Eur Radiol (2012) 22:908–914
DOI 10.1007/s00330-011-2320-2

BREAST

Increase in perceived case suspiciousness due to local contrast optimisation in digital screening mammography

Roelant Visser · Wouter J. H. Veldkamp · David Beijerinck · Petra A. M. Bun · Jan J. M. Deurenberg · Mechli W. Imhof-Tas · Klaas H. Schuur · Miranda M. Snoeren · Gerard J. den Heeten · Nico Karssemeijer · Mireille J. M. Broeders

Received: 6 June 2011 / Revised: 31 August 2011 / Accepted: 12 September 2011 / Published online: 10 November 2011

Abstract

Objectives To determine the influence of local contrast optimisation on diagnostic accuracy and perceived suspiciousness of digital screening mammograms.

Methods Data were collected from a screening region in the Netherlands and consisted of 263 digital screening cases (153 recalled,110 normal). Each case was available twice, once processed with a tissue equalisation (TE) algorithm and once with local contrast optimisation (PV). All cases had digitised previous mammograms. For both algorithms, the probability of malignancy of each finding was scored independently by six screening radiologists. Perceived case suspiciousness was defined as the highest probability of malignancy of all findings of a radiologist within a case. Differences in diagnostic accuracy of the processing algorithms were analysed by comparing the areas under the receiver operating characteristic curves (A_z). Differences in perceived case suspiciousness were analysed using sign tests.

Results There was no significant difference in A_z (TE: 0.909, PV 0.917, P=0.46). For all radiologists, perceived case suspiciousness using PV was higher than using TE more often than vice versa (ratio: 1.14–2.12). This was significant (P <0.0083) for four radiologists.

Conclusions Optimisation of local contrast by image processing may increase perceived case suspiciousness, while diagnostic accuracy may remain similar.

Key Points

- *Variations among different image processing algorithms for digital screening mammography are large.*
- *Current algorithms still aim for optimal local contrast with a low dynamic range.*

R. Visser · W. J. H. Veldkamp · P. A. M. Bun · K. H. Schuur · G. J. den Heeten · M. J. M. Broeders
National Expert and Training Centre for Breast Cancer Screening,
P.O. Box 6873, 6503 GJ, Nijmegen, the Netherlands

W. J. H. Veldkamp · P. A. M. Bun
Department of Radiology, Leiden University Medical Centre,
Albinusdreef 2,
2333 ZA, Leiden, The Netherlands

D. Beijerinck · J. J. M. Deurenberg
Screening Program Early detection of breast cancer
in the Centre/Mid-West Part of the Netherlands,
Utrecht, the Netherlands

M. W. Imhof-Tas · M. M. Snoeren · N. Karssemeijer
Department of Radiology,
Radboud University Nijmegen Medical Centre,
Nijmegen, the Netherlands

M. W. Imhof-Tas · M. M. Snoeren
Screening Program Early detection of breast cancer
in the Eastern Part of the Netherlands,
Nijmegen, the Netherlands

G. J. den Heeten
Department of Radiology,
Academical Medical Center Amsterdam,
Amsterdam, The Netherlands

M. J. M. Broeders
Department of Epidemiology, Biostatistics and HTA,
Radboud University Nijmegen Medical Centre,
Nijmegen, The Netherlands

W. J. H. Veldkamp (✉)
Department of Radiology, Leiden University Medical Centre,
PO Box 9600, 2300 RC, Leiden, The Netherlands
e-mail: w.j.h.veldkamp@lumc.nl

- *Although optimisation of contrast may increase sensitivity, diagnostic accuracy is probably unchanged.*
- *Increased local contrast may render both normal and abnormal structures more conspicuous.*

Keywords Mammography · Breast · Image processing · Screening · Recall rate

Introduction

Many studies have shown that conversion to digital mammography can increase screening sensitivity [1, 2]. Another known consequence is an increased recall rate [2, 3] especially in the first period after implementation [4, 5]. This increase can be explained partially by increased visibility of microcalcifications [3, 4], but differences in the appearance of digital and analogue mammograms may also be of influence.

Images acquired using a digital mammography system must be processed before they are suitable for display. Image processing converts the images so that they can be interpreted by radiologists. Because digital mammography is a relatively new technique, it is continuously being developed. Important factors like X-ray spectrum and image processing have not yet been fully optimised. As a consequence, the variations among different commercially available processing algorithms are large. When comparing image processing algorithms, one should concentrate on diagnostic accuracy rather than on the appealingness of the images. Because of the lack of easy and objective methods for measuring processed image quality however, we often have to rely on the impression experts have of the appearance of images to rate image processing [6–8].

Originally developments in image processing were mainly pushed by the need to decrease the image dynamic range, because the dynamic range of softcopy reading stations was much smaller than that of the films displayed on a light box [9, 10]. Although the dynamic range of modern display stations is increasing rapidly, current image processing algorithms still aim for a maximum (optimal) local contrast while decreasing the total image dynamic range. Such contrast optimisation techniques can have a large impact on the appearance of images. These techniques are aimed at increasing the diagnostic accuracy, although they could also influence the perceived suspiciousness of healthy breast tissue. In addition to this, some studies have shown that differences in image processing can influence both sensitivity and specificity [11–14].

The purpose of this study is to determine the influence that local contrast optimisation has on diagnostic accuracy and the perceived suspiciousness of digital screening mammograms.

Materials and methods

Study dataset

The data for this study were collected from a screening region in the eastern part of the Netherlands in the period April 2007 up to November 2007. This screening region had converted to digital mammography several months before this period. After digitisation, a temporal increase in recall rate equivalent to those described in recent studies [4, 5] was observed. Although the recall rate was not as high as during the first month after conversion (6%), during the full study period an increased recall rate was observed (3–4%) compared with the recall rate for the analogue screening (2%). The recall rate dropped back further (2–3%) after the study period.

The dataset for this study contains all recalled cases in the study period for which digitised previous mammograms of the previous screening round were available (153 studies). 43 of these cases were biopsy-confirmed true-positives (TP), 110 cases were negative (FP). For each negative, the last non-recalled case that was acquired before it and for which the previous mammograms were also present, was added to the dataset. This last group of cases are referred to as the normal cases (N). There was a total of 263 cases. The age range of women in the study was 51–86, and the median age was 60. Approval of the institutional review board was not required. Informed consent was obtained from the participants and all cases were anonymised.

All cases were acquired using the General Electric (GE) Senographe Essential (GE Medical Systems, Buc, France) digital mammography system. GE provides two processing algorithms for this system; the standard processing algorithm Tissue Equalisation (TE) and the local contrast optimisation algorithm Premium View (PV) which can be applied as an additional processing step after TE.

All previous mammograms had been routinely digitised using an Array 2,905 Laser Film Digitizer (Array Corporation Europe, Roden, the Netherlands) at a resolution of 100 µm, because an earlier study had shown this to be sufficient for comparison of previous mammograms [15]. For views consisting of multiple images (mosaics) only the image containing the largest part of the breast was digitised.

Postprocessing methods

Tissue Equalisation (TE) is a standard General Electric application that corrects for low frequency variations resulting from under- and over-penetration of X-rays (with the latter occurring for example at the breast edge). As a result the image dynamic range is reduced, enabling improved softcopy image display.

Premium View (PV) has been designed more recently to further improve the quality of the information presented to the radiologist for diagnosis as well as the reading speed by optimising the local contrast in breast structures. In short, PV works as follows [16]: low-frequency structures (i.e. large-scale structures) are obtained from the original image by low-pass filtering. High frequency structures (i.e. small-scale structures) are obtained by subtracting the low-pass filtered image from the original image. These low and high frequency images are both processed and weighted individually and then added together. The resulting image exhibits reduced contrast between different tissue types, but enhanced contrast of small scale anatomical architecture.

Observer study

Six screening radiologists read two versions of the study set processed with the algorithms TE and PV. All radiologists were familiar with the use of both types of post-processing due to their participation in activities at the national training center for breast cancer screening. Two of them used these types of processing in their daily practice. The two versions of the 263 cases were grouped in ten sessions: 5 sessions with TE processing and 5 sessions with PV; each session with 52 to 53 cases; for each TE session, there was a related PV session containing the same cases. The order of the cases within the sessions was randomised. All cases within a session were processed using the same algorithm. The time between reading two sessions with the same cases was at least one month. Digitised previous mammograms were available at each session. The sets were read independently by each radiologist. Radiologist experience varied and is summarised in Table 1.

The studies were displayed on Hologic SecurView DX diagnostic workstations (Hologic Inc., Danbury, CT, USA). All radiologists were familiar with this system before the study. The radiologists were allowed to use all viewing functionality (e.g. zooming, panning, inverting, adjusting brightness and contrast, hanging protocols) that is normally used while screening.

Radiologists were asked to use a low threshold for reporting lesions and could report up to three findings for each case on a printed form. For each finding the radiologist assigned a suspiciousness score by marking a point on a 10-cm strongly non-linear Visual Analog Scale (VAS) (Fig. 1). The scores were measured automatically after digitising the forms. Case suspiciousness was calculated as the maximum suspiciousness of all findings by a radiologist within a case. This study examines the impression radiologists get of the suspiciousness of cases when these are presented in different ways, while the raw data on which these presentations were based were identical. To emphasise this, case suspiciousness is referred to as perceived case suspiciousness in this study.

Statistical analysis

For each combination of radiologist and processing algorithm the diagnostic accuracy was measured as the area (A_z) under the maximum likelihood estimated binormal ROC curve [17, 18] based on the suspiciousness score using DBM MRMC (University of Chicago and University of Iowa, version 2.2, June 2008). Significance of the average difference in A_z between both algorithms was tested with the Dorfman-Berbaum-Metz method [19, 20] treating both readers and cases as random samples. The *P* value was tested against a significance threshold of 0.05.

The exact interpretation of a VAS by individual radiologists is unknown. Therefore, only the order of the suspiciousness scores for individual radiologists are relevant for analysis, the actual values along the VAS are not. Differences in perceived case suspiciousness were therefore analysed with two-tailed paired sample sign tests using SPSS (version 16.0.1, November 2007; SPSS, Chicago, IL, USA). The *P* values were tested against a significance threshold of 0.0083 (0.05/6) to compensate for applying the tests for six radiologists separately, according to the Bonferroni method.

Table 1 Radiologist experience at study initiation

Observer	Mammography		Digital mammography	
	Years of experience	Current yearly reading volume	Years of experience	Current yearly reading volume
1	20	17,000	2	8,000
2	34	5,000	1	2,000
3	21	40,000	6	10,000
4	21	40,000	6	10,000
5	1	7,000	0.5	3,500
6	12	12,000	4	5,000

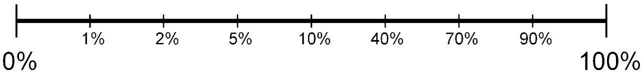

Fig. 1 Visual analogue scale used for scoring suspiciousness of individual findings within each case

Results

The six radiologists reported 1,565 findings in total for the TE cases and 1,683 for the PV cases. This corresponds to an average of 0.99 and 1.07 findings per case per radiologist respectively. An example of a finding in a normal case that was marked by four radiologists when using PV but by none when using TE is shown in Fig. 2. Suspiciousness scores for this particular finding varied from 0.9% to 39%.

Table 2 lists the diagnostic accuracies for the individual radiologists with both processing algorithms. The difference between the mean A_z values for the two algorithms was not significant (TE: 0.909, PV: 0.917, P=0.46).

Table 3 lists the results for the sign tests. For all radiologists, the perceived case suspiciousness for the full dataset was higher when using the PV algorithm. For four out of six radiologists this difference was significant. The table also indicates the results for the TP, FP and N subgroups and all negative cases (FP + N). The perceived case suspiciousness was higher with PV than with TE for nearly all combinations of radiologists and subgroups. The only exception was radiologist 5, who rated the FP cases slightly higher with TE. Because of the small numbers of cases in the subgroups, most of the corresponding P values are above the significance threshold.

Table 2 Diagnostic accuracy scores (A_z) for the ROC analysis

	A_z	
Observer	TE	PV
1	0.934	0.935
2	0.943	0.936
3	0.879	0.930
4	0.905	0.919
5	0.889	0.891
6	0.904	0.889
Mean	0.909	0.917

We assume a simple model in which cases are recalled when they contain a finding that exceeds a certain suspiciousness threshold. At a given threshold, the recall rate can be computed for both processing algorithms. In Fig. 3a the recall rates for TE and PV are compared for every possible recall threshold. The dataset for this study was an enriched set, where 58% of the cases (43 TP + 110 FP / 263 cases) was originally recalled. Our dataset contains all recalled cases from the data collection period and the recall rate during this period was up to three times the pre-digitisation recall rate. Before digitisation as few as 19% (58% / 3) of the cases in the dataset might have been recalled. Figure 3b is an excerpt of Fig. 3a showing only

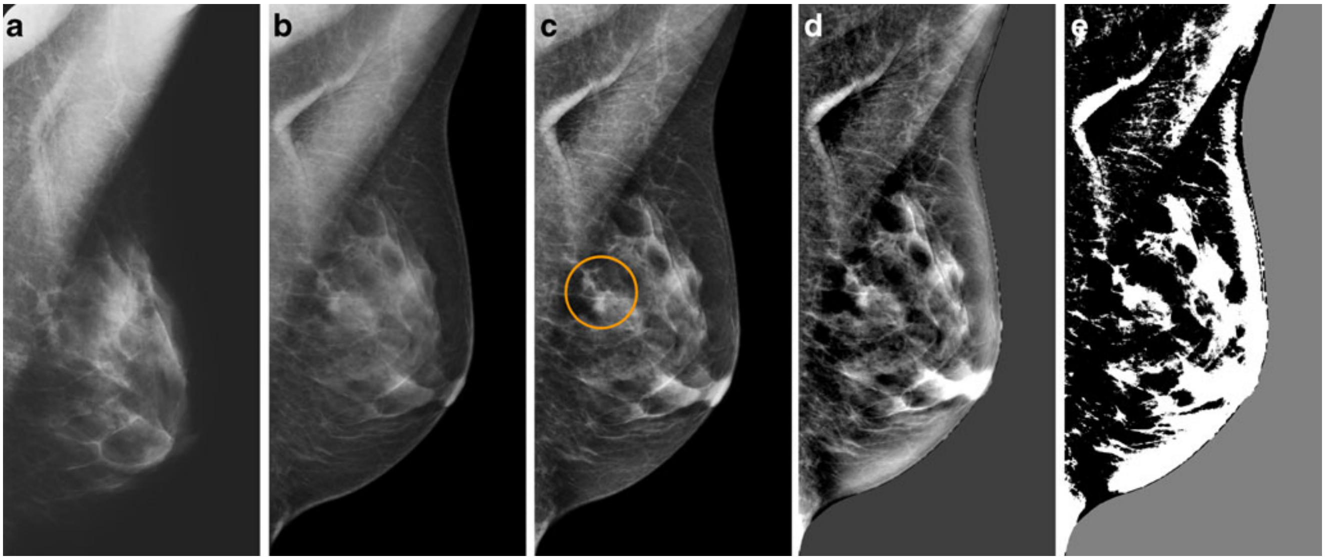

Fig. 2 Example of a finding in a left-sided mediolateral oblique view, reported by four radiologists when using Premium View (PV) only. **a** Digitised prior. **b** Tissue equalisation (TE) processed image. **c** PV processed image with the annotation. **d** is the result image of subtracting (TE) from (PV). **e** is the thresholded version of **(d)**. White areas indicate that pixels in the PV image have relatively higher intensity than the related pixels in the TE image whereas black areas indicate the opposite. It shows that in PV images low frequency trends are suppressed (no noticeable signal decrease in the breast edge in PV compared with TE) whereas higher frequency structures are emphasised (e.g. glandular structures)

Table 3 Comparison of perceived case suspiciousness

	Observer	1	2	3	4	5	6	Mean
TP (*n*=43)	Number TE > PV	15 (34.9%)	19 (44.2%)	18 (41.9%)	15 (34.9%)	17 (39.5%)	16 (37.2%)	16.7 (38.8%)
	Number PV > TE	25 (58.1%)	21 (48.8%)	24 (55.8%)	23 (53.5%)	24 (55.8%)	23 (53.5%)	23.3 (54.3%)
	Ratio*	1.67	1.11	1.33	1.53	1.41	1.44	1.41
	P value†	0.155	0.874	0.440	0.256	0.349	0.337	–
FP (*n*=110)	Number TE > PV	40 (36.4%)	39 (35.5%)	48 (43.6%)	40 (36.4%)	48 (43.6%)	41 (37.3%)	42.7 (38.8%)
	Number PV > TE	56 (50.9%)	61 (55.5%)	54 (49.1%)	58 (52.7%)	45 (40.9%)	56 (50.9%)	55.0 (50.0%)
	Ratio*	1.40	1.56	1.13	1.45	0.94	1.37	1.31
	P value†	0.126	0.036	0.621	0.086	0.836	0.155	–
N (*n*=110)	Number TE > PV	18 (16.4%)	35 (31.8%)	31 (28.2%)	20 (18.2%)	25 (22.7%)	22 (20.0%)	25.2 (22.9%)
	Number PV > TE	74 (67.3%)	54 (49.1%)	45 (40.9%)	32 (29.1%)	34 (30.9%)	41 (37.3%)	46.7 (42.4%)
	Ratio*	4.11	1.54	1.45	1.60	1.36	1.86	1.99
	P value†	**<0.001**	0.056	0.136	0.127	0.298	0.023	–
N + FP (*n*=220)	Number TE > PV	58 (26.4%)	74 (33.6%)	79 (35.9%)	60 (27.3%)	73 (33.2%)	63 (28.6%)	67.8 (30.8%)
	Number PV > TE	130 (59.1%)	115 (52.3%)	99 (45.0%)	90 (40.9%)	79 (35.9%)	97 (44.1%)	101.7 (46.2%)
	Ratio*	2.24	1.55	1.25	1.50	1.08	1.54	1.53
	P value†	**<0.001**	**0.004**	0.154	0.018	0.685	0.009	–
TP + FP + N (*n*=263)	Number TE > PV	73 (27.8%)	93 (35.4%)	97 (36.9%)	75 (28.5%)	90 (34.2%)	79 (30.0%)	84.5 (32.1%)
	Number PV > TE	155 (58.9%)	136 (51.7%)	123 (46.8%)	113 (43.0%)	103 (39.2%)	120 (45.6%)	125.0 (47.5%)
	Ratio*	2.12	1.46	1.27	1.51	1.14	1.52	1.50
	P value†	**<0.001**	**0.006**	0.092	**0.007**	0.388	**0.005**	–

* Ratio of the number of cases for which the suspiciousness was higher with PV over the number of cases for which the suspiciousness was higher with TE

† *P* values resulting from two-tailed paired sample sign tests. *P* values below the significance threshold of 0.0083 are printed in boldface

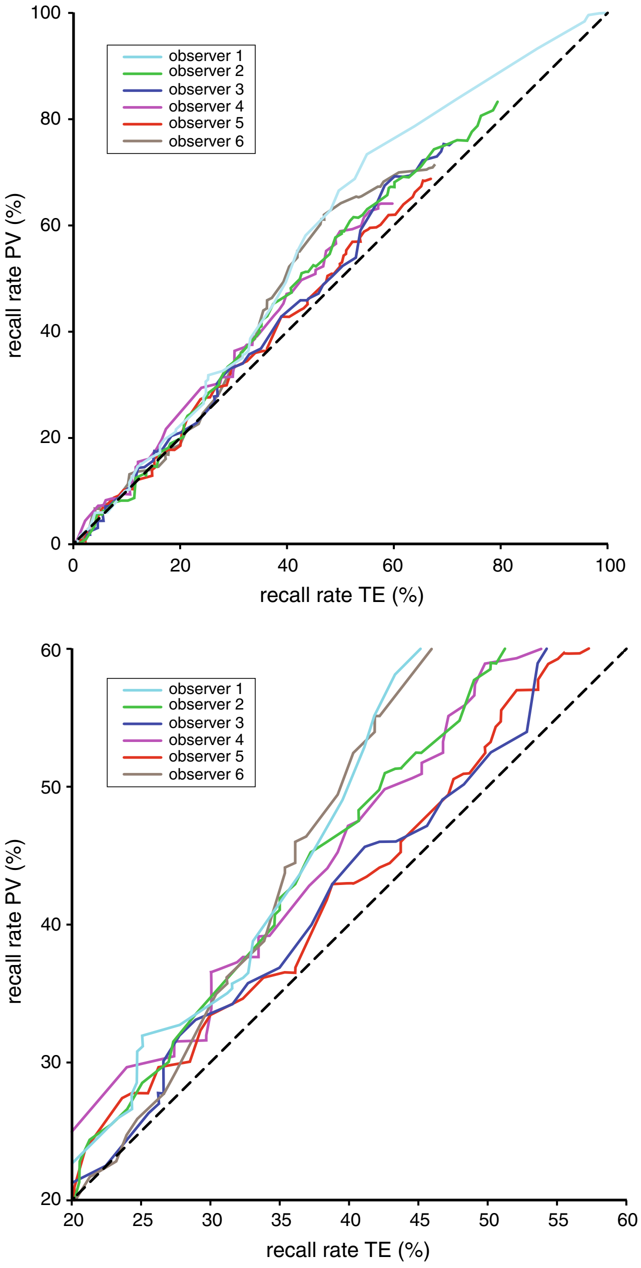

Fig. 3 **a** Recall rates for equal suspiciousness thresholds with TE and PV. **b** Excerpt of (**a**)

this relevant range from 19% (bottom left) to 58% (upper right). For practically every recall threshold in this range the calculated recall rate is higher for PV than for TE.

Discussion

We evaluated two commercially available image processing algorithms by comparing diagnostic accuracy and perceived case suspiciousness. The diagnostic accuracy was not significantly different. The perceived case suspiciousness averaged over all observers of all case types was higher when using PV.

The major difference between the processing algorithms used in our study is an additional local contrast optimisation step when PV is applied. PV is aimed at increasing the visibility and suspiciousness of malignant lesions, but in our study the perceived suspiciousness of benign lesions and normal cases is increased as well. An effect of local contrast enhancement could be that both normal (dense) structures and abnormal structures appear more suspicious due to their enhanced signal. An additional aspect may be the decreased similarity of the PV images to the digitised previous mammograms. Comparison of current and previous mammograms is very important for breast cancer screening, especially for discerning growing lesions from benign findings already present in the previous mammograms [21]. Preference studies using only malignant lesions may conclude that high contrast images are preferable because of the increased visibility of the lesions, while missing the effect that the algorithm could have on normal cases. In our study the perceived suspiciousness of the normal cases increased even more than that of the malignant cases. Even when diagnostic accuracy is not influenced by the choice of image processing, the image processing may still influence the recall rate. Earlier studies have shown an increase in recall rate during the first months after converting to digital mammography [4, 5]. It was proposed that this temporal increase could have been caused by a learning effect and/or by the previous mammograms being film-screen.

Comparability of currents and previous mammograms is not only an issue when converting from analogue to digital mammography. In a recent study, an increase in recall rate was found in a clinical setting after switching from TE to PV [16]. The increase was explained as a training effect, but the necessity of switching off the contrast optimisation for better similarity to archived comparison mammography was also recognised. Future studies should therefore investigate the influence of both the learning effect and the degree of similarity with previous mammograms on diagnosis with respect to the introduction of new postprocessing methods.

In conclusion, this study examines just two out of many possible combinations of appearances of currents and previous mammograms. For manufacturers of digital mammography systems, image appearance has become an important means of distinguishing themselves from each other. Previous studies have suggested that algorithms using contrast enhancement techniques may improve diagnostic accuracy [16, 22, 23]. This effect is not convincingly present in our study. Our study suggests that the introduction of new image processing algorithms is likely to influence the recall rate because of changes in perceived case suspiciousness while diagnostic accuracy may be similar.

References

1. Vinnicombe S, Pinto Pereira SM, McCormack VA et al (2009) Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data. Radiology 251:347–358
2. Skaane P (2009) Studies comparing screen-film mammography and full-field digital mammography in breast cancer screening: updated review. Acta Radiol 50:3–14
3. Karssemeijer N, Bluekens AM, Beijerinck D et al (2009) Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. Radiology 253:353–358
4. Bluekens AM, Karssemeijer N, Beijerinck D et al (2010) Consequences of digital mammography in population-based breast cancer screening: initial changes and long-term impact on referral rates. Eur Radiol 20:2067–73
5. Vernacchia FS, Pena ZG (2009) Digital mammography: its impact on recall rates and cancer detection rates in a small community-based radiology practice. AJR Am J Roentgenol 193:582–585
6. Pisano ED, Cole EB, Major S et al (2000) Radiologists' preferences for digital mammographic display. The International Digital Mammography Development Group. Radiology 216:820–830
7. Sivaramakrishna R, Obuchowski NA, Chilcote WA, Cardenosa G, Powell KA (2000) Comparing the performance of mammographic enhancement algorithms: a preference study. AJR Am J Roentgenol 175:45–51
8. Van Ongeval C, Van Steen A, Geniets C et al (2008) Clinical image quality criteria for full field digital mammography: a first practical application. Radiat Prot Dosimetry 129:265–270
9. Siegel E, Krupinski E, Samei E et al (2006) Digital mammography image quality: image display. J Am Coll Radiol 3:615–627
10. Pisano ED, Cole EB, Hemminger BM et al (2000) Image processing algorithms for digital mammography: a pictorial essay. Radiographics 20:1479–1491
11. Cole EB, Pisano ED, Zeng D et al (2005) The effects of gray scale image processing on digital mammography interpretation performance. Acad Radiol 12:585–595
12. Baydush AH, Floyd CE Jr (2000) Improved image quality in digital mammography with image processing. Med Phys 27:1503–1508
13. Diekmann F, Heinlein P, Drexl J et al (2001) Visualization of microcalcifications by full-field digital mammography using a wavelet algorithm. Int Congr Ser 1230:526–530
14. Zanca F, Jacobs J, Van Ongeval C et al (2009) Evaluation of clinical image processing algorithms used in digital in digital mammography. Med Phys 36:765–775
15. Roelofs AA, van Woudenberg S, Otten JD et al (2006) Effect of soft-copy display supported by CAD on mammography screening performance. Eur Radiol 16:45–52
16. Goldstraw EJ, Castellano I, Ashley S et al (2010) The effect of Premium View post-processing software on digital mammographic reporting. Br J Radiol 83:122–128
17. Metz CE, Pan XC (1999) "Proper" binormal ROC curves: theory and maximum-likelihood estimation. J Math Psychol 43:1–33
18. Pan XC, Metz CE (1997) The "proper" binormal model: parametric receiver operating characteristic curve estimation with degenerate data. Acad Radiol 4:380–389
19. Dorfman DD, Berbaum KS, Metz CE (1992) Receiver operating characteristic rating analysis - generalization to the population of readers and patients with the Jackknife method. Invest Radiol 27:723–731
20. Hillis SL, Berbaum KS, Metz CE (2008) Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. Acad Radiol 15:647–661
21. Roelofs AA, Karssemeijer N, Wedekind N et al (2007) Importance of comparison of current and prior mammograms in breast cancer screening. Radiology 242:70–77
22. Kamitani T, Yabuuchi H, Soeda H et al (2010) Detection of breast cancer by soft-copy reading of digital mammograms: comparison between a routine image-processing parameter and high-contrast parameters. Acta Radiol 51:15–20
23. Chen B, Wang W, Huang J et al (2010) Comparison of tissue equalization, and premium view post-processing methods in full field digital mammography. Eur J Radiol 76:73–80