Pergamon

Neural Networks, Vol. 8, No. 7/8, pp. 1201–1214, 1995

0893-6080/95 $9.50 + .00

0893-6080(95)00061-5

1995 SPECIAL ISSUE

Artificial Convolution Neural Network for Medical Image Pattern Recognition

SHIH-CHUNG B. LO,[1] HEANG-PING CHAN,[2] JYH-SHYAN LIN,[1] HUAI LI,[1] MATTHEW T. FREEDMAN[1] AND SEONG K. MUN[1]

[1] Georgetown University Medical Center and [2] University of Michigan Medical Center

(*Received* 1 *November* 1994; *revised and accepted* 4 *May* 1995)

Abstract—*We have developed several training methods in conjunction with a convolution neural network for general medical image pattern recognition. An unconventional method of using rotation and shift invariance is also proposed to enhance the neural net performance. The structure of the artificial neural network is a simplified network structure of the neocognitron. Two-dimensional local connection as a group is the fundamental architecture for the signal propagation in the convolution neural network. Weighting coefficients of convolution kernels are formed by the neural network through backpropagated training for this artificial neural net. In addition, radiologists' reading procedure was modelled in order to instruct the artificial neural network to recognize the predefined image patterns and those of interest to experts. Our training techniques involve (a) radiologists' rating for each suspected image area, (b) backpropagation of generalized distribution, (c) trainer imposed functions, (d) shift and rotation invariance of diagnosis interpretation, and (e) consistency of clinical input data using appropriate background reduction functions.*

We have tested these methods for detecting lung nodules on chest radiographs and microcalcifications on mammograms. The performance studies have shown the potential use of this technique in a clinical environment. We also used a profile double-matching technique for initial nodule search and used a wavelet high-pass filtering technique to enhance subtle clustered microcalcifications. We set searching parameters at a highly sensitive level to identify all potential disease areas. The artificial convolution neural network acts as a final detection classifier to determine whether a disease pattern is shown on the suspected image area.

Keywords—Neural network, Computer-assisted diagnosis, Classification invariance of operations, Output association fuzzy function, Trainer imposed function.

1. INTRODUCTION

As high speed computers become cost-effective tools, many scientists have started to investigate potential technologies for computer-assisted diagnosis (Doi, 1989; Doi et al., 1992). More and more digital imaging systems are available to radiology departments as well. It is known that conventional diagnostic procedures can be enhanced by various methods through computers. The applications in computer-assisted diagnosis will be much more meaningful when clinical images are fully computerized and networks are available in radiology departments. Medical diagnoses involve very sophisticated decision-making processes. Integration of the patient information in a Picture Archiving and Communication System (PACS) (Horii et al., 1990; Huang et al., 1990) and development of computer-aided diagnosis will provide radiologists with more relevant information to significantly improve patient care.

Skilled radiologists have a high degree of accuracy in diagnosis. However, there remain problems in the detection of some diseases, problems that cannot be corrected with current methods of training and high

Acknowledgements: This project was supported in part by a U.S. Army Grant DAMD17-93-J-3007 and an American Cancer Society Grant No. EDT-93. The reviews, opinion and/or findings contained in this paper are those of the authors and should not be construed as an official Department of Army position, policy or decision unless so designated by other documentation. The LABROC program was provided by Dr. Charles Metz of the University of Chicago. The authors are grateful to the reviewers' constructive suggestion as well as to Ms Susan Kirby and Dr. Walid Tohme for their editorial assistance.

Requests for reprints should be sent to Dr Shih-Chung B. Lo, Radiology Department, Georgetown University Medical Center, 2115 Wisconsin Avenue, N.W., Suite 603, Washington, DC 20007, USA.

levels of clinical skill and experience. These problems would cause for example the miss rate in the detection of small pulmonary nodules, the detection of minimal interstitial lung disease and the detection of changes in pre-existing interstitial lung disease. In this paper, we employed a convolution neural network and proposed several training methods to enhance the detection of small pulmonary nodules and microcalcifications on digital projection X-ray images. Both diseases are clinically important in diagnostic imaging and are relatively difficult to identify when they are superimposed on other anatomical structures.

Several image processing techniques have been proposed for the detection of lung nodule: (a) thresholding and circularity calculation (Giger et al., 1988), (b) morphological operation (Giger et al., 1990), and (c) 2-D sphere profile matching technique (Lo et al., 1993). With each of these methods there is a trade-off between increased sensitivity and decreased specificity. By setting less stringent criteria with the above algorithms, the sensitivity of the detection programs would be relatively high but false detection would also increase. On the other hand, a low sensitivity setting of the program would potentially miss many true positives. To use these methods for the detection of small lung nodules, additional techniques are needed to reduce the number of false positives and maintain high true positive detection. A similar situation was found in the detection of microcalcifications in mammography (Chan, Doi & Galhotra, 1987; Chan, Doi & Vyborny 1988, 1990). For this reason, several investigators have intended to use the neural network as a classifier to improve the detection rate (Lo et al., 1993, 1996; Wu et al., 1992).

The nets of the artificial neural network used in conventional backpropagation are fully and uniformly connected from one node of the upper layer to each node in the next lower layer. When applying this type of neural network to directly perceive image patterns, the performance seems rather limited (Lo et al., 1993). In some applications the features generated by image processing techniques were used for image pattern recognition. In observing clinical radiologists' work, it is clear that they use findings in the region surrounding the suspected area when identifying the presence of a true disease. We therefore believe that the neighborhood information rather than non-local information in the image must be taken into more serious consideration during the neural network training. For direct image input, we learned that the neocognitron has been successfully used in the recognition of characters and numbers of handwriting (Fukushima, 1980, 1989; Fukushima et al., 1983; Fukushima & Wake, 1991). The neocognitron also seems likely to be able to incorporate information of the area surrounding the suspected area into its processes and has the potential to deal with ambiguity in the information set. This is the motivation to incorporate artificial visual neural network in our research for medical image pattern recognition. In this paper we propose a convolution neural network structure and several associated algorithms for general medical image applications when an abnormality of a disease pattern can be shown in a small image area. The reduction of the image area for each training or testing is recommended for two reasons: (a) a large area demands a great deal of computation and (b) it potentially defocuses the features with which the user intends to train the neural network.

2. MATERIAL AND METHODS

2.1. Fundamental Approach

Radiographs for diagnostic medical imaging have been used for many years. The diagnostic results are based on the visual pattern recognition by trained radiologists. Throughout this study we tried to mimic the radiologists' diagnostic viewing routine. Typically radiologists scan the image, looking for potential abnormalities, then evaluate each suspected area. In detecting lung nodules, radiologists first search for suspected areas on the chest radiograph, looking for bright round objects within the rib cage boundary. Next, each suspected area is examined to compare the contrast information of the bright spot to the local background. Sometimes a radiologist uses several viewing positions to look at the area. When using a workstation, the radiologist may utilize zoom and "window and level" functions to get different views about roundness and contrast information for the suspected areas. The "window" function takes a given digital value range (e.g., 3000) and rescales onto a monitor gray value range (typically 256). The "level" function selects the middle digital value for the "window". Since both window range and level can be simultaneously operated, the radiologist is able to observe various contrasts. The differentiation between a nodule and an end-on vessel can be very difficult for the human eye to discern but is often based on the presence of projections from the round shape and its relative contrast compared to the background and to other vessels. For the detection of microcalcifications, radiologists use similar viewing steps. The main differences between the detection of lung nodules and microcalcifications are the disease patterns, clinical indications, and experience.

The radiologist diagnostic viewing steps described above were modelled and were converted to computer algorithms. The detail algorithms and techniques for the pre-scan were previously de-

scribed by Lo et al. (1993) for the detection of lung nodule on chest radiographs and by Chan and coworkers (1987, 1988, 1990, 1995) for the detection of microcalcifications on mammograms. In this paper, we concentrate mainly on the proposed convolution neural network and methods to adjust and arrange the input and output signals in order to achieve maximum efficiency.

2.2. The Convolution Neural Network

Based on the pre-scan methods, we set the computer programs to a highly sensitive level to extract possible objects which included all true-positive detections as well as a large number of false-positive detections. Differentiation of false positives from true positives is the remaining issue. We propose to use the trained convolution neural network (CNN) as the final classifier to carefully study each suspect area in the second phase of the diagnostic process. The proposed CNN can be considered a simplified vision machine designed to perform the second part of the disease detection study for the classification into disease and non-disease. This neural network is based on the network structure of neocognitron (Fukushima et al., 1983) which is designed to simulate the vision of vertebrate animals. We believe that the CNN should be suitable for general medical image pattern recognition.

Before entering an input matrix into the neural network, we employed a background reduction method (see Section 2.3.1) to mimic the function of "window and level". This image function has been widely used in clinical workstations. It is utilized to adjust the overall brightness of the image so that nodules of the same size would have similar contrast when compared to the background. In a way, it can help minimize the contrast variation of the disease pattern. In this study the background of all the suspected image blocks was reduced for the CNN training and testing. The purpose of using the two-dimensional convolution operation is to simulate radiologists' viewing of a suspected area. In other words, we instructed the neural net to utilize the information on both the center of the image block and its neighborhood and to train the neural network to extract necessary local features through the supervised backpropagation training. On the output side, we tried to educate the neural net by simulating the radiologists' decision making process. To model radiologists' interpretation of an image area with a certain probability of abnormality, a method of utilizing fuzzy output association is proposed in Section 2.3.3 to turn this kind of clinical measure into information readable by a neural network.

2.2.1. *The Structure of the Proposed Convolution Neural Network.* The CNN is a simplified version of the neocognitron. Since there is no theory to indicate what is the best neural network structure for medical image pattern recognition, we started our studies by using two-level and three-level neocognitron structures. We did not use complex-layer and did not extend our study beyond a three-level structure due to computation constraints. Nets between two adjacent levels (layers) are selectively interconnected across groups. The forward propagation algorithm was developed for non-supervised training in the original neocognitron method. The supervised training from one layer to the next (i.e., layer training) also proposed by Fukushima and Wake (1991) for handwriting recognition may be applied to the classification of disease patterns but will not be discussed in this paper. Instead we used a convolution constrained neural network with the well known backpropagation method for training. Figure 1 shows the global three-level structure of this neural network.

We group the operations of kernels and the image block in such a way that the center of the suspected nodule area is separated from the surrounding region. Basically each group in the receiving layer receives signals from two groups of weights (e.g., kernels). The kernels operating in the surrounding areas are referred to as peripheral kernels and the kernels operating in the central areas as inner kernels. This arrangement is specifically designed for image blocks containing a suspected tumor. In such a case, the bright spot is located relatively in the central area indicated by the pre-scan procedures. The purpose of using the dual-kernel is to instruct the peripheral and the inner kernels to learn different image patterns. However, for those tasks not involving the recognition of round objects, the use of a single kernel is recommended. In the following experiment, we use dual-kernel and single kernel for detection of lung nodules and microcalcifications, respectively. For the forward signal propagation, the resultant of the weighting factors of the kernel convoluting the element values of the front layer are collected into the corresponding matrix elements of the receiving layer. This operation accounts for the major difference between the convolution type neural network and a regular fully connected neural network. The collected value at each element is further operated with a sigmoid function in the forward propagation as it functions in an ordinary forward propagation neural network system.

Each suspected image block of 32×32 pixels indicated in the pre-scan program is extracted as an object for CNN classification. Due to the long training time of the computer using the CNN algorithm, every four pixels in a 2×2 square were averaged into one pixel so that each image block was reduced to 16×16 pixels. We used an array size of

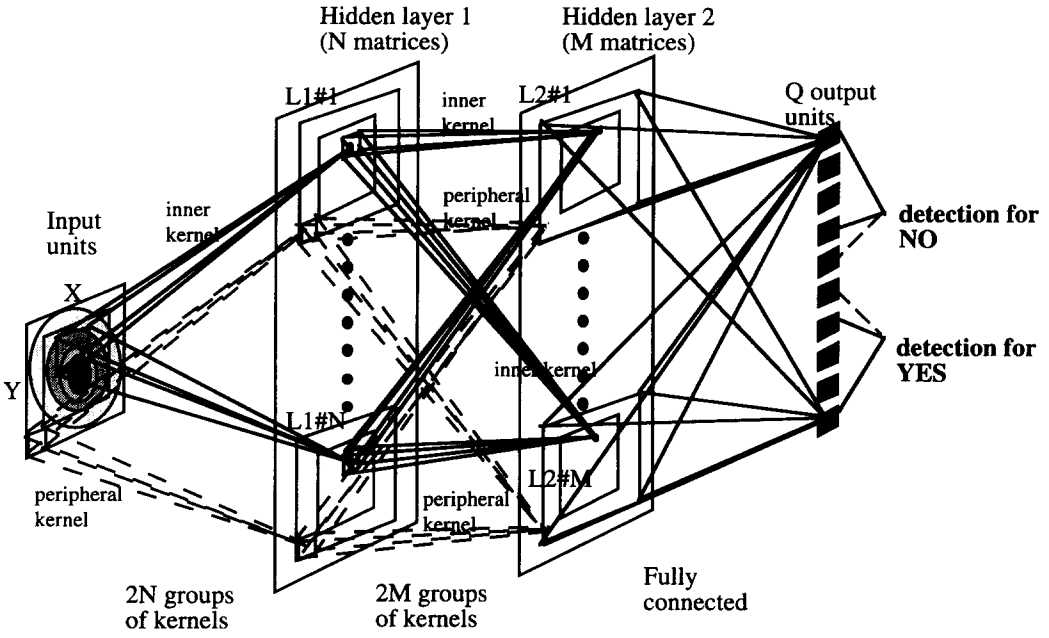

FIGURE 1. An artificial convolution neural network with dual-kernel for lung nodule detection.

5×5 for both inner and peripheral kernels between layers. The first hidden layer consists of 12 groups. Each group has 12×12 pixels formatted in a square array where the center 8×8 pixels and outer area covering by two pixels along the side are contributed by the inner and peripheral kernels, respectively. The second hidden layer also consists of 12 groups. Each group has 8×8 pixels where the center 6×6 area and outer area covering only one pixel along the side are contributed by the corresponding inner and peripheral kernels, respectively. The output layer has 10 nodes (groups) which are fully connected to the second hidden layer. However, in the experiment involving the detection of microcalcifications described later, no peripheral kernel was used. This is because most microcalcifications are concentrated on a few pixel regions while using a digitization pixel size of 105 μm.

It is important to realize that the total number of nodes needed in the hidden layers somewhat depends on the total number of training samples. Since we plan to expand our database and the use of rotated versions of an input matrix, we expect that our training samples will be very large in the future. The number of layers used should depend upon the sophistication of the features that the neural network is intended to perceive. The more complicated the disease patterns, the more layers are required to distinguish high order information of image structures. The convolution kernels are organized in such a way as to emphasize a number of image characteristics rather than those less correlated values obtained from feature spaces for input. These characteristics are: (a) the horizontal versus vertical information; (b) local versus non-local information; and (c) image processing (filtering) versus signal propagation.

2.3. Image Processing and Training Methods

An appropriate neural network structure is an important working base to form a signal propagation platform in a given recognition task. The training materials and methods, which provide intellectual information for the construction of the knowledge, are essential for the performance of the neural network. We believe that the success of using the neural network relies not only on the network structure but also on the sufficient training information and effective training methods. This study demonstrates our approaches to convert expert knowledge into computer readable information, which is the key issue in terms of training. In this experiment, we provided the network with all possible radiological diagnostic information and set up the studies by adding one method at a time to optimize the neural network performance.

2.3.1. *Background Reduction for Suspected Image Blocks.* We found that the consistency of input matrix contrast is an important factor in stabilizing the neural network learning. In our experiment, the neural network did not reach a solution for the image blocks provided in the training sets, even though all suspected nodule blocks were corrected for background trend and relatively centered in terms of brightness. Their contrasts (the difference between

the center and peripheral brightness) are unevenly distributed due to a variation of X-ray exposure and different sensitivity of the films. In addition, the image block may involve many superimposed background structures, namely vessels, ribs, and the heart. Separating nodules or suspected round objects from chest structures is not an easy task. We concluded that elimination of some background information and enhancement of contrast information are necessary procedures to assist the neural network in the recognition of disease patterns. We designed a background subtraction technique to simulate "window and level" function, which is clinically useful for enhancing disease patterns. In fact, we only used "level" function and ignored "window" function. A fixed "window" function may distort and mix the contrast information that exists between nodules and end-on vessels. For the "level" function, gray values in each image block are uniformly subtracted from the calculated background by averaging the outer ring area.

$$fs(x,y) = \begin{cases} f(x,y) - B \\ \quad \text{for } f(x,y) > B \text{ and } (x,y) \text{ is inside the circle} \\ 0 \\ \quad \text{for } f(x,y) < B \text{ or } (x,y) \\ \quad\quad \text{is on or outside the circle,} \end{cases} \tag{1}$$

where

$$B = \frac{\sum_{n \in \text{the ring}} f(x_n, y_n)}{C}$$

for circular object detection and C equals the number of pixels in the ring. Figure 2 shows that heavily-shaded pixels are used in the calculation for background averaging. Both heavily and lightly shaded areas are given in pixel value of 0. Our studies indicated that this ring area averaging method produces better results than the peripheral area averaging method. This may be due to the fact that the ring area is closer to the central area and possesses greater background information than the entire peripheral area.

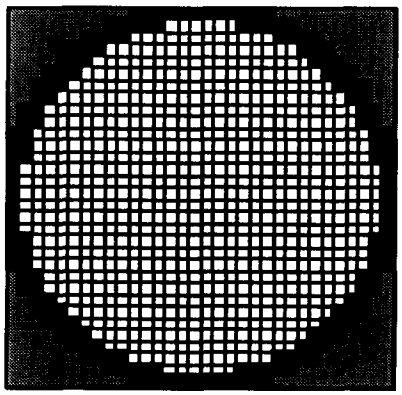

FIGURE 2. A 32 × 32 image block. The white area is the area of interest for image pattern recognition using convolution neural network. Original pixel values in the heavily shaded area are averaged as a background value.

For the detection of non-circular objects, the background value should be obtained by averaging pixel values on the frame. Each gray value from the background-reduced image block is one-to-one transferred to a node of the input layer for the neural net processing. These signals received at the input layer are equivalent to the light signals received by the retina as far as the vision type neural network is concerned.

2.3.2. *Backpropagation Training.* The main difference between conventional weights and kernel weights is that conventional weights are independent and kernel weights are constrained by grouping. We believe that the latter method is more powerful than the former method for direct image pattern recognition. In addition, the trained kernels can be analyzed to understand what features were learned during the training. This design would allow researchers to further investigate the artificial neural network learning. Training requires many iterations for the network to obtain solutions for all weights applied to the propagation while the error function reaches a minimum value.

By looking at the CNN processing, one may find that signals are filtered and modulated as in a complicated circuit system. Signal propagation from one layer to the next is composed of a two-step calculation: (a) adaptive convolution combiner and (b) an activation function (a sigmoid function is used in this study) which is given below:

$$S_x((i,j);n) = \frac{1}{1 + \exp\left\{ - \sum_m [k_x((u,v);n,m)) \otimes S_{x-1}((i,j);m))] \right\}} \tag{2}$$

or

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

where $S_x((i,j);n)$ represents the signal at node (i, j),

nth group, and x layer; $k_x((u,v);n,m)$ denotes the weighting factor value of net (u, v) in the nth group of the $x - l$ layer which connects the mth group of the x layer.

The error function which is expected to reach a local minimum through the error backpropagation training can be given as:

$$E = \frac{1}{2}\sum_{n_o=1}^{T}[y(n_o) - S_o(n_o)]^2, \tag{4}$$

where $y(n_o)$ and $S_o(n_o)$ are the target output and calculated output signals for output node n_o, respectively and T is the total number of output nodes. Based on eqns (3) and (4), the iterative version of kernel weights derived by the generalized delta rule is given as:

$$\begin{aligned} k_x((u,v);n,m)[t+1] &= k_x((u,v);n,m)[t] \\ &\quad + \eta\sum_{i,j}\delta_x((i,j);n)S_{x-1}((i-u,j-v);m) \\ &\quad + \alpha\Delta k_x((u,v);n,m)[t], \end{aligned} \tag{5}$$

where t is the iteration number during the training, η is the gain for the current weight changes, α is the gain for the momentum term received in the last learning loop, and δ is the weight-update function which is given as:

$$\delta_x(i,j);n) = S_x((i,j);n)[1 - S_x(i,j);n)]Q_x((i,j);n) \tag{6}$$

and

$$Q_x((i,j);n) = \sum_{u,v;m} k_{x+1}((u,v);n,m) \times \delta_{x+1}((i+u,j+v);m).$$

For the output layer,

$$\begin{aligned} \delta_o((i,j);n_o) &= \frac{\partial E}{\partial S_o(n_o)}S_o(n_o)[1 - S_o(n_o)] \\ &= [S_o(n_o) - y(n_o)]S_o(n_o)[1 - S_o(n_o)] \end{aligned} \tag{7}$$

where o denotes the output layer. In this study, all weighting factors including the kernels were initially given a normalized random number. The normalization is based on the number of nets connecting to a destination node in the next layer.

2.3.3. *Neural Network Output Assignment Using Radiological Diagnostic Rating.* The design of the output layer for the medical diagnostic decision is not as straightforward as in other applications. Our goal is to distinguish non-disease patterns from disease patterns. We can classify the data set in two categories. However, this is probably not an optimal design for the output layer. In some obvious cases, radiologists are able to make a clear diagnostic indication of a disease shown on an image. Often they work with different degrees of sensitivity (different levels of suspicion) depending on the clinical situation. Thus they estimate the likelihood that a radiograph or an area of a radiograph may possess a disease. For the neural network, it may be more realistic to define the output in terms of probability. Depending upon the number of output nodes used, the arrangement of output nodes and the probability associated with a score varies. Intuitively, one can proportionally scale scores onto node numbers.

Although the above output node assignment follows the general diagnostic decision rule used by many radiologists, one output node has no relation to any other. No output node relation will be passed to the neural network for the training. To circumvent this problem, we propose to use a narrow output distribution to establish a fuzzy association between the adjacent output nodes. 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 by a variation. We modelled this probability with a generalized distribution (Szepanski, 1980) in the score space.

$$G(\sigma, v, p) = Ce^{-|gv|^p} \tag{8}$$

where v is the distance from a given score,

$$C = \frac{pg}{2\Gamma\left(\frac{1}{p}\right)} \text{ and } g = \frac{\sqrt{\Gamma\left(\frac{3}{p}\right)}}{\sigma\sqrt{\Gamma\left(\frac{1}{p}\right)}}$$

The reason for modelling a generalized distribution is that we do not know exactly what kind of distribution can represent the radiologists' interpretation in various diseases. When $p > 2$, the distributions may be too flat which probably is not the case with highly experienced pulmonary radiologists. For simplicity, we use $p > 2$ for Gaussian distribution in the experiment.

In addition to the distribution function, the trainer can impose a driving function, $r(v,s)$, onto it to indicate the belonging of the determination category, where s denotes the strength of the repulsion introduced by the user. An example of the trainer

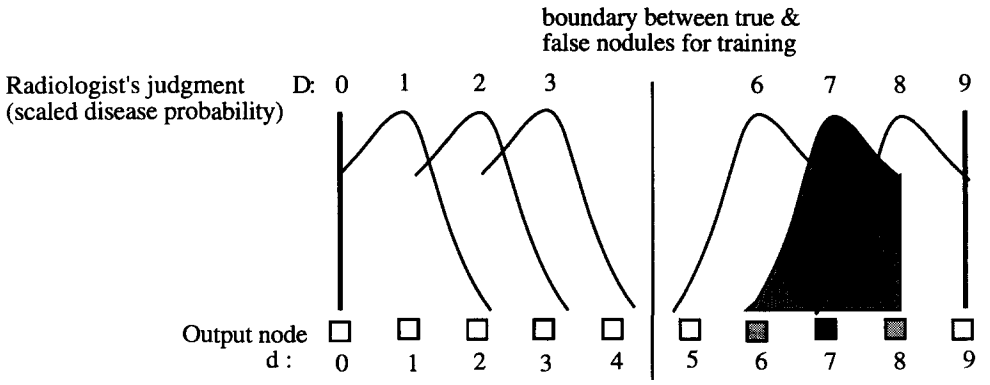

FIGURE 3. A fuzzy output association is constructed by Gaussian distribution and repulsive functions. (Note this drawing is not in scale. Only one curve is used for a training case.)

imposed driving function for scores indicating positive determination is given below:

$$r(v,s) = \begin{cases} 1 & \text{for } v \leqslant 0 \\ sv + 1 & \text{for } v > 0. \end{cases} \tag{9}$$

For those scores associated with negative determination, the repulsion of eqn (9) should be changed to the opposite direction. Therefore, the output association functions at a single node for a score indicating an image block involving disease and for a score indicating a disease-free image block are:

$$Ah = K \times G(\sigma, v, p) \times r(v,s) \tag{10}$$

and

$$Al = K \times G(\sigma, v, p) \times r(-v,s), \tag{11}$$

respectively. For the extreme scores in the score space (i.e., minimum and maximum scores), the use of a delta function is recommended.

In the lung nodule detection studies, we assigned scores for all image blocks for training. Based on the score which corresponds to an output node, a Gaussian distribution ($p = 2$) with a standard deviation of $\sigma = 0.55$, an output scaling constant of $K = 2.5$ and a repulsive strength of $s = 1.5$ for the asymmetric output association ($s = 0$ for the symmetric output association) were used for correlating the adjacent scores. For the neural network output, we used a discrete form of the score. We estimate that it would take a great deal of computation time for 100 nodes or more in the output layer. Realistically, 10 discrete output nodes are proposed for the classification. We assigned nodes 0–3 to correspond to definitely negative-possibly negative; detection nodes 6–9 correspond to possibly positive-definitely positive detection. Nodes 4 and 5 are not used for decision buffering. During the experiment, we collected all the suspected nodes in two categories (i.e., true nodule and non-nodule). In the course of rating for the training set, a senior radiologist scored each suspected nodule based on his clinical knowledge. Pathologically proven truth (either has a nodule or not) of training case was also provided to assist in the radiologist's rating.

Figure 3 shows all the asymmetric output association distributions corresponding to a radiologist's judgement. However, only one curve was used for each judgement with a suspected image block. Figure 3 also highlights a case when score 7 is determined. In this situation, output node 7 received the highest activation (1.0), node 8 received the second highest activation (0.5), node 6 receives some activation (0.2), and remaining nodes receive no activation.

Two examples of output assignments associated with probability in discrete form are given below:

(a) Symmetric output assignment:

$$A(d,D) = \begin{cases} 0.2 & \text{for } 9 > D > 5 \text{ and } d = D+1 \\ & \text{or for } 0 < D < 4 \text{ and } d = D-1 \\ 1.0 & \text{for } D = d \\ 0.2 & \text{for } 9 \geqslant D > 5 \text{ and } d = D-1 \\ & \text{or for } 0 \leqslant D < 4 \text{ and } d = D+1 \end{cases} \tag{12}$$

otherwise $A(d,D) = 0$.

(b) Asymmetric output assignment:

$$A(d,D) = \begin{cases} 0.5 & \text{for } 9 > D > 5 \text{ and } d = D+1 \\ & \text{or for } 0 < D < 4 \text{ and } d = D-1 \\ 1.0 & \text{for } D = d \\ 0.2 & \text{for } 9 \geqslant D > 5 \text{ and } d = D-1 \\ & \text{or for } 0 \leqslant D < 4 \text{ and } d = D+1 \end{cases} \tag{13}$$

otherwise $A(d,D) = 0$.

The use of asymmetric output assignments attempted to push (train) the non-disease pattern toward low score nodes and to instruct the disease pattern toward high score nodes. With this output assignment for the output node in the training, the adjacent node relation is also established. This supervised training can be generally applied to any situation where association of outputs is necessary.

2.4. Classification Invariance of Matrix Operations

The use of moment invariance via rotation and shift has been proposed for applications in graphic pattern recognition. The direct use of this method as a classifier may not be suitable for those image patterns possessing circular symmetric property (e.g., nodule) or lacking a fixed geometric pattern (e.g., calcification).

Often medical image pattern recognition does not concern "top-down" or "left-right" as classification criteria. In such a case we can take advantage of this characteristic as an invariance. In other words, we propose to rotate and/or to shift the input vector and maintain the same output assignments for the training. This method may affect the neural network in two ways: (a) by instructing the neural network that the rotation and shift of the input vector would receive the same classification result; and (b) by increasing the total number of training samples which is expected to enhance the performance of the neural network.

Using the center pixel as the origin, a standard rotation and shift of the image block was used;

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \tag{14}$$

where ϕ is rotation angle of the origin (center of a 32 $\times$ 32 image block); Δx and Δy are shifts in the x and y directions, respectively.

In this work, we only rotated each input matrix eight times to test our hypothesis. Four of the rotations are:

$$\phi \in \{0°, 90°, 180°, 270°\}. \tag{15}$$

We also flipped over (left-right) the original image matrix and used the above rotations again to obtain four additional rotations. This type of rotation would only reposition pixel values. No interpolation calculation of pixel values was involved. We believe that other rotations and minor shifts are also valid methods for the use of classification invariance of operations. Rotation may require interpolation which would slightly alter the pixel values and should be acceptable for the input of the CNN. However, the use of shifting can be complicated, because it involves (a) how important the center information for disease patterns are in the neural network learning and (b) how much shifting can be used without sacrificing critical portions of image information.

2.5. Classification of Output Values in the Testing

We assigned scores with a narrow asymmetric peak distribution on the output nodes for the training in order to associate each node with its adjacent node. We believe that the distribution assignment is not a unique method to link rating score relations. However, the output relation information must be passed to the neural network for learning. This relationship does not exist in recognition for characters or Arabic numbers. In those applications, each node is independent from others.

After the training a typical output pattern will be very close to the corresponding perfect pattern (the assigned narrow asymmetric peak distribution) for most of the training cases. In the case of testing, many of them have different output signal patterns. It is not a simple task to interpret what the representation of each output pattern means if the testing output does not follow an output pattern assigned to the training. Corresponding to the grading system arranged in the training, a polarized (linearly weighted) function is given as an indication. With this we can define a normalized disease detection index (NDDI) for the judgement of a suspected area:

$$\text{NDDI} = \frac{\sum_{n=N/2}^{N-1} [O_n \times (n - (N-1)/2)]}{\sum_{n=0}^{N-1} [O_n] \times (N-1)/2} \tag{16}$$

where n denotes the node in the output layer, 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 or near 0 indicates a definite non-nodule and a nodule detection index of 1 or greater implies a definite nodule case with the judgement of the neural network. The reason for the weighting is that the score line is centered at $(N-1)/2$ (i.e., 4.5 for 10 nodes in the output layer) and polarization of true and false depends on the position of the nodes. Equation (16) is the net effect of all output nodes.

We do not recommend using a detection index of 0 as a cut-off point to determine a disease or a disease-free image block using the trained neural network. The cut-off point may be shifted by the inevitable bias in the training cases. In practice the cut-off point is established by many clinical cases in a rigorous

evaluation study. In this paper a pre-clinical performance study was conducted and is discussed below.

2.6. Performance Evaluation of The Convolution Neural Network

Receiver operating characteristic (ROC) is an analytical method generally applied to the performance evaluation of a system. In an ROC analysis, the distributions of the normal and abnormal cases may be represented by binormal distributions (Swets & Pickett, 1982). When the two distributions overlap on the decision axis, a cut-off point can be made at an arbitrary decision threshold. The corresponding true-positive fraction (TPF) versus false-positive fraction (FPF) for each threshold can be indicated in Cartesian coordinates. By marking several points on the plot, curve fitting can be employed to construct an ROC curve. The area under the curve referred to as Az can be read as a performance index of the system using ROC analysis. In general the higher the Az, the better the performance. A computer program (LABROC) using two sets of data, one for true and the other one for false categories, is employed for the analysis of the NDDIs derived from neural net outputs.

3. EXPERIMENTS AND RESULTS

3.1. Detection of Lung Nodules on Digital Chest Radiographs

Chest radiographs in patients with primary and metastatic cancer and with one or several lung nodules are converted into digital form using a laser film digitizer (Konica Laser Film Scanner Model: KDFR-S; Tokyo, Japan). About one third of chest images were acquired from a computed radiographic system (AGFA ADC prototype computed radiography; Mortsel, Belgium). The digital data are transmitted and stored in our PACS until needed for the research project. The images were then retrieved to a high speed workstation and the computer searches were used sequentially: a thresholding evaluation, use of background reduction, a test of profile matching rate, and neural network classification.

The pre-scan process was performed first to locate the center of the island and isolate the image block for training. The pre-scan program was running in a highly sensitive mode with a matching rate (MR) of 0.7 for all images involved in the training. Suspected image blocks included various types of rib crossing, and various sizes of end-on vessels and vessel clusters. The true-positive nodules may also overlap with lung, vessels, and rib structures. Figure 4 shows some randomly sampled suspected image blocks which were background-reduced and contrast-balanced for display purposes. These image blocks were mirrored and rotated 90°, 180°, and 270° for the training. Note that each original and its seven "brother" image blocks share the same score vector (probability of a disease and output fuzzy association). During the training, the original and its seven "brother" image blocks as a group were entered in the same sequence.

During the training we found that the error-function did not monotonously decrease for each learning epoch. However, the overall errors decreased throughout many iterations. We did not completely

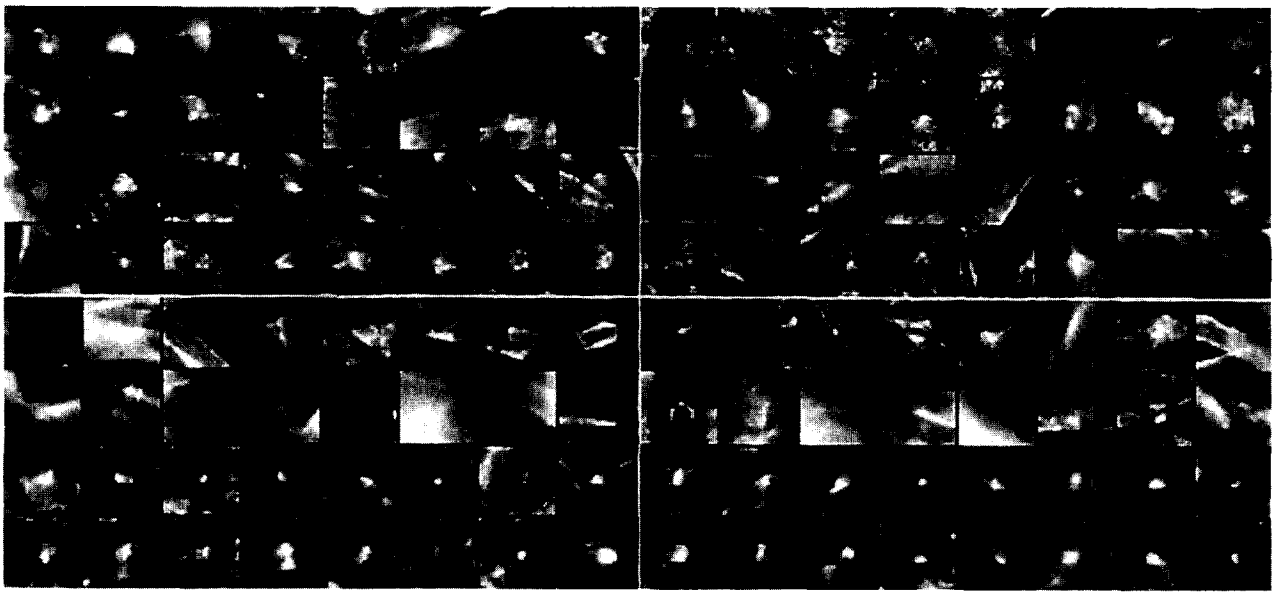

FIGURE 4. The upper four rows show 64 nodule blocks sampled from the database. Each image block on rows 5 and 6 contains no nodule but lung or rib structure. Each image block on the bottom two rows contains an end-on vessel.

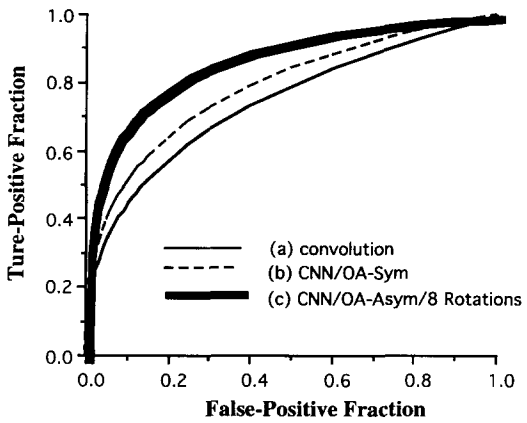

FIGURE 5. Three ROC curves representing the performance of (a) CNN without output association and rotation (plain line, the average $Az = 0.77$), (b) CNN using a Gaussian output association in the training (dashed line, the average $Az = 0.83$), and (c) CNN with asymmetric output association and eight rotated input matrices (bold line, the average $Az = 0.88$).

retrain the kernels for different assignments in the output layer. Instead, based on the trained kernels we continued to train the neural network with additional conditions. The sequence is: (a) CNN with symmetric output association, (b) use of trainer imposed driving function, and (c) rendering seven "brother" images for training.

The database had 55 chest radiographs and only 25 images contained at least one nodule. In the prescan, 52 nodules and 155 non-nodules were extracted from all 55 images. All cases were confirmed by biopsy or by follow-up showing growth of the nodule. In this study, we employed a grouped jackknife method (Fukunaga & Hayes, 1989) to evaluate the performance of the CNN. We randomly selected 28 images for training and the other 27 images for testing in the study. Final ROC curves were obtained by averaging the results from 30 grouped jackknife experiments. The results obtained from the tests were very encouraging. Figure 5 shows the improvement of using a convolution neural network and corresponding enhancement techniques using output association and classification invariance of matrix operation for the input.

In this experiment, we found that the average Az was 0.77 using the CNN with a delta function for output determination, and was 0.83 using the CNN with a narrow Gaussian distribution for output association. Using a Gaussian output association and eight types of rotated image blocks for input, we found that the Az was increased to 0.87. After a trainer imposed function was added, we obtained an insignificant increase of Az to 0.88. From the ROC curve corresponding to $Az = 0.88$, we found that the CNN reduced 79% of false-positive detections equivalent to 2–3 false nodule detections per image and preserved 80% of true-positive detections.

We also tested the same database using two nodes in the output layer. In such a case, no output association can be used. The CNN achieved an average Az of 0.83 when eight input matrices shared the same diagnostic interpretation (true or false).

3.2. Detection of Microcalcifications on Digital Mammograms

We also evaluated the use of CNN in the detection of subtle microcalcifications. A total of 68 mammograms (only 38 of them consisted of subtle

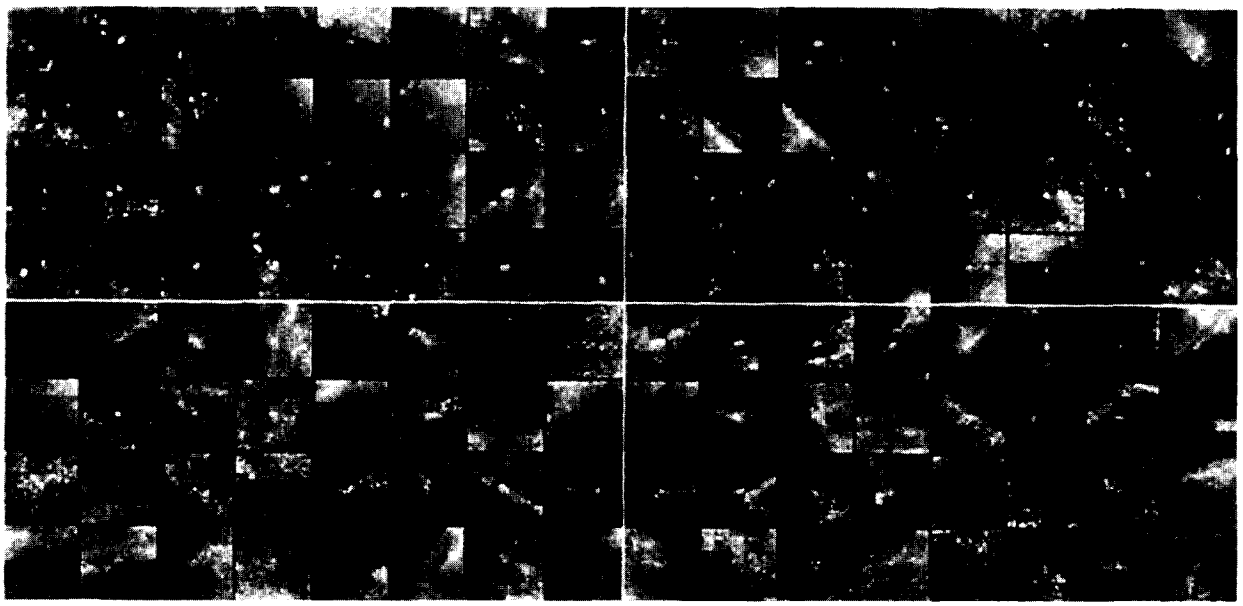

FIGURE 6. Each image block, extracted from the mammogram, on the upper four rows contains at least one calcification. Each image block on the bottom four rows contains at least a local maximum value of gray scale (bright spot) that is not a calcification. Each block at matrix elements (1,4), (5,4), 7,4),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

microcalcifications) were digitized by a laser scanner with a pixel size of 0.105 mm. The initial search prior to the final interpretation by the neural network follows the basic scheme which uses background removal and signal extraction methods to pre-scan the mammograms and to extract all possible suspected areas (Chan et al., 1988, 1990, 1991). After the pre-scan process by the computer program, the 68 digital mammograms provide 265 true and 1821 false subtle microcalcifications. Figure 6 shows some of the suspected regions which may or may not contain microcalcifications.

Prior to the CNN process, the background of all the image blocks were removed using a wavelet high-pass filtering technique instead of using the circular averaging method described in Section 2.3.1 where lung nodule detection was the objective. Specifically, after extracting each suspected region from the original digital mammogram, a three-level wavelet transform was used and only the lowest frequency was eliminated for high-pass filtering before image reconstruction. The high-pass filtered image blocks were used as the input of the CNN. For this study, we also employed the grouped jackknife method to evaluate the performance of the CNN. We did not ask radiologists to rate image blocks in the mammography training set. Only two output nodes with eight rotations for input were used. Neither output association nor trainer imposed function was employed.

In the first study, we randomly selected two sets of mammograms (i.e., 34 for training and 34 for testing) with variable sizes of kernels and image block. Figure 7 shows the *Az*s of various CNN structures used in the experiment with the same data set described above. In this figure, the CNN structures are indicated by *Sn/Hm/Kt* representing $n \times n$ pixels for input, m hidden layers, and with a kernel size of $t \times t$. The image blocks are centered on a suspected calcification indicated by the pre-scan method. This study indicated that significantly higher *Az*s were obtained when a square area of 1.7 mm (i.e., 16 × 16 pixels) region for the input and kernel size of 0.52 mm

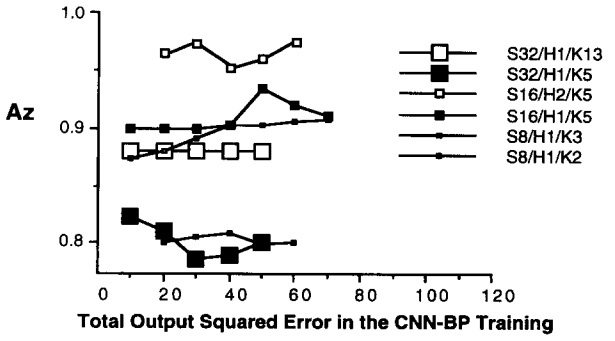

FIGURE 7. *Az*s in the detection of clustered microcalcifications using different CNN parameters.

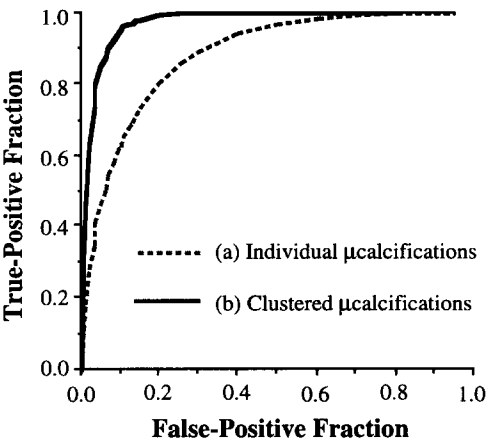

FIGURE 8. Two ROC curves representing the performance of (a) CNN using two outputs and eight types of rotation for input with the determination based on individual microcalcifications: the average *Az* = 0.89 and (b) CNN using two outputs and eight types of rotation for input with the determination based on clustered microcalcifications: the average *Az* = 0.97.

(i.e., 5 × 5 pixels) were used. In addition, the use of two hidden layers is better than the use of one hidden layer. We also found that the best results are obtained at a relatively large square error (i.e., cost function was 40–70 for 2104 cases) which suggests a fuzzy membership in the output or that more nodes in the output layer may be necessary for the optimization of the CNN in the detection of this database.

Based on the above initial study, we decided to use the CNN structure with the parameter of S16/H2/K5 for the grouped jackknife study of the CNN performance. Final ROC curves were obtained by averaging the results from 30 grouped jackknife experiments. Figure 8 shows the results of using the CNN and classification invariance of matrix operation for the input. In this experiment, the average *Az* was 0.89 when the determination was based on individual microcalcifications and was improved to 0.97 when the determination was based on the clustered microcalcifications. In the latter method, suspected clusters including one or two calcifications were rejected and the average NDDI taken from the clustered calcifications was used for the ROC evaluation. One must realize that the detection of clustered microcalcifications is more clinically significant than individual calcifications, since the clustered microcalcifications (three or more) are a strong indication of breast carcinoma in radiological diagnosis. The clustering procedure was done by grouping the detected microcalcifications in a 1 cm^2 region of the mammogram. Only a minimum of three clustered microcalcifications was considered a detection. The average ROC curve for the detection of clustered microcalcifications indicated that the CNN can eliminate 90% of false-positive detections, resulting

in 0.5 false clustered detection per image, and preserve a true-positive detection rate of 87%.

4. DISCUSSION

Medical image pattern recognition using feature extraction as an input has been proposed in the detection of disease patterns (Wu et al., 1991). Since only a small number of inputs are used (as compared to 16×16 input signals), less computation is necessary for training. As long as the features of a disease pattern are well defined and can be quantified as values or vectors, a nonconvolution neural network should be able to classify the features. However, both the proposed diagnosis invariance and the output assignment methods for the enhancement of disease detections may only be used in limited cases. On the other hand, the structure of the convolution neural network is complicated and requires more computations, particularly for the training. The CNN does not require the feature extraction of disease patterns from the image and is capable of distinguishing non-disease patterns from disease patterns. A potential advantage of using the proposed CNN is that feature extraction can be more specifically defined not only by the user's experience but also by the confirmation of the CNN when the function of each kernel is discovered. Some complementary features learned by the CNN may be able to contribute image information of a disease thereby assisting the radiologist in better understanding all the features of the disease. Further investigation of the CNN specified features, other than known features, should be very interesting to radiologists and imaging scientists.

In this work we used preliminary scanning methods to define suspected abnormal areas. The final disease classification was analyzed by using an artificial convolution neural network with back-propagation training. We proposed several methods to mimic the radiologists' reading patterns in detecting diseases on radiographs. Though conventional image processing techniques can capture true diseases, many false-positive detections are obtained. We found that the CNN substantially reduced the number of false-positive detections.

In this study, we designed the convolution neural network to focus on local information with expert-trained output distribution. The use of diagnosis invariance of rotation seems likely to enhance the performance of the CNN by virtually increasing the number of training cases. It is obvious that both the expert-trained output distribution and the classification invariance of matrix operations are not only applicable to CNN but also to a conventional neural network as long as an image (or an image associated vector which depends on image orientation) is used in the input layer.

Summarizing the failure cases in the study of lung nodule detection, we found that the majority of false-negatives related to nodules partially overlapped with rib and many false-positives related to end-on vessels. This is because our training database was small and did not have enough true cases to cover various situations in rib overlapping on nodules and did not have enough false cases to cover various contrasts of end-on vessels. We believe that the performance of the CNN will be greatly improved when the training cases are sufficiently expanded in the future study.

One may interpret eqn (16) as another network fully connected to a single output node. This subnetwork can be included in the backpropagation training with a linear activation function for the output node. However, this subnetwork does not ensure that the backpropagated signals on the previous layer (i.e., the output layer consisting of 10 nodes) are matched with the radiologists' scores. The use of 10 nodes in the output layer also provides flexibility for the researcher to investigate the migration of kernel changes when an additional training strategy is added. The fuzzification of the teaching signals and the use of a trainer imposed function are examples of the training strategies used in this paper. The kernel changes corresponding to the training can be important information for future optimization of the CNN algorithm associated with disease pattern recognition.

In this study, we learned that the background reduction was a necessary procedure for the detection of both lung nodules and mammographic microcalcifications otherwise the error function would not reach a minimum for the training data set. Several broad output distributions were also tested. The CNN performance (i.e., generation) of those tests were inferior to that of the narrow output distribution. A comparison experiment was also conducted to evaluate the difference between the training using image groups (the original and its seven "brother" image blocks as one group) and image blocks (all image blocks). We found that the CNN seems to perform better using image groups than randomizing each image block in the training. We also modified our neural network structure to one hidden layer. The CNN performance with one hidden layer was not as effective (the average $Az = 0.81$ for kernel size of 5×5 and the average $Az = 0.85$ for kernel size of 13×13) as when two layers were used for the detection of microcalcifications. However, the performance was about the same with one hidden layer and two hidden layers for the study involving lung nodules. We do not know whether this effect was due to the fine structure of microcalcifications or smaller

samples used in the experiment of lung nodule detection. We are currently testing three hidden layers to see if there is any improvement in the generalization. We are also working on 32×32 original image block and expect better outcomes. By increasing the sizes of image block and kernel, the computation time will increase 10–16 times as much as the training time needed for the CNN configuration described in Section 2.2.1. When the database is expanded, a higher power computer will be required for the training.

5. CONCLUSIONS

In this study, we have added several effective techniques to the convolution neural network for the enhancement of disease diagnosis: (a) development of a better background reduction method so that the neural network has a better "observation" of the image block, (b) providing radiologists' rating scale for the backpropagation training, (c) introducing the neural network with the classification invariance of input matrix operations, (d) use of output association functions to mimic the radiologists' interpretation and to establish the relationship between adjacent output nodes, and (e) rendering trainer imposed functions to enhance the performance of CNN. We found that the performance of the CNN in detecting disease was improved significantly by administering these training methods.

Studies in the use of chest radiographs for the detection of lung nodules (Stitik et al., 1985; Hellan et al., 1984) have demonstrated that even with highly skilled and highly motivated radiologists working with high quality chest radiographs, only 68% of all retrospectively detected lung cancers were detected prospectively when read by one reader, and only 82% were detected by two readers. Our studies did not have the same clinical setting as Stitik and Hellan's due to our smaller database, therefore, we could not compare our results with the radiologists' sensitivity of 68% mentioned above. However, we consider it likely that radiologists will benefit from the use of a nodule detection program such as this in one of two ways. First, the radiologist will use the program as a second reader, thus increasing the detection of lung nodules similar to the results seen in the study by Stitik. In the second method, the radiologist may call on the system as a consultant on an individual suspected area. The radiologist can point to the suspected area and ask for the interpretation from the CNN system. The CNN system, in fact, may be able to work as a trained referral system for the consultation of detecting lung nodules. Such a program is also readily available in our computer and clinical evaluation is in progress. A fully automatic lung nodule detection program takes 12–18 s for a 512×512 digital chest radiograph in a DEC Alpha workstation. To evaluate an identified area, it only takes the CNN program 0.2 s to respond.

This work has demonstrated two successful medical diagnostic applications using an artificial visual neural network and expert-trained computer procedures instead of a non-convolution neural network or other conventional classification method. This technique attempted to simulate the radiologists' reading pattern: pre-screen and classification for interpretation. We believe that the proposed convolution neural network and its associated training techniques can be extended to many diagnostic imaging areas such as the detection of low contrast mass in mammography and the pattern recognition of interstitial lung disease in chest radiography. In fact, the proposed CNN technique should be able to be trained to detect almost all disease patterns perceivable by a trained radiologist.

REFERENCES

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

Chan, H. P., Doi, K., & Vyborny, C. J. (1988). Computer-aided detection of microcalcifications in mammograms. *Investigative Radiology*, **23**, 664–671.

Chan, H. P., Doi, K., & Vyborny, C. J. (1990). Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer aided diagnosis. *Investigative Radiology*, **25**, 1102–1110.

Chan, H. P., Lo, S. C., Sahiner, B., Lam, K. L., & Helvie, M. A. (1995). Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network. *Medical Physics* (in press).

Doi, K. (1989). Feasibility of computer-aided diagnosis in digital radiography. *Japanese Journal of Radiological Technology*, **45**, 653–663.

Doi, K., Giger, M. L., & MacMahon, H. (1992). Potential usefulness of real-time computer output to radiologists' interpretations. Scientific Exhibit, Space 10-001. Presented at RSNA 1992, Chicago, Ill.

Fukunaga, K., & Hayes, R. R. (1989). Effects of sample size in classifier design. *IEEE Transactions on Pattern Analysis of Machine Intelligence*, PAMI-**11**, 873–885.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, **36**, 193–202.

Fukushima, K. (1989). Analysis of the process of visual pattern recognition by the neocognitron. *Neural Networks*, **2**, 413–420.

Fukushima, K., Miyake, S., & Ito, T. (1983). Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, **13**(5), 826–834.

Fukushima, K., & Wake, N. (1991). Handwritten alphanumeric character recognition by the neocognitron. *IEEE Transactions on Neural Networks*, **2**, 355–365.

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

Giger, M. L., Ahn, N., & Doi, K. (1990). Computerized detection of pulmonary nodules in digital chest images: Use of morphological filters in reducing false-positive detections. *Medical Physics*, **17**, 861–865.

Hellan, R. T., Flechinger, B. J., & Melamed, M. R. (1984). Non small cell lung cancer: Results of the New York screening program. *Radiology*, **151**, 289–293.

Horii, S. C., Mun, S. K., & Levine, B. A. (1990). PACS clinical experience at Georgetown University. *Computerized Medical Imaging and Graphics*, **15**(3), 183–190.

Huang, H. K., Kangarloo, H., & Cho, P. S. (1990). Planning a total digital radiology department. *American Journal of Radiology*, **54**, 635–639.

Lo, S. B., Freedman, M. T., & Lin, J. (1993). Automatic lung nodule detection using profile matching and back-propagation neural network techniques. *Journal of Digital Imaging*, **6**(1), 48–54.

Lo, S. B., Lou, S. L., & Lin, J. (1996). Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Transaction on Medical Imaging* (in press).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representation by error propagation. In D. E. Rumelhart & J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing*, (Vol. 1, pp. 318–362). Cambridge, MA, MIT Press.

Stitik, F. P., Tockman, M. S., & Khouri, N. F. (1985). Chest radiology. In A. B. Miller (Ed.), *Screening for cancer*, (pp. 163–191). New York: Academic Press.

Swets, J. A. & Pickett, P. M. (1982). *Evaluation of diagnostic systems*. New York: Academic Press.

Szepanski, W. (1980). Δ-entropy and rate-distortion bounds for generalized-Gaussian information source and their applications to image signals. *Electronics Letters*, **16**(3).

Wu, Y., Doi, K., Giger, M. L., & Nishikawa, R. M. (1992). Computerized detection of clustered microcalcifications in digital mammograms: Applications of artificial neural networks. *Medical Physics*, **19**, 555–560.