameter of the tumor and the length of the spicules might not exist. Besides, in early tumors, spicules might be poorly discernible. Stellate lesions may also lack a

distinct central mass (and might be only seen as an architectural distortion or asymmetry).

Despite the use of a good assessment regarding the criteria mentioned, a number of nodular lesions are difficult to be identified. A nodular lesion, specially when seen as a new or a growing lesion, may represent carcinomas.

Differentiation between malignant lesions and benign ones, such as cysts, fibroadenomas or hematomas, may become possible when clinical history is taken into account, or maybe only through further examinations as ultrasound and biopsy.

3.4 Calcifications

Calcifications exceeding $2mm$ are referred to as *macrocalcifications*, being 'almost always' benign. The ones less than $1mm$ are referred to as *microcalcifications*, and might indicate malignancy (even though not invariably caused by a malignant process).

Variation of density and size is a sign of malignancy. The differential diagnosis might be aided by the number of calcification when taken into account also size and shape. Newly developed group of microcalcifications, or a group with an increased number of calcifications, has an indication for malignancy. Furthermore, often also suspicious for malignancy are the microcalcifications of irregular shape (and different size and density) which appear crumbled, usually referred to as *granular* microcalcifications.

"Sometimes, calcifications can be restricted to a certain region of the breast, resulting in a segmental or lobar arrangement that raises the possibility of malignancy." Regularly dispersed calcifications over the major portions of the breast indicates a benign abnormality.

Depending on shape, size, orientation and distance between calcifications, a small clustered group of calcifications might indicate malignancy. Specially if multiple small groups are present in the same breast segment.

The presence of microcalcifications alone indicates a high probability of ductal carcinoma in situ (DCIS). Branching linear microcalcifications in a linear arrangement point to poorly differentiated ductal carcinoma. Coarse granular calcifications - in special when forming a single cluster - imply a higher probability of a poorly differentiated carcinoma than fine granular calcifications. Multiple clusters of fine granular microcalcifications indicate well-differentiated ductal carcinoma.

Benign-appearing macrocalcifications in a lesion with ill-defined margin usually require a histological analysis for a differential diagnosis.

3.5 Architectural distortion and asymmetry

"In addition to presenting as a solitary nodular or stellate lesion, breast cancer can also manifest itself as architectural distortion of glandular structures or as asymmetry."

Lesions are less obvious when the central tumor mass is small relative to the radiating structure (or spicules). Even less obvious still when that mass is absent (and only the radiating structure is visible). If the radiating structure is not visible as such, the only sign of malignancy may be a subtle architectural distortion or asymmetry of the fibroglandular tissue.

Asymmetry of the glandular structures might be the only sign of a invasive lobular carcinoma (ILC). The development of asymmetry (in a previously normal mammogram) is suspicious, specially when it is associated to an increase in tissue density. “Almost one-third of the cases of ILC are only detected by locally altered architecture or by asymmetry.”

Besides local shrinkage of breast tissue, ILC relatively often indicesthickening of the skin, and retraction of skin and mamilla. However, shrinkage with retraction can also be present in benign lesion such as radial scar.

3.6 General Rules

Abnormal region: appears different from other regions of the same breast or from the corresponding region in the other breast.

General rule for referral (Dutch screening program):

If a lesion is observed in both views then refer the case. Otherwise, examine other lesion characteristics and risk factors.

Specific rules for referral:

- If a region is very suspicious, e. g., highly-spiculated or newly developed, even when the region is present in one view only.
- If a finding is newly developed or grows in size compared to previous exam.
- If mammographic signs are subtle or occult but there is information for pain, palpable mass (lump), skin retraction or nipple discharge. This is especially applicable for highly dense breasts.

Rules for classifying a finding as malignant:

- If the finding is a mass with ill-defined or spiculated margin, irregular or lobular shape, average size of 3 cm, highly-dense, located mostly in the upper-outer quadrant, and have vertical orientation. Benign masses are most likely with round or oval shape, circumscribed margins, and low- or isodensity.
- If the finding is architectural distortion and there is no information for previous biopsy or operation.
- If the finding is asymmetry that appears to be concentrated toward its center, i.e., with high density and relatively small size (so-called focal asymmetry).

(Micro)calcifications are expected to be invariant from one view to the other - i. e., they should be observable in both views. And the same holds for certain spiculated masses. However, some other masses - e. g., lobular carcinoma (in situ) - might appear in only one view. One reason might be the position of the lesion. For instance, those ones which are located close to the muscle in the MLO view might not appear in the CC view. Another reason is that the breast tissue may also “hide” the lesion in one of the views.

4 CADs and Breast cancer screening

As explained in the previous section, an early diagnosis of breast cancer is important for an increase in the chance of curing the disease. In this section we discuss the application of computer-aided systems on the reading of mammograms, as a means of assisting with the identification and interpretation of breast abnormalities, and (potentially) decreasing the need of multiple readings - or missed lesions.

We overview CAD systems’ state-of-art and evaluate what remains to be done/improved.

4.1 CAD UMCN

First we present the general scheme of the CAD system developed at UMCN [9, 10]. It consists of four steps, shortly described below:

1. Segmentation of the mammogram. For MLO views, the mammogram is segmented into three regions - background area, breast, and pectoral muscle. For CC views, the mammogram is segmented two regions - background area and breast.
2. Initial detection of suspicious pixel-based locations. For each pixel in the breast area a number of features are computed, which are related to the detection of spiculation and a focal mass. Based on these features, a neural network classifier is used to compute a normalized likelihood for malignancy (so-called level) for each pixel. The level value ranges between 0 and 255. The corresponding image, called likelihood image, is then smoothed and a number of locations (pixels) with a level above a certain threshold are selected (so-called peaks).
3. Region extraction based on the detected locations. By using each of the detected peaks as a seed point, a number of regions with their features are segmented. These features are continuous and include: *local area features* (spiculation measures, focal mass measures, originally computed peak level and two normalized level measures, relative location, shortest distance to the skin line and the pectoral edge), *region features* (mean values of spiculation measures, mean values of focal mass measures, density, contrast, number of microcalcifications in the region, border measures, linear texture, gray level variance)

4. Region classification as true abnormalities and false positives. Based on a selected subset of region features, a classifier is used to compute a malignancy score, which indicates the likelihood that a region is malignant. Different classifiers are used: LDA, SVM, K-nn, NN. Training and testing a classifier is done by cross-validation scheme.

Data description. Every mammographic image corresponds to a breast and a particular view (MLO or CC). For each image there is a **ground-truth file (GTR)**, which provides information about: exam ID, view, exam date, exam type (normal, recall, clinical, prior1, prior2), digitizer, pixelsize, pathology, type of medical detection (screening, interval, diagnostic, other), breast density (fat, scattered densities, heterogeneously dense, extremely dense), total number of annotated regions within mammogram and feature description of each region (regionID, malignant or benign type, mass or non-mass, location, and contour annotation).

The name of a GTR file consists of the exam ID number, the view ('c' for CC and 'm' for MLO), and the breast ('r' for right and 'l' for left).

4.2 CAD systems for prediction of breast cancer risk by using Bayesian networks

In [6, 5], the authors propose a Bayesian network (MammoNet) to assist mammographic diagnosis of breast cancer. The network combines five patient-history features (age, age at menarche, age at first live birth, number of relatives with breast cancer, previous biopsy), two physical findings (pain, nipple discharge) and 15 mammographic findings grouped in mass, microcalcification, arch.distortion, asymmetry, and developing density. Mass and microcalcifications are further represented by a number of features. The knowledge base of the system is constructed by using peer-reviewed medical literature, census data, health statistics reports, and subjective probability estimates provided by expert mammographers.

Other two CAD systems, based on Bayesian networks and presented in [3, 2], are developed to determine the probability of breast cancer represented by mammographic microcalcifications. The systems combine patient-history features (age, family history and hormone replacement therapy) and mammographic findings. The structure of the networks is built by using the BI-RADS lexicon [1]. Instead of learning the network's parameters from data, the authors chose to use pre-existing knowledge about the probabilistic relationships among the variables on the basis of medical literature, census data, results from a large randomized trials and subjective probability estimates provided by expert mammographers.

5 What now

What directions we plan to consider in the development of our model:

- which risk factors and mammographic lesions will be considered
- intention to improve classification between mass versus architectural distortion/asymmetry
- accurate representation of expert knowledge.

This might be achieved through the *correct* representation of regions, findings, vertical + horizontal linking, breast structure comparison, context information, etc.

References

- [1] *Breast Imaging Reporting and Data System (BI-RADS)*. American College of Radiology, Reston, VA, 1993.
- [2] E. S. Burnside, D. L. Rubin, J. P. Fine, R. D. Shachter, G. A. Sisney, and W. K. Leung. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results. *Journal of Radiology*, 2006.
- [3] E. S. Burnside, D. L. Rubin, and R. D. Shachter. Using a bayesian network to predict probability and type of breast cancer represented by microcalcifications on mammography. *Medinfo*, pages 13–17, 2004.
- [4] R. H. D. J. Dronkers, J. H. C. L. Hendriks and G. Rosenbusch. *The practice of mammography: pathology, technique, interpretation, adjunct modalities*. Thieme: Stuttgart, New York, 2002.
- [5] C. E. Kahn, L. M. Roberts, K. A. Shaffer, and P. Haddawy. Construction of a bayesian network for mammographic diagnosis of breast cancer. *Computers and Biology and Medicine*, 27(1):19–30, 1997.
- [6] C. E. Kahn, L. M. Roberts, K. Wang, D. Jenks, and P. Haddaw. Preliminary investigation of a bayesian network for mammographic diagnosis of breast cancer. *Annual Symposium on Computer Applications in Medical Care*, 19:208–212, 1995.
- [7] J. L. Kelsey and M. D. Gammon. The epidemiology of breast cancer. *A Cancer Journal for Clinicians*, 41:146–165, 1991.
- [8] T. J. Key, P. K. Verkasalo, and E. Banks. Epidemiology of breast cancer. *Oncology*, 2:133–140, 2001.
- [9] S. Timp. *Analysis of temporal mammogram pairs to detect and characterise mass lesions*. PhD thesis, Radboud University Nijmegen, 2006.
- [10] S. van Engeland. *Detection of mass lesions in mammograms by using multiple views*. PhD thesis, Radboud University Nijmegen, 2006.