An improved method for simulating microcalcifications in digital mammograms

Federica Zanca,^{a)} Dev Prasad Chakraborty, Chantal Van Ongeval, Jurgen Jacobs, Filip Claus, Guy Marchal, and Hilde Bosmans

Department of Radiology, University Hospital Gasthuisberg, Herestraat 49, 3000 Leuven, Belgium

(Received 28 November 2007; revised 17 July 2008; accepted for publication 17 July 2008; published 12 August 2008)

The assessment of the performance of a digital mammography system requires an observer study with a relatively large number of cases with known truth which is often difficult to assemble. Several investigators have developed methods for generating hybrid abnormal images containing simulated microcalcifications. This article addresses some of the limitations of earlier methods. The new method is based on digital images of needle biopsy specimens. Since the specimens are imaged separately from the breast, the microcalcification attenuation profile scan is deduced without the effects of over and underlying tissues. The resulting templates are normalized for image acquisition specific parameters and reprocessed to simulate microcalcifications appropriate to other imaging systems, with different x-ray, detector and image processing parameters than the original acquisition system. This capability is not shared by previous simulation methods that have relied on extracting microcalcifications from breast images. The method was validated by five experienced mammographers who compared 59 pairs of simulated and real microcalcifications in a two-alternative forced choice task designed to test if they could distinguish the real from the simulated lesions. They also classified the shapes of the microcalcifications according to a standardized clinical lexicon. The observed probability of correct choice was 0.415, 95% confidence interval (0.284, 0.546), showing that the radiologists were unable to distinguish the lesions. The shape classification revealed substantial agreement with the truth (mean kappa=0.70), showing that we were able to accurately simulate the lesion morphology. While currently limited to single microcalcifications, the method is extensible to more complex clusters of microcalcifications and to three-dimensional images. It can be used to objectively assess an imaging technology, especially with respect to its ability to adequately visualize the morphology of the lesions, which is a critical factor in the benign versus malignant classification of a lesion detected in screening mammography. © 2008 American Association of Physicists in Medicine.

[DOI: [10.1118/1.2968334](https://doi.org/10.1118/1.2968334)]

Key words: microcalcifications, digital mammography, simulation, lesion characterization, morphology

I. INTRODUCTION

Observer performance studies are widely used to evaluate digital imaging systems. For example, with the increasing adoption of digital mammography there is much interest in evaluating the clinical performance for diagnostic or screening purposes in terms of detectability and characterization of lesions like microcalcifications. The morphologic and distributional characteristics of the microcalcifications in breast tissue are critical factors in assessing if the lesion is malignant and dictate whether or not to perform a biopsy on the woman.^{1,2}

Receiver operating characteristic (ROC) and free-response studies^{3,4} require a relatively large number of cases, typically 100 normal and 100 abnormal and 5–10 readers for ROC studies. In screening mammography abnormal cases are relatively rare since there are approximately 3–6 cancers per 1000 women.¹ Consequently, it is difficult to assemble an adequate number of images to conduct such studies.

There has been much interest in this problem and several simulation models for single microcalcifications or clusters

have been proposed.^{5–10} Basically, the methods involve digitally extracting lesion attenuation profiles from abnormal cases and superposing them on normal images to create hybrid abnormal images. While they covered a wide range of simulation related issues, most of the prior methods employed digitized film mammograms,^{5–9} whereas the current trend is toward fully digital mammography. More significantly, the simulations were based on lesions extracted from projection images of the breast including the embedded lesions. Since the net projection attenuation is the lesion attenuation plus the over- and underlying tissue attenuations, the contribution of the lesion by itself cannot be separated. In addition the extracted microcalcification will also depend on film, developer, and film digitizer. Therefore, none of these methods can be applied to simulate images acquired under different conditions (image processing, x-ray parameters, detector type, modulation transfer function, etc.) than those used by these authors. In other words, the simulations would have to be repeated for every new imaging technique that has to be evaluated.

FIG. 1. Illustrations of the morphologic characteristics of the microcalcifications according to Le Gal's classification system. Type 1 calcifications were defined as annular; type 2 defined as regular punctiform (point-like); type 3 defined as dusty; type 4 defined as irregularly punctiform, and type 5 defined as vermicular (worm-like) calcifications. The simulation procedure described in this article attempts to accurately preserve the shape information, which is critical in diagnosing malignancy.

To overcome this limitation, we previously proposed a method¹¹ which used microcalcifications extracted from digital images of biopsy specimens (not the whole breast). In other words, we physically separated the microcalcification from over- and underlying tissues. However, the prior approach has some limitations. While it starts from the actual gray values of the extracted lesion, it does not utilize them as such; rather the gray values are used to model the microcalcification, which is segmented in a series of annular regions with gray values assigned subjectively and manually by a method similar to that described by Lefebvre *et al.*⁵ The method was very time consuming and consequently only a limited number (33) of simulated microcalcifications could be segmented. The procedure was limited to microcalcifications with relatively simple shapes. As noted previously, it is critical to accurately mimic the morphologic and distributional characteristics of real calcifications, and this was not possible with the previous method.

The aim of this work was: (1) to facilitate observer performance studies in digital mammography by developing an improved method for simulating breast microcalcifications that would enable the creation of a large number of hybrid mammograms with realistic microcalcifications, for detec-

tion and discrimination studies, and (2) to validate the presented method using concept borrowed from current state-of-the-art observer performance techniques.^{12,13}

II. MATERIALS AND METHODS

II.A. Le Gal classification

Since lesion shape fidelity is critical to a good simulation method, a brief overview of the shape classification system used in the clinic follows. In Europe, the Le Gal classification¹⁴ system is widely used. There are five Le Gal classes: type 1 calcifications is defined as annular; type 2 is defined as regular punctiform (point-like); type 3 is defined as too fine to characterize the shape (dusty); type 4 is defined as irregularly punctiform; and type 5 is defined as vermicular (worm-like) calcifications. Figure 1 is an artist's rendition of the morphological characteristics of the microcalcifications according to the Le Gal classification. Because in the United States the breast imaging reporting and data system¹⁵ (BI-RADS) is used rather than Le Gal classification, the correspondence between the Le Gal terminology,¹⁴ approximate probability of malignancy,¹⁴ and BI-RADS is shown in Table I.¹⁶

II.B. Creation of template microcalcifications with known Le Gal types

Twenty-seven (27) stereotactically guided vacuum large-core needle biopsy specimens (11 gauge) were obtained with a digital biopsy table [Mammotest Plus/S (Fisher Imaging, Norderstedt)]. Digital images of the biopsied specimens, superposed on 40 mm poly(methyl methacrylate) (PMMA) (for simulation of a 45 mm breast, 50% glandularity), were acquired with a Siemens Novation DR (Erlangen, Germany) machine, at 27 kVp, 180 mA s (for maximum dose), molybdenum-molybdenum (Mo/Mo) anode-filter combination and magnification of 2 (for maximum resolution). Figure 2 shows a typical biopsy specimen used in our simulation methodology. It consists of breast tissue containing suspicious clusters of microcalcifications and extracted under stereotactic guidance (needle of 11 gauge).

A total of 97 regions of interest (ROIs), each containing a single microcalcification were digitally extracted from the 27

TABLE I. The correspondence between Le Gal terminology, approximate probability of malignancy and BI-RADS terminology is shown this table.

Type of Le Gal	% of malignancy ^a	BI-RADS terminology
(1) Annular	0	Lucent centered calcium deposit, "eggshell" or "rim" calcifications
(2) Regular punctiform	22	Round punctuate calcifications
(3) Dusty	40	Amorphous or indistinct calcifications
(4) Irregular punctiform	66	Pleomorphic or heterogeneous calcifications (granular)
(5) Vermicular	100	Fine linear or fine linear branching (casting) calcifications

^aReference 13.

FIG. 2. A photograph of a typical biopsy specimen used for our simulation methodology. It consists of breast tissue containing clusters of microcalcifications and extracted under stereotactic guidance (11 gauge needle).

specimen images. Each ROI contained a microcalcification surrounded by a relatively small amount of uniform anatomical background tissue and is referred to as the starting template T_0 . Figure 3 shows an image of a biopsy specimen superposed on 40 mm PMMA, and imaged at 27 kVp, Mo/Mo filter and 180 mA s. A Le Gal II microcalcification containing ROI is shown in the small box, and a magnified view is shown next to it. This is an example of the starting template T_0 .

The size of the ROI was chosen as a compromise between not too small, which would crop microcalcification pixels, and not too large, which would include excessive background tissue, resulting in more background nonuniformity and possibly create a visible artifact. The pixel values of each ROI were normalized to the average pixel value of several background regions chosen to be outside the microcalcification but inside the ROI. Due to the high microcalcification contrast, (maximum difference between gray values of microcalcification and averaged background, divided by the averaged background), this was a relatively unambiguous

FIG. 3. An x-ray image of a biopsy specimen superposed on 40 mm PMMA. A Le Gal II microcalcification containing ROI is shown in the small box, and a magnified view is shown next to it. This is an example of the starting template T_0 . Note the background inside the ROI is quite uniform compared to that observed in mammograms. This shows that this method of extracting the lesion is more accurate than previous methods that start with mammograms.

process, and the resulting background pixel value variation was less than 1.5%. Microcalcifications surrounded by anatomical background with larger variability were discarded. In the resulting template T_1 , the gray values of the background were close to 1, while the other values, corresponding to pixels within the microcalcifications, were smaller than 1.

The Le Gal classification of each template was established by an independent experienced mammographer, who did not participate in the subsequent observer study (CVO): 7 were classified Le Gal type 1, 24 as Le Gal type 2, 21 as Le Gal type 3, 36 as Le Gal type 4, and 9 as Le Gal type 5. Pathology results of the specimens were as follow: 19% benign, fibrosis; 22% fibrocystic disease with sclerosing adenosis; 11% fibrocystic disease with ductal hyperplasia; 48% atypical ductal hyperplasia [which turned out to be malignant ductal carcinoma *in situ* after surgery].

II.C. Simulation of microcalcifications for other chosen system parameters

To simulate the appearance of a specific microcalcification for another choice of system parameters one needs to correct for relative differences between the acquisition condition and the desired condition. To correct for kVp and beam quality differences the pixel values were converted to Al-equivalent thicknesses as follows.^{17,18} Aluminium was chosen as the reference element since it is close in atomic number to the composition of microcalcifications¹⁹ and is therefore expected to respond similarly to changes in kVp and beam quality. We exposed 0.1, 0.2, 0.3, and 0.4 mm Al of 99.9% purity (Nuclear Associates, NY) superposed on 40 mm PMMA at 27 kVp and 180 mA s, Mo/Mo filter, and calculated the average x-ray transmissions of the different sheets. This calibration measurement allowed us to linearly interpolate the gray values in the template T_0 to Al thicknesses for breast equivalent to 40 mm PMMA. Additional calibrations were performed with 0.1, 0.2, 0.3, and 0.4 mm Al superposed on different thicknesses of PMMA (20, 40, 50, 60, and 70 mm) and using also different anode-filter combination than Mo/Mo (molybdenum-rhodium, Mo/Rh, and tungsten-rhodium, W/Rh),²⁰ to allow us to convert the gray values to Al thicknesses appropriate to different breasts thicknesses and beam qualities. In this way, a template T_1 , for any specified kVp, beam quality and breast thickness can be calculated from the initial normalized template.

Template T_1 was corrected for the difference in the system modulation transfer functions (MTFs) between the original (MTF_1) and the desired system (MTF_2), yielding template T_2 . The MTF includes all presampling components, e.g., focal spot size and magnification, detector. The presampling MTFs along the horizontal and vertical direction scan are determined using the edge method.²¹ The template T_2 was resampled according to the new pixel size of the detector yielding the final template T_3 . The following equation summarized these operations:

$$T_2 = \text{FFT}^{-1} \left(\text{FFT}(T_1) \frac{MTF_2}{MTF_1} \right),$$

$$T_3 = \text{REBIN}(T_2).$$

Here, FFT represents the two-dimensional discrete Fourier transform, FFT^{-1} is the corresponding inverse transform and REBIN is the rebinning operator that resamples the image at a different pixel size using bilinear interpolation when magnifying or nearest-neighborhood averaging when minifying. In the present validation study we simulated microcalcifications in mammograms acquired with the same mammographic x-ray unit (Novation DR, Siemens, Erlangen, Germany) as used for the acquisition of the biopsied microcalcifications, so MTF_1 and MTF_2 are the same, except for the twofold magnification of the original image acquisition.

From the 97 templates, 166 artificial microcalcifications with different x-ray transmission were generated. The sizes of the simulated microcalcifications varied from 200 to 1238 μm , contrast varied from 1.3% to 25%, and Al-equivalent thicknesses varied from 20 and 325 μm .¹⁷ Out of the 166 artificial microcalcifications, 59 were selected for the validation experiment, in order to obtain a representative sample of the morphology.

Fifty-nine (59) mammograms acquired with the x-ray machine described above containing real microcalcifications were selected for the validation study and stored as raw data, i.e., the detector signal, prior to any image processing. These images were distinct from the ones used for the creation of the templates. For each image the experienced mammographer, referred to earlier, identified a location near a real microcalcification at which to insert a simulated microcalcification with the same Le Gal classification. The simulated microcalcification was inserted by multiplying the final template T_3 described above, with the raw data of the mammogram in a template-sized region centered on the location identified by the mammographer. The entire image then underwent the clinically used image processing algorithm, Op-View v1 (Siemens, Erlangen, Germany) for this machine yielding the final image, containing a real and an adjacent simulated microcalcification, which was viewed by the observers. Other real microcalcifications in the image were not used in the observer study.

II.D. Validation of the simulation

II.D.1. Observer study

An observer study employing the two-alternative forced choice (2AFC) paradigm was used to validate the simulation, i.e., tested whether simulated lesions could be distinguished from real lesions. A rectangle, which could be toggled on or off, was overlaid on each of the two microcalcifications to be scored (one real, one simulated). As multiple real microcalcifications often occur in the same image, the markers were necessary to identify the ones to be scored. The observers were told that one of the two microcalcifications was simulated and the other was real and they were asked to find the simulated one by clicking on it. The observers were also asked to classify each microcalcification according to the Le Gal classification system. Images were scored blindly by five

radiologists. All readers had at least 10 years experience in mammography (2 had 10 years, 3 more than 15 years experience).

The images were viewed on a soft-copy viewing station using a graphical user interface (OPerAA) (Ref. 17) that had been extended to include also the scoring of the morphology. A training session (eight images) preceded the evaluation, in order to familiarize the radiologists with the user interface. The radiologists were allowed to window and level the images as desired, as well as to magnify to a maximum of 200%. The study was conducted in a darkened room (ambient light <10 lux) on high quality 5 megapixel monitors intended for mammographic applications: Barco LCD 5MP MFGD5621HD, Barco CRT 5MP, and Barco LCD NIO 5MP MDNG 5121 BB (Barco, Kortrijk, Belgium), calibrated by the vendor to the DICOM grayscale standard display function.²² Each reader was assigned to one monitor for the whole study.

II.D.2. Statistical analysis

For a given reader the probability of correct decision (PC) was defined as the average of 1's and 0's for all the images, where 1=correct decision, the observer indicated the simulated lesion, and 0=incorrect, the observer indicated the real lesion. Under the null hypothesis (NH) that the real and simulated lesions cannot be distinguished, $\text{PC}=0.5$. The data from this multiple-reader multiple-case (MRMC) 2AFC study were analyzed by estimating a nonparametric bootstrap confidence interval for the reader-averaged PC.²³ Two thousand (2000) bootstrap samples were generated, where each sample consisted of 59 images sampled with replacement from the image pool (a particular image could occur 0 or multiple times) and five readers sampled with replacement from the reader pool. The PC averaged over the readers was calculated for the particular bootstrap sample. The histogram of the 2000 PC values was percentiled to determine the 95% confidence interval (CI). In other words we empirically determined the cutoffs at 2.5% and 97.5% percentile of this distribution.

Since shape is an important predictor of cancer, one of the aims of this study was to accurately simulate the morphology of the lesions. In addition to choosing the simulated lesion from the 2AFC pair, each radiologist also provided Le Gal classifications for each lesion in the pair. The commonly used Cohen's kappa statistic²⁴ is a measure of agreement beyond that which would be expected by chance. For each reader, the unweighted kappa statistic for the Le Gal classifications for each lesion in the pair, and its 95% confidence interval, was calculated using the online kappa calculator.³⁴ By convention²⁵ $\kappa < 0$ is interpreted as "poor" agreement, $0 < \kappa < 0.20$ as "slight" agreement, $0.21 < \kappa < 0.40$ as "fair" agreement, $0.41 < \kappa < 0.60$ as "moderate" agreement, $0.61 < \kappa < 0.80$ as "substantial" agreement and $0.81 < \kappa < 1.0$ as "almost perfect" agreement.

The agreement analysis regards the radiologist's Le Gal classification of the real lesion in the 2AFC pair as the truth, since the radiologist is considered as an expert. His classifi-

cation could be different from that of the independent expert who did not participate in the observer study. While experts do not always agree, disagreement greater than one Le Gal level separation is rare. The independent expert classifications were used to create the test images. The two lesions in a 2AFC pair, real and simulated, came from different physical lesions. The lesions were paired to have the same independent expert classifications (for the simulated lesion this was that of the template from which it was created). The agreement between the classifications of all lesions (two per image) versus those of the independent expert would not be a good measure of the quality of the simulation; rather, it measures the reliability of the Le Gal lexicon, which is not the aim of this work. Similarly, kappas between the radiologist's and the independent expert's classifications for real lesions or simulated lesions are not good measures of the quality of the simulation.

III. RESULTS

From the 2AFC study, the 95% confidence interval CI for the probability of correct decision (PC) was (0.284, 0.546). Since it includes the value 0.5, the NH hypothesis cannot be rejected. In other words the artificial microcalcifications could not be distinguished from the real ones.

In Fig. 4, second column, examples of simulated microcalcifications are indicated by arrows, while the other microcalcifications are real. An artist's rendition of the corresponding Le Gal type is shown in the first column, and the third column is a magnified view of the simulated microcalcification. In the observer study the mammographer saw the entire image, not just the ROI. Le Gal Type 1 microcalcifications (first row) can occur with or without a "hole;" the example shown happens to be of one without a hole. The correspondence between the simulated microcalcifications and the Le Gal types is difficult to appreciate in the printed image, but was much clearer to the mammographers who viewed them on high quality monitors.

Table II lists the results of the agreement analysis of the Le Gal classifications for each lesion in the 2AFC pair, real and simulated. Listed are κ values for each reader and the corresponding 95% confidence intervals, along with the conventional interpretations of the numbers. The mean kappa was 0.70 (substantial agreement) and the range was 0.53 (moderate agreement) to 0.82 (almost perfect agreement).

IV. DISCUSSION

The presented method is similar to our previous approach¹¹ in so far as it is based on microcalcifications extracted from digital images of biopsy specimens, but unlike the previous approach it uses the actual gray levels of the extracted microcalcifications, without recourse to modeling, thereby fully preserving the shape of the original microcalcifications.

As evidenced by the substantial agreement between the radiologists' assessments of shape for the simulated and real lesions, as measured by the kappa statistic (mean=0.70), we were able to accurately simulate the morphology. The 2AFC

FIG. 4. This figure shows typical simulated images. The second column, shows examples of simulated microcalcifications, pointed to by the arrows, the other microcalcifications are real. A drawing of the corresponding Le Gal type is shown in the first column, and the third column is a magnified view of the simulated microcalcification. In the observer study the mammographer saw the entire image, not just the ROI. Le Gal Type 1 microcalcifications (first row) can occur with or without a "hole;" the example shown happens to be of one without a hole. The correspondence between the simulated microcalcifications and the Le Gal types is difficult to appreciate in the printed image, but was much clearer to the mammographers who viewed them on high quality monitors.

observer study also showed that the simulated lesions could not be distinguished from real lesions, i.e., the confidence interval for the percentage correct in the 2AFC task could not rule out 50% correct. Therefore, subject to study limitations noted later, the goals of the study were met.

The validation of the simulations was based on current state-of-the-art observer performance techniques.^{12,13} Previous validation work involved relatively few radiologists (typically 2 or 3) and cases (typically 50, except⁵ which in-

TABLE II. The agreement, as measured by the kappa statistic, between the Le Gal classifications of the real and simulated lesions presented as a 2AFC pair in the same image. Listed are the values of the kappa statistic, the 95% confidence interval, and the conventional interpretations of the kappa values for each observer.

Radiologist	kappa, (95% confidence interval)	Interpretation
1	0.80 (0.66,0.94)	Substantial agreement
2	0.66 (0.5,0.81)	Substantial agreement
3	0.53 (0.35,0.71)	Moderate agreement
4	0.73 (0.58,0.88)	Substantial agreement
5	0.82 (0.69,0.94)	Almost perfect agreement

volved 100) and none of the methods employed statistical analysis that combined the measurements from different readers into a single statement of equivalence between simulated and real lesions that generalizes to the case and radiologist populations; in other words the intrinsic MRMC structure of the data was not exploited.¹²

The new method lends itself to automation (but achieving this was not a goal), therefore allowing for the creation of large databases, and involves minimal subjectivity and associated user variability, compared to the previously mentioned methods, that rely on manual intervention of the person performing the simulation.^{7,8,11} The fact that we were able to generate 166 lesions versus 33 lesions in our earlier study¹¹ shows the potential advantage of the present approach in terms of the ability to generate large numbers of simulated lesions.

At our institution we routinely perform about 5–6 biopsies per week. Consequently, the template database can be continually enriched as new specimens become available. Acquisition of the specimen image is immediately following biopsy; template extraction depends on the number of microcalcifications present in the specimen and it generally takes about 15 min from biopsy to create the final template. Extension of this procedure to clusters is possible. Specimens from core needle biopsy procedures acquired with a larger needle (7 gauge instead of 11) can include entire clusters, whose individual microcalcifications can be extracted, using the approach described in this article, while preserving the spatial distribution of the microcalcifications in the cluster. The dimension of the extracted cluster is obviously limited by the dimension of the needle and very large clusters cannot be simulated. Care has to be taken in selecting the starting templates where non-uniformity of the anatomical background is negligible, see Fig. 3. In principle the method could even be extended to three-dimensional (3D) breast images, such as are used in breast computed tomography and breast tomosynthesis, by imaging the specimens and determining the 3D distribution of attenuation values from the projections of microcalcifications into raw image series of the 3D modalities (prior to reconstruction).

Our method provides realistic microcalcifications with well-defined characteristics, representative for the whole spectrum of microcalcifications seen in mammography. Similar to other studies, the simulated microcalcifications can be used to investigate possible dose reduction methods;^{26,27} optimization of exposure settings in terms of kVp and/or anode/filter combination; effect of anatomical background;^{28–30} comparison of different displaying³¹ modalities; and evaluation of image processing algorithms.^{32,33} Likewise different image processing algorithms could be evaluated for their ability to enhance the detectability of the lesions and to reproduce the morphology of the lesion. We are currently working at a database of templates that could be used for standardized comparative studies of image postprocessing. We intend to make our template database available on the web. This could be used by software developers to optimize image processing algorithms, train computer aided design algorithms, etc.

The study has some limitations. (1) Creation of the template cannot entirely suppress the background anatomical “noise.” However, since the method starts with very thin specimen the problem is less than that with previous methods which had to contend with varying background from over and underlying tissue of the whole breast. Figure 3 shows that the background is much smoother than that of mammograms, which is why we believe that the extracted templates are closer to the truth than is possible with the other methods. (2) The simulation model is strictly valid for breasts with 50% fat–50% glandular composition. Consequently the x-ray transmissions of simulated microcalcifications will be slightly too low in glandular tissue and slightly too high in fatty tissue. Using tissue equivalent slabs with different glandular equivalence, such as are commercially available (CIRS, Norfolk, VA) the method can be extended to different glandular compositions along the lines indicated in the methods section. (3) The method requires cooperation from the manufacturer since one needs access to the raw data and the ability to reprocess the raw data with the added microcalcifications. (4) While single microcalcifications can be simulated, simulating whole clusters is more difficult, but as noted above, in principle it can be done. (5) While we have described how to simulate images appropriate to other imaging systems, we have not validated this aspect of the method. (6) Extension of the method to simulating masses is also desirable.

V. CONCLUSIONS

An improved method for simulating microcalcifications in digital mammograms has been described. Based on the observer study, simulated microcalcifications could not be distinguished from the real ones and the method preserved the shapes of the microcalcifications. The method is expected to be useful for generating large numbers of images with known truth for observer performance studies for imaging system optimization.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Marie-Christine Wagnon, Dr. Charlie Wagnon, Dr. Maximilian Kadivec, Dr. Nagoda-Niklewicz Halina, and Dr. Maria Bernathova, who participated as observers in the studies. This reading session was organized in the framework of a multicenter workshop within the Sentinel project. This is an EC supported project (FP6-012909) on safety and efficacy for new techniques and imaging using new equipment to support European legislation. We would like to thank Thomas Mertelmeier and Daniel Fisher from Siemens (Erlangen, Germany) for all practical help with the reprocessing of the Novation DR images. Herman Pauwels is greatly acknowledged for his kind help with all practical problems. One of the authors (D.P.C.) was supported by grants from the Department of Health and Human Services, National Institute of Health, Grant Nos. R01-EB005243 and R01-EB006388.

- ^{a)}Electronic mail: federica.zanca@uz.kuleuven.ac.be
- ¹R. R. Millis *et al.*, "The detection and significance of calcifications in the breast: A radiological and pathological sign," *Br. J. Radiol.* **49**, 12–26 (1976).
 - ²M. Lanyi, *Diagnosis and Differential Diagnosis of Breast Calcifications* (Springer, New York, 1986).
 - ³C. E. Metz, "Receiver operating characteristic analysis: A tool for the quantitative evaluation of observer performance and imaging systems," *J. Am. Coll. Radiol.* **3**, 413–422 (2006).
 - ⁴D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: modelling, analysis, and validation," *Med. Phys.* **31**, 2313–2330 (2004).
 - ⁵F. Lefebvre *et al.*, "Simulation model of clustered breast micro calcifications," *Med. Phys.* **21**, 1865–1874 (1994).
 - ⁶M. Lado *et al.*, "Real and simulated clustered microcalcifications in digital mammograms. ROC study of observer performance," *Med. Phys.* **24**, 1385–1394 (1997).
 - ⁷R. Saunders *et al.*, "Simulation of mammographic lesions," *Acad. Radiol.* **13**, 860–870 (2006).
 - ⁸M. Kallergi *et al.*, "Simulation model of mammographic calcifications based on the American College of Radiology Breast Imaging Reporting and Data System, or BIRADS," *Acad. Radiol.* **5**, 670–679 (1998).
 - ⁹A. E. Burgess and S. Chakraborty, "Producing lesions for hybrid mammograms: Extracted tumours and simulated microcalcifications," *Proc. SPIE* **3663**, 316–322 (1999).
 - ¹⁰A. E. Burgess *et al.*, "Lesion detection in digital mammograms," *Proc. SPIE* **4320**, 555–560 (2001).
 - ¹¹A. K. Carton *et al.*, "Development and validation of a simulation procedure to study the visibility of microcalcifications in digital mammograms," *Med. Phys.* **30**, 2234–2240 (2003).
 - ¹²D. D. Dorfman *et al.*, "ROC characteristic rating analysis: Generalization to the Population of Readers and Patients with the Jackknife method," *Invest. Radiol.* **27**, 723–731 (1992).
 - ¹³S. V. Beiden *et al.*, "Components-of variance models and multiple-bootstrap experiments: An alternative method for random-effects, receiver operating characteristic analysis," *Acad. Radiol.* **7**, 341–349 (2000).
 - ¹⁴M. Le Gal *et al.*, "Diagnostic value of clustered microcalcifications discovered by mammography," *Bull. Cancer (Paris)* **71**, 57–64 (1984).
 - ¹⁵American College of Radiology (ACR), *Breast Imaging Reporting and Data System Atlas (BI-RADS® Atlas)*, 4th ed. (American College of Radiology, Reston, VA, 2003).
 - ¹⁶J. Stines, "BI-RADS: Use in the French radiologic community How to overcome with some difficulties," *Eur. J. Radiol.* **61**, 224–234 (2007).
 - ¹⁷F. Zanca *et al.*, "Development and implementation of a user friendly and automated environment for the creation of databases of digital mammograms with simulated microcalcifications," *Proc. SPIE* **6146**, 614615–1–614615-8 (2006).
 - ¹⁸A. K. Carton *et al.*, "Quantification of Al-equivalent thickness of just visible microcalcifications in full-field digital mammograms," *Med. Phys.* **31**, 2165–2176 (2004).
 - ¹⁹A. Fandos-Morea *et al.*, "Breast tumors: Composition of micro calcifications," *Radiology* **169**, 325–327 (1988).
 - ²⁰P. Toroi *et al.*, "Experimental investigation on the choice of the tungsten/rhodium anode/filter combination for an amorphous selenium-based digital mammography system," *Eur. Radiol.* **17**, 2368–2375 (2007).
 - ²¹A. K. Carton *et al.*, "Validation of MTF measurement for digital mammography quality control," *Med. Phys.* **32**, 1684–1695 (2005).
 - ²²NEMA, *Digital Imaging and Communications in Medicine (DICOM) Part 14: Gray Scale Standard Display Function, NEMA PS 3* (NEMA, Rosslyn, VA, 2000).
 - ²³B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (Capital City, Montpelier, 1982).
 - ²⁴J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.* **20**, 37–46 (1960).
 - ²⁵J. R. Landis and G. G. Koch, "The measurements of observer agreement of categorical data: The kappa coefficient," *Biometrics* **33**, 159–174 (1977).
 - ²⁶E. Samei *et al.*, "Digital mammography: Effects of reduced radiation dose on diagnostic performance," *Radiology* **243**, 396–404 (2007).
 - ²⁷M. Ruschin *et al.*, "Dose dependence of mass and microcalcification detection in digital mammography: Free response human observer studies," *Med. Phys.* **34**, 400–407 (2007).
 - ²⁸F. Zanca *et al.*, "Evaluation of the global effect of anatomical background on microcalcification detection," *Proc. SPIE* **6515**, 651556-1–651556-12 (2007).
 - ²⁹E. Samei *et al.*, "Detection of subtle lung nodules: Relative influence of quantum and anatomic noise on chest radiographs," *Radiology* **213**, 727–734 (1999).
 - ³⁰J. Beutel, H. Kundel, and R. Van Metter, *Handbook of Medical Imaging, Vol. 1: Medical Physics and Psychophysics* (SPIE, Bellingham, 2000).
 - ³¹N. Buls *et al.*, "Influence of display quality on radiologists' performance in the detection of lung nodules on radiographs," *Br. J. Radiol.* **80**, 738–743 (2007).
 - ³²B. M. Hemminger *et al.*, "Improving the detection of simulated masses in mammograms through two different image-processing techniques," *Acad. Radiol.* **8**, 845–855 (2001).
 - ³³S. Suryanarayanan *et al.*, "Detection of simulated lesions on data-compressed digital mammograms," *Radiology* **236**, 31–36 (2005).
 - ³⁴Online kappa calculator available at <http://faculty.vassar.edu/lowry/VassarStats.html> on May 23, 2008.