Application of Artificial Neural Networks to Medical Image Pattern Recognition: Detection of Clustered Microcalcifications on Mammograms and Lung Cancer on Chest Radiographs

SHIH-CHUNG B. LO, JYH-SHYAN J. LIN*, MATTHEW T. FREEDMAN AND SEONG K. MUN
ISIS Center, Department of Radiology, Georgetown University Medical Center, 2115 Wisconsin Ave., Suite 603, Washington, DC 20007

Abstract. Three neural network models were employed to evaluate their performances in the recognition of medical image patterns associated with lung cancer and breast cancer in radiography. The first method was a pattern match neural network. The second was a conventional backpropagation neural network. The third method was a backpropagation trained neocognitron in which the signal propagation is operated with the convolution calculation from one layer to the next. In the convolution neural network (CNN) experiment, several output association methods and trainer imposed driving functions in conjunction with the convolution neural network are proposed for general medical image pattern recognition. An unconventional method of applying rotation and shift invariance is also used to enhance the performance of the neural nets.

We have tested these methods for the detection of microcalcifications on mammograms and lung nodules on chest radiographs. Pre-scan methods were previously described in our early publications. The artificial neural networks act as final detection classifiers to determine if a disease pattern is presented on the suspected image area. We found that the convolution neural network, which internally performs feature extraction and classification, achieves the best performance among the three neural network models. These results show that some processing associated with disease feature extraction is a necessary step before a classifier can make an accurate determination.

1. Introduction

Clinical studies in the use of chest radiographs for the detection of lung nodules including those reported by Stitik [1] and Heelan [2] have demonstrated that even highly skilled and highly motivated radiologists, task-directed to detect any finding of suspicion for a pulmonary nodule, and working with high quality chest radiographs, still fail to detect more than 30 percent of the lung cancers that can be detected retrospectively. In the series reported by Stitik, many of the missed lesions would be classified as T1NxMx lesions, the stage of non-small cell lung cancer that C. Mountain indicates has the best prognosis (42%, 5 year survival) [3]. This is the stage (nodules 0.3–2 cm in diameter, separate

Current address: Deus Technologies, Inc., Rockville, MD.

from the hilum) of lung cancer that a computer-assisted diagnostic program should tackle. Figure 1 shows a chest radiograph containing a nodule overlapped by a rib. This is a rather typical case, because 40% of the lungs are covered by posterior ribs or rib crossings.

Although mammography has a high sensitivity for detection of breast cancers when compared to other diagnostic modalities, studies indicate that radiologists do not detect all carcinomas that are visible in retrospective analyses of the images [4–6]. These missed detections are often a result of the very subtle nature of the radiographic findings. However, many missed diagnoses can be attributed to human factors such as subjective or varying decision criteria, distraction by other image features, or simple oversight [7, 8]. Early breast cancers are often characterized by masses and clustered microcalcifications [9]. It has been reported that

Figure 1.   A chest radiograph showing a nodule overlapped on a rib.

Figure 2.   A mammogram showing clustered microcalcifications. The cluster area in original size is enhanced by a local histogram equalization process for display purposes.

between 40% and 50% of breast carcinomas detected radiographically demonstrate masses on mammograms [10, 11]; 30–50% of breast carcinomas presented as microcalcifications, and 60–80% of breast carcinomas reveal microcalcifications upon histologic examinations [11–13]. Breast cancer patterns associated with masses will be discussed in our future papers. Breast cancer associated clustered microcalcifications are one of two disease objects studied in this paper. Typically, the sizes of microcalcifications vary from 0.16 mm to 1.0 mm. Figure 2 shows a mammogram containing clustered microcalcifications which are surrounded by dense glandular tissues.

Various computer-based image perception techniques have been proposed for the detection of disease patterns [14, 15]. With each of these methods there is a trade-off between increased sensitivity and decreased specificity. In general, by setting less stringent criteria on computer algorithms, the sensitivity of the detecting programs can be increased. However, when using any of these methods to detect subtle diseases, we must use additional methods to decrease the number of false positives. For this reason, several investigators have attempted to use various advanced image processing and artificial classifiers to improve disease detection [16–18].
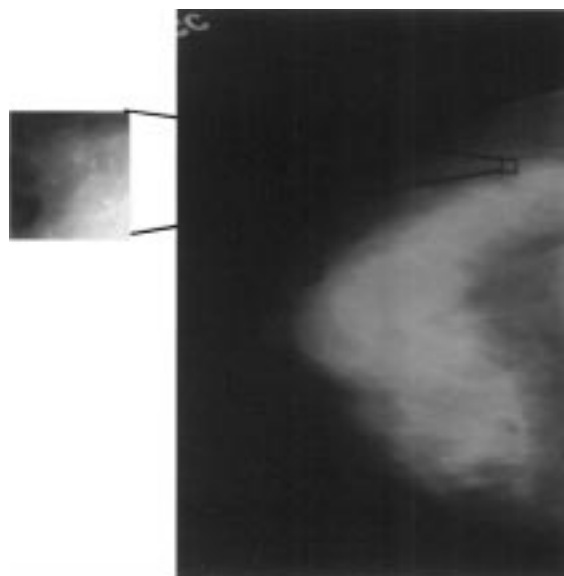
Many artificial neural network models have recently been applied to diagnostic imaging research [19, 20]. The main tasks of these research efforts are aimed at assisting radiologists either in the accuracy improvement of quantitative measures or in the improvement of sensitivity and specificity for a disease detection. In diagnostic imaging, the neural network techniques incorporated with image processing methods have become a major research trend in the field of computer-aided diagnosis. Medical diagnoses involve very sophisticated decision-making processes. We, therefore, limited our studies to the recognition of specific disease patterns. In this paper, we will also discuss characteristics of some disease patterns in clinical images and their implications on the neural network classifications.

2.   Materials and Research Objectives

2.1.   Disease Patterns on Projection X-Ray Images

Projection radiographs shown on films are generated by the transmission of X-ray beams through a patient. The resulting X-rays, of varying intensity, form a radiographic image. For many years, this technique has been used as a diagnostic procedure for screening or primary examination of a disease associated with physical tissue

changes. The major drawback of projection radiography is that X-ray beams project the original anatomical three-dimensional objects onto a two-dimensional image. In other words, each pixel intensity on the image represents a total X-ray attenuation integrated from a line passing through the patient. Bone and soft tissue, and abnormal changes of tissue can be distinguished from one another in an X-ray image because they attenuate X-rays differently. However, subtle abnormalities superimposed on various normal tissues and bones are difficult to discern. The degree of sophistication in recognition of disease patterns in these images, which requires professional training, differs significantly from that of the character recognition or other image pattern recognition. The degree of difficulty is not easy to measure. Qualitatively speaking, the ratio of signal and structure noise in the task of disease pattern recognition can be very small. Consider a local suspected area that may or may not contain a disease pattern, $s(x, y) \approx d(x, y) \in P_d$, where $d(x, y)$ represents an image patch that has been proven to be a disease pattern. The collected set of these proven patches is called P_d. This local area often contains some background information resulting from normal tissues, $b(x, y) \in B$. The total intensity function denoted as $f(x, y)$ is given by

$$f(x, y) = s(x, y) + b(x, y). \tag{1}$$

In general, four situations are possible in a suspected area:

(a) $s(x, y) \gg b(x, y)$ (i.e., high signal to background ratio) representing obvious true cases;
(b) $s(x, y) \ll b(x, y)$ representing subtle cases,
(c) $s(x, y) = 0$ and $b(x, y)$ is similar to one of $d(x, y)$, where $d \in P_d$, and
(d) $s(x, y) = 0$ and $b(x, y)$ is not similar to any disease patterns, representing obvious false cases.

Most cases falling in situation (b) result in true-negatives. Cases associated with situation (c) may produce a false-positive by a classifier.

Pattern match and backpropagation, two commonly used pattern classifiers, were employed to compare the performance in the detection of clustered microcalcifications selected from mammograms and the detection of lung nodules extracted from chest radiographs. Regions of interest (ROI), formatted at $32 \times 32 \times 12$ bit, normal or abnormal, were extracted by the corresponding methods previously described [16, 17]. Both geometrical pattern and relative intensity of a local area

on a radiographic image are important information in a radiographic reading. The background trend of each ROI was removed to eliminate low frequency variation [16]. However, the background structures (i.e., radiographic image of bone on chest image, vessels, and large soft tissue differences) remained in each ROI. No normalization procedure was taken, because normalization can mix a disease pattern with a non-disease pattern. For example, (a) small nodules and end-on vessels and (b) microcalcifications and film defects basically differ only in contrast. They would not be distinguishable if the feature of contrast is normalized in the pre-processing. Since many disease patterns are superimposed on background structures, we have not experienced a successful unsupervised training technique with our database. Three supervised training methods, however, achieved some success and are discussed in the following sections.

2.2. Disease Pattern Characteristics of Microcalcifications on Mammograms and Lung Nodules on Chest Radiographs

In general, the larger the nodule the higher the contrast of the nodule profile on the radiograph. Small rounded objects possessing high contrast are most likely end-on vessels. In addition, the size of end-on vessels is inversely proportional to their distance from the center of the heart. This is because anatomical distribution of larger arteries and vessels are closer to the heart. Clinical instruction indicates that faint tails of the vessel turned in a horizontal direction may be observable. A rib crossing, which sometimes look like an opaque round object, can also produce a false-positive detection. See Figure 3 for examples of end-on vessels, rib crossings, and true nodules.

On the other hand, the gray value differences (i.e., contrast) between the peak of microcalcifications and local background tissue are somewhat proportional to the size of the calcifications on mammograms. Film defects, caused by scratches of screen/film system or cold spots of film emulsion, are high contrast bright spots. The contrast of film defects is independent of size. Several image blocks shown in Figure 4 demonstrate the difference between microcalcifications and film defects. All image blocks were randomly selected from our database and processed by a histogram expansion for display purposes. It is essential to use a sufficiently small digitization to preserve the disease pattern. Potential problems of using a large digitization spot for
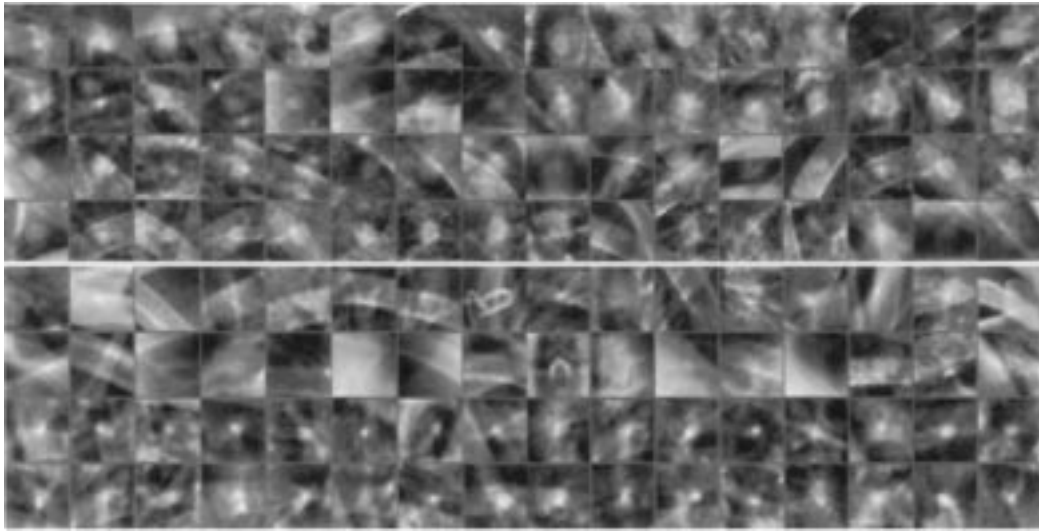
Figure 3.   The upper 4 rows show 64 nodule blocks sampled from the database. Each image block on rows 5 and 6 contain no nodule but a lung or rib structure. Each image block on the bottom two rows contains an end-on vessel.
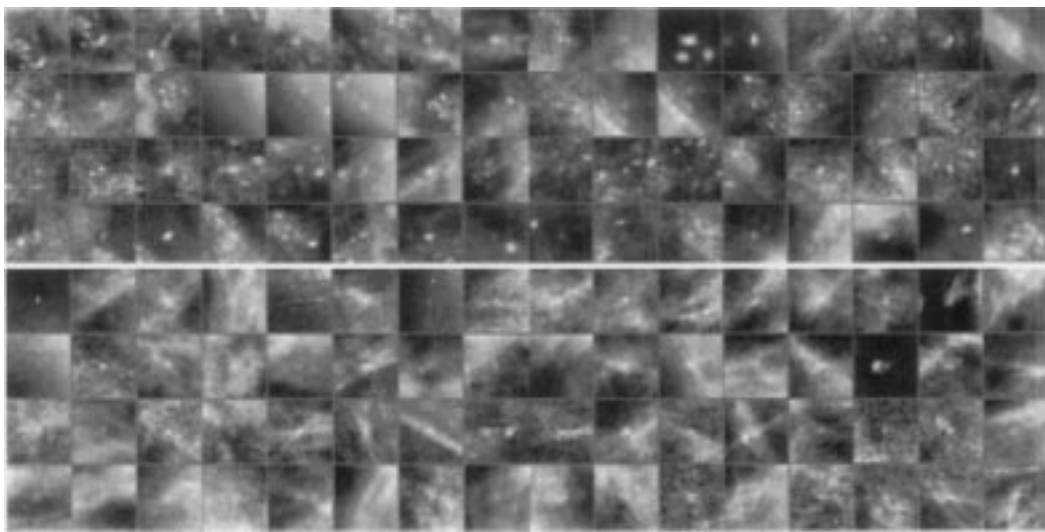

Figure 4.   Each image block, extracted from mammogram, on the upper 4 rows contains at least a calcification. Each image block on the bottom 4 rows contains at least a local maximum value of gray scale (bright spot) which is not a calcification. Each block, at matrix elements (1, 4), (5, 4), (7, 4), (9, 4), and (2, 6) contains a bright spot due to a film defect.

acquiring mammographic images are: (a) the edge of a small film defect can be blurred and (b) very small microcalcifications are not actually digitized. These problems are less pronounced with a digitization spot size of 0.1 mm which was the specification of the Lumysis laser film scanner (Lumiscan Model 150).

Chest images were digitized and reformatted (shrunk by using pixel averaging) with a matrix size of $512 \times 625 \times 12$ bits per image and each pixel rep-resents a 0.7 mm $\times$ 0.7 mm square area. Mammograms were digitized with a computer format of $2048 \times 2500 \times 12$ bits per image and each pixel represents 0.1 mm $\times$ 0.1 mm square area. The suspected microcalcification patches shown in Figure 4 are for display purposes. In the study of microcalcification detection, only the central region of 16×16 pixels (i.e., 1.6 mm $\times$ 1.6 mm) was used as input for the performance evaluation of the three neural network systems.

3. Comparative Studies Using Neural Networks

3.1. Associated Memory Based Pattern Match Neural Networks for Disease Detection

A classifier takes a feature vector and produces a classification. The core portion of the pattern match classifier searches for the most similar pattern in the memory. If no pattern match is found in the memory, a new pattern is created and stored for that particular classification in the memory during the training. Several neural networks belong to this type of pattern match: (a) adaptive resonance theory (ART) and its extensions (i.e., ART-2 [21], ARTMAP [22], etc.), (b) category learning originated by Reilly et al., 1992, known as RCE method [23], and (c) Dynamic Stable Associate Learning (DYSTAL) [24–26].

We used the processed image block (i.e., patch) as the input feature vector. Many feature vectors of this kind may be "contaminated" by original background structures, which are difficult to discern as disease patterns or background patterns. The authors are aware that it is important to extract features representing various aspects of disease patterns prior to the classification task. However, our goal was to compare which method better distinguishes disease patterns from non-disease image patterns using the image patches as input data. Since DYSTAL was originally designed to use image data as input for a classification task, it was selected as one of the methods for the study.

In DYSTAL, there are three rules for aggregating the input feature vector and propagating the signals:

(a) the aggregation rules are based on the correlation between the input feature vectors and learned patterns (the correlation measures the similarity between the inputs and learned patterns),
(b) the propagation rule depends on the maximum number of these resulting similarity values, and
(c) the learning rule permits the system to maintain learning patterns as needed.

The similarity measure is defined as the correlation of a learning pattern and the input feature vector [26]

$$S^j = CC(P^j, I)$$
$$= \frac{\sum_i \left(P_i^j - \bar{P}^j\right) \times (I_i - \bar{I})}{\sqrt{\left(\sum_i \left(P_i^j - \bar{P}^j\right)^2 \times \sum_i (I_i - \bar{I})^2\right)}}, \quad (2)$$

where P_i^j is the value of the ith element of the jth patch vector, $\bar{P}^j$ is the mean value of the elements of the patch P^j, I_i is the value of the ith element of the input feature vector, and $\bar{I}$ is the mean value of the elements of the inputs. This similarity measure uses the cosine of the angle between the two vectors I and P^j in the n dimensional hyper-space, where n is also the number of patch elements.

The DYSTAL also uses the winner-take-all approach of propagating maximum simliarity. If the maximum similarity is lower than a pre-defined value, the new feature vector, will be stored as a newly learned pattern in the memory. The learned pattern is then assigned to an associated class which is either a true or a false disease case.

3.2. Convolution Neural Network for Disease Pattern Recognition

The connection between nodes in the conventional backpropagation neural network (BPNN) uniformly spreads from a front layer to a back layer [27]. However, it is known that the neighborhood correlation is usually higher than that of the long distance correlation between two pixels on an image. It is conceivable that features associated with nearby pixels should be emphasized. In neural network terms, the local signal interactions rather than non-local interactions shall be established to instruct the neural network learning. A convolution neural network (CNN), whose nets are locally formed, is selected as one of the classification methods in the experiment. The structure of the CNN is a simplified version of the neocognitron [28, 29]. We used only a 2 hidden-layer structure and eliminated all the complex-cell layers. Nets between two adjacent layers were selectively interconnected across groups. We modified the neocognitron network structure and used a convolution constrained backpropagation method for the training. This modification is necessary because (a) the original neocognitron is designed for a binary image, (b) the original 9 hidden-layer structure is very computationally intensive for an iterative training method such as the BPNN, (c) a one or two hidden-layer structure is considered adequate for relatively simple image patterns such as lung nodules and microcalcifications. Figure 5 shows the fundamental structure of this neural network.

In the CNN signal processing, each group in the receiving layer gets signals from a group of weights (e.g., kernels). For the forward signal propagation, the resultant of the weighting factors of the kernel convoluting the element values of the front layer is collected
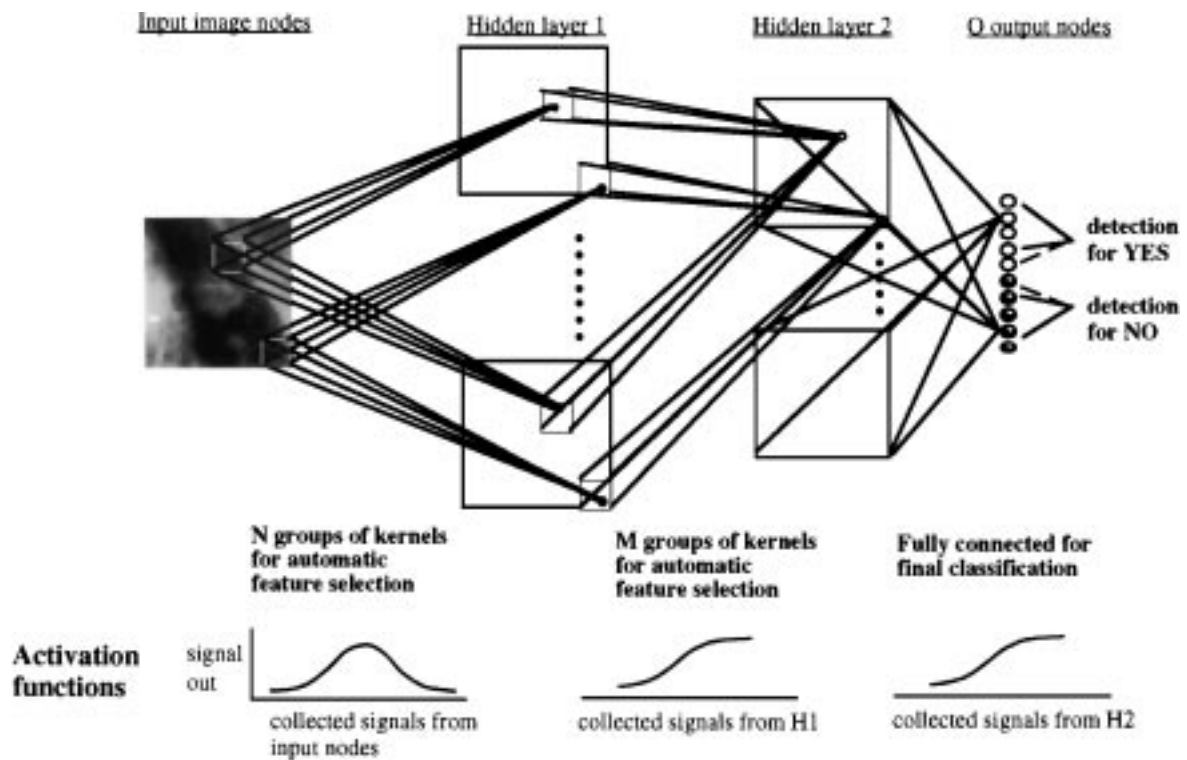
Figure 5.   Artificial convolution neural network for disease pattern recognition.

onto the corresponding matrix elements of the receiving layer. This operation accounts for the major difference between the convolution type neural network and regular fully connected neural network. In the lung nodule study, we used an image patch size of 32×32 (i.e., $21.4 \, \text{mm} \times 21.4 \, \text{mm}$) with a convolution kernel size of 7×7. In the study of microcalcification detection, the central region of 16×16 pixels (i.e., $1.6 \, \text{mm} \times 1.6 \, \text{mm}$) of the original image patch size of 32×32 with a convolution kernel of 5×5 was used. The choices of using 7×7 and 5×5 convolution kernels were based on extensive studies [30, 31] in lung nodule and microcalcification cases, respectively. One reason for using much smaller size kernel is that microcalcifications are very tiny compared to observable lung nodules. In addition, small kernels are appropriate for small objects for evaluating the difference between true and false microcalcifications. Each hidden layer consists of 10 groups. The output layer has 10 nodes (2 categories) which were fully connected to the second hidden layer.

3.3.   Training of Neural Networks

3.3.1. Classification Invariance of Matrix Operations.
In general, medical image patterns possess either a circular symmetric shape (e.g., nodules) or appear as small objects with a variety of geometric patterns (e.g., calcifications). In such cases, image pattern recognition does not call on top-down or left-right geometry as classification criteria. Therefore, we can take advantage of this characteristic as an invariance. In other words, we can rotate and/or shift the input vector two-dimensionally and maintain the same output assignments for the training. This method may have two effects on the neural network: (i) to instruct the neural network that the rotation and shift of the input vector would receive the same classification result; and (ii) to increase the total number of training samples which is expected to enhance the performance of the neural network. We only rotated each suspected image block 8 times for input to test our hypothesis. Four of the rotations are: $0°$, $90°$, $180°$, $270°$. In addition, we also flipped over (left-right) the original image matrix and used the same rotations again to obtain 4 additional rotations.

3.3.2. Modification of Backpropagation Training for the CNN.
As indicated in Section 2.2, a high signal of a feature can result from a negative object such as higher contrasts in end-on vessels than those in nodules
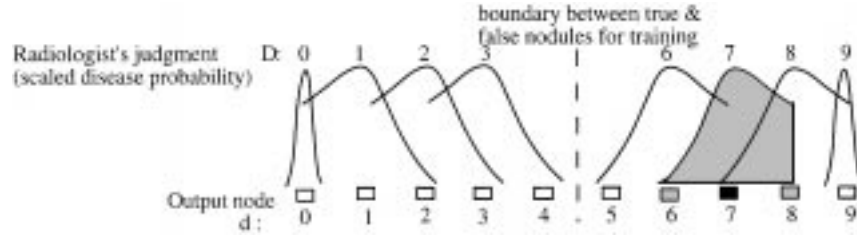
Figure 6. Fuzzy output association is constructed by a Gaussian and a trainer imposed repulsive function. Note that only one curve is used for a training case associated with an output target node (e.g., node 7 represents the activated fuzzy function).

and higher peak values in film defects than those of microcalcifications. Therefore, we used a Gaussian-like activation function for the cumulated signal propagation between input layer and the first hidden layer. The purpose of this activation function is to treat both low and high cumulated signals as false features that would eventually facilitate the classification process in the following layers. This Gaussian-like activation function would also be appropriate for the BPNN using an image block as the vector described in Section 3.5. In the conventional BPNN, fully connected rather than locally connected networks were implemented.

We used the sigmoid activation function for the forward signal propagation for all layers other than the first hidden layer and applied backpropagation training for the adjustment of weights between any two adjacent layers. The main difference between conventional weights and kernel weights is that the former are independent and the latter are constrained by grouping. By looking at the CNN processing, one may find that signals are filtered and modulated as in a circuit system. Signal propagation from one layer to the next is composed of: (a) an adaptive convolution combiner and (b) activation functions (Gaussian-like—Eq. (3)—and sigmoid—Eq. (4) functions for the first hidden layer and for other layers, respectively. See Fig. 5) which are given below:

$$S_x(i, j; n)$$
$$= \frac{4 \times \exp\left\{-\sum_{u,v,m \Leftrightarrow n}[k_x(u, v; n) \times S_{x-1}(i-u, j-v; m)]\right\}}{1 + \exp\left\{-\sum_{u,v,m \Leftrightarrow n}[k_x(u, v; n) \times S_{x-1}(i-u, j-v; m)]\right\}} \quad (3)$$

and

$$S_x(i, j; n)$$
$$= \frac{1}{1 + \exp\left\{-\sum_{u,v,m \Leftrightarrow n}[k_x(u, v; n) \times S_{x-1}(i-u, j-v; m)]\right\}} \quad (4)$$

where $S_x((i, j); n)$ represents the signal at node (i, j), nth group, and x layer. $k_x((u, v); n)$ denotes a weight-

ing factor value at net (u, v), nth group, and connecting from $x - 1$ to x layer. $m \Leftrightarrow n$ represents those in group m that connect to group n.

3.3.3. Backpropagation Neural Network Trained by Radiologists. We modeled radiologists' diagnostic rating (i.e., the probability of a disease existing in a suspected area) and incorporated it into the neural network training. In fact, when a radiologist determines a specific probability of a disease pattern in an image area based on his/her training and experience, this probability would be accompanied with a variation (or a standard deviation). An asymmetric output association distribution is shown in Figure 6. The use of asymmetric fuzzy assignment attempted to direct non-disease cases toward low value nodes and to push disease cases toward high value nodes. With this fuzzy assignment for the output nodes in the training, the relation between adjacent nodes was established. This supervised training can be generally applied to any situation where an association of outputs is necessary.

3.4. Classification of Output Values in the Testing

Corresponding to the grading system arranged in the training, a polarized (linearly weighted) function is given as an indication. In practice, we can define a normalized disease detection index (NDDI) for the judgment of a suspected area:

$$\text{NDDI} = \frac{\sum_{n \in \text{true nodes}}\left[O_n \times \left(n - n_0 + \frac{1}{2}\right)\right]}{\sum_{n=0}^{N-1}[O_n] \times \frac{N-1}{2}}, \quad (5)$$

where n denotes the node in the output layer, n_0 is the node number of the least likely true node, O_n is the output value at node n, and N is the total number of output nodes. Hence, a nodule detection index of 0 indicates a definite non-nodule and a nodule detection index of 1 or greater implies a definite nodule case determined by the neural network. The calculated
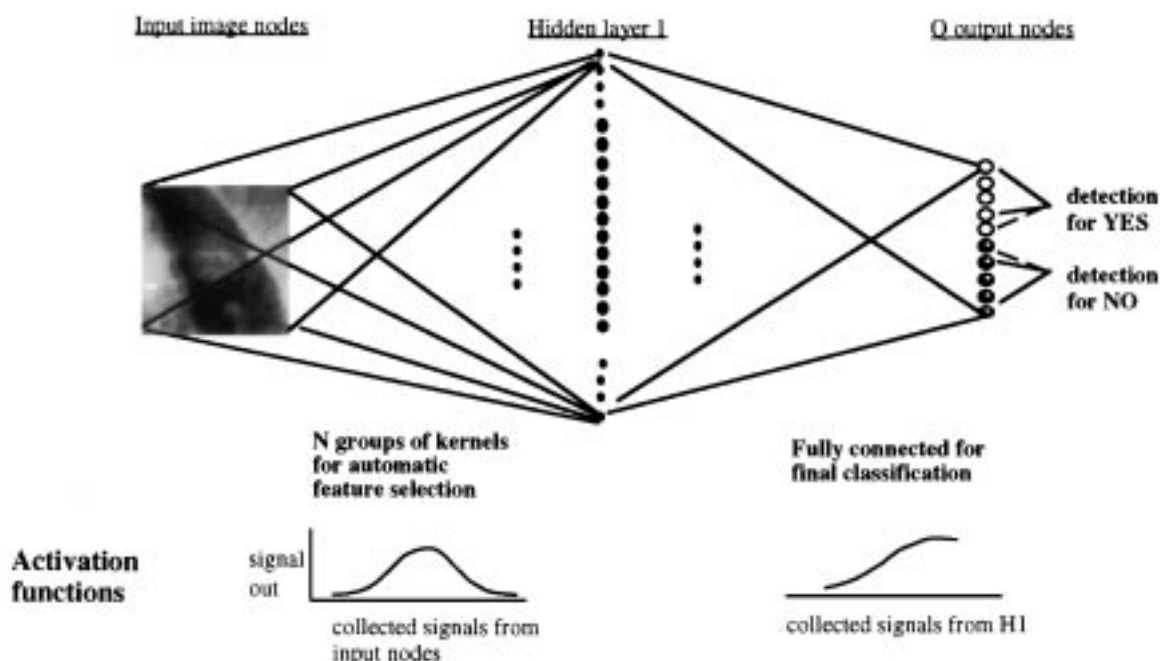
Figure 7.   Artificial backpropagation neural network with fully connected nodes for disease pattern recognition.

NDDIs were evaluated by the receiver operating characteristic (ROC) analysis to measure the performance of the neural network. In general, A_z, representing the area underneath the ROC curve, is an index which signifies the performance of a system. The ROC curve is formed with the true-positive rates versus the false-positive rates of a system. We also used a performance measure—relative detection accuracy—converted from the curve to compare the results of the neural network systems.

3.5.   Backpropagation Neural Network Technique for Disease Pattern Recognition

We have also investigated the performance of the conventional backpropagation (BP) neural network with (BP/1H) and without (BP/0H) a hidden layer. In other words, the background reduced image pixel values were used as input signals for the input layer. We expected that the hidden layer would serve as a feature extractor. The same training and testing data sets, which again were "contaminated" (i.e., very strong background sructures such as rib ovelapping a nodule) and used in the pattern match neural network, were entered into the BP neural network. Basically, we set up the experiment as described in Section 3.2 with one exception: the fully connected neural nets are used to

compare the effectiveness of the neural network architecture designs. The structure of the BPNN with one hidden layer is shown in Figure 7. We tried several arrangements for the number of nodes used in the hidden layer. Our experiment indicated that approximately 200 nodes and 450 nodes in the hidden layer would be appropriate for nodule and microcalcifcation studies, respectively. These numbers may be altered according to the size of the image block used.

4.   Results

4.1.   Detection of Clustered Microcalcifications

After the pre-scan process by the computer program, 38 digital mammograms provided 220 true and 1132 false subtle microcalcifications. For the neural network studies, we divided the mammograms into two sets: (A_m) 19 images (containing 108 true and 583 false image blocks) and (B_m) another set of 19 images (containing 112 true and 549 false image blocks). We did not ask radiologists to rate image blocks in the training set. Therefore, only 2 output nodes with 8 rotated input patches were used. Neither output association nor a trainer imposed function was employed. We found that the use of a small image block of 16×16 resulted in

Table 1. Performance of neural networks in the detection of clustered microcalcifications using group A_m as training set and group B_m as testing set.

Neural networks	DYSTAL	BP/0H	BP/1H	CNN
A_z (Area under the ROC curve)	0.78	0.75	0.86	0.97
Detection accuracy				
(% true-positive detection)	70	70	75	90
(# false-positive per image)	4.3	4.5	3.5	0.5

Table 2. Performance of neural networks in the detection of clustered microcalcifications using group B_m as training set and group A_m as testing set.

Neural networks	DYSTAL	BP/0H	BP/1H	CNN
A_z (Area under the ROC curve)	0.76	0.77	0.84	0.97
Detection accuracy				
(% true-positive detection)	70	70	75	90
(# false-positive per image)	4.3	4.2	3.7	0.5

the best performance in the detection of single microcalcification [31].

Tables 1 and 2 show the performance resulting from the three neural network systems. DYSTAL and BP/0H, acting as classifiers, receive the lowest performance. The best performance index (A_z) was 0.90 when the determination was based on individual microcalcifications and was improved to 0.97 when the determination was based on the clustered microcalcifications using CNN. In the latter evaluation, suspected clusters including 1 or 2 spots within a 1 cm area were

rejected and the average NDDI taken from the clustered spots was used for the ROC evaluation. This is because the detection of clustered microcalcifications is more clinically significant than individual calcifications, since the clustered microcalcifications (3 or more) are a strong indication of breast carcinoma in radiological diagnosis. The comparative study was based on detection strategies: (i) to first detect suspected individual microcalcifications and (ii) then to cluster them as group when possible; otherwise rejected the detection.

4.2. Detection of Lung Nodules

The first group (A_1) of image blocks were extracted from 31 chest radiographs containing multiple nodules. A senior radiologist selected 91 true nodules and 247 non-nodules areas. The second group (B_1) was collected from 31 images containing 95 nodules and 258 non-nodules and was confirmed by biopsy or by follow-up showing growth of the nodule. The pre-scan process was performed first to locate the center of the high intensity island and isolate the image block for training. For the training, each original and its 7 "brother" image blocks shared the same score vector (probability of a disease and output association) pre-determined by the radiologist. During the training, the original and its 7 "brother" image blocks were entered as a group in the same sequence. Tables 3 and 4 show the performance of using different neural network techniques and corresponding enhancement methods (i.e., fuzzy output training).

Table 3. Performance of neural networks in the detection of lung nodules using group A_1 as training set and group B_1 as testing set.

Neural networks	DYSTAL	BP/0H	BP/1H	CNN	CNN/FUZZY
A_z (Area under the ROC curve)	0.56	0.58	0.68	0.82	0.89
Detection accuracy					
(% true-positive detection)	60	60	70	80	80
(# false-positive per image)	7	6.6	5	4	2.5

Table 4. Performance of neural networks in the detection of lung nodules using group B_1 as training set and group A_1 as testing set.

Neural networks	DYSTAL	BP/0H	BP/1H	CNN	CNN/FUZZY
A_z (Area under the ROC curve)	0.57	0.61	0.70	0.83	0.88
Detection accuracy					
(% true-positive detection)	60	65	70	80	80
(# false-positive per image)	7	6.5	4.8	4	2.5

These comparison studies of both diseases imply that pattern classifiers such as DYSTAL and BP/0H cannot function alone to analyze image blocks (patches) with substantial background structures. Once the feature extraction procedure was added, the performance of the neural network increased as evident in the results of BP/1H, CNN, and CNN/FUZZY in Tables 1–4. We also learned that the convolution for two-dimensional feature extraction and fuzzy training guided by radiologists' determination were successful methods to improve the disease detection. With the neural network used in these studies, we could not isolate which procedure, the feature extraction or the final classification, was improved by the CNN training.

5. Discussion and Conclusions

Medical image pattern recognition using extracted features for input has been proposed in the detection of disease patterns [20]. Since only a small number of inputs are used, the computational time can be much less than that of the CNN for the training. As long as the features of a disease pattern are well-defined and can be quantified as values or vectors, many neural network techniques should be able to classify them. On the other hand, the CNN can internally extract features of disease patterns and is capable of distinguishing non-disease from disease patterns. A potential advantage of using the CNN is that once trained kernels are analyzed, feature extraction can be specifically defined not only by the users' experience but also by the confirmation of the CNN.

In this study, we have utilized the CNN in conjunction with several effective training methods: (i) providing a radiologists' rating scale for the training of neural nets, (ii) introducing the neural network with the classification invariance of input matrix operations, (iii) using output association functions to fuzzify the radiologists' determination and to establish the relationship between adjacent output nodes, and (iv) rendering trainer imposed functions to enhance the performance of the neural network. We found that the performance of the CNN in detecting both diseases improved significantly by administering these training methods.

Considering the convolution operation as a feature extraction processing from the input layer to the hidden layer in the CNN, we found that feature extraction is an important procedure to assist the classifier (e.g., conventional BP) in performing the recognition task. Pattern classifiers, including those newly developed neural

networks, would not be able to distinguish "highly contaminated" feature vectors. Approximately 40% of lung nodules are superimposed on a posterior rib with various of orientations. However, less than 10% of microcalcifications in our database are obstructed by other abrupt breast tissues. In addition, the probability of having end-on vessels, which resemble nodules, on a chest radiograph is higher than that of film defects, which resemble microcalcifications. In other words, chest radiographs contain much more background structures than mammograms do. These background structures will contaminate the feature vector and lead to degradation of the machine observers' performance. Clinical studies also indicated that highly experienced human observers can detect only 68% of lung nodules [1] and 95% of clustered microcalcifications [5].

References

1. F.P. Stitik, M.S. Tockman, and N.F. Khouri, "Chest radiology," in *Screening for Cancer*, A.B. Miller (Ed.), Academic Press, New York, pp. 163–191, 1985.
2. R.T. Heelan, B.J. Flechinger, M.R. Melamed et al., "Non small cell lung cancer: Results of the New York screening program," *Radiology*, Vol. 151, pp. 289–293, 1984.
3. C.F. Montain, "Value of the new TNM staging system for lung cancer," *5th World Conference on Lung Cancer*, CHEST, Vol. 96, No. 1, pp. 47s–49s, 1989.
4. P.J. Haug, I.M. Tocino, P.D. Clayton, and T.L. Bair, "Automated management of screening and diagnostic mammography," *Radiology*, Vol. 164, p. 747, 1987.
5. C.J. Baines, A.B. Miller, C. Wall et al., "Sensitivity and specificity of first screen mammography in the Canadian national breast screening study: A preliminary report from five centers," *Radiology*, Vol. 160, p. 295, 1986.

6. L.W. Bassett, D.H. Bunnell, R. Jahanshahi, R.H. Gold, R.D. Arndt, and J. Linsman, "Breast cancer detection: One versus two views," *Radiology*, Vol. 165, p. 95, 1987.

7. J.E. Martin, M. Moskowitz, and J.R. Milbrath, "Breast cancer missed by mammography," *AJR*, Vol. 132, p. 737, 1979.

8. L. Kalisher, "Factors influencing false negative rates in xero-mammography," *Radiology*, Vol. 133, p. 297, 1979.

9. L. Tabar and P.B. Dean, *Teaching Atlas of Mammography*, 2nd edition, Thieme, NY, 1985.

10. F.M. Hall, J.M. Storella, D.Z. Silverstone, and G. Wyshak, "Nonpalpable breast lesions: Recommendations for biopsy based on suspicion of carcinoma at mammography," *Radiology*, Vol. 167, p. 353, 1988.

11. E.A. Sickles, "Mammographic detectability of breast microcalcifications," *AJR*, Vol. 139, p. 913, 1982.

12. E.A. Sickles, "Mammographic features of 300 consecutive nonpalpable breast cancers," *AJR*, Vol. 146, p. 661, 1986.

13. W.A. Murphy and K. DeSchryver-Kecskemeti, "Isolated clustered microcalcifications in the breast: Radiologic-pathologic correlation," *Radiology*, Vol. 127, p. 335, 1978.

14. K. Doi, "Feasibility of computer-aided diagnosis in digital radiography," *Japanese Journal of Radiological Technology*, Vol. 45, pp. 653–663, 1989.

15. K. Doi, M.L. Giger, H. MacMahon et al., Potential usefulness of real-time computer output to radiologists' interpretations. Scientific Exhibit, Space 10-001, Presented at Raiological Society of North America, 1992 Chicago Ill.

16. H.P. Chan, K. Doi, and S. Galhotra, "Image feature analysis and computer-aided diagnosis in digital radiography: 1. Automated detection of microcalcifications in mammography," *Medical Physics*, Vol. 14, pp. 538–548, 1987.

17. M.L. Giger, K. Doi, and H. MacMahon, "Image feature analysis and computer-aided diagnosis in digital radiography: 3. Automated detection of nodules in peripheral lung field," *Medical Physics*, Vol. 15, pp. 158–166, 1988.

18. M.L. Giger, N. Ahn, K. Doi, H. MacMahon, and C.E. Metz, "Computerized detection of pulmonary nodules in digital chest images: Use of morphological filters in reducing false-positive detections," *Medical Physics*, Vol. 17, pp. 861–865, 1990.

19. S.-C.B. Lo, M.T. Freedman, J. Lin, and S.K. Mun, "Automatic lung nodule detection using profile matching and back-propagation neural network techniques," *J. Digital Imaging*, Vol. 6, No. 1, pp. 48–54, 1993.

20. Y. Wu, K. Doi, M.L. Giger, and R.M. Nishikawa, "Computerized detection of clustered microcalcification in digital mammograms: Applications of artificial neural networks," *Medical Physics*, Vol. 19, pp. 555–560, 1992.

21. G.A. Carpenter, S. Grossberg, and J.H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Networks*, Vol. 4, pp. 565–588, 1991.

22. G.A. Carpenter, S. Grossberg, and D.B. Rosen, "ART-2 A: An adaptive resonance algorithm for rapid category learning and recognition," *Neural Networks*, Vol. 4, pp. 565–588, 1991.

23. D.L. Reilly, L.N. Cooper, and C. Elbaum, "A neural model for category learning," *Biol. Cybern.*, Vol. 45, pp. 35–41, 1982.

24. D.L. Alkon, K.T. Blackwell, G.S. Barbor, A.K. Rigle, and T.P. Vogl, "Pattern-recognition by an artificial network derived from biologic neuronal systems," *Biol. Cybern.*, Vol. 62, p. 363, 1990.

25. J.M. Irvine, K.T. Blackwell, D.L. Alkon, and T.P. Vogl, "Angular separation in neural networks," *J. Artificial networks*, Vol. 1, No. 1, pp. 167–180, 1994.

26. K.T. Blackwell, T.P. Vogl, D.S. Hyman, G.S. Barbour, and D.L. Alkon, "A new approach to hand-written character recognition," *Pattern Recognition*, Vol. 25, pp. 655–666, 1991.

27. D.E. Rumelhart and J.L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, 1986.

28. K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition," *IEEE Trans. on Systems, Man, and Cyber.*, Vol. 13, No. 5, pp. 826–834, 1983.

29. K. Fukushima and N. Wake, "Handwritten alphanumeric character recognition by the neocognitron," *IEEE Trans. on Neural Networks*, Vol. 2, pp. 355–365, 1991.

30. S.-C.B. Lo, H.P. Chan, J.S. Lin, H. Li, M.T. Freedman, and S.K. Mun, "Artificial convolution neural network for medical image pattern recognition," *Neural Networks*, Vol. 8, Nos. 7/8, pp. 1201–1214, 1995.

31. J.S. Lin, "Convolution neural network architecture with application for lung nodule detection in digital chest radiography," Ph.D. Dissertation 1994, Department of Electrical Engineering, University of Maryland, College Park, Maryland.

Shih-Chung B. Lo received his B.S. in Physics from National Cheng-Kung University, Taiwan, R.O.C. in 1975 and his Ph.D. in Medical Physics from UCLA in 1986. He was an MRI system engineer at Phillips Medical Systems, Shelton, Connecticut for about two years. He joined the faculty at Georgetown University in 1987 where he is currently an Associate Professor of Radiology. He has published numerous papers on computer-aided diagnosis, radiological image compression, computed tomography and digital radiography. He was the program chairman of The Workshop of Computer-Aided Diagnosis in Medical Imaging 1994 and Biomedical Application Co-Chairman of The World Congress of Neural Networks 1995. Since 1995, he has been a program committee member of the Image Processing Session of the Medical Imaging Symposium sponsored by SPIE. His current research interests are in computer-aided diagnosis, medical image processing, computed tomography, and magnetic resonance functional imaging.

Jyh-Shyan Lin received his B.S. degree in Electrical Engineering from Tankang University in 1983 and his M.S. degree in Communication Engineering in 1995 from National Chiao-Tung University, both in Taiwan, R.O.C. He received his Ph.D. degree in Electrical Engineering from the University of Maryland at College Park in 1994. He was a research assistant from 1992–1994 and became a research associate from 1994 to 1996 at the Center for Imaging Science and Information Systems (ISIS), Radiology Department, Georgetown University Medical Center. Later he served as a research scientist at Neuromedical Systems, Inc., Suffern, New York. Currently he is a senior scientist at Deus Technologies, Inc., Rockville, Maryland. His research interests include artificial neural networks, image processing, and pattern recognition.

Matthew T. Freedman received his A.B. in General Science in 1963 from the University of Rochester, Rochester, New York and received his doctorate in medicine in 1967 from the State University of New York, Brooklyn. Currently, he is Associate Professor of Radiology at the Georgetown University Medical Center, Washington, DC. He is a general radiologist with clinical and research activities in chest, musculoskeletal, and breast imaging. He is Director of Mammography Research and Clinical Director of the Center for Imaging Science and Information Systems. He has been involved in research in digital radiography, image processing and computer aided diagnosis since 1991.

Seong K. Mun received his B.S. in Physics in 1969 from University of California, Riverside, and his Ph.D. in Physics in 1979 from the State University of New York, Albany. He was an Assistant Professor of Radiation Medicine at Georgetown University Hospital from 1981 to 1983. He became Director of Imaging Physics, Department of Radiology, Georgetown University Hospital from 1982 to 1983. From 1983 to 1984, he was Managing NMR Physicist at Columbia University, Neurological Institute of New York, Columbia Presbyterian Hospital. Currently, he is Professor of Radiology at the Center for Imaging Sciences and Information Systems, Georgetown University Medical Center. He was a co-organizer of the National Forum: Telemedicine On-Line Today and the Strategic Defense Initiative Technology Application Symposium. He is the founder of the International Image Management and Communication Systems Conference; President of the Board of Scientific Counselors, National Library of Medicine, NIH; and is the Editor of *Medical Advances through Technology*.