Advances in digital chest radiography: impact on reader performance

De Boo, D.W.

Publication date
2012

Gray-scale Reversal for the Detection of Pulmonary Nodules on a PACS workstation

3

Diederick W De Boo
Martin Uffmann
Shandra Bipat
Eelco FA Boorsma
Maeke J Scheerder
Michael Weber
Cornelia M Schaefer-Prokop

American Journal of Roentgenology 2011; 197:1096-1100

Abstract

Objective
The purpose of this article is to evaluate the impact of gray-scale reversal on the detection of small solid pulmonary nodules in two-view chest radiography (CXR).

Material and methods
One hundred and twenty-eight patients (mean age, 62y) who underwent CT and chest radiography within 6 weeks were retrospectively selected for this study. Seventy-three percent of patients showed variable degrees of radiographic findings of a 'dirty lung'. A total of 129 solid pulmonary nodules were present in 74 patients (nodule diameter range, 5-30mm; mean 13mm). The remaining 54 patients served as negative control subjects. Six readers of varying experience levels evaluated the images without and with the availability of gray-scale reversal in two separate reading sessions. Figure of merit (FOM), sensitivity per lesion, mean number of false positive marks per image and accuracy were calculated.

Results
Five of the six readers showed a slight increase of sensitivity with the use of gray-scale reversal, but on average, the difference was not significant (48% vs. 50%; $p>0.05$). The mean number of false positive marks per image also nonsignificantly increased from 0.20 to 0.23. The increases in both sensitivity and mean number of false positive marks translated into nonsignificant decrease in averaged FOM (0.79 vs. 0.77) and accuracy (72% vs. 71%). Data analysis of subgroups of nodules or different reader groups, dependant on level of experience, did not reveal significant differences.

Conclusion
Using PACS display of digital chest radiographs, gray-scale reversal does not help the radiologists in detecting pulmonary nodules.

Introduction

Modern digital radiography offers a number of processing tools to improve the detection of focal lung lesions. Signal normalization, gradation adjustment and multi-frequency edge enhancement are performed automatically in the background without further interaction from the radiologist. They represent steps within the process which are designed to produce images of constant quality and optimal display; their parameter sets have been separately optimized. Other more elaborate processing tools, such as temporal or energy subtraction and computer-aided detection, are increasingly available but demand specific software and represent tools specifically required by the radiologists. In addition to online window and level adjustments, gray-scale reversal represents a rather simple tool that is a built-in feature on all PACS workstations. In the literature, little evidence is available for gray-scale reversal in chest radiography[1-5]. All studies applied techniques that are obsolete today, such as hard-copy radiographs, low-resolution soft-copy display, or secondary digitized images and low resolution digital radiographs. It is known from optical physiology that optical contrast perception is increased when a dark object is presented on a white background[6]. Consistent with that idea, we found in clinical practice that evaluation of chest radiographs in the positive mode (“bones black”) in addition to the usual negative mode (“bones white”) is helpful for the detection of focal lung densities. We therefore decided to reevaluate with modern digital image and monitor equipment whether inversion of grey-scale values would facilitate detection of nodular lung lesions. To assess the effects of reader experience and increased interstitial markings in the images, we included six readers with varying experience and a study group of elderly patients, the majority of whom were smokers.

Material and Methods

Study group

We retrospectively selected 128 patients from our institution's data archive who had undergone both a posteroanterior (PA) and lateral chest radiograph and a chest CT within 6 weeks for clinical purposes. Approval for this study was obtained by our institution's ethic committee (registration number 08170465). The study group included 74 patients with nodules and 54 without intrapulmonary nodules. Patients' mean age was 62 years (range, 15-89 years), 74 were males and 54 were females. Patient records showed a smoking history in 74 patients (58%) and no prior or current tobacco use in 24 patients (19%). For the remaining 30 patient (23%) no data were available. Both smoking and increase of age led to a variable increase of parenchymal markings on the chest radiograph, also known as "dirty lung"[7]. Its presence was subjectively scored on the chest radiographs by an experienced chest radiologist and the researcher, ranging from none to mild, moderate and severe. Hereafter, we will refer to this score as "anatomic noise" (AN).

Intrapulmonary lesions

CT findings were used as the reference standard for the definition of nodule size, type and location. A total of 129 CT-proven nodular opacities with diameters between 5 and 30mm (mean, 13 mm; median, 11 mm) were present in 74 patients. Eighty-six nodules were round, had smooth margins on CT, and had well-defined contours on the chest radiographs. None of these nodules were calcified. Forty-three nodules were patchy with spiculated margins on CT and ill-defined contours on the CXR.

Lesion conspicuity

Lesion conspicuity was subjectively graded on the conventional negative mode chest radiography ("bones white") by a board-certified chest radiologist (with > 15 years of experience) and by the researcher after the reading had been completed and ranged from high to moderate, low and very low.

Chest radiography

Chest radiographs were obtained in digital technique using a dedicated chest stand (Thoravision, Philips Healthcare). Images were processed using nonlinear multifrequency processing (Unique, Philips Healthcare). Processing parameters were chosen according to the recommendations of the manufacturer and represent the standard processing used in routine clinical studies.

Gray-Scale Reversal

Gray-scale reversal was accomplished by inverting the slope of the lookup table and transcribing the original gray-scale values (bones white) to their inverted counterparts. No additional processing was applied. To produce gray-scale reversed images, we used the facilities of the workstation (a single mouse click).

Image evaluation

Images were evaluated using a dedicated PACS system (Impax version 4.5) equipped with high-resolution (1.2K x 1.9K) liquid crystal display monitors (MDCG 2121-CB, Barco). Three radiology residents (first to third year training) and three board-certified radiologists (all with > 10 years of experience) independently interpreted the images. The 128 PA and lateral radiographs were evaluated twice in two separate reading sessions, once without and once with the availability of gray-scale reversal. Approximately half of the cases were seen first without gray-scale reversal, and the other half was interpreted first with the availability of gray-scale reversal. During both sessions, readers were allowed to use processing tools, such as windowing or magnification, according to their preferences. There was an interval of at least 6 weeks between the two reading sessions, and images were evaluated in different random orders. The presence or absence of an intrapulmonary opacity was scored using a 5-point scale of confidence, as follows: 5, definite pulmonary lesion; 4, probable pulmonary lesion; 3, unequivocal; 2, probably no lesion; and 1, definitely no lesion. For the confidence ratings 3, 4 and 5, readers were asked to indicate the anatomic location of the suspected lesion on a separate data sheet. This was done separately for each patient; per patient more than one lesion could be marked. The readers were informed that images could contain none, a solitary, or multiple lesions but did not know the percentage of each subgroup. They were instructed to ignore lesions smaller than 5 mm. Reading time was documented per reader and reading session.

Statistical analysis

The patients with and patients without intrapulmonary focal lesions were compared with regard to gender and smoking history by chi-square test, with regard to age by Student t-tests because the data were normally distributed, and with regard to anatomic noise by chi-square test for trend. For smoking history, the missing data (19%) were excluded. All reader markings, without and with gray-scale reversal, were determined to be true or false positive by comparing the markings on the separate data sheet with the original chest radiograph and the corresponding CT. This was done in consensus by the researcher and an experienced chest radiologist after all readings had been completed. The data were analyzed using the jackknife free-response receiver operating characteristic (JAFROC) method[8-10]. JAFROC software was used to calculate a figure of merit (FOM). The FOM is defined as the probability on a scale from 0 to 1.0 that true positive marks for lesions are rated with higher confidence than false positive marks (nonlesions) on control chest radiographs[9]. An FOM of 1.0 describes the ideal situation in which all nodules are correctly marked with high confidence and there are no false positive marks on the control images; an FOM of 0.5 means that the confidence for true positive marks for lesions on abnormal chest radiographs is equal to the confidence for false positive marks on negative control chest radiographs. Additionally, we calculated descriptive statistical measures, such as sensitivity, mean number of false positive marks per image, and accuracy. Sensitivity was calculated on a per-lesion basis. Reader ratings 4 and 5 were considered true positive if made for correctly localized lesions. The mean false positive marks per image was calculated by dividing the total number of false positives per reader by the number of study cases (n = 128). Reader ratings 4 and 5 were considered false positive if made for nonexisting lesions. Accuracy was calculated per image: an image was considered true positive if at least one nodule was correctly identified and was considered true negative if there was no false positive mark on a negative control image. Sensitivity, the mean number of false positive marks per image, and accuracy were calculated for all observers, for experienced, and for inexperienced observers. For all observers, the data of all six observers were counted as separate observations. Similarly, for the subgroups of experienced or inexperienced readers, data of the respective group were also counted as separate observations. Differences in sensitivity, mean number of false positive marks per image and accuracy were compared with the McNemar test. All analyses were performed in SPSS software (version 15.0.1, SPSS) Significance was assumed at p less than 0.05.

Results

Study groups and lesion characteristics

An increase in anatomical noise was scored present in 93 patients (73%). The severity of anatomic noise varied from mild in 43% of patients, to moderate in 20% of patients, and severe in 10% of patients. Forty patients had a solitary nodular lesion, 20 patients had two lesions and 14 patients had three or more lesions. The lesions were located in the upper lobes in 57%, in the middle lobe in 9% of patients, and in the lower lobes in 35% of patients.

Table 1
Reader outcome by figure of merit, sensitivity, mean number of false positive marks per image, and accuracy.

Readers	**Reading**	**Figure of merit, mean (95%, CI)**	**Sensitivity, % (95%, CI)**	**Number of false positives per image, mean (total no. of false positive marks by readers / no. of study cases)**	**Accuracy, % (95%, CI)**
All	Baseline	0,79 (0,66-0,92)	48% (45%-52%)	0,20 (156/768)	72% (68%-75%)
	Grey-scale reversal	0,77 (0,61-0,92)	50% (46%-54%)	0,23 (179/768)	71% (68%-74%)
Inexperienced	Baseline	0,72 (0,36-1,0)	42% (37%-47%)	0,25 (97/384)	66% (61%-71%)
	Grey-scale reversal	0,66 (0,34-0,99)	45% (40%-50%)	0,30 (115/384)	64% (59%-69%)
Experienced	Baseline	0,85 (0,78-0,92)	55% (49%-59%)	0,15 (59/384)	77% (73%-82%)
	Grey-scale reversal	0,87 (0,73-1,0)	55% (50%-60%)	0,17 (64/384)	79% (75%-83%)

Baseline reading was performed without gray-scale reversal.

Lesion conspicuity was rated very low in 24% of patients, low in 19% of patients, moderate in 37% of patients, and high in 20% of patients. More patients with nodules had a positive smoking history compared to the patients without nodules ($p < 0.001$). There were no differences with regard to age ($p = 0.83$), gender ($p = 0.78$) and AN ($p = 0.14$).

Reader performance without gray-scale reversal

The average FOM was 0.79, with a mean FOM of 0.72 for the inexperienced readers and a mean FOM of 0.85 for the experienced readers. The mean sensitivity for the inexperienced readers was 42%, with a mean number of false positive marks per image of 0.25, whereas the mean sensitivity for the experienced readers was 55% with a mean number of false positive marks per image of 0.15. For both reader groups, there was no significant association between the mean number of false positive marks per image and the degree of anatomic noise. The accuracy varied widely among the readers, with a mean of 72% (Table 1). The group of missed lesions comprised 74% of the lesions of low to very low conspicuity and 75% of the lesions smaller than 10 mm (Table 2).

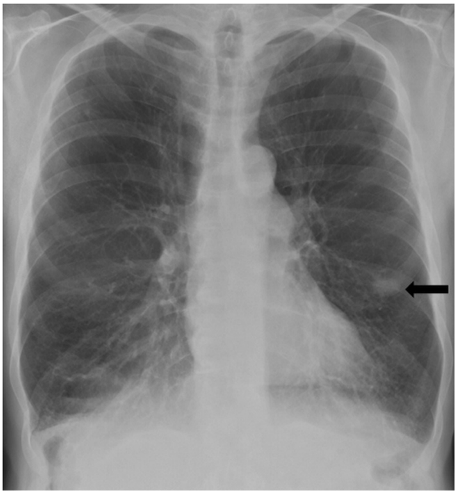
Reader performance with the availability of gray-scale reversal

For the inexperienced readers, the average FOM decreased from 0.72 to 0.66 with the availability of gray-scale reversal; for the experienced readers, the mean FOM slightly increased from 0.85 to 0.87, though both differences did not reach significance. Five out of six readers showed an increase in sensitivity with the use of gray-scale reversal, but differences were not significant (43% vs. 43%; 40% vs. 44%; 45% vs. 47%; 51% vs. 47%; 55% vs. 57% and 57% vs. 61%). The mean sensitivity of the inexperienced readers nonsignificantly increased from 42% to 45%; the mean sensitivity of experienced readers remained unchanged at 55%. Both inexperienced and experienced readers showed a nonsignificant increase in mean number of false positive marks per image (0.25 vs. 0.30 and 0.15 vs. 0.17, respectively). Mean accuracy decreased from 66% to 64% for the inexperienced readers and slightly increased from 77% to 79% for experienced readers. Again, differences did not reach statistical significance (Table 1). Examples are given in Figure 1 - 3.

Reading time

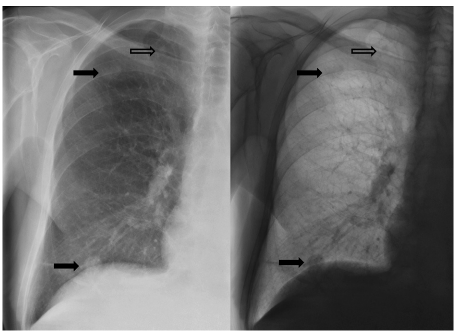
Reading time was documented per reading session. On average, reading time per image increased by 5 seconds (8%) with the availability of gray-scale reversal (59 vs. 64 seconds per image).

Figure 1



Example of a patient with moderate increased parenchymal markings. The nodule in the left lower lobe (arrow) was detected by all readers without and with gray-scale reversal.

Figure 2



Three nodules, two in the upper lobe and one in the lower lobe. None of the readers detected the two larger nodules (black arrows) when using gray-scale reversed images complimentary. One reader saw the small nodule (open arrows) only with the use of the gray-scale reversed image.

Table 2

Sensitivity of readers without and with gray-scale reversal for subgroups of lesions, by conspicuity and size.

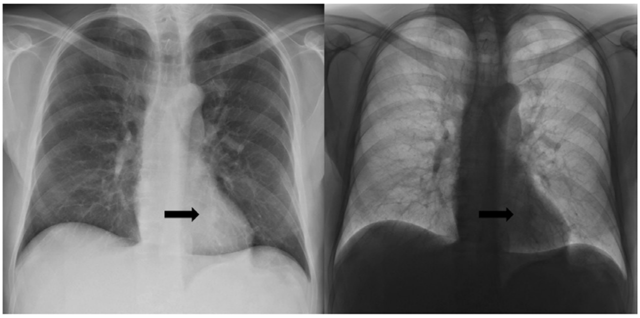
Reader group	**reading**	**Conspicuity**		**Size**	
		High or moderate	Low or very low	< 10 mm	≥ 10 mm
All					
	Baseline	79 (77 - 81)	26 (24 - 28)	25 (22 - 28)	60 (57 - 62)
	Gray-scale reversal	82 (80 - 84)	26 (24 - 28)	25 (23 - 28)	62 (60 - 64
Inexperienced					
	Baseline	72 (68 - 75)	21 (18 - 23)	19 (16 - 23)	54 (51 - 57)
	Gray-scale reversal	78 (74 - 81)	21 (18 - 23)	20 (16 - 23)	57 (54 - 60)
Experienced					
	Baseline	86 (83 - 89)	31 (28 - 34)	31 (27 - 35)	67 (63 - 68)
	Gray-scale reversal	86 (83 - 89)	32 (28 - 35)	32 (27 - 35)	67 (64 - 70)

Data are percentage (95% CI). Baseline reading was performed without gray-scale reversal.

Discussion

To our knowledge, only few studies evaluating gray-scale reversal have been published so far(1-5). They all are more than 15 years old and used image and display techniques that are outdated with respect to contrast and spatial resolution and are not applied any more. As known from optical physiology, the contrast sensitivity of the human eye is greater for the detection of dark structures on bright background than vice versa(6). We therefore wanted to reevaluate the usefulness of gray-scale reversal for the detection of nodular opacities using modern state-of-the-art equipment. Gray-scale reversal can be performed in different ways, dependant on whether the slope of the lookup table that determines the transfer of pixel values into gray levels is preserved or changed to shift contrast resolution from the high to the low absorption area or vice versa. In our study, the gray levels were transcribed to their inverted counterparts, preserving a higher contrast in the area of the lung (originally black) compared with the area of the mediastinum (originally white). No further processing was obtained. We chose so to maintain contrast characteristics more closely to the original display. The gray-scale reversed images were used as an adjunct to the normal negative displays. In this way, the gray-scale reversed images served as a complimentary display, similar to when radiologists change the window or level settings to focus on the detection of subtle density differences. Images were evaluated side-by-side on two monitors or by toggling from one to the other on a single monitor, dependant on the reader's preference. Though there was a tendency for an improved sensitivity for the detection of nodular opacities (e.g. five out of six readers showed an increase of sensitivity with gray-scale reversal), the differences failed to reach statistical significance. The increase of sensitivity

Figure 3



One nodule in the left lower lobe (arrow). Three readers missed the nodule and detected it with the complimentary use of the gray-scale inverted image. One reader detected the nodule without gray-scale and missed it with the availability of gray-scale reversal. For the remaining two readers there was no difference without or with the availability of gray-scale reversal.

was counteracted by loss of specificity and therefore did not lead to a change of FOM or accuracy. Also, when comparing subgroups of nodules as a function of nodule size or subgroups of readers as function of experience, differences did not reach significance. These results somewhat contradict our clinical experience that suggested a more positive impact of gray-scale reversal on the detection of focal lesions. A possible explanation for our results refers to the overall small diameter of our test lesions with a median of 11 mm and mean of 13 mm. In addition, most chest radiographs showed a varying degree of increased parenchymal markings, which further impaired detection. Both aspects certainly contributed to the loss of specificity due to misinterpretation of focal densities or vascular markings that became more pronounced in the gray-scale reversed images. Lack of familiarity with gray-scale reversed images might also have played a role. Sheline et al. found an overall improved detection of pulmonary nodules using secondary digitized radiographs as compared to conventional screen-film radiographs but failed to prove a significant difference comparing "bones white" and "bones black" soft-copy display(2). Three other studies evaluating gray-scale reversal in chest radiography all reported a decrease in performance when gray-scale reversal was applied(3-5). It should be mentioned, however, that all studies evaluated the gray-scale reversal as stand-alone display and not as an adjunct to the normal negative ("bones white") display, as we did in our study. Kheddache et al. were the only ones to report improved detection of simulated nodules projected over the area of the mediastinum of an anthropomorphic chest phantom applying the same type of gray-scale reversal we used(1). However, the tested study images had been obtained with an image intensifier with a much lower spatial and contrast resolution as compared with the chest radiographs we evaluated. In our clinical study group, the number of nodules projecting over the mediastinum and the retrocardiac area on PA radiographs was too low to detect significant differences. Besides, both PA and lateral images were available for evaluation. Our study has some limitations. We evaluated a selected group of patients with a higher prevalence of lesions than normally seen in clinical routine. In addition readers were specifically asked to look for small nodular densities, which probably lowered their overall specificity. Images were evaluated without and with gray-scale reversal in two reading sessions, introducing the additional factor of intrareader variability. We chose to do this to keep reading conditions possibly realistic and equal for both conditions and not to interrupt the readers' visual analysis by requiring lesion documentation without and also with gray-scale reversal. Intra-reader variability may have contributed to the nonsignificant performance differences, however, this again appears to reflect more closely clinical reality. We conclude that, regardless of experience of radiologist or nodule characteristics, complimentary use of gray-scale reversal does not improve the detection of radiologists for small nodules in chest radiographs. Further studies are needed to assess whether gray-scale reversal is advantageous for the detection of larger ill-defined lesions.

References

1. Kheddache S, Månsson LG, Angelhed JE, Denbratt L, Gottfridson B, Schlossman D. Digital chest radiography: should images be presented in negative or positive mode? Eur J Radiol 1991; 13:151-155
2. Sheline ME, Brikman I, Epstein DM, Mezrich DM, Kundel HL, Arenson RL. The diagnosis of pulmonary nodules: Comparison between standard and inverse digitized images and conventional chest radiographs. AJR 1989; 152:261-263
3. MacMahon H, Metz CE, Doi K, Kim T, Giger ML, Chan HP. Digital chest radiography: effect on diagnostic accuracy of hard copy, conventional video and reversed gray scale video display formats. Radiology 1988; 168:669-673
4. Oestmann JW, Kusher DC, Bourgouin M, Llewellyn HJ, Mockbee BW, Greene R. Subtle lung cancers: Impact of edge enhancement and gray scale reversal on detection with digitized chest radiographs. Radiology 1988; 167:657-658
5. Schaefer CM, Greene R, Hall DA, et al. Mediastinal abnormalities: detection with storage phosphor radiography. Radiology 1991; 178:169-173
6. Blackwell H. Contrast thresholds of the human eye. J Opt Soc Am 1946; 36:624-643
7. Gückel C, Hansell DM. Imaging the 'dirty lung' –has high resolution computed tomography cleared the smoke? Clin Radiology 1998; 53:717-722
8. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys 2004; 31:2313-2330
9. Chakraborty DP. Analysis of location specific observer performance data: validated extensions of the jackknife free-response (JAFROC) method. Acad Radiol 2006; 13:1187-1193
10. Vikgren J, Zachrisson S, Svalkvist A, et al. Comparison of chest radiography for the detection of pulmonary nodules: human observer study of clinical cases. Radiology 2008; 249:1034-1041