Performance of Radiologists in Detection of Small Pulmonary Nodules on Chest Radiographs: Effect of Rib Suppression With a Massive-Training Artificial Neural Network

Seitaro Oda1
Kazuo Awai1
Kenji Suzuki2
Yumi Yanaga1
Yoshinori Funama3
Heber MacMahon2
Yasuyuki Yamashita1

OBJECTIVE. A massive-training artificial neural network is a nonlinear pattern recognition tool used to suppress rib opacity on chest radiographs while soft-tissue contrast is maintained. We investigated the effect of rib suppression with a massive-training artificial neural network on the performance of radiologists in the detection of pulmonary nodules on chest radiographs.

MATERIALS AND METHODS. We used 60 chest radiographs; 30 depicted solitary pulmonary nodules, and 30 showed no nodules. A stratified random-sampling scheme was used to select the images from the standard digital image database developed by the Japanese Society of Radiologic Technology. The mean diameter of the 30 pulmonary nodules was 14.7 ± 4.1 (SD) mm. Receiver operating characteristic analysis was used to evaluate observer performance in the detection of pulmonary nodules first on the chest radiographs without and then on the radiographs with rib suppression. Seven board-certified radiologists and five radiology residents participated in this observer study.

RESULTS. For all 12 observers, the mean values of the area under the best-fit receiver operating characteristic curve for images without and with rib suppression were 0.816 ± 0.077 and 0.843 ± 0.074; the difference was statistically significant (p = 0.019). The mean areas under the curve for images without and with rib suppression were 0.848 ± 0.059 and 0.883 ± 0.050 for the seven board-certified radiologists (p = 0.011) and 0.770 ± 0.081 and 0.788 ± 0.074 for the five radiology residents (p = 0.310).

CONCLUSION. In the detection of pulmonary nodules, evaluation of a combination of rib-suppressed and original chest radiographs significantly improved the diagnostic performance of radiologists over the use of chest radiographs alone.
Materials and Methods

Our institutional review board approved the use of the database. Informed consent for the inclusion of cases was waived because the database is open to public inspection [21]. The institutional review board also approved the participation of the radiologists in the observer performance test. Informed consent for the observer performance study was obtained from all participants.

Rib-Suppressed Chest Radiography With Massive-Training ANN

Use of a massive-training ANN for rib suppression on chest radiographs involves two steps: a training step (Fig. 1A) and an application step (Fig. 1B). In the training step, we used chest images obtained with a dual-energy radiography system. Dual-energy subtraction, a technique for separating bone from soft tissue on chest radiographs, can produce two tissue-selective images, that is, a bone image and a soft-tissue image. Standard chest radiographs are used as input images to the massive-training ANN, and the corresponding dual-energy bone images are used as the teaching images. We did not directly use dual-energy soft-tissue images as the teaching images because the massive-training ANN trained with dual-energy soft-tissue images produced results slightly inferior to the massive-training ANN trained with dual-energy bone images [20].

In this study, we trained the massive-training ANN with five sets of chest radiographs obtained with a dual-energy radiography system. The training set consisted of four cases of nodules (average diameter, 20 mm) and one case without nodules. The nodule cases were selected with the criterion that at least 20% of the nodule area overlapped ribs. The nodule-free case represented a mean or typical case in terms of rib contrast and size. An important property of the massive-training ANN is its ability to be trained with a very small number of cases. Massive training with a large number of training samples (subregions) obtained from a small number of cases through a training-sample enrichment process allows the massive-training ANN to avoid the overfitting problem of an ANN [20].

In the application step, when an input chest image obtained with a standard radiographic system was entered into the trained massive-training ANN, it output a bone-image-like image on which the bones were isolated and enhanced. A soft-tissue-image-like image on which ribs were suppressed was produced by subtraction of the bone-image-like image from the input chest image. Thus, our massive-training ANN technique produced bone-image-like and soft-tissue-image-like images from a single chest image obtained with a standard radiographic system. The technical aspects of image processing with a massive-training ANN have been detailed previously [20].

Database

To investigate the effect of rib suppression on the performance of radiologists in the detection of pulmonary nodules on radiographs, we conducted an observer performance study. One chest radiologist with 22 years of experience in reading chest radiographs, who did not participate in the observer performance study, selected 60 chest radiographs from the Japanese Standard Digital Image Database developed by the Japanese Society of Radiologic Technology [21]. He then selected 30 patients with only one malignant solid pulmonary nodule not larger than 20 mm without calcification and 30 patients without nodules. The presence or absence of a pulmonary nodule was confirmed by CT examination, and the nodule diagnosis was made by histologic and cytopathic examination. No patient had other CT findings of pleural or parenchymal abnormalities, such as consolidation, ground-glass opacity, or pulmonary fibrosis.

The 60 patients were 36 men and 24 women with a mean age of 59.4 years (range, 27–89 years). The mean age of the patients with pulmonary nodules was 60.4 years (range, 39–78 years), and that of patients without nodules was 58.3 years (range, 27–89 years). According to results of the two-tailed Student’s t test, there was no significant difference between the age of patients with and that of patients without pulmonary nodules ($p = 0.555$). Twenty-seven of the 30 patients with pulmonary nodules had primary lung cancer, and three had solitary pulmonary metastatic lesions. The mean diameter of the pulmonary nodules was $14.7 \pm 4.1$ (SD) mm.

Observer Performance Study

We used a sequential test method for receiver operating characteristic (ROC) analysis to evaluate the diagnostic performance of radiologists in detecting pulmonary nodules on chest images without and with rib suppression. The observers were seven board-certified radiologists with 5–16 years of experience (mean, 8.4 years) and five radiology residents with 2–3 years of experience (mean, 2.4 years).
Chest Radiography of Pulmonary Nodules

TABLE 1: Areas Under the Curve for Board-Certified Radiologists and Radiology Residents in Detection of Pulmonary Nodules Without and With Rib-Suppressed Images

<table>
<thead>
<tr>
<th>Observer</th>
<th>Area Under Curve</th>
<th>Without Rib Suppression</th>
<th>With Rib Suppression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Board-certified radiologists</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.859</td>
<td>0.962</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.936</td>
<td>0.880</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.856</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.896</td>
<td>0.821</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.773</td>
<td>0.820</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.841</td>
<td>0.913</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.775</td>
<td>0.890</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>0.848 ± 0.059</td>
<td>0.833 ± 0.050</td>
<td></td>
</tr>
<tr>
<td>Radiology residents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.834</td>
<td>0.810</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.669</td>
<td>0.868</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.872</td>
<td>0.848</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.741</td>
<td>0.844</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.735</td>
<td>0.751</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>0.770 ± 0.081</td>
<td>0.788 ± 0.069</td>
<td></td>
</tr>
<tr>
<td>All radiologists</td>
<td>0.816 ± 0.077</td>
<td>0.843 ± 0.074</td>
<td></td>
</tr>
</tbody>
</table>

years). All board-certified radiologists specialized in body imaging and read chest radiographs on a regular basis. They were allowed to change the level and width of the window on the monitor; reading time was not limited.

All observers used a continuous rating scale and a line-marking method to rate their confidence level by placing marks on a 7-cm-long line on a recording form. The left end of the line indicated complete confidence that the chest radiographs without and with rib suppression showed no nodule; the right end indicated complete confidence that the images depicted a nodule. Intermediate levels of confidence were indicated by the position of the mark between the two line termini. Marks close to the right and left ends indicated greater and lesser degrees of confidence that a nodule was present. One author measured the distance between the left end of the line and the mark and converted the distance to an ordinal confidence rating ranging from 0–100. A continuous rating scale containing a pair of horizontal lines was used in the sequential test. Observers first recorded their rating of chest radiographs without rib suppression on the upper line. They then recorded their rating of rib-suppressed images on the lower line. They entered their results for each case on a record form.

To become familiar with the observer study, each observer first received training that involved reading images of five training cases not among the 60 cases used in the observer performance study. The observers read the training images in the 15 minutes immediately before the observer performance test. The five training cases consisted of two cases with and three without a nodule. The observers were instructed to use the rating scale consistently. Before training and test-taking, the participating radiologists were informed that the purpose of the experiment was to evaluate whether rib suppression did or did not enhance the detection of pulmonary nodules.

### Statistical Analysis

We used ROC analysis to compare the radiologists’ performance in detecting pulmonary nodules on chest images without and with rib suppression. A binormal ROC curve was fit to each radiologist’s confidence rating data acquired under the two reading conditions by use of quasi–maximum likelihood estimation [22]. A computer program (Rockit, Charles E. Metz) was used for obtaining binormal ROC curves from the ordinal-scale rating data [22]. The area under the best-fit ROC curve (AUC) plotted in the unit square was calculated for each fitted curve. The statistical significance of the difference in AUC values between the ROC curves obtained without and with rib suppression was tested for all board-certified radiologists and all radiology residents. The paired Student’s t test was performed with a statistical software package (SPSS version 15.0, SPSS), and \( p < 0.05 \) was considered to indicate a significant difference.

### Results

Six of the 30 nodules (20.0%) completely overlapped bone shadows of the ribs or a clavicle, 17 (56.7%) overlapped partially, and seven (23.3%) did not overlap on chest radiographs. The AUC values for all 12 observers were significantly higher with than without rib suppression (Table 1). Analysis of the overall performance of the 12 observers in the detection of pulmonary nodules (Fig. 2) indicated that the mean AUC values for all observers increased from 0.816 ± 0.077 without rib suppression to 0.843 ± 0.074 with rib suppression. The difference was statistically significant (\( p = 0.019 \)). For the board-certified radiologists (Table 1, Fig. 3), the mean AUC values obtained without and with rib suppression were 0.848 ± 0.059 and 0.883 ± 0.050. These values were significantly different (\( p = 0.011 \)). For the group of residents (Table 1, Fig. 4), however, the mean AUC values obtained without and with rib suppression were 0.770 ± 0.081 and 0.788 ± 0.069. This difference was not significant (\( p = 0.310 \)). Representative cases are shown in Figures 5 and 6.

### Discussion

Massive-training ANN is a nonlinear pattern recognition tool for suppression of rib opacity on chest radiographs while soft-tissue contrast is maintained [20]. We found that when board-certified radiologists evaluated a combination...
of rib-suppressed images and original chest radiographs, their diagnostic performance in the detection of pulmonary nodules improved to a statistically significant degree. In this study, we selected nodules smaller than 20 mm for the observer performance study. Our results suggested that the rib suppression technique helps skilled radiologists in the detection of relatively small pulmonary nodules.

In the group of radiology residents, improvement in diagnostic performance was not statistically significant. Because the radiology residents had only 2–3 years of experience, we presume that their interpretation skills were still developing; thus variations in their diagnostic performance were relatively large. A statistically significant difference was not found in this group, probably because the statistical power was not strong enough owing to such variations. We expect that a statistically significant difference would be found if the number of observers or cases were increased.

The poorer performance of the residents also may relate to the type of error found in the two observer groups. Experience with the interpretation of chest radiographs appeared to decrease the false-positive fraction but had less effect on reducing the false-negative fraction, suggesting that readers with less experience would report more false-positive findings. Use of a massive-training ANN lowers the false-negative rather than the false-positive rate by enhancing visualization of pulmonary nodules on soft-tissue-selective images. Therefore, a massive-training ANN may be more effective for experienced board-certified radiologists than for less experienced radiology residents.

Because of its low cost, simplicity, and low radiation dose, chest radiography continues to be the most widely used imaging technique for detection of chest disease. However, the false-negative rate for detection of pulmonary nodules is relatively high, reported to range from 12% to 90% [3–6]. Malignant lung tumors missed on chest radiographs reportedly share the following characteristics: they are visually subtle [8, 23], small [8], and located predominantly in the upper lobes [23]. Importantly, 82–95% of malignant lung tumors missed by radiologists were partly obscured by overlying bones such as the ribs and clavicle [5]. This finding suggests that suppression of these structures on chest radiographs may improve detection accuracy. In our database, more than 75% of the nodules overlapped with ribs or clavicles. In this study, however, we did not compare detection of nodules overlaid by bony structures with identification of nodules without bone overlay on rib-suppressed images because our series contained an insufficient number of nodules in each category.

Because the massive-training ANN suppresses bone attenuation, areas of calcification in a nodule or calcified nodules can become indistinct on rib-suppressed images. A solution consists in careful comparison of the original chest radiographs and the rib-suppressed images. Bone-selective imaging is a counterpart of rib-suppressed imaging, and the addition of bone-selective images to interpretation of chest radiographs may help to avoid missing nodules with calcifications.

The dual-energy subtraction technique of chest radiography can remove overlying bony structures to generate soft-tissue-selective images [10, 12], and it enhances visualization of pulmonary nodules overlaid by bones. Single-exposure [10, 11] and double-exposure [12, 24] dual-energy systems are available. Chest radiography with either system improves nodule visibility and has been found to improve accuracy of detection [24, 25]. Despite this advantage, radiography systems...
with dual-energy subtraction are used at few institutions because they require specialized equipment and because they may deliver a higher radiation dose than standard chest radiography does [24]. A major advantage of the massive-training ANN over dual-energy subtraction is that the former requires no specialized equipment for generating dual-energy x-ray exposures because only software is needed. There also is no increase in radiation dose exposure because chest radiographs acquired with standard radiographic systems are used [20].

In the temporal subtraction technique, a previously acquired image is subtracted from a current chest radiograph. Use of this method enhances detection of changes that occurred during the interval between imaging examinations [13]. This technique has effectively improved the performance of radiologists in detection of pulmonary nodules [13] and is commercially available in Japan [26]. It cannot be used, however, unless a previously acquired radiograph is available. In contrast, the massive-training ANN can be used with only one radiograph.

Studies have shown that use of CAD systems with automated detection methods for lung nodules significantly improves diagnostic accuracy in identification of pulmonary nodules on chest radiographs. Kakeda et al. [27], who applied CAD to chest radiographs of 45 patients with solitary pulmonary nodules measuring up to 25 mm in diameter, reported a sensitivity of 73% with 3.15 false-positive findings per image. They also found that 75% of false-positive findings were attributable to the presence of normal anatomic structures such as ribs and clavicles. In a more recent study, Li et al. [23] also applied a commercially available CAD tool to chest radiographs. They reported an average of 5.9 false-positive findings per image and a 96% false-positive rate attributable to interference by anatomic structures. We propose that adding CAD to a massive-training ANN may have greater sensitivity and lead to fewer false-positive results than use of either method alone. Studies are under way in our laboratory to investigate whether the combined use of CAD and a massive-training ANN improves the detection of pulmonary nodules.

According to Ide et al. [10], who evaluated the detectability of lung cancer on chest radiographs subjected to single-exposure dual-energy subtraction, the method did not improve the detectability of non-solid nodules, also known as pure ground-glass-opacity nodules. These nodules have a high probability of being lung cancer, especially bronchioloalveolar carcinoma [28]. Because it is based on a dual-energy subtraction technique, a massive-training ANN also may not be useful for detection of non-solid nodules. Studies are under way to examine the detectability of non-solid nodules with a massive-training ANN.

There were several limitations to our study. First, we extracted chest radiographs from the Japanese Standard Digital Image Database [21]. Because images in that database were derived from many institutions and many were digitized from conventional chest radiographs, image quality varied from image to image. We are investigating nodule detectability on radiographs obtained with more recent digital radiographic systems. Second, central hilar nodules were not included because detection of those nodules remains more difficult than that of peripheral nodules on standard chest radiographs. In future studies we will address the effect of our technique on the detection of central peripheral nodules. Third, we did not compare the effectiveness of the massive-training ANN technique with that of the clinically used dual-energy subtraction technique, and such an investigation is needed.

We conclude that the massive-training ANN technique of rib suppression in combination with inspection of original chest radiographs significantly improved the diagnostic performance of radiologists in the detection of small pulmonary nodules. We strongly suggest that use of this technique has potential for reducing the number of lung nodules missed in routine clinical practice.

References
thin-section CT. *Radiology* 2002; 224:139–144