Development of a Digital Image Database for Chest Radiographs With and Without a Lung Nodule:

Receiver Operating Characteristic Analysis of Radiologists' Detection of Pulmonary Nodules

Junji Shiraishi[1]
Shigehiko Katsuragawa[2]
Junpei Ikezoe[3]
Tsuneo Matsumoto[4]
Takeshi Kobayashi[5]
Ken-ichi Komatsu[6]
Mitate Matsui[7]
Hiroshi Fujita[8]
Yoshie Kodera[9]
Kunio Doi[10]

OBJECTIVE. We developed a digital image database (www.macnet.or.jp/jsrt2/cdrom_nodules.html) of 247 chest radiographs with and without a lung nodule. The aim of this study was to investigate the characteristics of image databases for potential use in various digital image research projects. Radiologists' detection of solitary pulmonary nodules included in the database was evaluated using a receiver operating characteristic (ROC) analysis.

MATERIALS AND METHODS. One hundred and fifty-four conventional chest radiographs with a lung nodule and 93 radiographs without a nodule were selected from 14 medical centers and were digitized by a laser digitizer with a 2048 × 2048 matrix size (0.175-mm pixels) and a 12-bit gray scale. Lung nodule images were classified into five groups according to the degrees of subtlety shown. The observations of 20 participating radiologists were subjected to ROC analysis for detecting solitary pulmonary nodules. Experimental results (areas under the curve, A_z) obtained from observer studies were used for characterization of five groups of lung nodules with different degrees of subtlety.

RESULTS. ROC analysis showed that the database included a wide range of various nodules yielding A_z values from 0.574 to 0.991 for the five categories of cases for different degrees of subtlety.

CONCLUSION. This database can be useful for many purposes, including research, education, quality assurance, and other demonstrations.

Received January 22, 1999; accepted after revision June 14, 1999.

Presented at the annual meetings of the Radiological Society of North America, Chicago, November 1997 and November 1998.

Fully supported by the Japanese Society of Radiological Technology in cooperation with the Japanese Radiological Society.

[1]Department of Radiology, Osaka City University Hospital, 1-5-7 Asahi-machi, Abeno-ku, Osaka, Osaka 545-8586 Japan. Address correspondence to J. Shiraishi.

[2]Department of Radiology, Iwate Medical University, 19-1 Uchimaru, Morioka, Iwate 020-8505 Japan.

[3]Department of Radiology, Ehime University, Shizukawa, Shigenobu-cho, Onsen-gun, Ehime 791-0295 Japan.

[4]Department of Radiology, Yamaguchi University, 1144 Ogushi, Ube, Yamaguchi 755-8506 Japan.

[5]Department of Radiology, Kanazawa University, 5-11-80, Kotateno, Kanazawa, Ishikawa 920-0942 Japan.

[6]Medical Engineering Division, Toshiba, 1-1-1 Shibaura, Minato-ku, Tokyo 105-8001 Japan.

[7]Medical Imaging Division, Konica, 1 Sakura-cho, Hino, Tokyo 191-8511 Japan.

[8]Department of Information Science, Faculty of Engineering, Gifu University, 1-1, Yanagido, Gifu, Gifu 501-1193 Japan.

[9]Department of Radiological Technology, Nagoya University School of Health Sciences, 1-1-20 Taikouminami, Higashi-ku, Nagoya, Aichi 461-8673 Japan.

[10]Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, The University of Chicago, 5841 S. Maryland Ave., MC2026, Chicago, IL 60637.

AJR 2000;174:71–74

0361–803X/00/1741–71

An image database is important for research on digital imaging, such as image processing, image compression, image display, picture archiving and communication systems, and computer-aided diagnosis. Because investigators have generally used their own databases for evaluation of their techniques and methods, comparing results obtained with different databases can be difficult [1, 2]. Therefore, researchers anticipate the development of a common image database. A number of researchers are developing digital image databases for mammography (e.g., The Digital Mammography Home Page [www.rose.brandeis.edu/users/mammo/digital.html]) [3, 4].

When the image database is established, one of the most important properties associated with the database is the level of radiologists' performance for detection of abnormalities included in the database. If users of this database know the average level of radiologists' performance, it would be valuable information to radiologists who can compare their performance with the average performance of other radiologists. However, to our knowledge, an image database that includes the radiologists' performance is not available. Therefore, in this study, a digital image database (www.macnet.or.jp/jsrt2/cdrom_nodules.html) for chest radiographs with and without solitary pulmonary nodules was developed for common use by radiologists and researchers. In addition, the radiologists' performance for detection of pulmonary nodules included in the database was evaluated using receiver operating characteristic (ROC) analysis.

Materials and Methods

Original posteroanterior chest films (34.6 cm × 34.6 cm [14 inch × 14 inch]) for this database were collected from 13 medical centers in Japan and one institution in the United States under the following conditions: only one nodule on an image for nodule cases; confirmation of presence or absence of a lung nodule by CT examination; and nodule classification as malignant based on histologic and cytologic examination or as benign based on histology, definitive isolation of a pathogenic organism, shrinkage and disappearance with the use of antibiotics, or no change observed during a follow-up period of 2 years. All of the original radiographs were digitized using an LD-4500 or an LD-5500 laser film digitizer (Konica, Tokyo, Japan). Digitized images had a 2048 × 2048 matrix, 0.175-mm pixel size, and 12-bit gray levels corresponding to a 0.0–3.5 optical density range.

Three experienced chest radiologists excluded images with multiple pulmonary nodules or of poor image quality. The database included 247 posteroanterior chest images, which consisted of 154 images with and 93 images without a nodule. One hundred nodules were malignant and 54 were benign. Of the patients with nodules, 68 were men and 86 were women, whereas patients without nodules included 51 men and 42 women. The average age of patients with nodules was 60 years old. Two patients were 21–30 years old, seven were 31–40 years old, 24 were 41–50 years old, 37 were 51–60 years old, 53 were 61–70 years old, and 29 were 71 years old or older. Two patients' ages were unknown.

The database included 31 nodules that were 0–10 mm, 52 that were 11–15 mm, 36 that were 16–20 mm, 14 that were 21–25 mm, 17 that were 26–30 mm, and four that were 31–60 mm. The average size of all nodules included in the database was 17.3 mm.

To investigate the distribution of anatomic locations of nodules, right and left lungs on the radiographs were divided into upper, middle, and lower regions. Forty-two nodules were found in the upper right lung, 30 in the middle right lung, 23 in the lower right lung, 20 in the upper left lung, 21 in the middle left lung, and 18 in the lower left lung.

The database contains a text file including the following information about each patient: nodule size (mm), age, sex, final diagnosis, degree of subtlety, anatomic location of the nodule, *x*- and *y*-coordinates of the center of the nodule in a digital image (upper left corner was defined as the origin of *x*- and *y*-coordinates), and nodule classification as malignant or benign. Furthermore, a simple schematic diagram that indicates each location of a lung nodule is contained in a tagged image file format file for visual assistance. The database does not include information about calcification of nodules.

The images in the database were grouped according to the degree of subtlety of the lung nodule based on the consensus of three chest radiologists. The levels of subtlety in detecting a lung nodule were defined as follows: level 1, extremely subtle (detection is extremely difficult because of low contrast, small size, or overlap with a normal structure); level 2, very subtle (detection is very difficult); level 3, subtle (detection is difficult); level 4, relatively obvious (detection is relatively easy); and level 5, obvious (detection is easy). Twenty-five cases were categorized as level 1, 29 as level 2, 50 as level 3, 38 as level 4, and 12 as level 5.

An ROC study of detection of lung nodules was performed to evaluate the radiologists' performance and to characterize this database. All images in this database were printed on film (34.6 cm × 34.6 cm [14 inch × 14 inch]) with a Li-7DD laser printer (Konica). Twenty radiologists (nine chest and 11 general radiologists) from four institutions participated in this study. Their experience ranged from 2 to 22 years (average, 12.1 years). All images were presented in a randomized sequence for detection of lung nodules. For ROC experiments, we used a continuous rating scale and the curve fitting program LABROC5 (Metz CE, The University of Chicago) [5, 6].

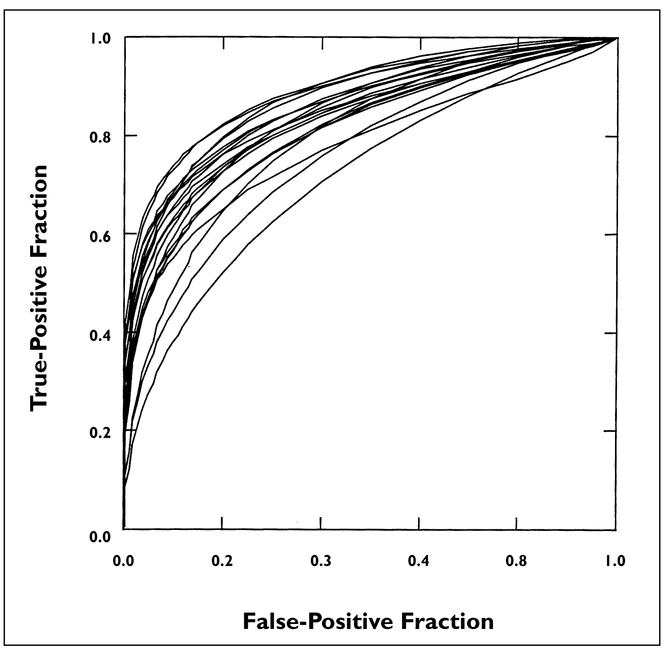

Fig. 1.—Receiver operating characteristic curves for all observers in detecting lung nodules on laser-printed films. Note large variation in radiologists' detection of nodules.

Results

ROC curves showed a large variation in the detection of nodules included in the database for all radiologists (Fig. 1). The average area under the curve (A_z) for all radiologists was 0.833 (Table 1). The average A_z value (0.849) for the chest radiologists exceeded the average A_z value (0.820) for the general radiologists. However, we found no statistically significant difference between these groups, with a *p* value of 0.066 determined by a one-tailed, unpaired *t* test.

ROC curves obtained from radiologists at four institutions were similar (the numbers of observers at institutions A, B, C, and D were three, eight, four, and five) (Fig. 2). We found no statistically significant differences among the average A_z values of the four institutions.

ROC curves were calculated for detection of patients with nodules in each group of different degrees of subtlety as positive cases and all patients without nodules as negative cases. These curves were clearly distinguished as five curves corresponding to degree of subtlety (Fig. 3). The relationship between the degree of subtlety and the aver-

TABLE 1 Area Under the Curve (A_z) Values for 20 Radiologists

Type and No. of Radiologist	Years of Experience	A_z
Chest radiologists[a]		
1	22	0.881
2	20	0.893
3	14	0.835
4	14	0.841
5	12	0.808
6	10	0.879
7	9	0.860
8	9	0.838
9	8	0.808
General radiologists[b]		
10	20	0.841
11	15	0.761
12	14	0.889
13	14	0.848
14	13	0.870
15	12	0.799
16	11	0.859
17	10	0.832
18	7	0.819
19	5	0.719
20	2	0.781

Note.—A_z of all radiologists (mean ± SD) = 0.833 ± 0.045.
[a]Mean A_z = 0.849.

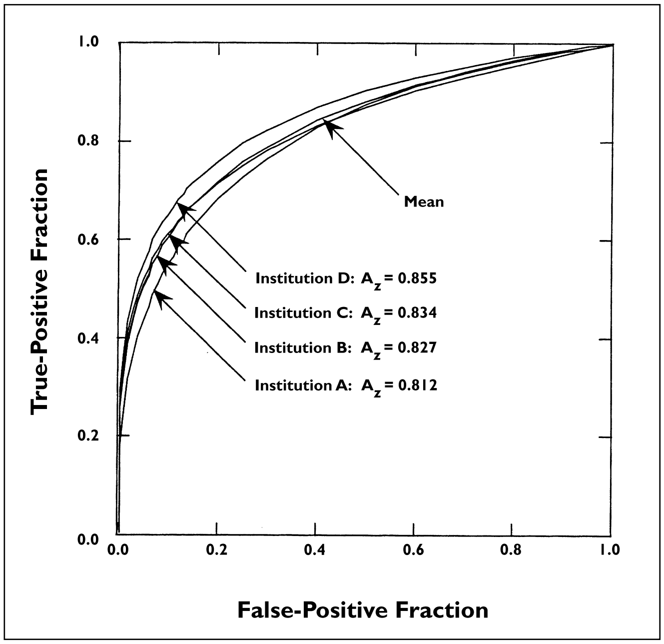

Fig. 2.—Comparison of average receiver operating characteristic curves from four institutions. Note that four curves were similar.

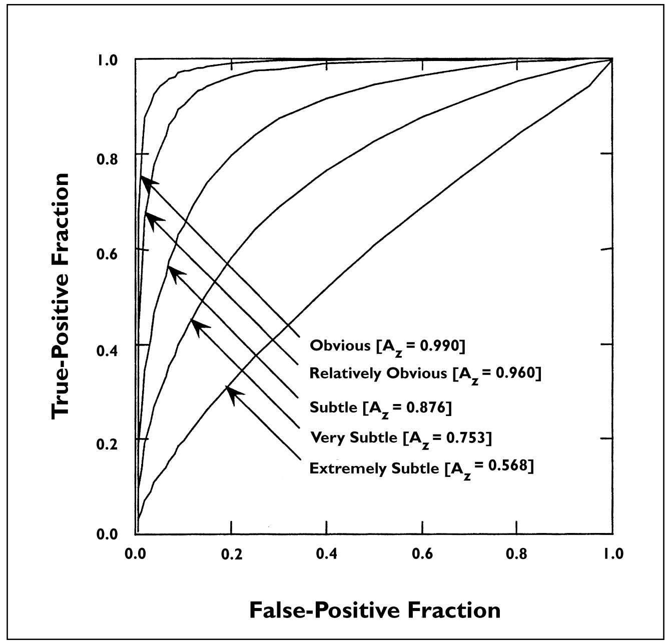

Fig. 3.—Receiver operating characteristic curves for five groups of cases with different degrees of subtlety. Note that they were clearly distinguished as five curves, corresponding to their degrees of subtlety.

age confidence rating score (maximum = 50), which were obtained in ROC, experiments indicated a high correlation coefficient, r = 0.999 (Fig. 4). These results indicate that the images showing nodules in each group of the database are distinctly different and cover a wide range of nodules that differ in their visual appearance.

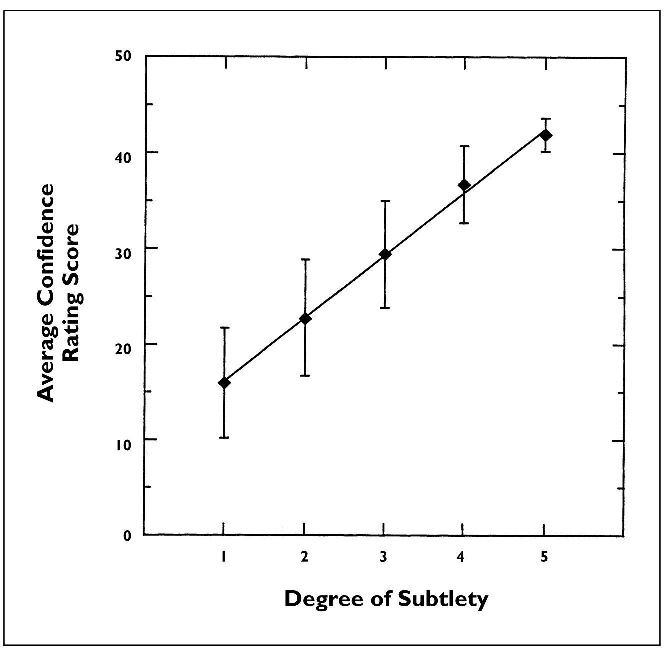

Fig. 4.—Relationship between degree of subtlety and average confidence rating score used by observers in receiver operating characteristic experiment for presence of nodule. High correlation coefficient, r = 0.999, was obtained.

Discussion

We believe that the basic requirements for a clinical image database are a large number of images, acceptable image quality for radiologic diagnosis, a definitive diagnosis established by another examination, images with a wide range of subtlety in detecting an abnormality involved, and confirmation of these subtleties by objective methods. We believe that the database developed in this study meets all of these requirements. Furthermore, this database includes a wide range of nodules with different sizes and different degrees of subtlety.

The proper choice of digital image parameters is important for constructing a digital image database. Small pixels provide high spatial resolution but require large amounts of storage space and increase the computation time for image processing. The parameters used for images in this database may not be optimal for chest images of patients with interstitial disease, but they are appropriate for detection of lung nodules [7, 8]. The pixel size and the gray-scale levels for the computed radiography system that has been used widely for chest radiography are similar to those used for the images in this database. Therefore, we believe that the image quality of this database is acceptable for evaluation of lung nodules.

Although this database was constructed for research on digital imaging, it might be useful for many different purposes, including training and education. One finding in this study indicated that the degree of subtlety correlated with the size of nodules and the age of patients as shown in Table 2 (two patients were not included because of unknown ages). Table 2 shows that the smaller the nodule size and the younger the patient, the more subtle the pulmonary nodule appears, and thus the more difficult nodules are to detect on chest images. Although the average detectability increased with increased average nodule size up to 20 mm, the

TABLE 2 Mean and SD of Nodule Sizes, Patients' Ages, and Sensitivities of 20 Radiologists in Detection of Nodules for Each of Five Groups Categorized by Degrees of Subtlety

Degree of Subtlety	Nodule Size (mm)		Patient Age (yr)		Sensitivity (%)[a]	
	Mean	SD	Mean	SD	Mean	SD
Obvious	23.0	9.03	65.1	10.10	99.58	1.86
Relatively obvious	17.9	6.05	61.2	9.95	92.63	13.11
Subtle	17.2	8.69	59.3	12.62	75.70	22.35
Very subtle	16.4	5.89	59.1	13.64	54.66	23.98
Extremely subtle	14.6	6.69	58.0	11.54	29.60	15.90

[a] The sensitivity was calculated as correct detection when the subjective rating score was ≥25 for nodule cases.

TABLE 3 Relationship Between the Average Sensitivity and Specificity in Detecting Pulmonary Nodules of Different Sizes

Type of Radiologist	A_z	Sensitivity (%)[a] for Various Sizes of Nodules					Specificity (%)
		1–10 mm	11–15 mm	16–20 mm	≥21 mm	All	
Chest	0.849	66.0	78.0	88.9	76.5	77.8	80.4
General	0.820	56.3	62.9	74.2	62.6	64.2	81.3
All	0.833	60.4	69.4	80.5	68.6	70.0	80.9

[a] The sensitivity was calculated as correct detection when the subjective rating score was ≥25 for nodule cases.

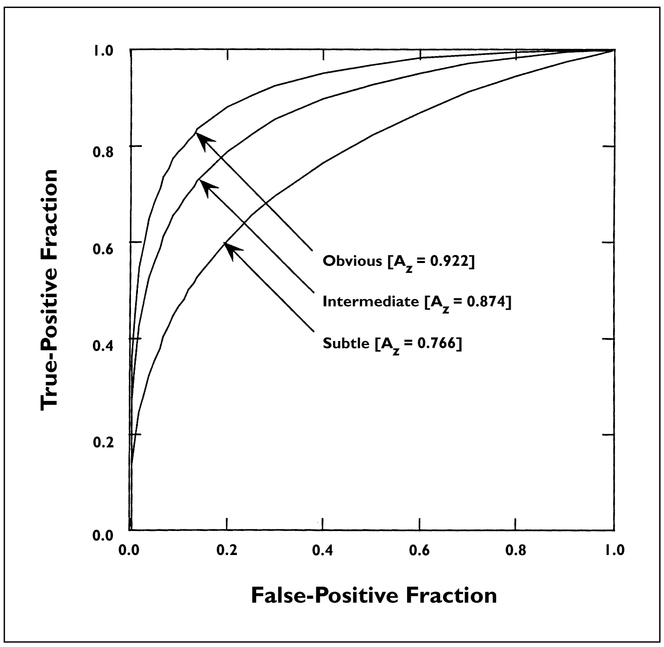

Fig. 5.—Receiver operating characteristic curves obtained by use of three groups of cases with different degrees of subtlety. Note that these subsets can be used for testing various techniques and methods.

average detectability decreased in nodules larger than 20 mm probably because the large nodules included in this database tend to overlap normal anatomic structures and thus are difficult to detect (Table 3).

If this database is used for ROC analysis or other observer performance studies, using all 247 cases may be difficult in some experiments. However, subsets of this database can be used for various studies intended for particular purposes. For example, a subset may be selected for a small number of cases or combinations of cases with several different degrees of subtlety. To illustrate this flexibility, we created three subsets of images using different hypothetic combinations of degrees of subtlety. The three subsets consisted of levels 1–3 as subtle, levels 2–4 as intermediate subtlety, and levels 3–5 as obvious. ROC curves show the expected detection performance using these subset data (Fig. 5). Therefore, subsets such as these can be used for testing various techniques and methods such as different display devices and data compression techniques.

In conclusion, we have developed a common image database for chest radiographs with and without a lung nodule. Twenty radiologists' performance in detecting solitary pulmonary nodules included in the database was evaluated using an ROC analysis. The results of our ROC analysis for detection of lung nodules indicated that this database includes a wide range of nodules with different degree of subtlety. Therefore, we believe that this database will be useful for many purposes, including research, education, quality assurance, and demonstration.

Acknowledgments

We thank Takahiro Kozuka, Ryusaku Yamada, and Akiyoshi Ohtsuka, for their encouragement and support of this work; Kazuhisa Kosakai and Akane Utsunomiya, Osaka City University Hospital, for their technical support; 20 radiologists at four institutions for their participation in the ROC observer studies; and 14 institutions for their kind cooperation in providing clinical images for the creation of this database.

References

1. Nishikawa RM, Giger ML, Doi K, et al. Effect of case selection on the performance of computer-aided detection schemes. *Med Phys* **1994**;21:265–269
2. Nishikawa RM. Design of a common database for research in mammogram image analysis. *Proc Society of Photo-Optical Instrumentation Engineers* **1993**;1905:548–549
3. Suckling J, Parker J, Dance DR, et al. The Mammographic Image Analysis Society digital mammogram database. In: Gale AG, Astley S, Dance DR, Cairms AY, eds. *Digital mammography: proceedings of the 2nd international workshop on digital mammography.* Amsterdam: Elsevier Science, **1994**:375–378
4. Dean PN, Mascio L, Ow D, et al. Proposed standard for image cytometry data files. *Cytometry* **1990**;11:561–569
5. Rockette HE, Gur D, Metz CE. The use of continuous and discrete confidence judgments in receiver operating characteristic studies of diagnostic imaging techniques. *Invest Radiol* **1992**;27:169–172
6. Metz CE, Herman BA, Shen JH. Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Stat Med* **1998**;17:1033–1053
7. Lams PM, Cocklin ML. Spatial resolution requirements for digital chest radiographs: an ROC study of observer performance in selected cases. *Radiology* **1986**;158:11–19
8. MacMahon HM, Vyborny CJ, Metz CE, et al. Digital radiography of subtle pulmonary abnormalities: an ROC study of the effect of pixel size on observer performance. *Radiology* **1986**;158:21–26