The Diagnosis of Pulmonary Nodules: Comparison Between Standard and Inverse Digitized Images and Conventional Chest Radiographs

Martin E. Sheline[1]
Inna Brikman
David M. Epstein
Jonathan L. Mezrich
Harold L. Kundel
Ronald L. Arenson

We compared plain chest radiographs, standard (bones white) digitized images, and inverse-intensity (bones black) images to determine their ability to identify pathologically confirmed malignant pulmonary nodules. The images were digitized by using a photo-optical laser scanner and were displayed on a 1024 × 1024 × 8 bit system capable of operator-controlled magnification (2× or 4×) and nonlinear (logarithmic/exponential) contrast transformation in both standard and inverse-intensity modes. Receiver-operator curve analysis was used to study the detection performance of six observers who viewed 40 images obtained in 15 normal subjects and 25 abnormal subjects. There was no statistically significant difference in the area under the ROC curve between the standard digital images and the plain chest radiographs. However, ROC areas were significantly greater ($p \leq .05$) for inverse-intensity digital images when compared with either standard-intensity digital images or plain chest radiographs.

These results suggest that inverse-intensity images may have some advantages in the detection of pulmonary nodules.

The purpose of this project was to compare plain chest radiographs, standard (bones white) digitized images, and inverse-intensity (bones black) digitized images to determine their usefulness in detecting pulmonary nodules. Several studies have been done recently comparing digitized images and digitizing systems with conventional plain radiographs [1–6]. The systems studied have used multiple features, including edge enhancement, inverse intensity, magnification, contrast transformation, and scrolling. Some studies have shown that conventional radiographs are superior in detecting pneumothoraces and mild interstitial disease, probably reflecting the higher resolution of plain films [3, 4]. When nodules are located in the mediastinum or behind the heart, digitized images may be better than plain radiographs [3, 5] because the local contrast can be adjusted by the use of either window and level controls or by a suitable intensity transformation.

Until recently, the inverse-intensity digital display mode had not been tested and compared with conventional radiographs and standard digitized images. However, MacMahon et al. [7] found that, for a variety of radiographic findings including pneumothorax, interstitial disease, bone lesions, and pulmonary nodules, diagnostic accuracy was greater with conventional films than with video display, and standard digitized images were more accurate than inverse-intensity images. We specifically concentrated on a comparison of these three image-display techniques to assess any inherent advantages in the detection of pulmonary nodules. We designed the study, both by selection of cases and by selection of readers, to isolate any possible differences between the three imaging methods.

Materials and Methods

Of the 40 posteroanterior chest radiographs that were selected, 15 were normal and 25 showed single or multiple pulmonary nodules. The 25 positive films were obtained in patients with pathologically proved pulmonary malignancy; each of these patients had had a chest

Received July 25, 1988; accepted after revision October 3, 1988.

This work was supported in part by grant #HL33332 from the National Institute of Heart, Lung, and Blood Diseases, United States Public Health Service.

[1] All authors: Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce St., Philadelphia, PA 19104. Address reprint requests to D. M. Epstein.

***AJR* 152:261–263, February 1989**
0361–803X/89/1522–0261

radiograph on which the pulmonary tumor had been missed when the films were interpreted routinely. The films chosen were the earliest radiographs on which the presence and location of a nodule could, in retrospect, be confirmed unanimously by a panel of three radiologists. This sample was biased toward difficult cases under the assumption that if image processing is to be useful, it must help with diagnosis of the cases that are ordinarily missed. The normal cases were chosen from a series of healthy people who had had follow-up for several years to establish the absence of a tumor. The age range of the normal subjects (27–94 years) was similar to that of the patients with malignancy. The radiologists who selected the films did not participate in interpreting the studies.

Digitized images were made of all 40 films by using a photo-optical laser scanner with 200-μm resolution (Film Digital Radiograph System [DRS-1], Du Pont, Wilmington, DE). Images were displayed on a Ramtek RM9460 display station (Ramtek, Santa Clara, CA). The digitizer and the display station have been described in detail elsewhere [1–3, 8]. We displayed digitized images of 1024 × 1024 pixels with an 8-bit gray scale (256 shades of gray). Inverse images were obtained by subtracting each pixel's gray scale from 255. Normal- and inverse-intensity images were treated as separate categories, and the reviewers were not allowed to use the operator-controlled inverse capabilities of the system. The reviewers were allowed to use the two- or four-power magnification software and to perform nonlinear (logarithmic/exponential) contrast transformation of the images. Six radiology residents-in-training independently reviewed the 40 cases in each of three modes: radiographs, normal-intensity digitized images, and inverse-intensity digitized images. They were allowed to examine the images at their own pace. In order to prevent any bias or learning, groups of eight studies from each of the three categories (plain film, standard digitized, and inverse digitized) were reviewed in separate, randomized blocks. The six radiology residents saw the films in different randomized sequences.

So that the residents would have some familiarity with interpreting digital images, several cases not used in the study were reviewed before starting the study. Plain radiographs, standard digital images, and inverse images were displayed side-by-side. For this study, the radiologists recorded their confidence level as either low, medium, or high for detecting the presence or absence of at least one pulmonary nodule for each of the four quadrants of the image (the readers were asked to make incidental note of granulomas, calcified lymph nodes, and nipple shadows but not to include them as nodules). The scoring was done on the basis of quadrants. There were a total of 40 quadrants containing at least one nodule and 120 quadrants without nodules. The ROC curves were compared for each of the three image-display categories to determine which method was superior in detecting lung nodules.

The ROC area (A) and the standard error (SE) were calculated for each viewing mode by using the jackknife method of Dorfman and Berbaum [9]. This method pools the response data from a group of observers. Two viewing techniques were then compared by using the approach of Hanley and McNeil [10]. The critical ratio, z, was calculated as

$$z = \frac{A_1 - A_2}{(SE_1^2 + SE_2^2 - 2rSE_1SE_2)^{1/2}} \quad (1)$$

where A_1 and A_2 are the ROC areas for two viewing techniques, SE_1 and SE_2 are the standard errors, and r is the correlation between the two techniques. This correlation is derived from the average ROC area and from the correlation coefficient for paired observer responses. Kendall's tau was used as the correlation coefficient. The average value for the six observers was used, and the value of r was taken from the table published by Hanley and McNeil [10].

Results

The pooled results from the RSCORE-J jackknife program for each of the three viewing modes are given in Table 1. A composite ROC curve for each of the three techniques is shown in Figure 1. The difference between pairs of viewing modes is given in Table 2. The difference between inverse video images and plain radiographs for pooled data was significant, with $p = .008$. The difference between inverse and normal video images had a p of .055, which is just at the borderline of significance when $p = .050$ is used as a criterion. The difference between normal video images and plain radiographs was not found to be significant ($p = .18$).

TABLE 1: Area Under the Receiver Operating Characteristic Curves and Population Standard Error for Three Display Techniques

Technique	Area Under ROC Curve	Standard Error
Film	0.72	0.05
Normal digital video	0.75	0.06
Inverse digital video	0.86	0.06

Note.—Area and standard error were estimated by the jackknife procedure RSCORE-J [9]. ROC = receiver operating characteristic.

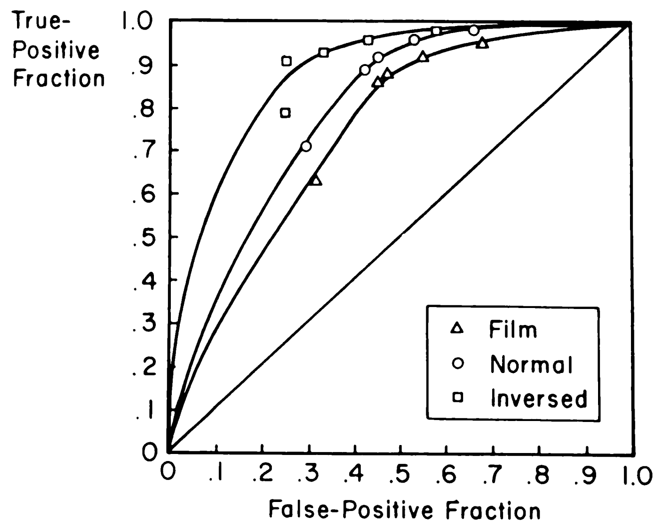

Fig. 1.—Receiver operating characteristic curves of pooled data for inverse images, normal images, and plain radiographs.

TABLE 2: A Comparison of Performance Between Different Display Techniques by Calculating the Difference in the Areas of Their Receiver Operating Characteristic Curves

Comparison	ΔA	r	z	p
Film vs normal	0.05	.42	0.93	.18
Normal vs inverse	0.09	.47	1.60	.055
Film vs inverse	0.14	.40	2.39	.008

Note.—Numbers in ΔA column show the difference in the areas of ROC curves, r is the average correlation coefficient, z is the critical ratio calculated by using equation 1 (see Materials and Methods), and p is the corrected ratio derived from the normal distribution. ROC = receiver operating characteristic.

Discussion

It appears that for the detection of nodules, there may be an advantage in displaying images in inverse intensity. In our study, every attempt was made to eliminate bias toward normal-intensity images (digitized and plain films). We used radiology residents, because we thought that more experienced radiologists, with much more experience with plain films, might be biased in favor of normal images. No attempt was made to simulate a "real practice"; reviewers were allowed to read films at their own pace.

By using pathologically confirmed lung tumors that were missed during regular interpretation, we were able to include cases in which the plain film findings were questionable. In some of these cases, the findings may have been more visible on inverse-display images. The inverse- and normal-intensity images were obtained from the same digitizer data and were displayed on the same display station. Both the magnification and the nonlinear contrast transformation were available for either technique. Therefore, any difference in observer performance when viewing normal-display and inverse-display images presumably represents a real difference between the two techniques. The difference between inverse and normal video images was more than twice that between normal video and plain radiographs, indicating that the digitization of images was a less significant factor than the mode in which they were displayed—that is, as inverse- or normal-intensity images. Although the difference between the inverse display and the normal display was just at the borderline of statistical significance ($p = .055$), we believe that inverse images may have an intrinsic advantage over normal images, especially in light of the many biases toward normal-intensity images (such as more familiarity with normal images).

The fact that observer performance was better with both types of digitized images than with plain radiographs may reflect the ability to manipulate contrast on the digital system. Other authors [3-5] have reported that digitized images are useful in identifying pulmonary nodules. A recent study by MacMahon et al. [7] evaluated hard-copy, standard, and inverse-intensity digitalized images but did not permit observer manipulation of the video controls, window width, or brightness. This may account for the fact that their degree of accuracy in detecting pulmonary nodules was lower with inverse-intensity images than with standard digitized images and hard copy. Our study suggests that, although digitized images clearly do not contain any more information than plain radiographs, the format and image manipulation possible with digitized systems may present pulmonary nodules in a more recognizable fashion, especially those obtained from under- or overexposed areas of radiographs.

ACKNOWLEDGMENTS

We are indebted to the radiology residents at the University of Pennsylvania for their time and effort in reviewing the images used in this study: Phil Bergey, Stuart Bobman, Theresa Chan, Elliott Lerner, and Robert Lipman. We thank Du Pont and Digital Equipment for generous gifts of equipment and Karen L. Weiss and John Kahler for help in preparing the manuscript.

REFERENCES

1. Kundel HL, Mezrich JL, Brikman I, et al. Digital chest imaging: comparison of two film digitizers with a classification task. *Radiology* **1987**;165: 747-752
2. Kundel H. Visual perceptions and image display terminals. *Radiol Clin North Am* **1986**;24:69-77
3. Goodman LR, Foley WD, Wilson CR, Rimm AA, Lawson TL. Digital and conventional chest images: observer performance with film digital radiography system. *Radiology* **1986**;158:27-33
4. MacMahon H, Vyborny CJ, Metz CE, Doi K, Sabeti V, Solomon SL. Digital radiography of subtle pulmonary abnormalities: an ROC study of the effect of pixel size on observer performance. *Radiology* **1986**;158:21-26
5. Chakraborty DP, Breatnack ES, Yester MV, Soto B, Barnes GT, Fraser RG. Digital and conventional chest imaging: a modified ROC study of observer performance using simulated nodules. *Radiology* **1986**;158: 35-39
6. Armstrong JD II, Sorenson JA, Nelson JA, et al. Clinical evaluation of unsharp masking and slit scanning techniques in chest radiography. *Radiology* **1983**;147:351-356
7. MacMahon H, Metz CE, Doi K, Kim T, Giger ML, Chan H-P. Digital chest radiography: effect on diagnostic accuracy of hard copy, conventional video, and reversed gray scale video display formats. *Radiology* **1988**;168:669-673
8. van der Voorde F, Arenson RL, Kundel HL, et al. Development of a physician-friendly digital image display console. *PACS IV*, SPIE **1986**;626:541-548
9. Dorfman DD, Berbaum KS. RSCORE-J: Pooled rating method data: a computer program for analyzing pooled ROC curves. *Behav Res Methods Instruments Comput* **1986**;18:452-462
10. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **1983**;148:839-843