

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Metrological Assessment of a CAD System for the Early Diagnosis of Breast Cancer in Digital Mammography

Arianna Mencattini and Marcello Salmeri
*Dept. of Electronic Engineering, Univ. of Rome Tor Vergata
Italy*

1. Introduction

Based on recent statistics from the International Agency for Research on Cancer (Ferlay et al., 2010), breast cancer account for 10.9% of all cancers diagnosed and ranks as the fifth cause of cancer death in the world. Although incidence rates are increasing, mortality rates are stable, representing an improved survival rate. This improvement can be attributed to effective means for the early detection as well as to significant improvement in treatment options, exposure, etc. Mammography is, at present, the only viable method for detecting most of tumors early enough for effective treatment, without unnecessary biopsies, or other invasive procedures. Therefore, screening mammography in women aged 40 to 70 years is currently the effective strategy to reduce breast cancer mortality. Early detection of invasive breast cancers is associated with better prognosis than waiting for women to become symptomatic. However, detecting the early signs of breast cancer is challenging because the cancerous structures have many features in common with normal breast tissue. Moreover, the accuracy of interpretation of screening mammograms is affected by several factors, such as image quality, the radiologist's level of expertise, and the high volume of cases. Recent statistics show that in current breast cancer screenings 10% – 25% of the tumors are missed by the radiologists (Burrell et al., 2001; Cheng, 2003; Nishikawa, 2007). Missed cancers are due to many reasons: low disease prevalence, breast structure complexity, finding subtleties, and radiologist fatigue. To accomplish with these difficulties, different methods have been analyzed: first of all, double reading, which provides either double perception or double interpretation of lesions. It has been demonstrated that a single radiologist is more accurate when reading mammograms methodically than quickly and that two observers achieve an improvement in detection rate of 5% – 15% (Mazzarella & Bazzocchi, 2007). Obviously, this procedure is too expensive, complex, and time consuming especially in screening programs where a huge number of mammographic images have to be read. The development of computerized systems as second readers represents an alternative. Researchers have been developing algorithms to detect mammographic abnormalities for more than 30 years with the aim of either automating mammographic interpretation or providing a tool that could enhance human film-reading accuracy. Computer-aided detection (CADe) and diagnosis (CADx) systems are widely used in mammography, where signs of breast cancer are often very subtle. Both systems involve the use of computer algorithms to detect patterns in images associated with signs of disease and to assign them a malignancy index. This result

should attract the clinicians’ attention to potentially abnormal regions in mammograms. In the recent years, the authors have developed a CADe-CADx system (CAD in the following), called Assisted Breast Cancer Diagnosis Environment (ABCDE) (Salmeri et al., 2009), to assist radiologists in the early detection of breast cancer. Beside the development of specific algorithms for the enhancement of abnormalities and the identification of pathological structures, the performance evaluation of each block of the whole CAD system has been also performed. In order to understand the actual complexity of the complete validation of a CAD system in mammography, consider the left-hand part of the flowchart reported in Fig. 1 where all the steps involved in the early breast cancer detection and diagnosis are considered.

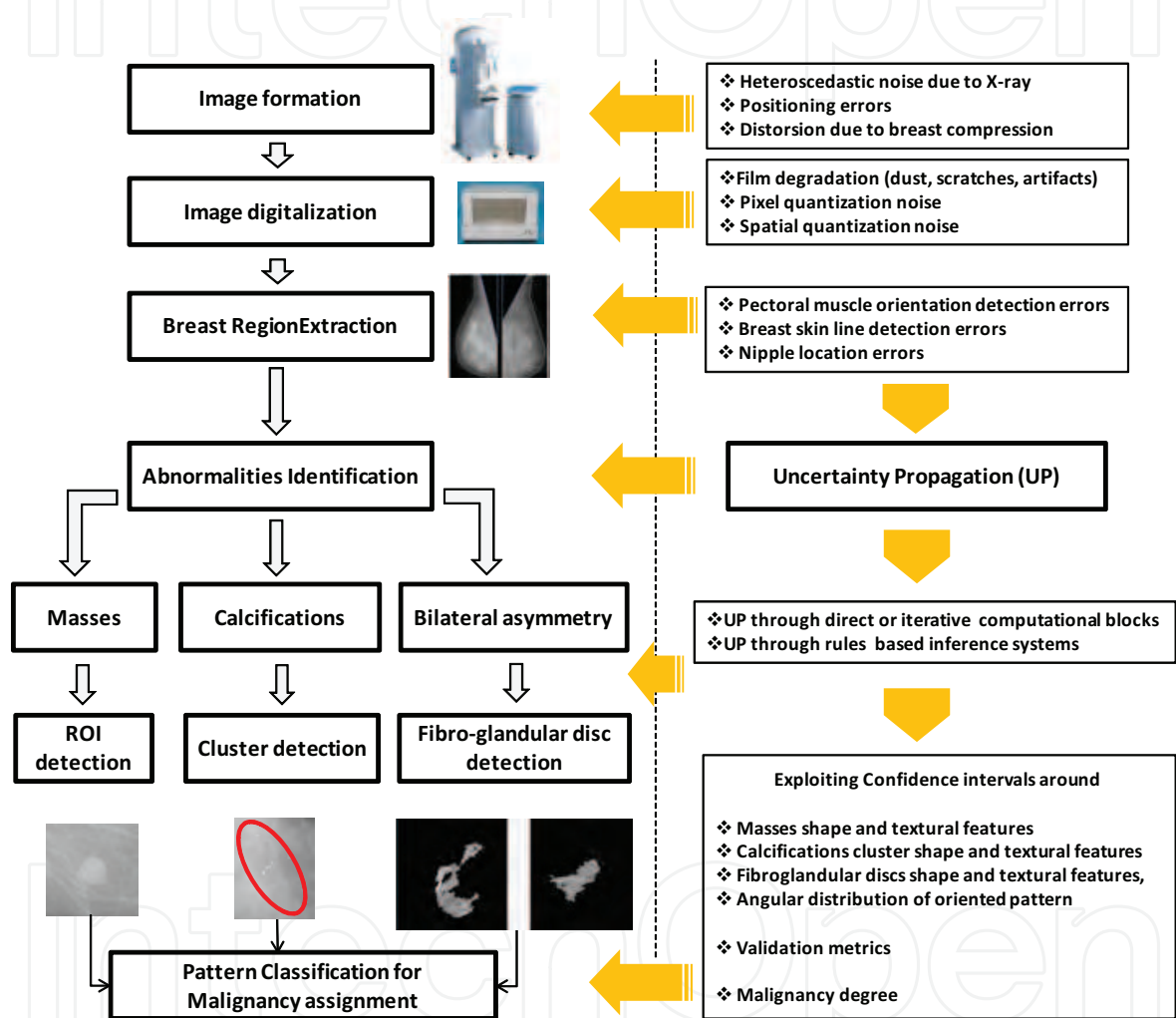


Fig. 1. Flowchart of the whole CAD ABCDE system. The right part shows the issues related to the uncertainty contributions handling and propagation through the blocks.

2. Motivations and objectives

This work stems from considerations of Wirth (Wirth, 2006) and Nishikawa (Nishikawa, 2007) regarding the performance evaluation of a CADe system. One of the major limitations in the design of novel computerized algorithms for breast cancer detection is related to the inherent difficulty of proving their effectiveness and improvement with respect to existing methods. In the context of performance evaluation a sequence of unresolved issues can be identified.

- The assessment of the whole CAD system does not coincide with the testing of individual components of the CAD system. The latter requires specific indicators according to each block and its functionalities, while the former coincides with testing classification ability of the system. In CADe systems, it means to test the capability of the system of discriminating between normal and pathological tissue, while in CADx systems, it is related to the ability of discriminating between malignant and benign abnormalities.
- In the literature, there are well-accepted characteristics needed to evaluate the performance of each algorithm in the CAD system: sensitivity, robustness, adaptability, accuracy and precision, reliability (or reproducibility), and efficiency. Each of them is defined in compliance with the definitions contained in reference documents such as (*International vocabulary of metrology. Basic and general concepts and associated terms (VIM)*, 2008). Unfortunately, exact limits beyond which the performance of the system is acceptable lack and are often verbally reported.
- Receiver Operating Characteristic (ROC) and Free-Response ROC (FROC) analysis are used to assess the classification performance of the whole CAD system. They plot sensitivity versus specificity or false positives per image/case. Optimal operating points can be derived based on the Youden-index, the distance from the ideal ROC curve, the optimal clinically acceptable false-detection rate. Unfortunately, often this is a subjective choice and depends on who is making the evaluation.
- Problems in the ROC analysis occur when more than two classes are used in the classification process, as for example in the classification of the lesion margins and shapes, or in algorithms used to automatically assess the breast tissue density.
- Standard ROC analysis does not implement the so called *failure assessment*, that is the analysis of the circumstances under which an algorithm fails and the correction of the failure causes. In some works, the authors verbally describe the specific cases where the algorithm fails, as for example in (Ferrari et al., 2004), examining in details the False Negatives produced by the algorithms and the possible causes.
- Datasets available is another important problem. Public databases such as MiniMIAS and DDSM are quite old and do not contain all the informations needed to evaluate the performance of a innovative CAD. In fact, modern issues in breast cancer detection involves important aspects such as bilateral asymmetry, architectural distortion, prior mammograms, interval cancers. The cited databases do not contain this kind of abnormalities, except for those reported in case there is also a lesion. Moreover, a few images are present with this kind of abnormalities and this limitation does not inspire full confidence in the final results produced by the CAD.
- The comparison of the various existing methods and possibly the introduction of improvements to one of them is extremely difficult if not impossible for two main reasons: the first one is that the algorithms are not completely described or documented so that it is not possible for authors to make a comparative analysis; secondly, the algorithms are often black-boxes so that testing involves only inputs and outputs, but not the internal decision-making process.

Many of the above mentioned problems can be partially solved or at least addressed considering a further important aspect in the CAD performance evaluation and algorithms testing, that is uncertainty sources modeling and propagation. In fact, considering all the uncertainty sources that can be present in the considered context of mammography,

performing a detailed modelization of each of them, constructing a dedicated propagation procedure through each block of the CAD, according to the uncertainty nature, will allow us to address the following topics:

- An analysis of sensitivity could be performed in order to extract the influence of each uncertainty source on each algorithm parameter and on the final output value. In this way, both the design and the testing steps will be improved and completed, allowing to introduce corrective factors or change the algorithm, test the robustness of more systems.
- Extended ROC and FROC analysis could be performed, providing, for each decision point (corresponding to a pair of values for the sensitivity and the specificity in the ROC curve for example), a confidence interval given a certain coverage probability. In this way, a direct failure analysis can be guided by this kind of analysis as will be shown in the following sections.
- According to the context, screening programs or diagnostic mammography, requested coverage probabilities can be different. Moreover, it can be interesting to assign a confidence interval to each output of the CAD and of each block, for any coverage probability, from an high sensitive choice (95%) to a more specific or operating case (80%). The application of a Monte Carlo analysis, as that described in (JCGM, 2008), allows us to simulate every kind of sensitivity analysis, starting from very few assumptions on the nature of the uncertainty sources able to affect the CAD output. Specific estimation procedures should be implemented to preliminary provide accurate modeling of the uncertainty sources.
- Novel inference mechanisms, such as those based on fuzzy logic theory, would allow us to transform a black-box system into a white-box system, where interaction with the physician and with the developers is simple and constructive.

In light of this, we have already partially addressed some of the above cited issues: breast masses automatic segmentation (Mencattini et al., 2011d; Rabottino et al., 2008) and the uncertainty propagation through the algorithm (Mencattini et al., 2010b), microcalcifications classification (Ferrero et al., 2010), performance evaluation of an automatic tumoral masses identification algorithm (Mencattini et al., 2010a), uncertainty propagation through the classification of breast masses (considering also the features extraction and the feature selection procedures), the uncertainty propagation by Random Fuzzy Variables (RFVs) through a denoising and enhancement procedure for the detection of microcalcifications (Mencattini et al., 2009b), and the noise variance estimation results under the assumption of signal-dependent noise (Mencattini et al., 2007; Salmeri et al., 2008).

3. Materials

The digital mammographic image can be obtained using Full-Field Digital Mammography (FFDM) or it can be obtained by digitizing a Screen Film Mammogram (SFM). The two kinds of images exhibit very different properties in linearity (relationship between pixel intensity and exposure to X-ray detector), contrast (contrast in SFM at low and high exposures is reduced with respect to FFDM), spatial resolution (40 – 50 μm in SFM, 100 μm in FFDM), and noise: in linear FFDM noise is proportional to the square root of the X-ray exposure, while at low exposures electronic noise is dominant; conversely, for logarithmic FFDM noise is proportional to the inverse of the square root of the X-ray exposure to the detector. The same holds for SFM, but changed according to the characteristic curve. Owing to these differences,

robust validation and testing procedures should consider images either from SFM and from FFDM. In order to map gray-levels to optical density the calibration curve of each acquisition system is used. Each acquisition system In particular, in this work we will consider three databases: the Digital Database for Screening Mammography (DDSM) (Heath et al., 1998), the Mammographic Image Database (MIAS) (Suckling et al., 1994), and a FFDM database from the San Paolo Hospital in Bari (Mencattini et al., 2011d). In the following, we briefly describe the characteristics of the three databases.

3.1 MiniMIAS database

The Mammographic Image Analysis Society (MIAS) is an organization of UK research groups interested in the understanding of mammograms and has generated a public database of digital mammograms available via the Pilot European Image Processing Archive (PEIPA) of the University of Essex, at: <http://peipa.essex.ac.uk/info/mias.html>. Films taken from the UK National Breast screening Programme have been digitized to 50 μm pixel edge with a Joice-Loebl scanning microdensitometer, a device linear in the optical density range $[0 - 3.2]$ and representing each pixel with an 8-bit word. The original database has been reduced to a 200 μm pixel edge and padded/clipped so that all the images are 1024×1024 and the mammogram is centered in the matrix. Fig. 2 reports four examples of images taken from MiniMIAS database.

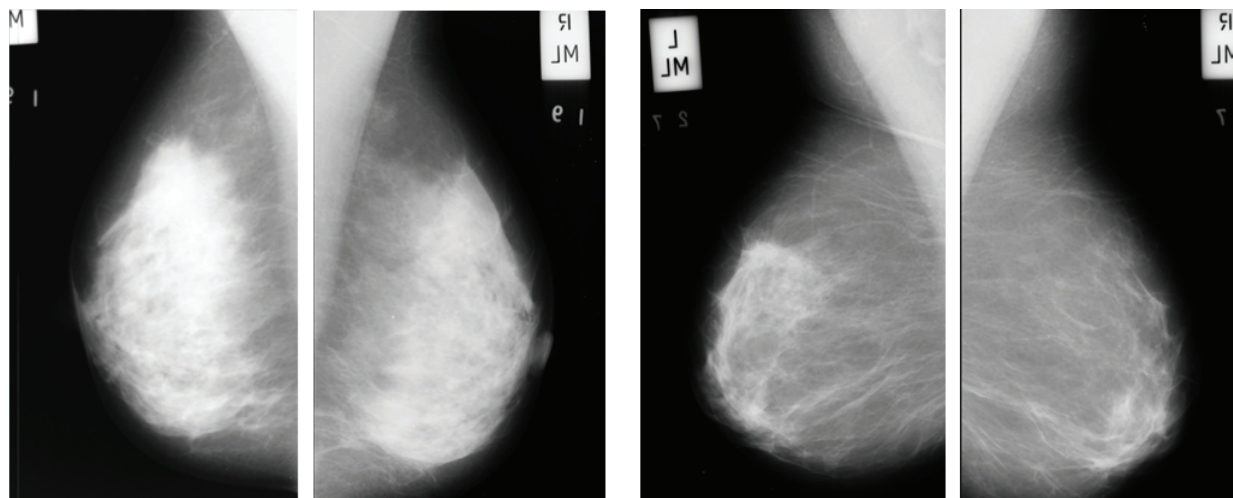


Fig. 2. Examples of mammographic images from the MiniMIAS database. (Figures-left) Normal case, right (mdb053) and left (mdb054) mammograms. (Figures-right) bilateral asymmetry case, right (mdb081) and left (mdb082) mammograms.

3.2 DDSM database

The Digital Database for Screening Mammography (DDSM) (Heath et al., 1998) was completed in 1999 and contains a total of 2620 cases. Primary support for this project was a grant from the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command. The DDSM project is a collaborative effort involving co-p.i.s at the Massachusetts General Hospital (D. Kopans, R. Moore), the University of South Florida (K. Bowyer), and Sandia National Laboratories (P. Kegelmeyer). Additional cases from Washington University School of Medicine were provided by Peter E. Shile, MD,

Assistant Professor of Radiology and Internal Medicine. Additional collaborating institutions include Wake Forest University School of Medicine (Departments of Medical Engineering and Radiology), Sacred Heart Hospital and ISMD, Incorporated. Each study includes two images of each breast, along with some associated patient information (age at time of study, ACR breast density rating, subtlety rating for abnormalities, ACR keyword description of abnormalities) and image information (scanner, spatial resolution, bpp). Images containing suspicious areas have associated pixel-level “ground truth” information about the locations and types of suspicious regions. Unfortunately, lesion boundary drawn by radiologists are not always reliable and often include also pixels from normal tissue. The images are digitized by four different scanners (HOWTEK-A/D, LUMISYS, and DBA) using spatial resolution equal to $[43.5, 50, 42]$ μm and a pixel resolution equal to 12 bpp for scanners (HOWTEK-A/D and LUMISYS) and equal to 16 bpp for scanner DBA. Each study contains four screening views, left and right mediolateral oblique (MLO) and cranio-caudal views (CC). Figure 3 reports three pairs of mammograms from DDSM, one for each scanner.

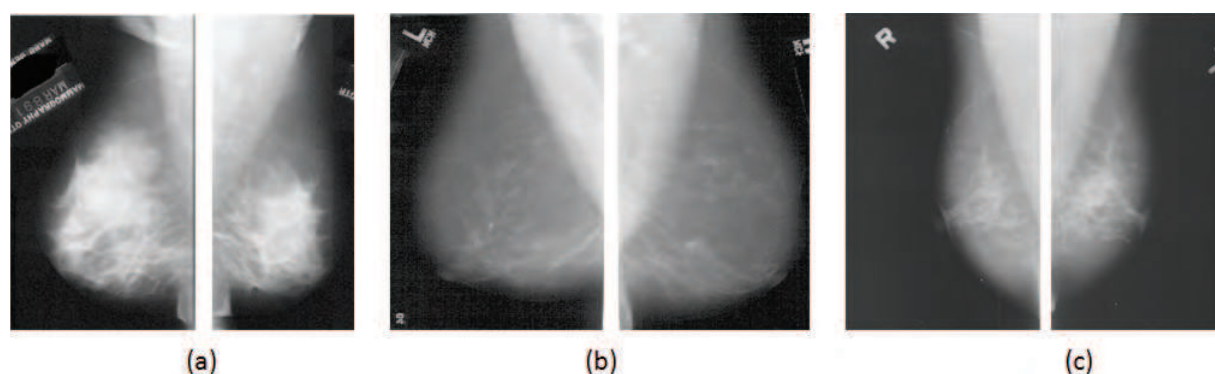


Fig. 3. Three pairs of DDSM images from three different scanners (a) DBA, (b) HOWTEK-A, (c) LUMISYS.

Images from scanner HOWTEK-D presents the same characteristics of scanner HOWTEK-A and are omitted.

3.3 FFDM database

Thanks to a collaboration with the San Paolo Hospital in Bari and with the Dept. of Mathematics at the University of Bari, we have access to Full Field Digital Mammographic (FFDM) images acquired using the Senograph 2000D ADS 17.3, GE Medical Systems, at a spatial resolution of $94 \mu\text{m}$ and a pixel resolution of 12 bpp. Each study contains two relevant series: a series including the four standard views (CC and MLO) for presentation state and a series including, if present, a screen save image reporting the lesion boundary manually drawn by radiologists. Currently, these images are used for the development and testing of an algorithm for the automatic extraction of the lesions boundary in order to assign a malignancy degree to each lesion detected by a radiologist. At the moment, the dataset includes 196 mammograms containing one or more benign or malignant massive lesions. The malignancy assessment is a part of the CADx section in ABCDE and is currently based on fractal analysis (Giuliato & Rangayyan, 2011; Mencattini et al., 2011d; Raguso et al., 2010). Figure 4 reports four FFDM images from San Paolo Hospital.

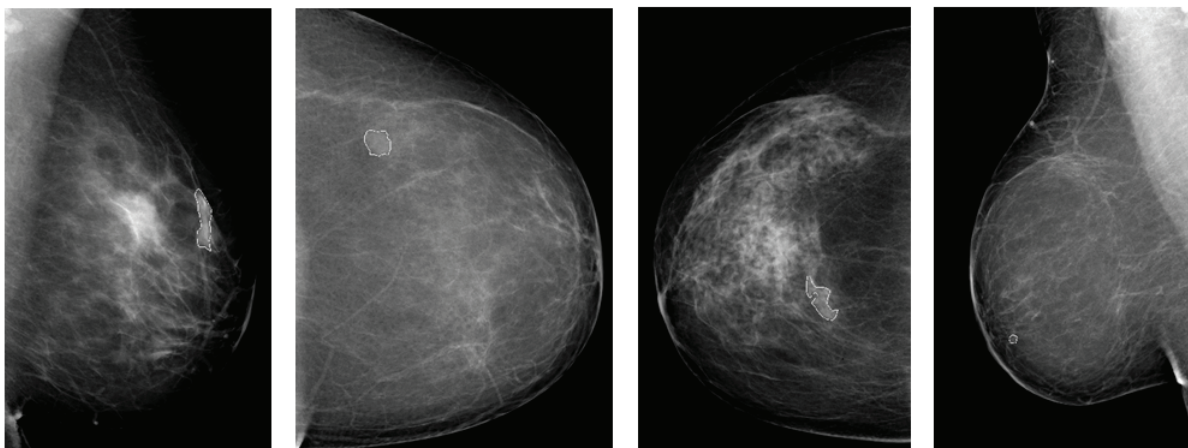


Fig. 4. Four FFDM images containing a massive lesion whose boundary has been drawn by a radiologist.

4. A Monte Carlo approach for CAD performance assessment

Recent reference documents (JCGM, 2008) revised almost totally the formal and practical definitions concerning measurement uncertainty, its modelization and propagation, as a part in the measurand estimation. There are two types of measurement error quantity that can occur during a measurement process: *systematic* and *random*. A systematic error (an estimate of which is known as a measurement bias) is associated with the fact that a measured quantity value contains an offset (that can be unpredictable and uncontrollable). A random error relies on the fact that when a measurement is repeated it will generally provide a measured quantity value that is different from the previous value. It is random in that the next measured quantity value cannot be predicted exactly from previous such values. The GUM (JCGM, 2008) provided a different way of thinking about measurement and in how to express the perceived quality of the result of a measurement. Rather than express the result of a measurement by providing a best estimate of the measurand, along with information about systematic and random error values (in the form of an “error analysis”), the GUM approach aims at expressing the result of a measurement as a best estimate of the measurand, along with an associated *measurement uncertainty*.

Let us denote with X_i any input quantities (measurands) and with Y the output quantity about which information is required. The output Y should be related to the inputs X_1, \dots, X_N by a measurement models that can be explicitly formulated $Y = f(X_1, \dots, X_N)$ or implicitly defined $h(X_1, \dots, X_N, Y) = 0$. Functions $f(\cdot)$ and $h(\cdot)$ should be estimated. The main stages of uncertainty evaluation are formulation, propagation, and summarizing.

- The *formulation* stage consists on defining the output quantity Y , identifying the input quantities on which Y depends (X_i), developing a measurement model relating Y to the input quantities (functions $f(\cdot)$ and $h(\cdot)$), and on the basis of available knowledge, assigning probability distributions to the input quantities (or a joint probability distribution to those input quantities that are not independent).
- The *propagation* stage consists of propagating the probability distributions for the input quantities through the measurement model to obtain the probability distribution for the output quantity Y . The propagation of distributions can be implemented in several ways: analytical methods, i.e. methods that provide a mathematical representation of the

probability density function (pdf) for Y ; uncertainty propagation based on replacing the model by a first-order Taylor series approximation (also called the law of propagation of uncertainty); numerical methods that implement the propagation of distributions, specifically using Monte Carlo method, that will be described in the following.

- The *summarizing* step is then performed to extract from the Y probability distribution the expectation of Y (the best estimation of Y), the standard deviation of Y (a measure of dispersion of Y), a confidence interval containing Y with a specified coverage probability.

Monte Carlo method provides a general approach to obtain an approximate numerical representation of the distribution function F_Y for Y . The heart of the approach is repeated sampling from the pdfs for the X_i . The implementation of the method is summarized below:

- a) select the number M of Monte Carlo trials to be made;
- b) generate M vectors, by sampling from the assigned pdfs, as realizations of the (set of N) input quantities X_i ;
- c) for each such vector, form the corresponding model value of Y , yielding M model values;
- d) sort these M model values into strictly increasing order, using the sorted model values to provide the estimation G of F_Y ;
- e) use G to form an estimate y of Y and the standard uncertainty $u(y)$ associated with y .
- f) use G to form an appropriate confidence interval for Y , for a stipulated coverage probability p .

Step f) is crucial because it can be formulated as a constrained minimization problem. In fact, given a coverage probability p , the associated confidence interval $[y_1, y_2]_p$ is obtained as the minimum interval $[a, b]$ such that $G^{-1}(b) - G^{-1}(a) = p$. When the pdf of Y is symmetric the solution is trivial and equal to $G^{-1}((1+p)/2) - G^{-1}((1-p)/2)$. Some relevant features of Monte Carlo method are: a reduction in the analysis effort required for complicated or non-linear models, especially since the partial derivatives of first- or higher-order of the model are not needed; generally improved estimate of Y for non-linear models; improved standard uncertainty associated with the estimate of Y for non-linear models, especially when the X_i are assigned non-Gaussian (e.g. asymmetric) pdfs, without the need to provide derivatives of higher order¹; provision of a confidence interval corresponding to a stipulated coverage probability when the pdf for Y cannot adequately be approximated by a Gaussian distribution or a scaled and shifted t-distribution, i.e. when the central limit theorem does not apply.

Being the measurement uncertainty modeling a central part in the assessment of the CAD performance, in the following section, we will describe the major sources of uncertainty that can be identified in a CAD for mammography.

5. Uncertainty modeling in a CAD

Numerous sources of uncertainty have to be taken into account in a medical image processing system for breast cancer detection and diagnosis, both in the system development step and in the operating conditions. Keeping in mind the two different situations, here below, we recall the most relevant sources of uncertainty subdividing them into five groups: instrumental

¹ This point is crucial when an evident asymmetric pdf should be assigned to the input quantity such as for photon noise.

uncertainty, uncertainty sources related to the patient, model uncertainty related to the CAD parameters selection, uncertainty sources related to the input provided by the radiologist during the system development and validation steps (reference values etc.), the uncertainty sources related to the possible interaction among the CAD and the physician who is manually setting the algorithm parameters during the system operation.

- GROUP 1: image acquisition and digitalization process (i.e., spatial resolution, geometric distortion, noise, pixel quantization, film degradation, artifacts introduction, etc.);
- GROUP 2: biological and patient variability (related to both pathological or normal cases); intrinsic and unpredictable data variability (e.g., patient movement);
- GROUP 3: model uncertainty due to the imperfect or incomplete knowledge about algorithm parameters values and their influence on the final results. Major concern should be paid to algorithm with a random initialization, where some parameters are initially set to a value randomly selected in a certain interval.
- GROUP 4: human observer subjective interpretation of the image according to his/her experience, in providing reference boundary, malignancy degree, reference region of investigation, useful for the CAD validation.
- GROUP 5: interaction between the human observer and the image data, in manually enhancing the image contrast, selecting Region Of Interest, manually sectioning the breast region for inspection etc., during the operation of the system.

Some of these sources of uncertainty (GROUPS 1 and 3) can be reasonably modelled and considered in the uncertainty estimation process (Mazzarella & Bazzocchi, 2007), while some other (GROUPS 2, 4, and 5) have unknown nature and are difficult to be embedded, even if noteworthy. As a consequence of the large subjectivity in using the CAD system in this context, a metrological validation of medical image processing approaches is required to highlight the intrinsic characteristics and behavior of a method, to evaluate its performance and limitations, and to compare the method with different existing approaches.

In the following sections, we will describe with more details a possible modelization of the uncertainty contributions summarized above, providing examples taken from ABCDE's functionalities.

5.1 GROUP 1: Noise contribution in mammographic images

Each step that contributes to the final digital mammographic image influences the global uncertainty in a different way: the formation process, based on X-ray exposure, introduces a noise contribution that is intensity dependent (e.g., photon noise), the microscopic structure of the impressed film introduces the so-called *film grain noise*, finally the digitalization process when performed through scanner devices, introduces an uncertainty contribution that is related to the number of bit per pixel (bpp) used in the conversion, or an uncertainty term that presents systematic periodic patterns (Mencattini et al., 2009b) superimposed to the image. Photon noise is the dominant random contribution in this context, since the bpp is usually 12 – 16, but other relevant noise contributions can be present at low and high exposures. Consequently, as suggested in (Jain, 1989), we consider an *heteroscedastic noise model* for the random contribution assuming that the noise variance depends on the intensity of each pixel, and it is not constant within the image. The noise model can be represented as follows:

$$\tilde{I}(n, m) = I(n, m) + \eta_1(n, m), \quad (1)$$

where $\tilde{I}(n, m)$ is the noisy image, $I(n, m)$ is the noise-free image, $\eta_1(n, m)$ is a zero mean non stationary random process given by $\eta_1(n, m) = \eta \cdot \sigma(n, m)$ with $\eta \sim \mathcal{N}(0, 1)$ a normal random process with zero mean and unitary variance, and $\sigma^2(n, m)$ the spatial-dependent variance of the process $\eta_1(n, m)$.

The assumption of a normal random process with non constant variance is fully justified (Jain, 1989) for luminance values in the subrange $[0.2 - 0.8]$ that corresponds to what we are interested in; actually, dark or very bright pixels are corrupted by a saturated gaussian noise with cut tails. This problem could occasionally produce an underestimation of the noise variance we perform, but involving regions far from the regions of interest.

It is well known that the noise variance $\sigma^2(n, m)$ depends on the intensity $I(n, m)$. In particular, we get that $\sigma^2(n, m) = \alpha I(n, m)^\beta$ for scanner devices and $\sigma^2(n, m) = \gamma \cdot \log_{10}(I(n, m)) - d_0$ for photographic films. Consequently, an estimation procedure is needed in order to estimate the unknown dependence of the noise variance $\sigma^2(n, m)$ on the intensity $I(n, m)$. To do this, we implement the noise estimation introduced in (Gravel et al., 2004), later extended and improved in (Mencattini et al., 2007; Salmeri et al., 2008). The strength of the estimation algorithm is that it can be applied to very different kind of medical images (MRI, ultrasound images, etc.) where noise is signal dependent following different probability distribution (Rician noise, Rayleigh, etc). In case of homoscedastic noise (where noise variance is constant through the image) the estimation algorithm still provides an accurate estimation of the noise variance, needed in many algorithms for image denoising and enhancement (Mencattini et al., 2010b; 2008). Here below, we only report a sketch of the whole estimation algorithm.

- a) Extract low-frequency components from the image containing homogeneous regions. This step is performed by applying a low pass gaussian filter to the original noisy image obtaining a smoothed original image.
- b) Evaluate the high frequency components of the image by the subtraction of the smoothed image to the original one. Obviously this image contains both small details, boundaries, and noise.
- c) Eliminate edges by applying a robust edge detector to the original image. This further step is needed in order to eliminate edges from the estimation procedure, that would decrease noise variance estimation accuracy. Then, by thresholding we obtain a binary mask of principal edges.
- d) Build an histogram of relating each bin to the intensity of image considering pixels at the same position in the smoothed image and in the noise image.
- e) Evaluate the standard deviation of each bin by a Median Absolute Deviation (MAD) estimator.
- f) Perform a robust regression analysis by Cubic Smoothing Spline in order to fit the data extracted at step (e).

A block diagram of the whole algorithm is reported in Fig. 5.

Unfortunately, uncertainty contribution do not embed only random effects, related to noise. In fact, in medical diagnosis, it is often crucial to take into account also non random contributions that can occur during the exposure, the acquisition and the digitalization of the image. We refer for example to physiological movements of the patient (a few

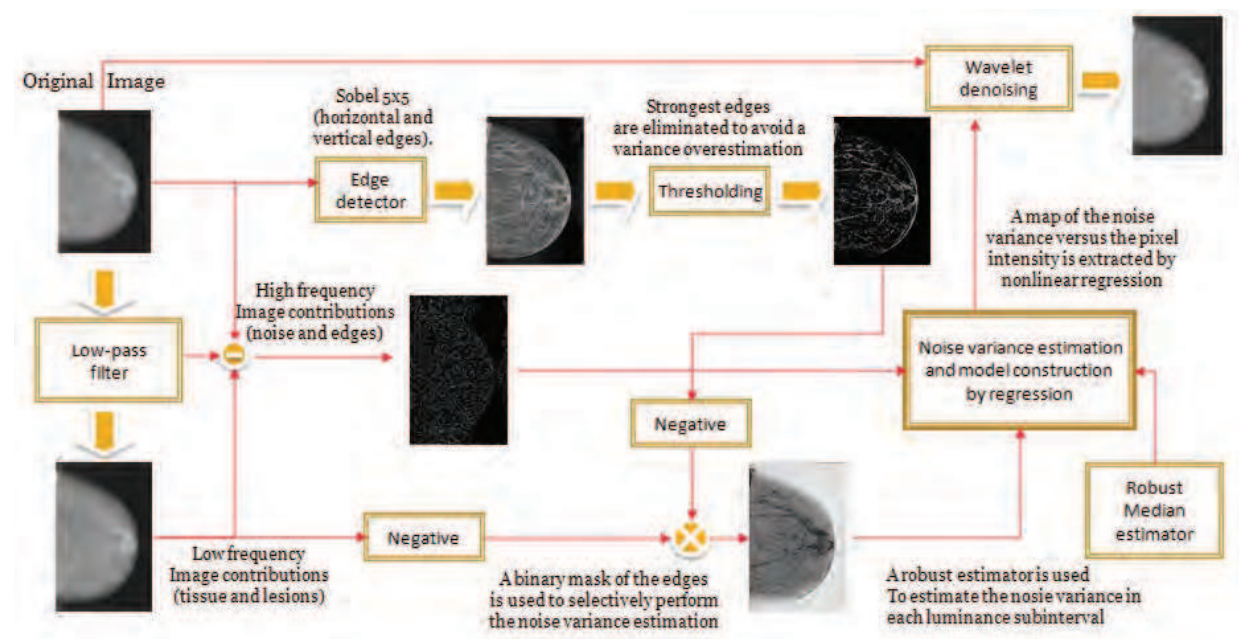


Fig. 5. The noise variance estimation algorithm.

seconds are needed in order to form a single mammographic image), to artifacts that can be present in the final image, due to dust, fingerprints, scratch, and to the non random contribution introduced by the scanner itself, such as the spatial non uniformity introduced in the luminance. For example, the images database we consider, e.g., the Digital Database for Screening Mammography (DDSM) (Heath et al., 1998) uses the three scanners DBA, HOWTEK, and LUMISYS to digitalize more than 4000 mammographic images. We can notice that the final images are corrupted by a periodic pattern contribution that we assumed to be systematic. An example is shown in Fig. 6(a)-(b) where two regions are extracted from the same mammographic image. Region A is dark and it is evident that there is an overlying periodic pattern that here is also emphasized by an histogram stretching. On the contrary, region B is corrupted by noise but the periodic pattern is much more subtle. In Fig. 6(c) the two-dimensional Fourier Transforms of the two regions are also shown proving that there is a unidirectional periodic pattern overlying region A and that the same does not hold for region B with the same evidence.

It is well known that systematic effects have to be corrected before performing a reliable uncertainty propagation of random effects (Santo et al., 2004). Unfortunately, as a consequence of the previous considerations, this kind of contributions is both luminance and spatial dependent so that it is very difficult to evaluate and correct it. Moreover, from both medical and visual viewpoints, it could be dangerous to preliminarily correct the images from non random contribution thus altering the final aspects of the images and falsifying the correct diagnosis of microcalcifications. In this regard, the new version of International Vocabulary of Metrology (VIM) (see (*International vocabulary of metrology. Basic and general concepts and associated terms (VIM)*), 2008) paragraph 2.26) states that *Sometimes, estimated systematic effects are not corrected for but, instead, associated measurement uncertainty components are incorporated.* Finally, scanner devices used in medical laboratories are not completely characterized from a metrological point of view thus making the exact evaluation of these effects for correction impossible. As a possible solution, we assume to treat non random contributions as systematic

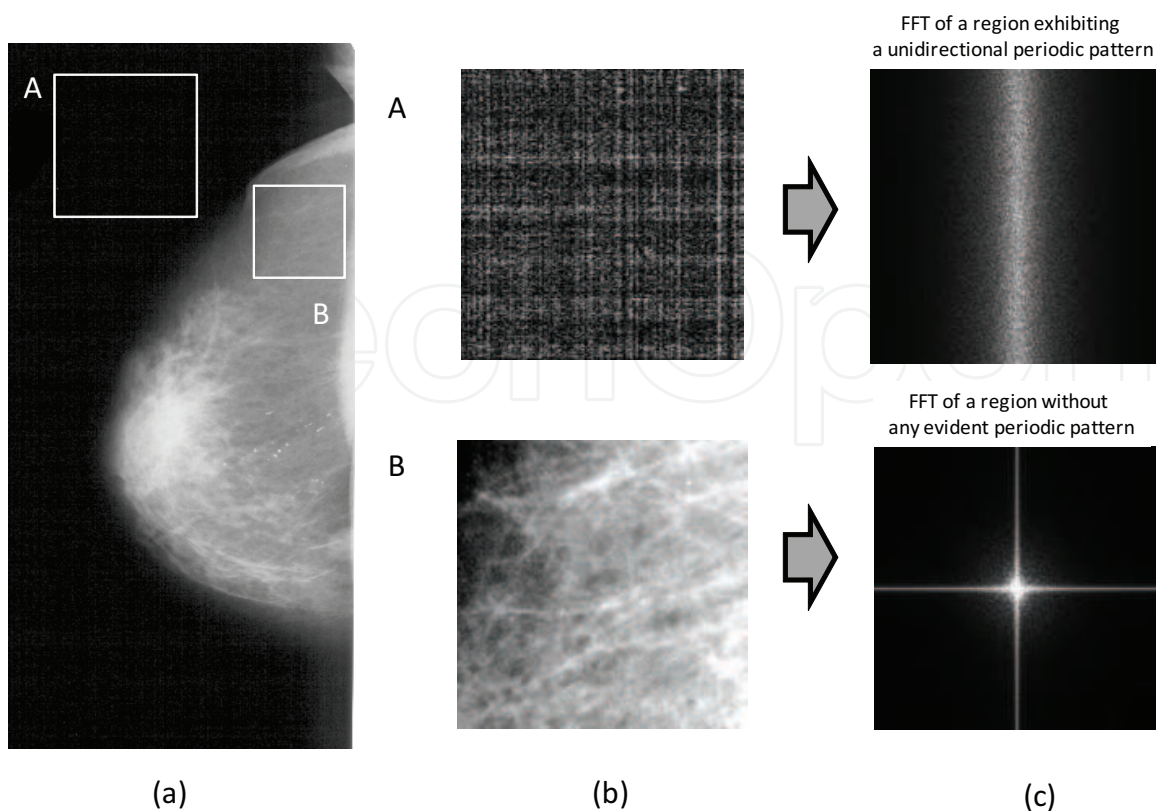


Fig. 6. (a) A mammographic image with a dark region (A) exhibiting a periodic pattern and a bright region (B) in which the same effect is much more subtle. (b) A zoom of the two regions A and B. (c) The 2D-FFT of the regions A and B respectively.

effects and to embed them in a unified uncertainty model, implementing an uncertainty propagation through a suitable method.

A viable solution is to use information extracted by the estimation procedure performed for the noise variance modeling. In this way, by merging different images coming from the same scanner or digitalization device, we are able to provide more accurate limits for the systematic contribution. In the following, we will provide an example for the whole noise estimation procedure considering the three different databases: MiniMIAS, DDSM, and San Paolo FFDM images.

Fig. 7 reports the noise variance modeling considering DDSM images for the three different scanners HOWTEK-A, LUMISYS, and DBA at different spatial and pixel resolution. Note that for low luminance values electronic noise in HOWTEK-A and in DBA is the dominant noise contribution, while photon noise due to X-ray exposure has a small variance at high and low exposure. Noise variance related to photon noise has a maximum value in the range $[0.4 - 0.6]$ for scanners LUMISYS, HOWTEK-A, and DBA. Fig. 8(a) shows the noise variance estimation for images from MiniMIAS database. Photon noise is low at high and low exposure and the noise variance has a maximum value in the same range as above. Both LUMISYS, HOWTEK-A, and MIAS scanning device have a linear characteristic curve. Conversely, scanner DBA has a logarithmic response. Finally, Fig. 8(b) reports the noise variance estimation performed on the FFDM images from the San Paolo Hospital. Also in

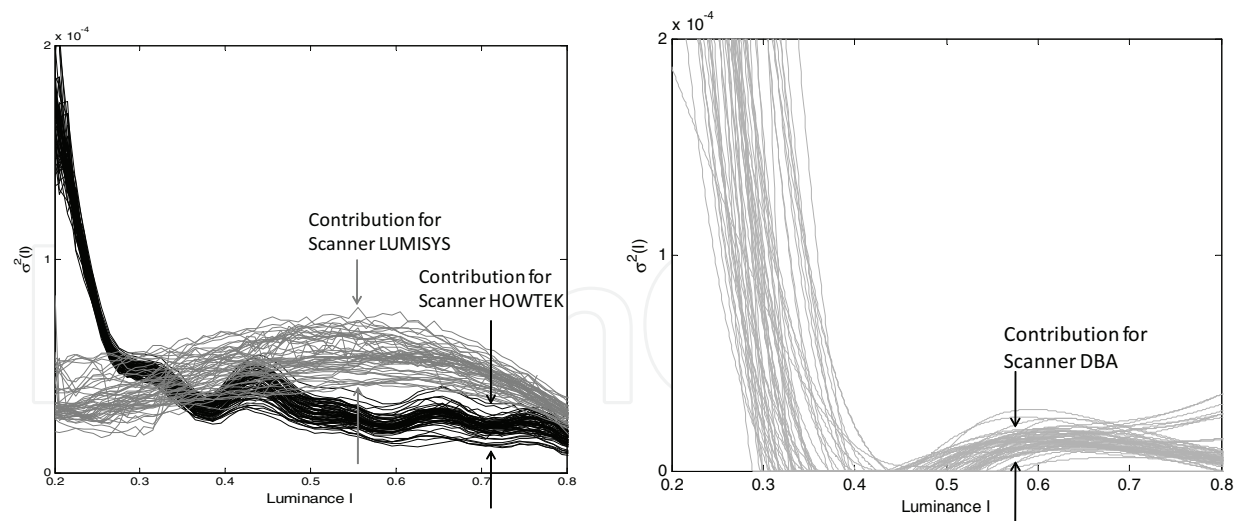


Fig. 7. Noise variance modeling for scanners HOWTEK-A and LUMISYS (left) and DBA (right). A Cubic Smoothing Spline regression is applied with a smoothing parameter equal to 0.9999 for scanners HOWTEK-A and LUMISYS and to 0.999 to scanner DBA. The systematic noise variation is marked by the arrows.

this case, photon noise is the dominant contribution. The digital mammography device has a logarithmic characteristic curve.

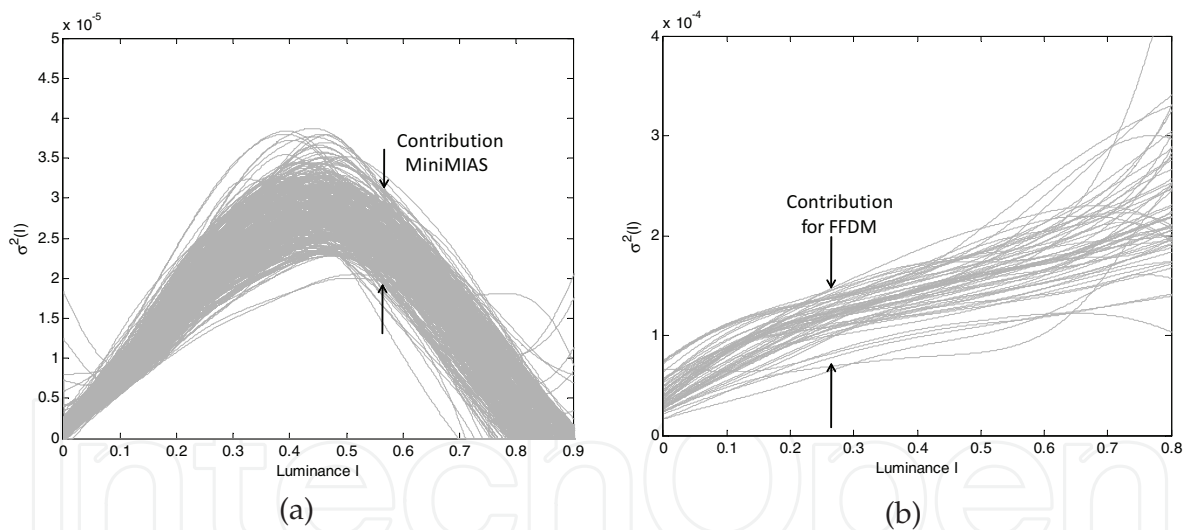


Fig. 8. Noise variance modeling for Joice-Loebl scanning microdensitometer (a) and for GE Senograph 2000D ADS 17.3 (b). A Cubic Smoothing Spline regression is applied with a smoothing parameter equal to 0.999. The systematic noise variation is marked by the arrows.

It can be noted that the noise dependence from the pixel intensity strongly varies across different digitalization devices, according to pixel quantization, spatial resolution, scanner calibration curve, etc. Also, the systematic noise variation is different across the devices.

5.2 GROUP 2: biological and patient variability

Uncertainty sources due to biological and patient variability are probably the most difficult kind of uncertainty to consider because it groups together unpredictable variations in patient

physiological status and position. Fortunately, the protocol that regulates mammography provides a set of rules that allow you to control many of these sources of uncertainty: the breast is compressed and locked during exposure to X-rays, reducing the possible involuntary movements also due to the simple respiratory motion; the hormonal changes that may occur during the patient's mammography are also controlled as much as possible through detailed recommendations about the optimal time to perform the examination in relation to the menstrual cycle. In any case, a complete and accurate CAD system should also provide the possibility to receive and process information relating to the phase of the woman's life: at puberty, lactation, or menopause, in order to adapt algorithm parameters such as for example those related to fibro-glandular disk extraction in bilateral asymmetry detection.

5.3 GROUP 3: model uncertainty

A CAD for mammography consists of several blocks, as already shown in the introduction. The correct functioning and adaptation of each block involve the use of inline tuned parameters that are initialized to a certain value (randomly or not) and can be changed during the system functioning, according to the image characteristics (adaptation) or according to pre-specified rules. The CAD developers identify these rules during the system development and validation steps, using large datasets of mammographic images, but a complete knowledge of the behavior of the selected parameters is not realistic. In light of this, a formal way to deal with this incomplete knowledge is to perform an analysis of sensitivity, that, considering a parameter at a time, introduces small perturbation of it and evaluate the final effect on the block output. By an iterative procedure, the developer can assign an interval of values to each block output with a dual effect: quantifying the most critical parameters that most influence the output and modify the block in order to improve the robustness of the whole system to unpredictable parameters variations.

The CAD elaboration units can be divided into three categories:

- direct computational blocks involving constant multiplicative coefficients and/or nonlinear operators;
- iterative blocks involving adaptive thresholds and signal-dependent parameters that change during the block functioning;
- conditional statements involving the comparison of input variables with thresholds (fixed or changing during the functioning).

According to the typology, different kind of uncertainty modelling should be performed. Here below, we provide some examples already investigated by the authors in previous works. In Section 6 will be described in more detail a recent study in the context of the identification of bilateral asymmetry.

5.3.1 Direct elaboration units

Context Handling and propagation of random and systematic uncertainty contributions due to image noise through a wavelet-based enhancement and denoising procedure for microcalcification detection (Mencattini et al., 2009b).

Materials DDSM Images containing malignant and benign calcifications.

Method Random Fuzzy Variables able to simultaneously represent and propagate through dedicated mathematical operators random and systematic terms. The method has been

developed by Ferrero and Salicone (Ferrero & Salicone, 2009; Salicone, 2007) and it is still under investigation the possible application of the method to iterative and more complex elaboration units.

5.3.2 Iterative elaboration units

Context Handling and propagation of random uncertainty contributions due to image noise through the automatic segmentation and classification of massive lesions in mammographic images (Mencattini et al., 2009a; 2010a; Rabottino et al., 2011).

Algorithm Segmentation is performed by an iris detection algorithm, an iterative region-growing procedure is applied to extract the mass boundary, geometric and textural (Haralick) features are computed for each segmented lesion, and finally a Bayes linear classifier is applied to assign a malignancy degree to each lesion (Mencattini et al., 2009a).

Materials DDSM Images containing malignant and benign massive lesions.

Method Monte Carlo simulations using 50 trials on ROIs 1000×1000 . The AUC-based performance assessment has been performed. Fig. 9 reports the cumulative distribution function and the histogram of the AUC obtained after the trials. Note that the AUC parameter

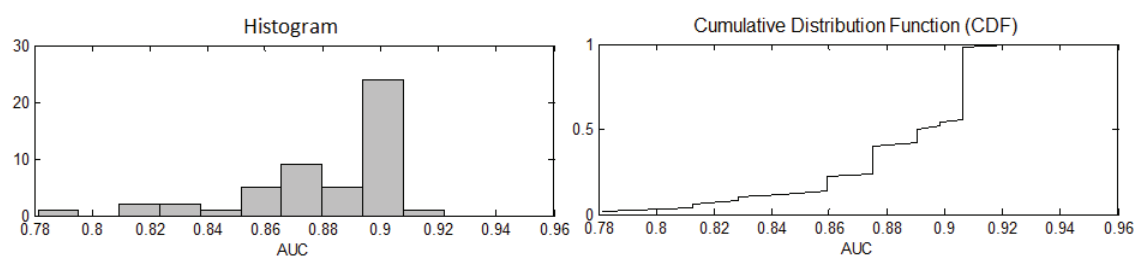


Fig. 9. AUC cumulative distribution function and histogram.

has a clear non-Gaussian probability distribution, such that the confidence interval for $p = 0.9$ is very different than the interval obtained under the assumption of normality.

Context and Algorithm An analogous study has been developed in collaboration with the San Paolo Hospital and the Dept. of Mathematics University of Bari, using the active contour algorithm developed by Chan and Vese (Chan & Vese, 2001) for the segmentation step, the fractal analysis and support vector machines for the malignancy assessment. Preliminary results can be found in (Mencattini et al., 2011d).

Materials San paolo FFDM Images and DDSM images containing malignant and benign massive lesions histologically proven.

Method Monte Carlo simulations using 100 trials.

5.3.3 Conditional elaboration units

Context Conditional statements in a fuzzy logic inference system for pattern classification (Ferrero et al., 2010; Mencattini & Salmeri, 2011a). The pattern classification step is an important section of a CAD because it is use in many functionalities: in the automatic identification of lesions a classifier is used to reduce false positive lesion candidates discriminating them from true lesions; in the malignancy assessment a classification step

is used to assign the malignancy degree to each considered suspicious lesion; lastly, in the identification of bilateral asymmetry the classifier assigns the asymmetry degree to each pair of mammograms under test. Many different classifiers exist in the literature and have been used in developing a CAD for mammography: Bayesian classifier, decision trees, artificial neural networks, logistic regression, support vector machine, fuzzy logic inference systems are the most popular approaches. Each of them has its own peculiarity. For example, SVM is excellent for small training dataset, while artificial neural network has great flexibility to very large datasets. However, most of them can be considered as completely black boxes able to learn from examples and to produce a final class membership. None of them, excluding fuzzy inference systems and decision trees, produces simple classification rules that can be easily understood by physicians. Finally, only fuzzy inference systems can be directly modified by physicians who can add classification rules, according to his/her experience. *Algorithm* Since the method uses a non-standard fuzzy logic system, in the following we will provide a brief introduction of the method. A Fuzzy Inference System (FIS) for medical diagnosis is a nonstandard FIS firstly proposed by Andersson in (Andersson, 2007). The core of such a system is a simple structure composed of three agents: the patient (the mammogram in our context), the symptoms (in our context the features extracted to characterize the object to be classified), and finally the diagnosis (in our context the class, normal or abnormal tissue, benign or malignant lesions, normal or asymmetric breast, etc.). The architecture of such a FIS involves three kind of relations: two input relationships (the relation Mammogram-Features and the relation Features-Diagnosis) and a third relation (the Mammogram-Diagnosis relation) that is derived by the first two relations using a specific inference process. The first relation is built by physicians in the patient-history or anamnesis process, while is learnt by the CAD during the features extraction and training steps. The second relation is derived by the physician according to his/her experience, retrospective study and clinical trials in which a symptom has been confirmed to be significant for a certain diagnosis (e.g., spiculated margins of a mass represent the infiltrating properties of the lesion and a sign of possible malignancy). Finally, the third relation assigns a final class membership to a mammogram or simply to a region of interest (P in the following), according to the features extracted on it and on their significance for a certain class. Such a system has been already applied in different situations: the malignancy assessment of calcifications clusters (Ferrero et al., 2010), the false positive reduction in the automatic mass identification (Mencattini & Salmeri, 2011b), the development of a novel scoring system to measure the severity of illness of patients admitted in Intensive Care Units (Mencattini et al., 2011e). In this section, we will consider in details the malignancy assessment of calcifications.

A FIS rule has an IF-THEN structure, an example of two distinct rules is reported below.

IF number of calcifications in a cluster is HIGH in mammogram P AND number of calcifications is HIGHLY significant for diagnosis MALIGNANT THEN mammogram P contains a MALIGNANT cluster of calcifications

IF mean entropy of calcifications is HIGH in mammogram P AND mean entropy of calcifications is HIGHLY significant for diagnosis BENIGN THEN mammogram P contains a BENIGN cluster of calcifications

Each rule is preliminarily evaluated computing the two antecedents according to specific mathematical fuzzy operators and then the rules are aggregated through a weighted average

operator (Ferrero et al., 2010). The evaluation of the two antecedents, *number of calcifications in a cluster is HIGH* and *number of calcifications in a cluster is HIGHLY significant for diagnosis MALIGNANT* involves the definition of the fuzzy variable *number of calcifications*, and consequently of the fuzzy sets *HIGH number of calcifications* and *LOW number of calcifications*. This definition consists in the construction of two trapezoidal Membership Functions (MFs) as those reported in Fig. 10 for the two symptoms *number of calcifications* (S_j) and *mean entropy of calcifications* (S_k). The two MFs are superimposed on the corresponding histograms. Note that,

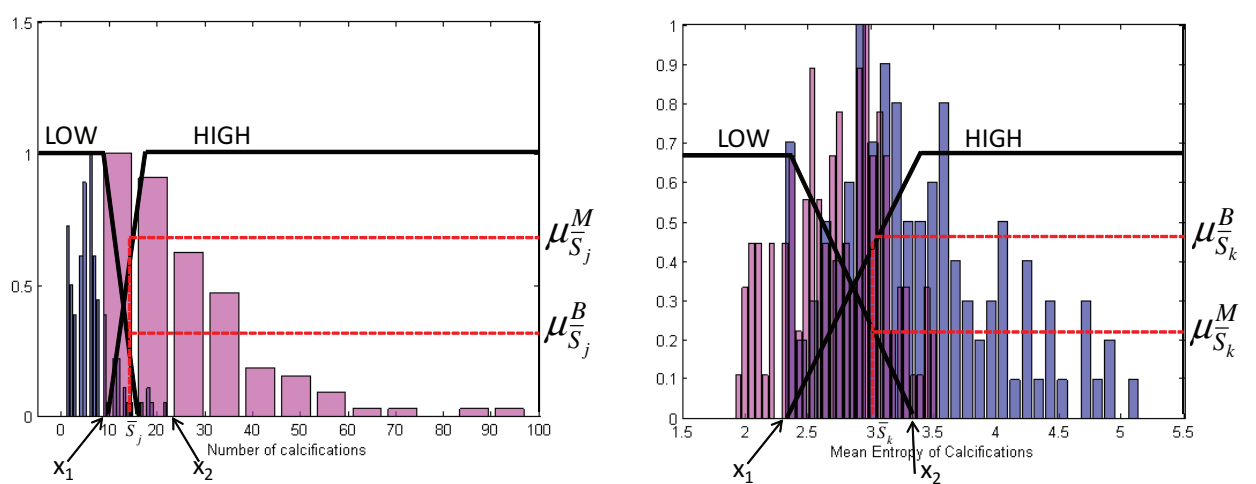


Fig. 10. Examples of MF construction for symptom S_j (number of calcifications) and S_k (mean entropy of calcifications). The magenta bars identify the class M while the blue bars represent the class B .

for a certain value \bar{S}_j of the first symptom, the system assigns two degrees of membership $\mu_{\bar{S}_j}^M$ and $\mu_{\bar{S}_j}^B$ to the class MALIGNANT (M) and BENIGN (B) respectively. As long as the value of S_j increases from x_1 to x_2 then $\mu_{\bar{S}_j}^M$ increases toward the unity and $\mu_{\bar{S}_j}^B$ decreases toward zero. The opposite happens when S_j decreases from x_2 to x_1 . It is evident that each of the two MFs depends on the two parameters x_1 and x_2 which are extracted from the distribution of the feature for the different classes. In particular, x_1 is the maximum value of S_j for benign cases such that all the S_j for malignant cases are greater than x_1 . Conversely, x_2 is the minimum value of S_j for malignant cases such that all the S_j values for benign cases are smaller than x_2 . The evaluation of x_1 and x_2 is performed after outliers elimination on each feature S_j . Parameters x_1 and x_2 influences the final class membership function and an analysis of sensitivity should be implemented for them.

The second antecedent *number of calcifications is HIGHLY significant for diagnosis MALIGNANT* involves the computation of what we called *Incidence Level* $IL_{S_j}^{D_i}$ of a symptom S_j for a diagnosis D_i , where in the considered example D_i is M or B . It represents the degree of influence that a symptom S_j can have on a certain diagnosis D_i and a degree of correlation among them. This correlation is evaluated according to physician's experience or again from the distribution of the symptom's values for each class. Two examples of this calculation are reported in Fig. 11 for the number of calcifications in a cluster and for the mean entropy of calcifications for the two classes Malignant and Benign. In particular, we get for symptom S_j

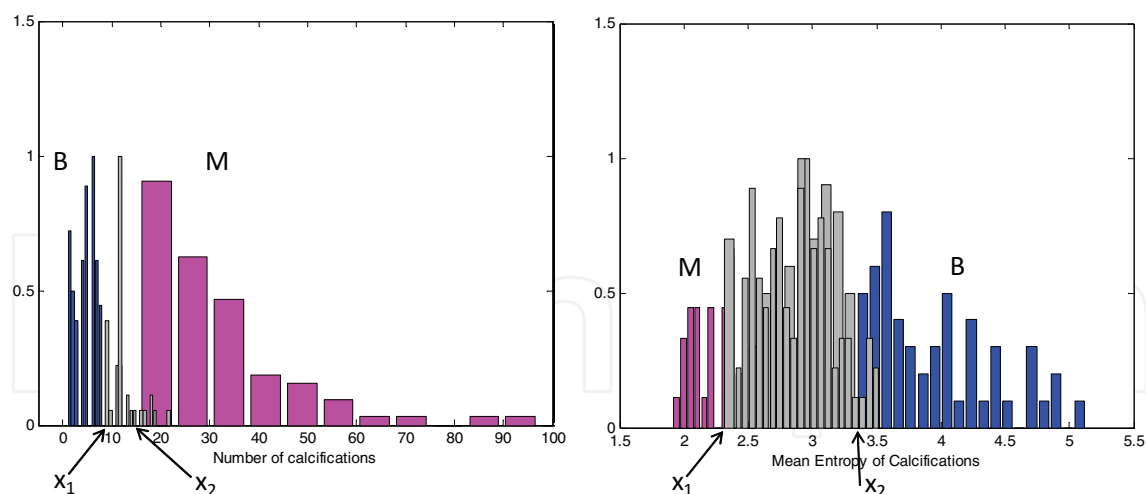


Fig. 11. Examples of evaluation of the incidence level IL for symptom S_j (number of calcifications) and S_k (mean entropy of calcifications). The gray bars identify those cases that do not contribute to the calculation of the incidence levels.

(number of calcifications) that

$$IL_{S_j}^M = \frac{\#\{S_j^M | S_j^M > S_j^B\}}{\#\{S_j^M\}}, \quad IL_{S_j}^B = \frac{\#\{S_j^B | S_j^B < S_j^M\}}{\#\{S_j^B\}} \tag{2}$$

From equations (2) it is fully justified that the incidence level has a strong influence on the final class membership and again a dedicated analysis of sensitivity should be performed. In the considered context, seven features are used to assess the malignancy of the calcifications: number of microcalcifications (S_1), minimum area of the microcalcifications (S_2), area of the convex hull containing microcalcifications (S_3), entropy of the cluster of microcalcifications (S_4), mean contrast of the microcalcifications (S_5), mean entropy of the microcalcifications (S_6), Haralick parameter H8 (Sum Entropy) (S_7).

Materials DDSM images, 119 benign and 122 malignant calcifications clusters.

Methods Recently, an innovative approach to propagate random and systematic uncertainty contributions through a FIS has been implemented (Ferrero et al., 2010). In light of this, within a collaboration with the Dept. of Electrical Engineering Politecnico of Milan, the method has been applied to our context. Basically, the method assigns a Random Fuzzy Variable (RFV) to each variable in order to model simultaneously random and systematic uncertainty terms. From Fig. 12 (figure top-left) it can be noted that an RFV is an extended Fuzzy Variable, whose Membership Function is composed of an inner and an outer standard MF. The inner MF represents the systematic uncertainty contribution, usually it is a rectangular MF, as in standard interval analysis; the outer part represents the random contribution and a direct approach can be implemented to derive it from the pdf of the variable (Dubois et al., 2004). Fig. 12 (figures right) reports the Random Fuzzy Variables (RFVs) for each feature used for calcifications malignancy assessment considering one of the images taken from DDSM. In this context, we assumed only systematic uncertainty contribution for discrete symptoms S_1, S_2, S_3 and only random contributions for symptom S_4, \dots, S_7 ². The corresponding

² The random contributions in symptoms S_1, \dots, S_3 are considered to be small and totally embedded in the systematic terms. On the contrary, we assumed that the systematic contribution in symptoms S_4, \dots, S_7 was negligible with respect to random variations.

histograms for features S_4, \dots, S_7 are reported on the left. Figure top-left reports the two RFVs representing the benign and the malignancy degrees associated to the considered cluster. The inner part results from the propagation of the systematic effects, while the outer part results from the propagation of the random effects through the FIS.

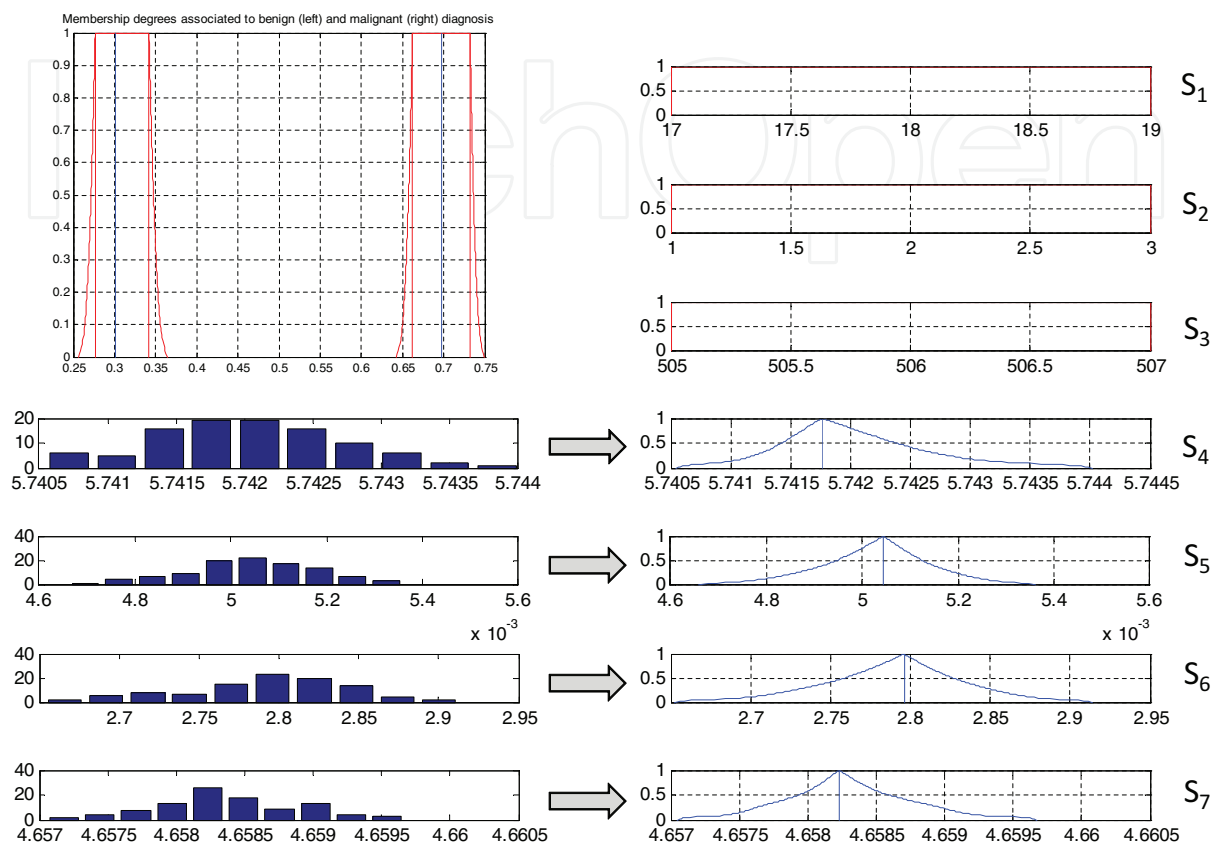


Fig. 12. Figure Top-left: RFVs of the malignancy and the benign degrees for one test image. Figure bottom-left: histogram of features $S_4 - S_7$. Figure right: RFVs of the features extracted. Red plot: inner MFs of the RFVs; blue plot: outer MFs of the RFVs.

5.4 GROUP 4: radiologist-dependent CAD inputs

During the development step and the CAD validation procedure, one or more radiologists should provide the ground truth needed as a reference to set the algorithm parameters.

- i) During the validation of the breast region segmentation, radiologists provide the actual breast skin line, the nipple position, the pectoral muscle profile. In particular, in case of minoris, very obscure nipple, very dark uncompressed tissue, this reference strongly depends on radiologist experience and more than one radiologist should be interpellated using a voting procedure to select a final robust reference.
- ii) During the validation of the automatic lesions identification, the radiologist is called to identify the possible lesion, whose position is then confirmed by biopsy.
- iii) Also, the boundary of the lesion can be drawn by the radiologist, or, more rarely, a core and an external boundary are drawn, as in DDSM database. The core identifies the internal

portion of the lesion, while the boundary includes also the spicules and the structures that radiate from the core.

- iv) In some situations, the radiologist draws the boundary of a region representing a bilateral asymmetry or locate the center of an architectural distortion. This situation is much more radiologist's dependent, because there is not any histological test that can confirm this reference.
- v) Finally, less frequently, the radiologist is called to provide additional information about the breast tissue such as ACR tissue density or ACR subtlety rating. Both these parameters are fundamental for the correct functioning of some algorithms such as bilateral asymmetry identification (Casti, 2011). A recent preliminary study proved the effectiveness of a procedure for the extraction of the fibro-glandular disk and of the oriented patterns in the breast, tuned according to the ACR tissue density.

All of the information provided by the radiologist influence both the CAD learning capability and the effectiveness of its operation. Each of them should be perturbed in order to evaluate the actual influence on the final results produced by the CAD. In particular, in Section 6 we will describe recent simulation results, obtained assuming a systematic perturbation of the ACR density rating assessed by a single radiologist. Using a leave one out cross-validation procedure, considering a pair of mammograms at a time, we assumed that the radiologist assessed a density rating higher than the actual density. Such a bias, is then propagated through the whole algorithm for the identification of bilateral asymmetry on a dataset including 23 pairs of normal mammograms and 23 pairs of asymmetric mammograms taken from MiniMIAS database. The same procedure has been implemented assuming that the radiologist assigns to the pair of mammograms under test a density rating smaller than the actual one. Doing so, the whole algorithm produces for each pair of mammograms an asymmetry degree in absence of perturbation and two asymmetry degrees in presence of a positive and a negative perturbation respectively. Further details will be provided later.

5.5 GROUP 5: CAD-radiologist interaction

A recent study by Nishikawa (Nishikawa, 2007) reports important considerations on the evaluation of a CADe system. He recalls that a CADe system does not need to detect cancers, but it needs to assist radiologist in doing this. This means that the real help is in detecting tumors missed by the radiologist increasing the overall sensitivity of the radiologist-CADe system. It is possible to have a CADe scheme with a sensitivity lower than 50% and still be a useful aid. So, one of the crucial point is the interaction CADe-radiologist. If cancers missed by the computer is too high then the radiologist will lose confidence in the ability of the computer to detect cancers. On the other hand, false detections reduce radiologist's productivity because the radiologists must spend time reviewing all computer detections.

Conversely, a CAD could over-learn from available samples produced by the same apparatus and reported by the same radiologist. Nishikawa observes that a positive bias can occur even if training and test set are the same or if the dataset used for features selection and the training set are the same. To reduce this bias and to reduce over-learning, more than one radiologist should report the images used for development, validation, training, and test, using a voting procedure to select one report for each case. Moreover, a separate dataset for each of the above phase should be used, preferably extracted from different databases.

6. A numerical example of uncertainty propagation through ABCDE: bilateral asymmetry identification

For almost thirty years, methods have been developed to try to use computers as second readers to make up for this potential loss of reliability. Almost all fields of image processing and analysis have been explored but little attention has been paid to the identification of bilateral asymmetry on mammograms. Asymmetric breast tissue is usually benign, but an asymmetric area may indicate a developing mass or an underlying cancer. Thus, the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS), which has developed a standardized method for breast imaging reporting, describes bilateral asymmetry as one of the four signs of breast cancer that radiologists have to detect. Some commercial CAD systems have obtained the Food and Drug Administration (FDA) approval and are beginning to be applied widely for the detection of masses and microcalcification on mammograms, but none of them has been developed for bilateral asymmetry identification. The paucity of works related to this topic may be due to the fact that bilateral asymmetries are difficult to be identified for the following reasons:

1. it needs a comparison between the left and right views which is a difficult task owing to the natural asymmetry of the breasts, and to the absence of good corresponding points to perform matching;
2. distortion inherent to the manual position of the breast during X-ray exposure;
3. the wide range of appearance associated with differences between women in the amount and distribution of fibroglandular tissue, which is also influenced by age and hormonal status and whose asymmetry could also represents simply a physiological sign.

The authors are currently working on the development of an automatic bilateral asymmetry identification procedure, one of the most innovative part of ABCDE (Casti, 2011; Mencattini et al., 2011c). The aim is to improve the early diagnosis of breast cancer, especially in screening test, providing clues about the presence of early signs of tumors like parenchymal distortion, small asymmetric bright spots and contrast, that are not detected by other methods. Asymmetric breasts could be reliable indicators of future breast disease in women and this factor should be considered in a woman's risk profile. The Breast Imaging Reporting and Data System (BIRADS) defines bilateral asymmetry as an area of fibroglandular tissue that is more extensive in one breast as compared to the contralateral one. Thus an asymmetric finding can be represented by a difference in shape and distribution of the fibroglandular tissues of the two breasts but also by a difference in the oriented pattern of the two disks. The oriented features detectors used in this work is a set of real Gabor filters oriented at different angles (spaced at angles of 18 degrees) (Rangayyan et al., 2008). Real Gabor filters, in fact, yield good performances in terms of capability to detect the presence of oriented features as well as in terms of accuracy in the estimation of the angle of the directional components and are fundamental tools in image understanding where the information of interest is displayed in the form of oriented features. The foregoing considerations guided the choice to segment the fibroglandular tissue from the rest of the breast parenchyma using Gaussian mixture modelling (Ferrari et al., 2004). Differences between the left and right breasts in terms of directional and morphological features are then extracted and used to assess the asymmetry degree of the patient. The idea is that the integration of measures of shape and distribution of the fibroglandular tissue with measures of the oriented pattern through a Gabor filtering analysis, can provide a better and accurate characterization of the

tissue, allowing the detection of the asymmetric abnormalities eventually present on the mammographic images. Once a numerical representation of the pairs of mammographic images has been obtained, bilateral asymmetry identification has been performed by pattern classification, based on a Support Vector Machine (SVM) classifier. Fig. 13 shows the flowchart of the whole procedure. The highlighted block denotes the most innovative part of the algorithm. It involves the tuning of the parameters of the algorithm according to the ACR breast tissue density. In particular, the thickness parameter in the Gabor filters and the

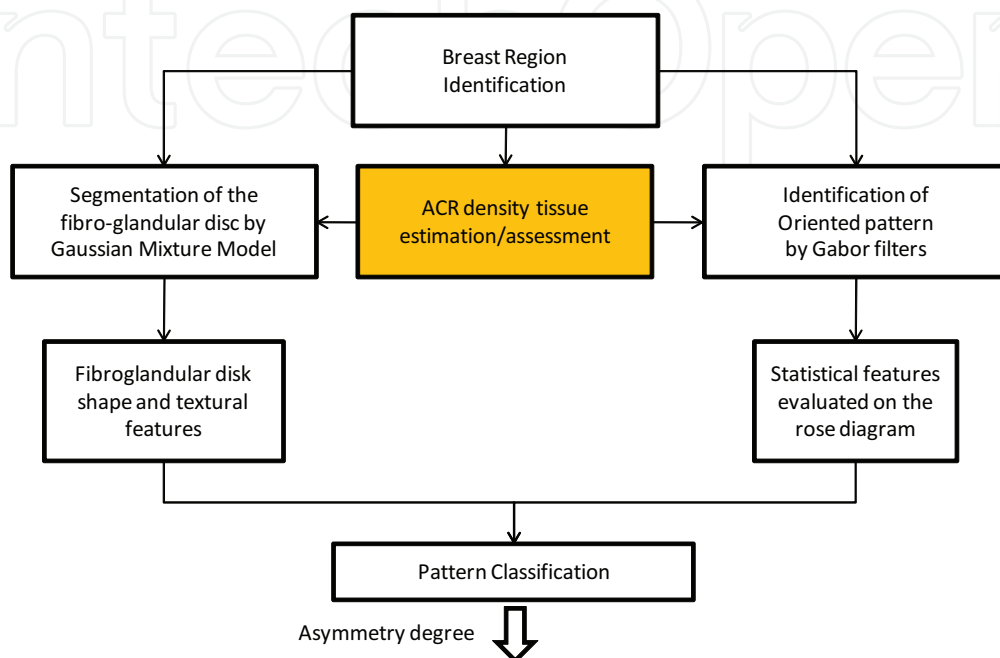


Fig. 13. A flowchart of the whole algorithm for the identification of bilateral asymmetry.

clustering procedure inside the gaussian mixture model estimation depend on the tissue density. This approach leads inevitably to a strong dependence on the density assigned to the final result, i.e. the degree of asymmetry. Fig. 14 reports two examples taken from DDSM images for an asymmetric and a normal pair of mammograms. The fibroglandular disks are also included along with the oriented pattern derived using Gabor filters. Finally, the rose diagram of the angular distribution of the left and right breast is reported. The considered asymmetric case has a density equal to 4 (at least 90% of the breast is composed of dense tissue), while the normal case has an assigned density equal to 3 (a portion in the range of 49% – 90% of the whole breast region is occupied by heterogeneously dense tissue).

The recent work has been devoted to the identification of the most relevant sources of uncertainty in the described algorithm. Two uncertainty contributions, respectively random and systematic, have been investigated:

- 1) the image power noise, estimated as described in previous sections;
- 2) subjective radiologist's bias in assessing the ACR density category, simulated by the assignment of a density value higher or lower than that assigned by the first radiologist.

The two different sources of uncertainty have to be separately considered and simulated, in order to avoid the interactions of them, possible compensation effects, in order to perform an individual failure analysis.

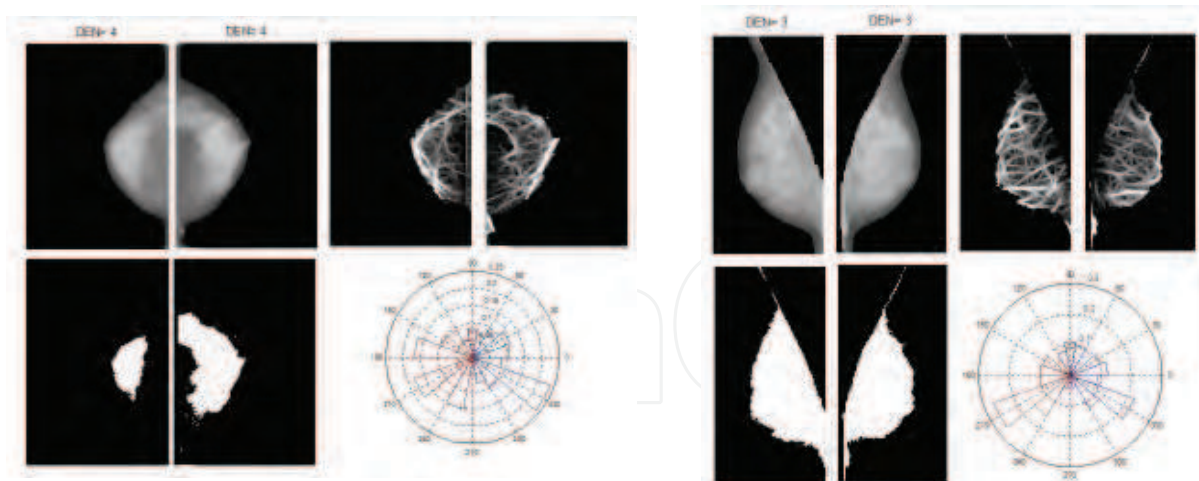


Fig. 14. Figure left: a pair of mammograms from DDSM with an assigned bilateral asymmetry ($B - 3056$ CC views). Figure right: a pair of normal symmetric mammograms from DDSM ($A - 0048$ MLO views).

6.1 Random uncertainty contribution: image noise

The Monte Carlo Method (MCM) has been implemented to evaluate and propagate the uncertainty contribution due to noise on the mammographic images. The implemented procedure is described as follows:

- i) represent the luminance of each pixel by a pdf of a normal random variable with mean value taken from a smoothed version of the given image and standard deviation given by the noise variance estimation;
- ii) select the number M of Monte Carlo trials;
- iii) generate M different noisy images, by sampling from the assigned pdf using the model in (1), thus producing M realizations of the input mammographic image;
- iv) apply the CADe algorithm to each of M image realizations and form the corresponding M different asymmetry degrees for each image;
- v) sort these M values into strictly increasing order and construct a discrete representation G of the cumulative distribution function for the asymmetry degree;
- vi) use G to derive an appropriate confidence interval for the asymmetry degree, for a given coverage probability p .
- vii) Repeat steps i)-vi) for any image in the test set, considering a smoothed version of the remaining $N - 1$ original images for training, in a leave one out cross-validation testing procedure, thus producing $N \times M$ different asymmetry degrees.
- viii) Using the values in vii) compute two vectors of M sensitivities and M specificities, thus producing extended ROC curves for each coverage probability.

Fig. 15 reports the asymmetry degrees for the asymmetric and normal cases taken from MiniMIAS database. The confidence interval associated is for a coverage probability equal to $p = 0.90$. The green circles identify the most critical pair of mammograms in terms of noise influence among the asymmetric and the normal cases. Fig. 16(a) reports the extended ROC curve accounting for the confidence intervals around the sensitivity and the specificity computed by steps: i)-viii).

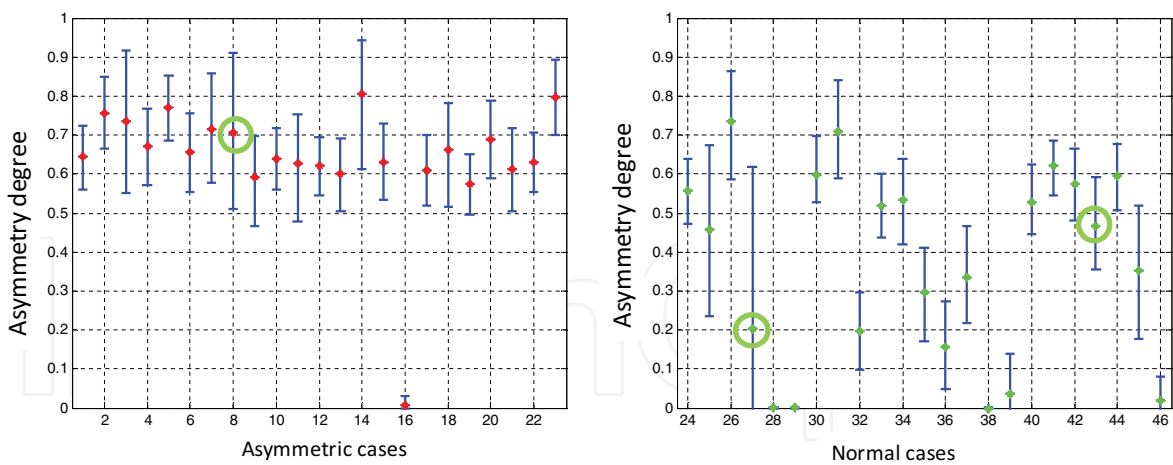


Fig. 15. Asymmetry degree for each case in the data set and the relative confidence interval, due to the noise uncertainty contribution, with 90% coverage probability. (a) Asymmetric cases. (b) Normal cases.

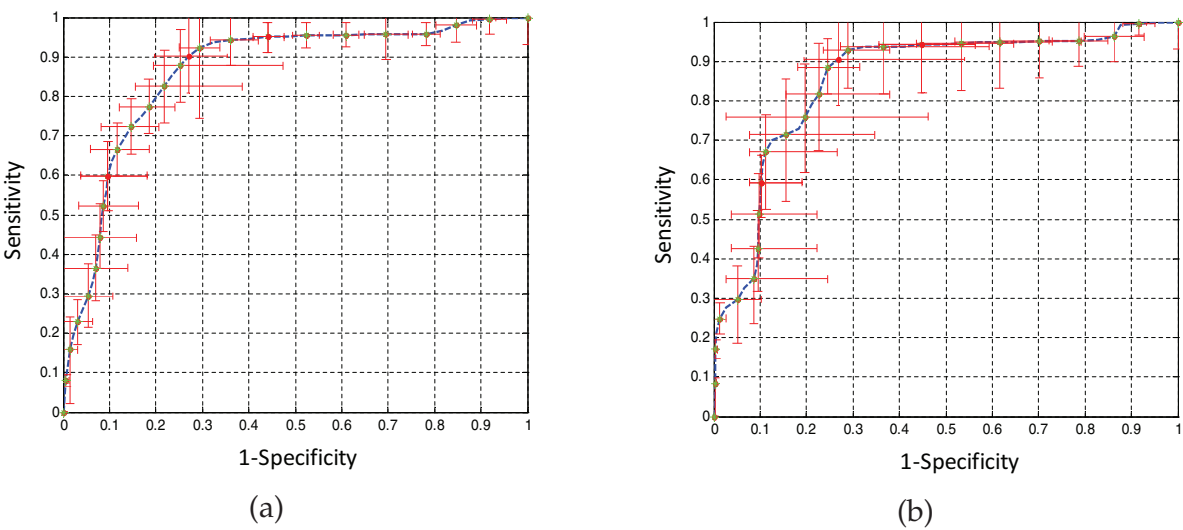


Fig. 16. ROC curve of the classifier (mean value in blue) with the uncertainty contribution due to (a) image noise and (b) ACR density assessment bias propagated to the False Positive Rate and to the True Positive Rate.

6.2 Systematic uncertainty contribution: density category assessment

As seen before, the assessment of the density ACR category plays an important role in the whole algorithm. As a consequence, a different value assessed to the test image by the radiologist could lead to changes in the system performance, which instead was calibrated using tree expert radiologists criterions and their experience. This kind of uncertainty contribution however is not random but systematic. In fact, a radiologist can introduce a subjective bias in assessing the ACR category to the mammographic image, which depends on his/her personal experience or on the particular mammographic equipment he/she is using. Therefore the CAdE system has been tested using to different kind of bias:

- 1) a negative bias modeled as³: $ACR_{biased} = ACR_{unbiased} - 1$;

³ Obviously, in this case density rating equal to 1 remains unchanged.

2) a positive bias modeled as ⁴: $ACR_{biased} = ACR_{unbiased} + 1$.

The following procedure has been implemented to perform the uncertainty propagation of the two kinds of bias:

- i) for each image of the dataset add a negative bias;
- ii) test the CADe algorithm on each image obtained in i) and form the corresponding model value of the output, yielding N output values (where N is again the number of images of the dataset);
- iii) for each image of the dataset add a positive bias;
- iv) test the CADe algorithm on each image obtained in iii) and form the corresponding model value of the output, yielding other N output values;
- v) sort these $2N$ values into strictly increasing order, using the sorted values to provide a discrete representation G of the distribution function for the asymmetry degree;
- vi) use G to form an appropriate confidence interval for the asymmetry degree, for a given coverage probability p .

Fig. 17 reports the asymmetry degrees for the asymmetric and normal cases taken from MiniMIAS database. The confidence interval associated is for a coverage probability equal to $p = 0.90$. The green circles identify the most critical pair of mammograms in terms of density bias influence among the asymmetric and the normal cases. Fig. 16(b) shows the extended

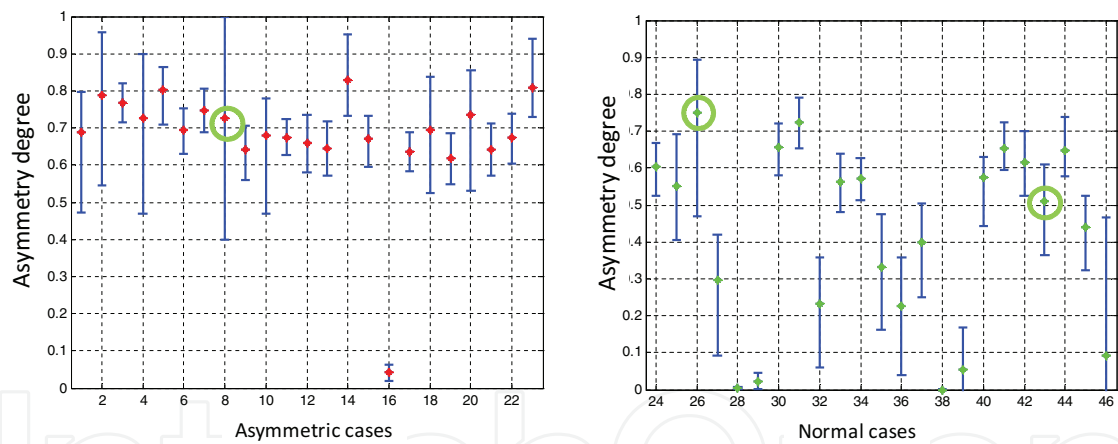


Fig. 17. Asymmetry degree for each case in the data set and the relative confidence interval, due to the noise uncertainty contribution, with 90% coverage probability. (a) Asymmetric cases. (b) Normal cases.

ROC curve accounting for the propagation of the ACR density assessment bias through the whole identification algorithm. A failure analysis could be guided from this kind of validation procedure.

7. Conclusions

This chapter has addressed the problem of metrological validation of a CAD system (both CADe-CADx) in mammography. Guided by the work of Wirth and that of Nishikawa, we

⁴ Obviously, in this case density rating equal to 4 remains unchanged.

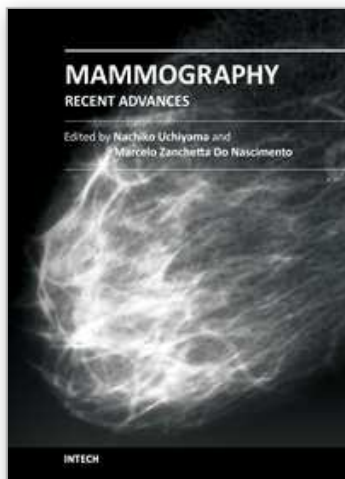
realized that there are still several open issues concerning the validation of CAD in this context. Based on the experience gained in the context of modeling and propagation of measurement uncertainty in the medical field, we have prepared this review providing some guidelines to expand and improve the validation and performance evaluation of CAD. Several examples have been inserted drawing inspiration from the CAD ABCDE that our research team is developing, especially showing those points that are distinctive compared to other commercial CAD already on the market. These points relate mainly to the metrological assessment, to the adaptability to different databases (either analog and digital), to the presence of new functional blocks relating to the identification of bilateral asymmetry, and last but not least to the ability to interact with the radiologist who can add rules in the CAD decision making process. This last functionality is allowed by the use of Fuzzy Inference Systems. Finally, taking a cue from the considerations of Wirth, we have also provided some hints on how to implement a failure assessment of the resulting False Negatives/Positives produced by a CAD.

8. References

- Andersson, E. (2007). *Fuzzy and rough techniques in Medical Diagnosis and Medication*, Studies in Fuzziness and soft computing, Springer.
- Burrell, H., Evans, A., Wilson, A. & Pinder, S. (2001). False-negative breast screening assessment: What lessons we can learn?, *Clin. Radiol.* 56(5): 385 – 388.
- Casti, P. (2011). *Development and validation of a computer-aided detection system for the identification of bilateral asymmetry in mammographic images*, Master's thesis, Medical Engineering, Faculty of Engineering, University of Rome Tor Vergata.
- Chan, T. & Vese, L. (2001). Active contours without edges, *IEEE Trans. Image Proc.* 10(2).
- Cheng, H. (2003). Computer-aided detection and classification of microcalcifications in mammograms: a survey, *Pattern Recognition* 36(12): 2967 – 2991.
- Dubois, D., Foulloy, L., Mauris, G. & Prade, H. (2004). Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities, *Reliable Computing* 10(4): 273 – 297.
- Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C. & Parkin, D. (2010). Estimates of worldwide burden of cancer in 2008: Globocan 2008, *Int. J. of Cancer* 127: 2893 – 2917.
- Ferrari, R., Rangayyan, R., Borges, R. & Frere, A. (2004). Segmentation of the fibro-glandular disc in mammograms using gaussian mixture modeling, *Med. Biol. Eng. Comp.* 42: 378 – 387.
- Ferrero, A. & Salicone, S. (2009). The construction of random-fuzzy variables from the available relevant metrological information, *IEEE Trans. on Instr. and Meas.* 58(2): 365 – 374.
- Ferrero, A., Salicone, S., Mencattini, A., Rabottino, G. & Salmeri, M. (2010). Uncertainty evaluation in a fuzzy classifier for microcalcifications in digital mammography, *IEEE Inst. and Meas. Tech. Conf. (IMTC '10)*, Austin, TX, USA.
- Giuliato, D. & Rangayyan, R. (2011). Modeling and analysis of shape with applications in computer-aided diagnosis of breast cancer, *Synthesis Lect. on Biom. Eng.* .
- Gravel, P., Beaudoin, G. & Guise, J. (2004). A method for modeling noise in medical images, *IEEE Trans. on Medical Imaging* 23(10): 1221 – 1232.
- Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, W., Moore, R., Chang, K. & MunishKumaran, S. (1998). *Digital Mammography*, Kluwer, chapter Current status of the Digital Database for Screening Mammography, pp. 457 – 460.

- International vocabulary of metrology. Basic and general concepts and associated terms (VIM)* (2008). Technical report.
- Jain, A. (1989). *Fundamentals of digital image processing*, Prentice Hall.
- JCGM (2008). Evaluation of measurement data - an introduction to the guide to the expression of uncertainty in measurement 104:2009 and supplement 1 to the guide to the expression of uncertainty in measurement - propagation of distributions using a monte carlo method, *Technical Report 100:2008*, JCGM.
- Mazzarella, F. & Bazzocchi, M. (2007). Cad systems for mammography: a real opportunity? a review of the literature, *La radiologia Medica* 112(3): 329 – 353.
- Mencattini, A., Rabottino, G., Salicone, S. & Salmeri, M. (2009). Uncertainty modeling and propagation through RFVs for the assessment of CADx systems in digital mammography, *IEEE Trans. on Instr. and Meas.* 58(11).
- Mencattini, A., Rabottino, G., Salmeri, M. & Lojacono, R. (2009). An iris detector for tumoral masses identification in mammograms, *IEEE Int. Work. on Med. Meas. and App. (MEMEA '09)*, Cetraro, Cosenza, Italy.
- Mencattini, A., Rabottino, G., Salmeri, M. & Lojacono, R. (2010). Assessment of a breast masses identification procedure using an iris detector, *IEEE Trans. on Instr. and Meas.* 59(10): 2505 – 2512.
- Mencattini, A., Rabottino, G., Salmeri, M., Sciunzi, B. & Lojacono, R. (2010). Denoising and enhancement of mammographic images under the assumption of heteroscedastic additive noise by an optimal subband thresholding, *Int. J. of Wavelets, Mult. Inf. Proc.* 8(5).
- Mencattini, A. & Salmeri, M. (2011a). Breast masses detection using phase portrait analysis and fuzzy inference systems, *International Journal of Computer Assisted Radiology and Surgery*. DOI: 10.1007/s11548-011-0659-0.
- Mencattini, A. & Salmeri, M. (2011b). Breast masses detection using phase portrait analysis, intrinsic coherence, *Computer Assisted Radiology and Surgery (CARS '11)*, Berlin, Germany.
- Mencattini, A., Salmeri, M., Caselli, F., Sciunzi, B. & Lojacono, R. (2008). Subband variance computation of homoscedastic additive noise in discrete dyadic wavelet transform, *Int. J. of Wavelets, Mult. and Inf. Proc.* 6(6): 895–906.
- Mencattini, A., Salmeri, M. & Casti, P. (2011). Bilateral asymmetry identification for the early detection of breast cancer, *IEEE Int. Work. on Med. Meas. and App. (MEMEA '11)*, Bari, Italy.
- Mencattini, A., Salmeri, M., Casti, P., Raguso, G., L'Abbate, S., Chieppa, L., Ancona, A., Mangieri, F. & Pepe, M. (2011). Automatic breast masses boundary extraction in digital mammography using spatial fuzzy c-means clustering and active contour models, *IEEE Int. Work. on Med. Meas. and App. (MEMEA '11)*, Bari, Italy.
- Mencattini, A., Salmeri, M., Lojacono, R. & Arnò, M. (2007). Noise estimation in mammographic images for adaptive denoising, *EFOMP Eur. Conf. on Med. Phys. (EFOMP '07)*, Castelveccchio Pascoli, Italy.
- Mencattini, A., Salmeri, M., Lordi, A., Casti, P., Ferrero, A., Salicone, S., Natoli, S., Leonardis, F. & P.David (2011). A study on a novel scoring system for the evaluation of expected mortality in icu-patients, *IEEE Int. Work. on Med. Meas. and App. (MEMEA '11)*, Bari, Italy.

- Mencattini, A., Salmeri, M., Rabottino, G. & Salicone, S. (2010). Metrological characterization of a CADx system for the classification of breast masses in mammograms, *IEEE Trans. on Instr. and Meas.* 59(11): 2792 – 2799.
- Nishikawa, R. (2007). Current status and future directions of computer-aided diagnosis in mammography, *Comp. Med. Imag. and Graph.* 31: 224 – 235.
- Rabottino, G., Mencattini, A., Salmeri, M., Caselli, F. & Lojacono, R. (2008). Mass contour extraction in mammographic images for breast cancer identification, *IMEKO TC4 Symposium (IMEKOTC4 '08)*, Firenze, Italy.
- Rabottino, G., Mencattini, A., Salmeri, M., Caselli, F. & Lojacono, R. (2011). Performance evaluation of a region growing procedure for mammographic breast lesion identification, *Computer Standard and Interfaces (Elsevier)* 33(2): 128 – 135.
- Raguso, G., Ancona, A., Chieppa, L., L'abbate, S., Pepe, M., Mangieri, F., Palo, M. D. & Rangayyan, R. (2010). Application of fractal analysis to mammography, *IEEE Int. Conf. of the IEEE Engineering in Med. and Biol.*
- Rangayyan, R., Prajna, S. & Ayres, F. (2008). Detection of architectural distortion in mammograms acquired prior to the detection of breast cancer using gabor filters, phase portraits, fractal dimension, and texture analysis, *Int. J. Comput. Assist. Radiol. Surg.* 2(6): 347 – 361.
- Salicone, S. (2007). *Measurement uncertainty. An approach via the mathematical theory of evidence*, Springer, NY, USA.
- Salmeri, M., Mencattini, A., Rabottino, G., Accattatis, A. & Lojacono, R. (2009). Assisted breast cancer diagnosis environment: A tool for DICOM mammographic images analysis, *IEEE Int. Work. on Med. Meas. and App. (MEMEA '09)*, Cetraro, Cosenza, Italy.
- Salmeri, M., Mencattini, A., Rabottino, G. & Lojacono, R. (2008). Signal-dependent noise characterization for mammographic images denoising, *IMEKO TC4 Symposium (IMEKOTC4 '08)*, Firenze, Italy.
- Santo, M. D., Liguori, C., Paolillo, A. & Pietrosanto, A. (2004). Standard uncertainty evaluation in image-based measurements, *Measurements* 36(3/4): 347 – 358.
- Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D. & Savage, J. (1994). The mammographic image analysis society digital mammogram database, in E. A. G. Gale et al. (ed.), *Proc. 2nd Int. Workshop Digital Mammography*, York, U.K., pp. 375 – 378.
- Wirth, M. (2006). *Recent Advances in Breast Imaging, Mammography, and Computer Aided Diagnosis of Breast cancer*, SPIE Press, Bellingham, Washington USA, chapter Performance Evaluation of CADE algorithms in Mammography, pp. 639 – 699.



Mammography - Recent Advances

Edited by Dr. Nachiko Uchiyama

ISBN 978-953-51-0285-4

Hard cover, 418 pages

Publisher InTech

Published online 16, March, 2012

Published in print edition March, 2012

In this volume, the topics are constructed from a variety of contents: the bases of mammography systems, optimization of screening mammography with reference to evidence-based research, new technologies of image acquisition and its surrounding systems, and case reports with reference to up-to-date multimodality images of breast cancer. Mammography has been lagged in the transition to digital imaging systems because of the necessity of high resolution for diagnosis. However, in the past ten years, technical improvement has resolved the difficulties and boosted new diagnostic systems. We hope that the reader will learn the essentials of mammography and will be forward-looking for the new technologies. We want to express our sincere gratitude and appreciation to all the co-authors who have contributed their work to this volume.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Arianna Mencattini and Marcello Salmeri (2012). Metrological Assessment of a CAD System for the Early Diagnosis of Breast Cancer in Digital Mammography, Mammography - Recent Advances, Dr. Nachiko Uchiyama (Ed.), ISBN: 978-953-51-0285-4, InTech, Available from:
<http://www.intechopen.com/books/mammography-recent-advances/metrological-assessment-of-a-cad-system-for-the-early-diagnosis-of-breast-cancer-in-digital-mammogra>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen