Application of Artificial Neural Networks for Quantitative Analysis of Image Data in Chest Radiographs for Detection of Interstitial Lung Disease

Takayuki Ishida, Shigehiko Katsuragawa, Kazuto Ashizawa, Heber MacMahon, and Kunio Doi

The authors have developed an automated computer-aided diagnostic (CAD) scheme by using artificial neural networks (ANNs) on quantitative analysis of image data. Three separate ANNs were applied for detection of interstitial disease on digitized chest images. The first ANN was trained with horizontal profiles in regions of interest (ROIs) selected from normal and abnormal chest radiographs for distinguishing between normal and abnormal patterns. For training and testing of the second ANN, the vertical output patterns obtained from the 1st ANN were used for each ROI. The output value of the second ANN was used to distinguish between normal and abnormal ROIs with interstitial infiltrates. If the ratio of the number of abnormal ROIs to the total number of all ROIs in a chest image was greater than a specified threshold level, the image was classified as abnormal. In addition, the third ANN was applied to distinguish between normal and abnormal chest images. The combination of the rule-based method and the third ANN also was applied to the classification between normal and abnormal chest images. The performance of the ANNs was evaluated by means of receiver operating characteristic (ROC) analysis. The average A_z value (area under the ROC curve) for distinguishing between normal and abnormal cases was 0.976 ± 0.012 for 100 chest radiographs that were not used in training of ANNs. The results indicate that the ANN trained with image data can learn some statistical properties associated with interstitial infiltrates in chest radiographs.

KEY WORDS: interstitial infiltrate, computer-aided diagnosis, artificial neural network, chest radiograph.

INTERSTITIAL LUNG DISEASE is one of the most common findings in abnormal chest radiographs.[1] However, the diagnosis of interstitial lung disease is considered a difficult task for radiologists because no quantitative criterion exists for distinction between normal patterns and subtle abnormal infiltrate patterns on chest radiographs, which results in considerable subjectivity in the radiologists' interpretation of interstitial disease.[2] Therefore, we have been developing computer-aided diagnostic (CAD) schemes for quantitative analysis of interstitial infiltrates to improve diagnostic accuracy and reproducibility. Two different CAD schemes for detection and characterization of interstitial lung disease have been developed by Katsuragawa et al,[3-8] based on texture analysis by use of Fourier transform and geometric-pattern analysis. In these schemes, image features associated with interstitial infiltrate patterns were extracted from digitized chest radiographs. These features are the root mean square (rms) variation and the first moment of the power spectrum obtained by texture analysis, which correspond to the magnitude and the coarseness (or fineness) of the infiltrates. In addition, by geometric-pattern analysis, the total area of the area components and the total length of the line components are obtained, which are related to nodular and linear opacities (respectively) in interstitial infiltrate patterns. Although the performance of these schemes is generally very good, there are still some cases in which they cannot distinguish correctly between normal and abnormal cases.[9] Therefore, we have developed an alternative approach based on radiographic image data.

Artificial neural network (ANN) has been applied to several chest CAD schemes, including the differential diagnosis of interstitial lung diseases[10] on the basis of clinical and radiographic information, and the elimination of false positives in the detection of lung nodules.[11] ANNs used in these schemes were trained with some extracted features associated with lesions and/or false positives. In this study, we attempted to detect interstitial infiltrates by using ANNs trained with digital chest image data.

MATERIALS AND METHODS

Materials

The database used in the study consisted of 100 normal and 100 abnormal posteroanterior (PA) chest radiographs obtained

From the Kurt Rossmann Laboratories for Radiologic Image Research and the Department of Radiology, The University of Chicago, Chicago, IL.

This study was supported by USPHS Grants CA24806 and CA62625.

K. Doi and H. MacMahon are shareholders of R2 Technology, Inc, Los Altos, CA. It is the policy of the University of Chicago that investigators disclose publicly actual or potential significant financial interests that may appear to be affected by the research activities.

Address reprint requests to Kunio Doi, PhD, Department of Radiology, The University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637.

0897-1889/98/1104-0004$8.00/0

with Lanex Medium screens and OC film (Eastman Kodak, Rochester, NY). The normal cases were selected by four experienced chest radiologists on the basis of unequivocally normal radiographs in terms of clinical data and follow-up chest radiographs. The abnormal cases of interstitial lung disease with varying severity were selected based on radiological findings, computed tomography results, clinical data, and/or follow-up radiographs, by consensus of the same radiologists. The abnormal cases contained interstitial infiltrates in more than 20% of the area of both lung fields. Twenty-six cases had focal interstitial infiltrates containing 20% to 50% of the lung area. The other abnormal cases had diffuse interstitial infiltrates involving more than 50% of the lung area.[7] Chest radiographs were digitized using a laser scanner (KFDR-S; Konica Corp, Tokyo, Japan) with a 0.175-mm pixel size and 1,024 gray levels. An IBM RISC/6000 Powerstation (model 590) was used for development of our CAD scheme.

For evaluation of the classification performance, our database of 100 normal and 100 abnormal cases with interstitial infiltrates was divided randomly into two groups: a data set for training and another for testing. Each data set included 50 normal and 50 abnormal cases. We prepared four different pairs of data sets for the evaluation of overall performance.

Overall Scheme of the ANN Analysis

The schematic diagram of our ANN analysis method is shown in Fig 1. First, the regions of interest (ROIs) are selected automatically from normal and abnormal chest radiographs; our method for automated selection of ROIs has been described in detail elsewhere.[5] The matrix size of the ROI is 32 × 32 (5.6 × 5.6 mm^2). The two-dimensional distribution of pixel values in an ROI is considered to consist of 32 one-dimensional horizontal profiles, each of which includes 32 pixel values along the horizontal direction. We used two different ANNs to detect the interstitial infiltrates. The horizontal profiles in each ROI were subjected to a background trend correction[3] based on the surface fitting technique, and then entered into the first ANN for distinction between normal and abnormal profiles. The output value obtained from the first ANN ranges from 0 to 1, which represents the degree of normality/abnormality (0 = normal; 1 = abnormal) for each horizontal profile. Because each ROI has 32 horizontal profiles, 32 outputs corresponding to the 32 profiles for each ROI are obtained from the first ANN. The sequence of these 32 outputs, which will be called the "vertical output pattern," is entered into the second ANN. The output value of the second ANN, which represents the degree of normality/abnormality (0 = normal; 1 = abnormal) for each ROI is used to distinguish between normal and abnormal ROIs. Finally, the classification for each chest image is performed for distinguishing between normal and abnormal cases with interstitial infiltrates.

First ANN for Horizontal Profiles

A three-layer feed-forward neural network with a back-propagation algorithm is employed for the first ANN. The ANN has 32 input units, 16 hidden units, and 1 output unit. For

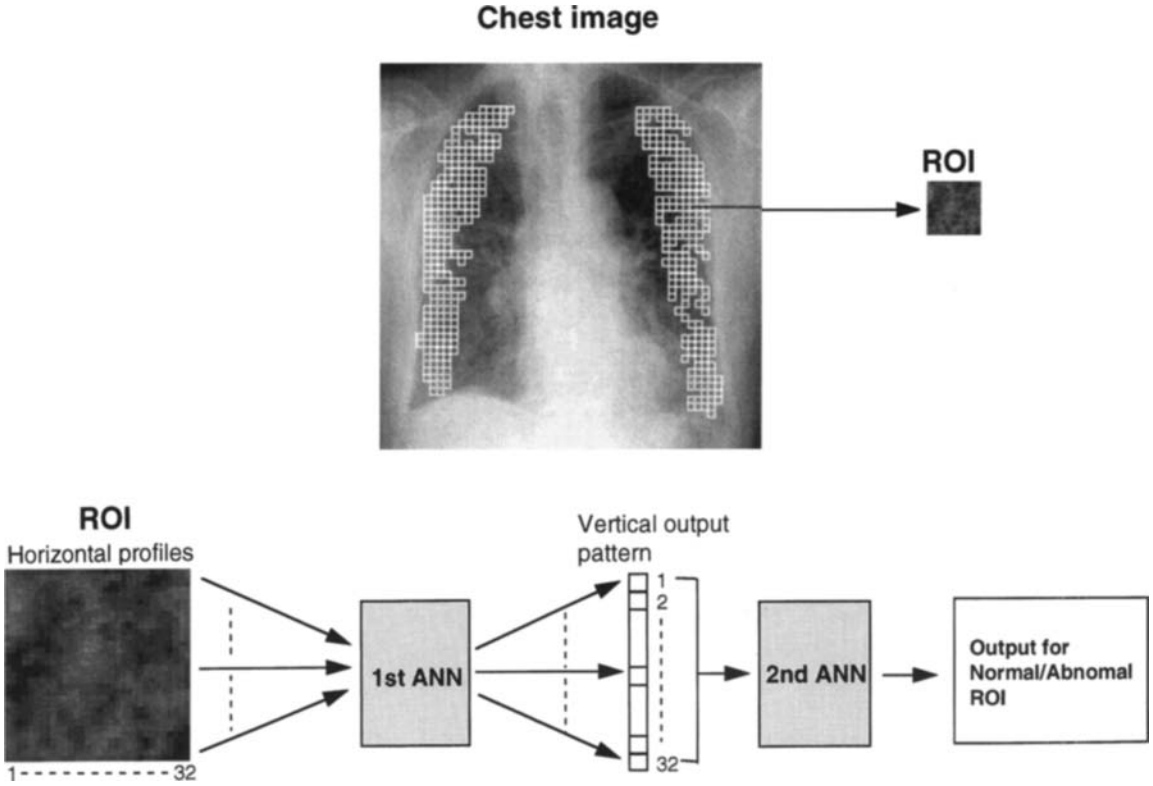

Fig 1. Schematic diagram of an ANN method for analysis of interstitial infiltrates in chest images.

training of the first ANN, we initially selected "representative" ROIs for normal and abnormal cases based on single-texture indices obtained by texture analysis.[4] The single texture indices indicate the magnitude and the coarseness (or fineness) of a texture pattern by combining the rms variation and the first moment of the power spectrum. The normal ROIs for this study were selected randomly from ROIs in normal chest images in our database where texture indices of ROIs are in the mid-30% range (from 35% to 65%) of all data in each case. The abnormal ROIs also are randomly selected from ROIs in abnormal chest images, in which the texture indices of ROIs are in the upper 30% range (from 5% to 35%) of all data in each case. Although "representative" ROIs were selected using a texture index, it should be noted that some "normal" ROIs include horizontal profiles that may appear "abnormal" mainly because of rib edges and/or sharp edge vessels. In addition, "abnormal" ROIs may include some horizontal profiles with very weak interstitial infiltrates that thus may appear "normal." Therefore, we attempted to reduce such "contaminated" training data by using the rms value of each profile. For the final selection of training data for normals, horizontal profiles with large rms values in the upper 25% of normal cases were eliminated. For the training data of abnormals, we eliminated horizontal profiles with small rms values in the lower 25% of abnormal cases.

The horizontal profiles of a normal and an abnormal case are shown in Fig 2A. It is apparent in Fig 2A that the variation of pixel values in the abnormal profile is larger than that in the normal case. The average pixel values and the rms values of all input data at each input unit for the first ANN are shown in Figs 2B and 2C, respectively. Because the trend correction technique is applied to all the ROIs, the average values are very close to zero at all input units. It is evident that the rms values of abnormal cases are greater than those of normals at all input units because the variation of the horizontal profiles with interstitial infiltrates usually is larger than that of normals. The rms values at the left and the right three units increased slightly, as shown in Fig 2C, because the fitting error for trend correction tends to be large near the edges of the ROIs. Figure 2D shows histograms of all input data (normals and abnormals) used for a unit of the first ANN. The distribution for abnormal cases was broader than that of normal cases. The two distributions are very similar to gaussian distributions.

Second ANN for ROIs

The second ANN has the same structure as the first. The second is trained with vertical output patterns that correspond to a series of the output from the first ANN for all horizontal

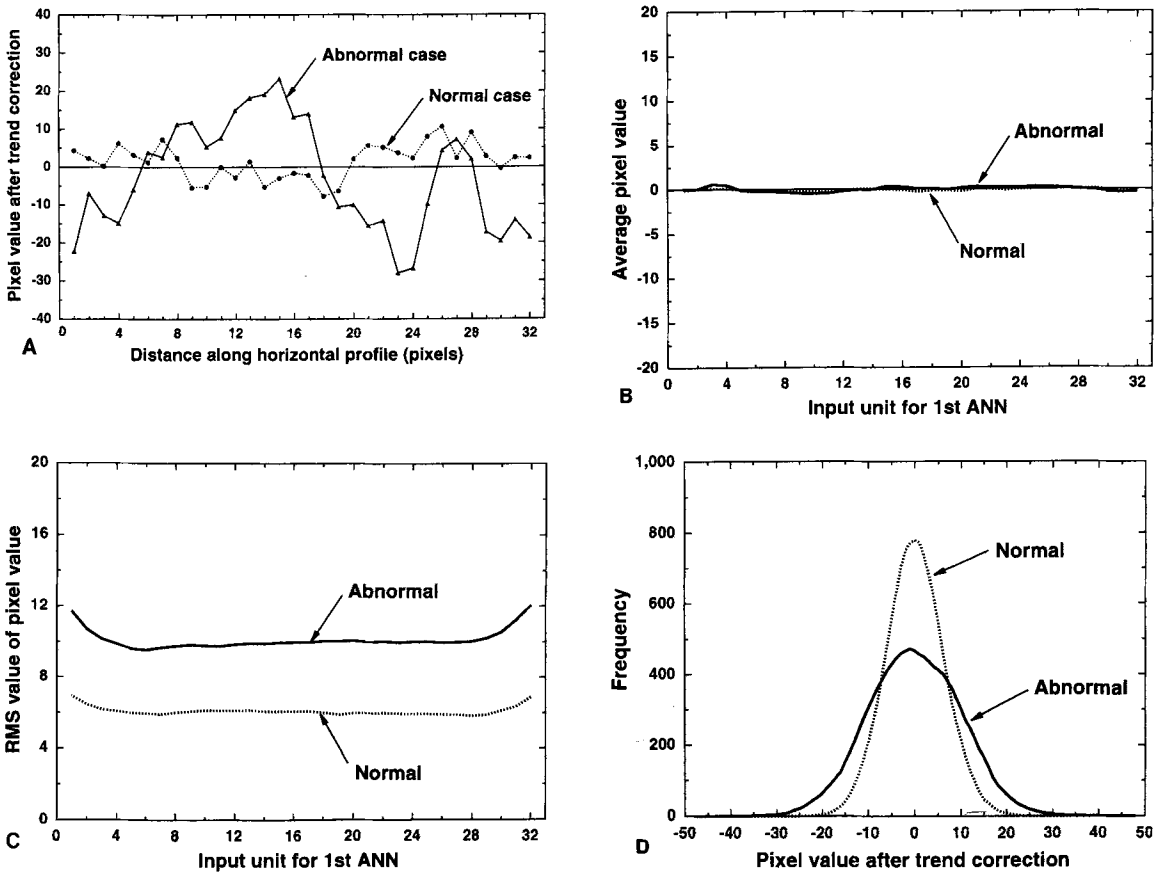

Fig 2. (A) Horizontal profiles obtained from normal and abnormal chest images. (B) Average pixel value of all input data at each input unit. (C) rms value of all input data at input unit. (D) Histogram of all input data for a unit of the first ANN.

profiles included in each ROI. For training of the second ANN, "representative" ROIs for normals and abnormals are selected initially in the same way as those used for training of the first ANN. However, the ROIs for the second ANN did not overlap with those used for the first ANN. To reduce the effect of "contamination" in the training data, we determined the average value of vertical output patterns as a measure. For training data of normal ROIs, we eliminated ROIs with large average values (upper 10% of normal ROIs) of vertical output patterns. For training data of abnormal ROIs, we eliminated ROIs with small average values (lower 10% of abnormal ROIs). The output of the second ANN was used to distinguish between normal and abnormal ROIs with interstitial infiltrates.

Vertical output patterns for a normal and an abnormal ROI are shown in Fig 3A. It is obvious that the output value from the first ANN of the abnormal ROI usually is larger than that of the normal ROI. The average values and the rms values of all input data at each input unit for the second ANN are shown in Figs 3B and 3C, respectively. Because the outputs from the first ANN for the abnormal ROI tend to be large, the average output value for abnormal cases is greater than that for normal cases at all input units. However, the rms value of the output from the first ANN for abnormal ROIs is very similar to that of normal ROIs.

Classification Method

The overall classification for a chest image between normal and abnormal cases is performed initially by a rule-based method alone and later by a rule-based + ANN method. If the output of the second ANN is greater than a predetermined threshold value, the ROI is considered abnormal. If the ratio of the number of abnormal ROIs to the total number of all ROIs in a chest image is greater than a certain threshold value, the chest image is classified as abnormal. This classification scheme is called a rule-based method.

In another classification scheme, the third ANN was applied to distinguish between normal and abnormal chest images. This classification scheme, with a rule-based + third ANN method, is shown in Fig 4. First, the rule-based method is employed for determination of "obviously normal" and "obviously abnormal" lungs. If the ratio of the number of abnormal ROIs to the total number of ROIs in a chest image is below the minimum "abnormal" ratio that can be obtained from all abnormal cases in a training data set, the chest image is classified as "obviously normal." On the other hand, the chest image is classified as "obviously abnormal" if the ratio is above the maximum "normal" ratio that can be obtained from all normal cases in the

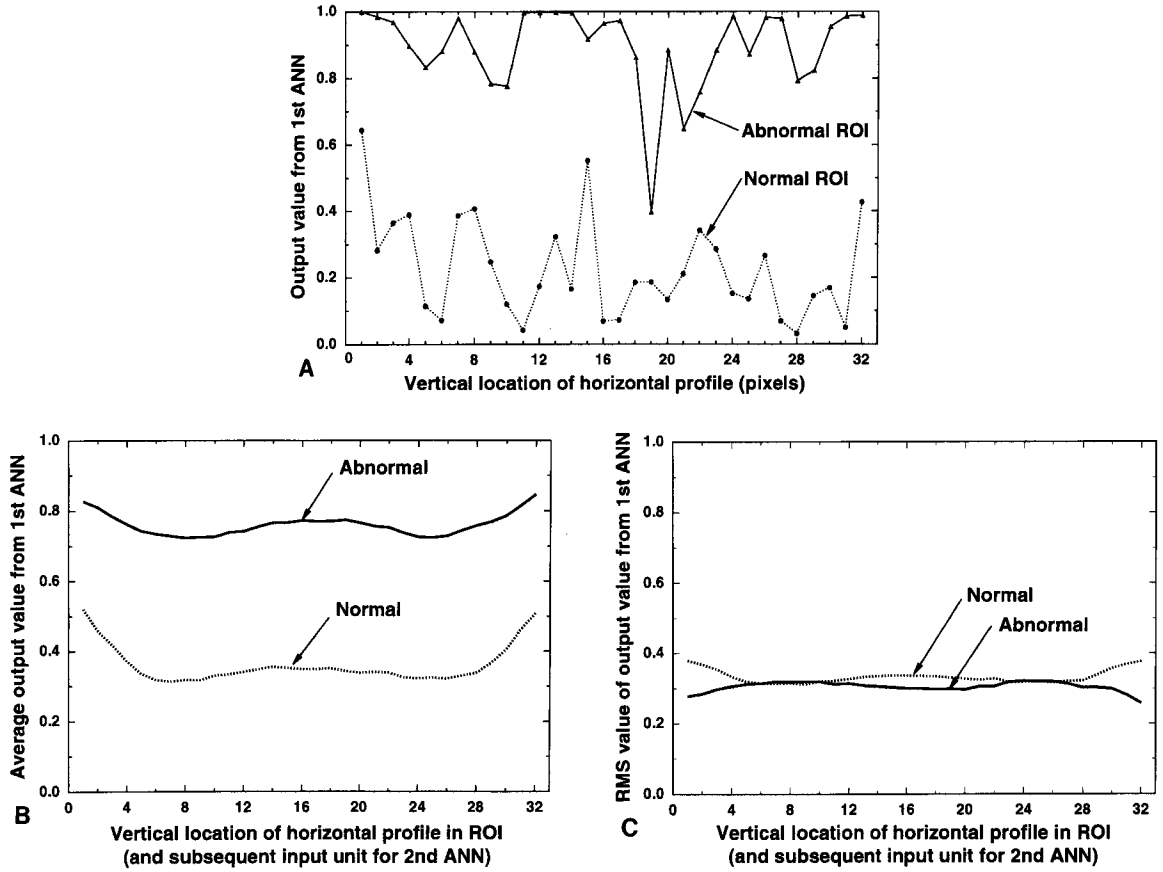

Fig 3. (A) Vertical output patterns obtained from normal and abnormal ROIs. (B) Average output value from the first ANN, and subsequent input data for each input unit of the second ANN. (C) rms values of output values from the first ANN for each vertical location of horizontal profiles in ROI.

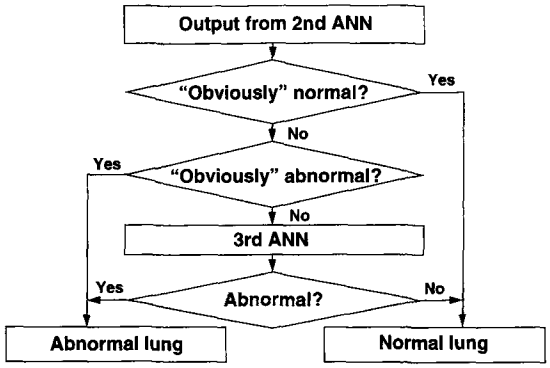

Fig 4. Classification scheme of rule-based + ANN method.

training data set. The remaining chest images are classified by applying the third ANN, which consists of 10 input units, 5 hidden units, and 1 output unit. The output value from the third ANN ranges from 0 (normal) to 1 (abnormal). The input data for the third ANN are determined from a histogram of the output values from the second ANN for each chest image. Figure 5 shows histograms of the output values for a normal case and an abnormal case, respectively. Ten input values correspond to the 10 frequency values at 10 bins, from 0-0.1 to 0.9-1.0. These 10 input values represent the output distribution of the histogram of the output from the second ANN.

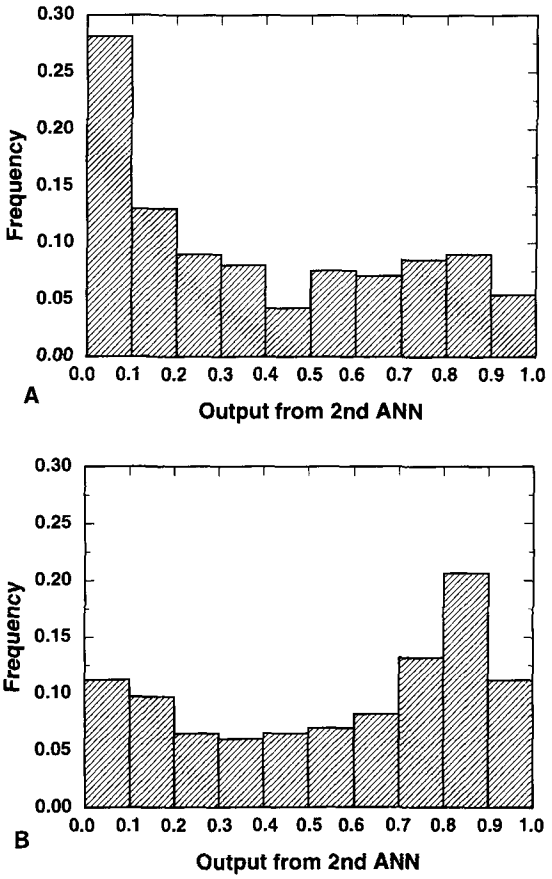

Fig 5. Histograms of the output values obtained from the second ANN from (A) a normal chest radiograph and (B) an abnormal chest radiograph with interstitial infiltrates.

RESULTS

Effect of the Number of Training Data for the First ANN

To investigate the effect of the total number of input data sets used for training of the 1st ANN, we randomly selected ROIs from 10 normal cases and 10 abnormal cases with severe interstitial infiltrates. The number of input data ranges from 320 to 32,000 horizontal profiles (10 to 1,000 ROIs). Half the input data were selected from normal cases and half from abnormal cases. For the validation test of the first ANN, however, a total of 12,800 horizontal profiles (200 normal and 200 abnormal ROIs) were selected randomly from 20 normal and 20 abnormal cases with various severities of interstitial infiltrates, which were not used as training cases. The effect of the total number of training data for the first ANN is shown in Fig 6. The A_z values (area under the ROC curve) increased gradually as the number of input data sets increased from 320 to 10,240. The classification performance with more than 10,240 input data sets was increased only slightly. The result indicates that the number of training data sets for the first ANN should be more than 10,000 horizontal profiles.

Overall Performance of the ANN Scheme

The overall performance with the rule-based classification scheme, and the individual perfor-

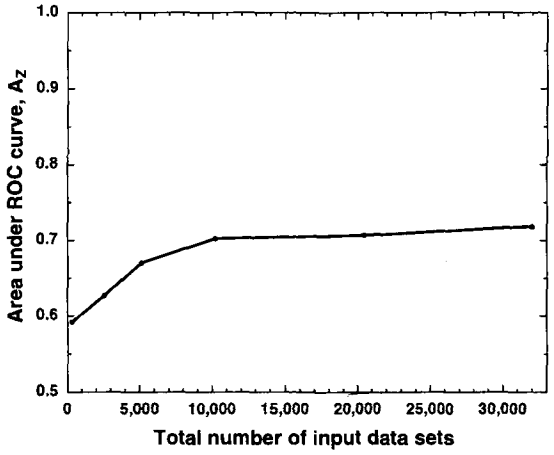

Fig 6. Effect of the total number of input data sets used for training of the first ANN.

mance of the first and of the second ANN, which were obtained by ROC analysis with a continuous rating scale (LABROC4 program),[12-14] are shown in Fig 7. The ROC curve of the overall scheme was high. However, the classification performance on normal and abnormal profiles by the first ANN was not high. This poor performance probably was attributable to the fact that some normal ROIs include "abnormal-like" profiles caused by rib edges and/or sharp edge vessels; also, some abnormal ROIs include "normal-like" profiles owing to extremely weak interstitial infiltrates.

The performance of the second ANN was greater than that of the first ANN. If the majority of horizontal profiles in a normal ROI are "normal" patterns, the second ANN can recognize the ROI as normal, even if the ROI includes a few contaminated horizontal profiles, ie, a few outputs from the first ANN can have large values, as shown in Fig 3A. In a similar way, when the majority of horizontal profiles in an abnormal ROI contain interstitial infiltrate patterns, the second ANN can recognize the ROI as abnormal, even if the ROI has a few contaminated horizontal profiles. Therefore, the performance of the second ANN was better than that of the first.

We found that the result of the ANN scheme tends to provide a number of false-positive ROIs at high optical densities and a number of false-negative ROIs at low optical densities in digitized radiographs. These results are caused by the optical density dependence of the film gradient. Therefore, we employed an optical-density correction technique[6] using a gradient curve of the OC film used. With this correction method, pixel values that are roughly proportional to the optical density are divided by the gradient of the film. The average A_z values with and without the density correction are 0.934 ± 0.004 and 0.906 ± 0.021, respectively. The overall performance with and without the density correction was evaluated by use of four independent data sets. It is obvious that the performance of the ANN scheme is improved considerably by applying the optical-density correction technique.

An abnormal chest image with and without the results of the ANN analysis are shown in Fig 8. Pluses and circles indicate normal and abnormal ROIs, respectively. The larger the circle, the greater the ANN output, which corresponds to the greater abnormality. This chest image shows localized interstitial infiltrates in the lower right and lower

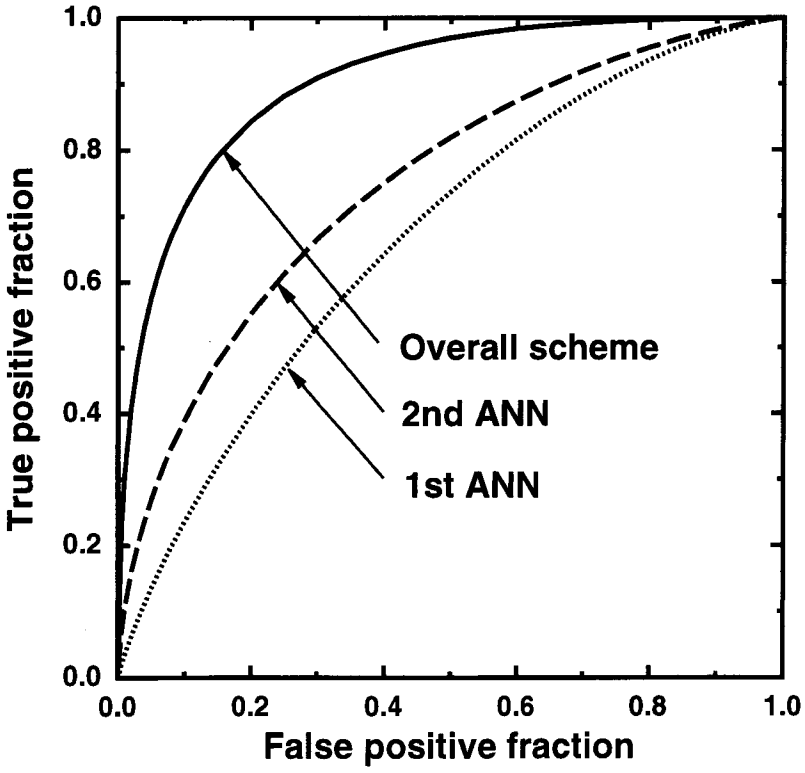

Fig 7. ROC curves for the first ANN, the second ANN, and the overall scheme in distinguishing between normals and abnormals.

Fig 8. (A) Chest image with localized interstitial infiltrates.

left lungs. The distribution of abnormal ROIs (indicated by circles) appears to correspond well to the areas of interstitial infiltrates.

The average A_z value according to the rule-based + ANN classification scheme was greatly improved, from 0.934 ± 0.004 to 0.976 ± 0.012, as shown in Fig 9. However, when the third ANN alone was applied to the overall classification without removal of obvious cases by the rule-based method, the A_z value was 0.938 ± 0.002, which is low and comparable to that obtained with the rule-based classification scheme alone. The sensitivities at 90% specificity obtained with the three classification methods are shown in Table 1. It is apparent that the best classification method is the rule-based + ANN.

DISCUSSION

In the initial phase of the study, we applied a three-layer, feed-forward ANN for distinguishing between normal and abnormal ROIs, instead of horizontal profiles, with interstitial infiltrates. The ANN consisted of 1,024 (32 × 32) input units, 128 hidden units, and 1 output unit. First, we randomly selected 400 ROIs for training and another 400 for testing of the ANN from 10 normal and 10 abnormal cases with severe interstitial lung disease. The ANN was trained by trend-corrected pixel values. On the consistency test, most ROIs were correctly identified by the ANN. However, according to the result of the validation test, the ROIs could not be recognized accurately in terms of normal and abnormal cases. We believe that the results indicate

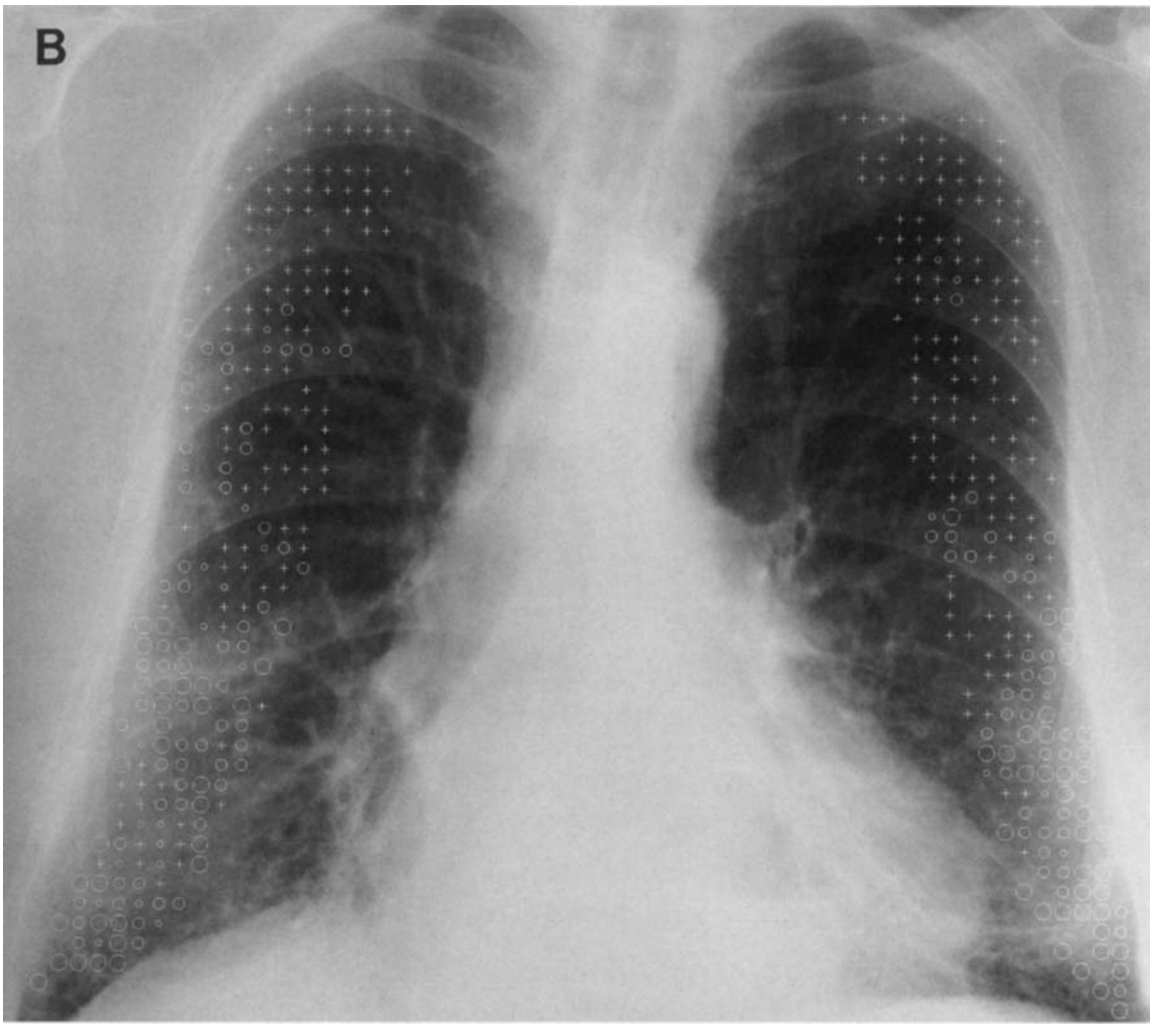

Fig 8 (cont'd). (B) The result of the ANN analysis for chest image (A).

that the number of training data sets was not adequate for learning the statistical property of interstitial infiltrate patterns. However, if the number of training ROIs is increased, the computational time for training the ANN with a large number of input units will become extremely long, and thus it becomes impractical to use the ANN. Therefore, we decided that the training data should first be reduced to one-dimensional data, ie, to the horizontal profiles, which can be used easily for training of the ANN, rather than use all pixel data in the ROIs. Second, the output of the first ANN for all profiles in an ROI can be analyzed by the second ANN, as described in this study.

To investigate the effect of the reduction of the contaminated training data, three sets of ROIs are selected randomly from 10 normal and 10 abnormal cases with severe interstitial infiltrates. Two sets were used for training and testing of the 1st ANN, and the third was used for the validation test. The first ANN was trained initially with the first training data set. Then, for testing, we entered the second set into the trained first ANN, and subsequently we eliminated the horizontal profiles in the second data set, which provided either large output values from the first ANN for normal horizontal profiles or small output values for abnormal profiles. The fraction of the number of profiles eliminated varied from 25% to 75%. After elimination of the contaminated profiles, we retrained the first ANN by the "noncontaminated" profiles in the second data set. Although the validation test data also include some contaminated horizontal profiles, we did not eliminate the contaminated horizontal

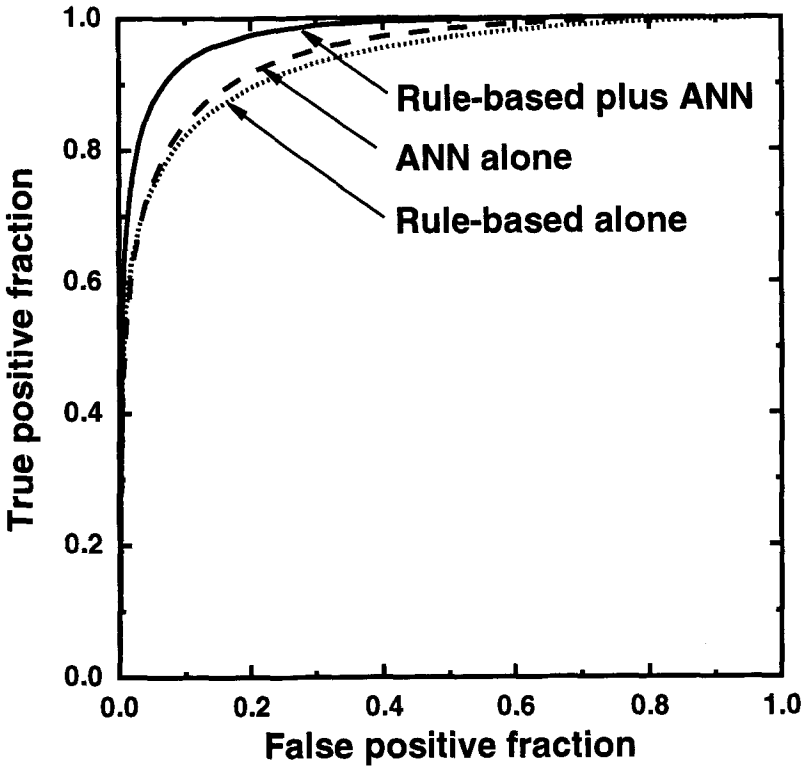

Fig 9. Comparison of ROC curves between the rule-based classification method and the rule-based + ANN classification method.

profiles for final testing in the validation test. If we had eliminated the contaminated horizontal profiles from the testing data, the classification performance would have been greatly improved. However, the improved performance obtained only with "noncontaminated" testing data would not be applicable to actual chest images, because chest images generally contain "contaminated" data.

As more training data are eliminated, consistency test results improve. In the validation test, however, it should be noted that the best performance was obtained when 25% of the training data were eliminated. When 75% of the training data were eliminated, the performance decreased slightly. This is probably because only obvious patterns would remain in the training data sets after the 75% removal, and thus the ANN would not be trained adequately because of lack of a proper mix of training data.

Table 1. Sensitivity (at 90% Specificity) Obtained With the Three Classification Methods Distinguishing Normal and Abnormal Chest Images With Interstitial Infiltrates

Method	Sensitivity
Rule-based alone	78.85% ± 1.86%
ANN alone	83.13% ± 2.10%
Rule-based + ANN	93.20% ± 3.85%

As described above, the rms values of abnormal profiles generally are greater than and different from those of normal profiles. Therefore, one may assume that the first ANN could have learned only the difference in the rms values of the horizontal profiles between normal and abnormal patterns. However, it is possible that the first ANN has learned the difference between normal and abnormal patterns in terms of more than the rms value; namely, some other statistical property associated with patterns of interstitial infiltrates. To examine this assumption, we determined the performance by using the rms values of horizontal profiles alone for distinction between normal and abnormal cases. The ROC curve obtained from the rms values alone was very low, as shown by the lower dotted line in Fig 10, and was below the ROC curve obtained from the first ANN. The ROC curve obtained from the rms values for ROIs was less than that obtained from the second ANN, as shown by the thin dashed line in Fig 10. For overall performance, the ROC curve obtained with the ANN scheme is clearly greater than that obtained with the rule-based

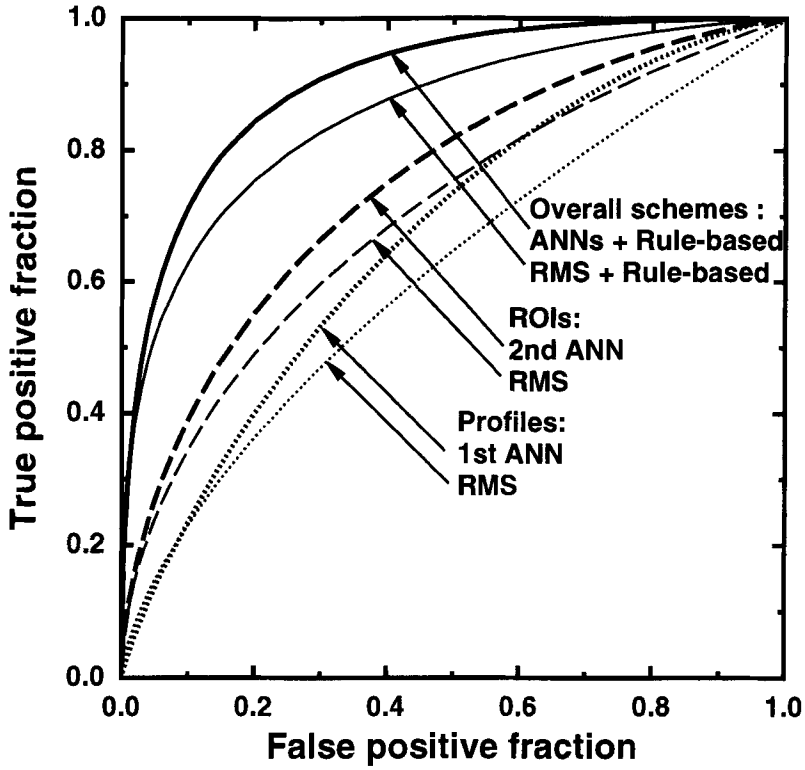

Fig 10. Comparison of ROC curves between the ANN method and the rule-based method by use of rms values.

method by use of rms values. This appears to indicate that the ANN can learn some statistical properties of normal and abnormal patterns caused by interstitial infiltrates.

The high performance level ($A_z = 0.976$) obtained with this new ANN approach was greater than the average A_z value of 0.949, which was obtained by 15 radiologists for distinguishing between normal and abnormal chest radiographs with interstitial infiltrates.[15] Because only a subset of our database has been used for this observer test, two A_z values could not be compared directly. However, the potential usefulness of our ANN approach is clear.

ACKNOWLEDGMENTS

The authors are grateful to E. Lanzl for editing the manuscript and to E.A. Ruzich for secretarial assistance.

REFERENCES

1. MacMahon H, Liu KJM, Montner SM, et al: The nature and subtlety of abnormal findings in chest radiographs. Med Phys 18:206-210, 1991

2. Herman PG, Gerson DE, Hessel SJ, et al: Disagreement in chest roentgen interpretation. Chest 68:278-282, 1975

3. Katsuragawa S, Doi K, MacMahon H: Image feature analysis and computer-aided diagnosis in digital radiography: Detection and characterization of interstitial lung diseases in digital chest radiographs. Med Phys 15:311-319, 1988

4. Katsuragawa S, Doi K, MacMahon H: Image feature analysis and computer-aided diagnosis in digital radiography: Classification of normal and abnormal lungs with interstitial diseases in chest images. Med Phys 16:38-44, 1989

5. Chen X, Doi K, Katsuragawa S, et al: Automated selection of regions of interest for quantitative analysis of lung textures in digital chest radiographs, Med Phys 20:975-982, 1993

6. Morishita J, Doi K, Katsuragawa S, et al: Computer-aided diagnosis for interstitial infiltrates in chest radiographs: Optical-density dependence of texture measures. Med Phys 22:1515-1522, 1995

7. Monnier-Cholley L, MacMahon L, Katsuragawa S, et al: Computerized analysis of interstitial infiltrates on chest radiographs: A new scheme based on geometric-pattern features and Fourier analysis. Acad Radiol 2:455-462, 1995

8. Katsuragawa S, Doi K, MacMahon H, et al: Quantitative analysis of geometric-pattern features of interstitial infiltrates in digital chest radiographs: Preliminary results. J Digit Imaging 9:137-144, 1996

9. Abe K, Doi K, MacMahon H, et al: Computer-aided diagnosis in chest radiography: Preliminary experience. Invest Radiol 28:987-993, 1993

10. Asada N, Doi K, MacMahon H, et al: Potential usefulness of an artificial neural network for differential diagnosis of interstitial diseases: Pilot study. Radiology 177:857-860, 1990

11. Xu XW: Computerized analysis for automated detection of lung nodules in digitized chest radiographs. PhD dissertation, University of Chicago, Chicago, IL, 1996; abstract in Med Phys 23:1467, 1996

12. Metz CE: ROC methodology in radiographic imaging. Invest Radiol 21:720-733, 1986

13. Metz CE: Some practical issues of experimental design and data analysis in radiographical ROC studies. Invest Radiol 24:234-245, 1989

14. Metz CE, Shen JH, Herman BA: New methods for estimating a binormal ROC curve from continuously distributed test results. Invited presentation at the 1990 Joint Statistical Meeting of the American Statistical Society and the Biometric Society, Anaheim, CA, 1990

15. Monnier-Cholley L, Katsuragawa S, Morishita J, et al: Objective evaluation of computerized scheme for detection of interstitial infiltrates on chest radiographs. Radiology 193P:146, 1994 (suppl)