

# Computerized classification of suspicious regions in chest radiographs using subregion Hotelling observers

Alan H. Baydush,<sup>a)</sup> David M. Catarious, Jr., and Joseph Y. Lo  
*Digital Imaging Research Division, Department of Radiology, Duke University Medical Center and  
Department of Biomedical Engineering, Duke University, Durham, North Carolina 27710*

Craig. K. Abbey,<sup>b)</sup>  
*Department of Medical Physics and Imaging, Cedars-Sinai Medical Center*

Carey E. Floyd, Jr.  
*Digital Imaging Research Division, Department of Radiology, Duke University Medical Center and  
Department of Biomedical Engineering, Duke University, Durham, North Carolina 27710*

(Received 22 May 2001; accepted for publication 26 September 2001)

We propose to investigate the use of subregion Hotelling observers (SRHOs) in conjunction with perceptrons for the computerized classification of suspicious regions in chest radiographs for being nodules requiring follow up. Previously, 239 regions of interest (ROIs), each containing a suspicious lesion with proven classification, were collected. We chose to investigate the use of SRHOs as part of a multilayer classifier to determine the presence of a nodule. Each SRHO incorporates information about signal, background, and noise correlation for classification. For this study, 225 separate Hotelling observers were set up in a grid across each ROI. Each separate observer discriminates an 8 by 8 pixel area. A round robin sampling scheme was used to generate the 225 features, where each feature is the output of the individual observers. These features were then rank ordered by the magnitude of the weights of a perceptron. Once rank ordered, subsets of increasing number of features were selected to be used in another perceptron. This perceptron was trained to minimize mean squared error and the output was a continuous variable representing the likelihood of the region being a nodule. Performance was evaluated by receiver operating characteristic (ROC) analysis and reported as the area under the curve ( $A_Z$ ). The classifier was optimized by adding additional features until the  $A_Z$  declined. The optimized subset of observers then were combined using a third perceptron. A subset of 80 features was selected which gave an  $A_Z$  of 0.972. Additionally, at 98.6% sensitivity, the classifier had a specificity of 71.3% and increased the positive predictive value from 60.7% to 84.1%. Preliminary results suggest that using SRHOs in combination with perceptrons can provide a successful classification scheme for pulmonary nodules. This approach could be incorporated into a larger computer aided detection system for decreasing false positives. © 2001 American Association of Physicists in Medicine. [DOI: 10.1118/1.1420402]

Key words: CAD, pulmonary nodules, classification, image processing, artificial neural networks

## I. INTRODUCTION

It is well established that cancer is one of the devastating and major diseases of our time. Cancer is the second leading cause of death in the United States (US).<sup>1</sup> In 1999 alone, over 1.2 million persons in the US were diagnosed with cancer and approximately 563 100 persons will perish.<sup>1</sup> Of all the types of cancer, lung cancer is the most common cause of death and accounts for about 28% of all cancer deaths. Estimates for 1999 indicate that about 158 900 persons will die from lung cancer in the US and about 171 600 new cases will be diagnosed.<sup>1</sup>

The prime method for cancer detection is through radiological imaging exams.<sup>2</sup> Of these exams, the simplest and most commonly performed diagnostic test is the chest radiograph or chest x ray. It is well established that radiologists may miss up to 30% of pulmonary nodules in a chest x ray. Early detection of these suspicious regions would certainly help improve patient outcome as early detection is key to patient care. With the current advent of digital radiography

systems, we firmly believe that development and application of artificial intelligence techniques for the automated detection of lung cancer will have a great impact on early detection of lung cancer and disease diagnosis.<sup>3</sup> Several studies have found that a computer aided diagnostic (CAD) system, when used in conjunction with a radiologist, does in fact improve<sup>4,5</sup> the detection performance of lung cancer.

In the past, many researchers<sup>6-15</sup> have investigated the use of artificial neural networks (ANNs) and other types of image processing for detection and classification of pulmonary nodules. The overall approach for CAD has been to either use image pixel data directly or to make measurements of pertinent image properties and to combine this information using neural network techniques. Most CAD systems can be viewed as a two-stage approach. Typically, the first stage uses some type of initial linear processing, which has high sensitivity and low specificity, to detect a set of potential nodules. The second stage consists of classifying these

suspicious regions using predictive modeling techniques (ANNs, cluster analysis, etc.) to reject a large number of the false positives. Using this approach, investigators have developed systems which have been able to detect obvious nodules with a receiver operating characteristic (ROC) area of 0.93 and to detect subtle nodules with a ROC area of 0.66.<sup>11</sup> Similar systems are also used to determine if a suspicious nodule is cancerous or benign, with a performance of 0.85.<sup>9</sup>

These previous results have been extremely positive and have helped to improve radiologist performance for diagnostic accuracy.<sup>16</sup> We wish to investigate an innovative new approach based on models from the human vision system for the classification stage of the CAD system. It is well known that most human sensory processes, including the visual system, are understood to work via a linear mechanism followed by a nonlinear integration of features to perform basic decision tasks. In the case of the visual processing system, the linear step is the receptive fields which process basic visual stimuli and are used to reduce data complexity. This linear step is followed by a nonlinear combination of the important data to determine decisions. This multilayered process is what we have chosen as our model.

For the study presented here, we propose to investigate using subregion Hotelling observers in conjunction with simplified artificial neural networks for the automated classification of regions that are suspicious of being solitary pulmonary nodules from a database of "suspicious" regions on chest radiographs.

## II. METHOD

For the study presented here, we are choosing to classify a positive region as any region that the radiologist feels should go to follow-up for additional information. We will not be attempting to classify nodules as benign or malignant. Radiologists do not diagnose cancer versus benign nodules, they detect suspicious regions and send them for additional work up. This classification would not return a positive for regions that were clearly benign, such as rib crossings, vessels on end, clearly benign calcified granulomas and would return a positive for any other region which is suspicious of being a nodule and should be checked with follow-up. In essence, we will be training our classifier as a decision on "suspicious nodule, needs follow-up" or not. This type of system is not a cancer detector. However, it will mirror how a radiologist makes a decision so it can act as a second reader acting on the same logic as the radiologist when coupled with the first stage of a CAD system.

The classification process that we are introducing here reduces to a three-layer scheme. Layer 1 models the linear portion of the visual system. We have chosen to use a grid of subregion Hotelling observers (SRHOs) for this layer. Layer 2 models the data reduction in the visual process and will be performed by examining the rank ordering of the weights of a perceptron. Layer 3 combines the reduced data set to determine final classification via an additional perceptron.

## A. Background

First we would like to present some background material on Hotelling observers (HOs) and perceptrons before we get into the specifics of this study.

### 1. Hotelling observers

Most human sensory processes, including the visual system, are understood to work via a linear mechanism followed by a nonlinear integration of features to perform basic decision tasks. In the case of visual processing, this linear mechanism is the linear receptive fields, which process the basic visual stimuli. For our proposal, we will be using HOs as our linear mechanism. HOs have been shown to be effective in tracking the performance of human observers for detection.<sup>17-22</sup> Many researchers have also used HOs as a means of measuring image quality or as an imaging metric.<sup>23-27</sup>

The Hotelling observer is the optimal linear detector for a known signal, known background, and known covariance matrix when statistics are approximately Gaussian. In the covariance matrix, each element is the covariance between two pixels and the diagonal is the variance for each pixel. For real medical images, where we do not know the exact signal or the distribution of the background, we use estimates of the signal to be detected, the general background, and the image covariance matrix to calculate the set of linear weights for the suboptimal observer. The weights or template for the HO are defined as

$$W = [\langle S+B \rangle - \langle B \rangle] / K, \quad (1)$$

where  $\langle \rangle$  represents the mean,  $S$  is the signal,  $B$  is the background,  $S+B$  is the signal in the background, and  $K$  is the covariance matrix. Multiplying these weights by the image data,  $I$ , and summing over all the pixels gives the test statistic,  $L$ ,

$$L = \sum W * I. \quad (2)$$

This test statistic can be used as a decision variable. If the observer is working properly, the test statistic will be higher in value when the signal is present and lower when it is absent. In white noise, the HO is a matched filter; however, in correlated noise, such as in medical images, this observer estimates a template that decorrelates the noise.<sup>28</sup>

### 2. Perceptrons

The methods of developing the perceptron models which we use have been described in previous studies from our lab and will only be summarized here. A perceptron is an extremely simplified ANN which has no hidden layers and only one processing unit. The perceptron we used is a feed-forward, error-backpropagation processing unit with a sigmoidal activation functions. Each perceptron is presented with the input findings for each case and the corresponding known "truth" outcome. The perceptron merges all the findings to generate a single output value between zero and one corresponding to its prediction of the likelihood of a nodule

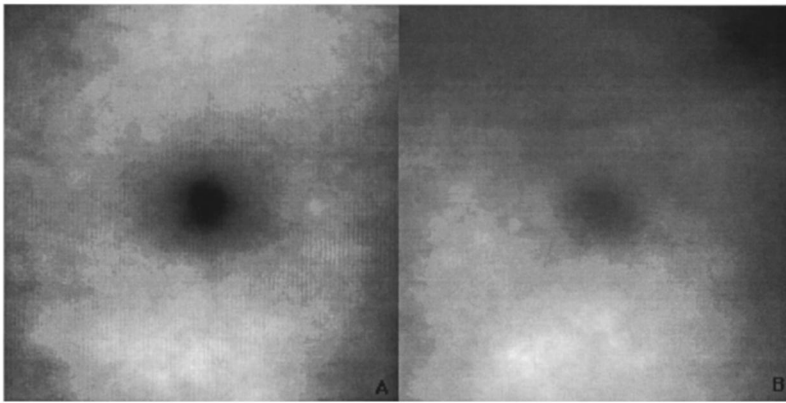


FIG. 1. Average image of the (a) positive and (b) negative ROIs. The difficulty of this database can be seen by observing the obvious signal at the center of the negative ROIs.

being present for that case. The perceptron is trained and learns iteratively under this supervised training process in order to improve its performance.

The model parameters (including training rates, momentum constants, and the number of training iterations) are all optimized empirically. Additionally, a “round robin” or “leave one out” sampling scheme is utilized in order to use all cases for training and testing while still maintaining independence between the training and testing sets. Network training can be halted when the ROC area under the curve,  $A_z$ , is maximized over the testing cases. Our custom ANN and perceptron software was written in C and runs on a Sun Ultra 60 workstation (Sun Microsystems Inc., Mountain View, CA). Initial training requires up to several minutes for each new combination of parameters, but a finalized perceptron can evaluate each new case within a fraction of a second.

As stated previously, for each case the model produces a prediction number between zero and one. To use the perceptron as a diagnostic aide, one could select a certain threshold value such that those cases with output values below the threshold would be considered unlikely nodules. The remainder of cases with values exceeding the threshold would be considered suspicious for pulmonary nodules. The sensitivity is the number of correctly classified nodules divided by the number of all actual nodules; the specificity is the number of correctly diagnosed negative lesions out of all actual negative lesions. Varying this threshold value results in a trade off between sensitivity and specificity and will generate an ROC curve for analysis.

## B. Image database

We have previously<sup>29</sup> collected a database of 239 regions of interest (ROIs) for nodule classification and detection studies. Each ROI was 256 by 256 pixels and was spatially averaged to make the ROIs 128 by 128 pixels. For the purposes of this database, a nodule was defined as any lesion that should be sent to follow-up to rule out being a nodule. Negatives are regions that contain no nodules, obviously benign calcified nodules, rib crossings, and vessels. The database breaks down into two distinct groupings of regions. The first group is collected from patients who were identified

with areas which were “suspicious” for having a pulmonary nodule and underwent fluoroscopy (142 regions). The second grouping consists of patients who had obvious nodules on chest radiographs and underwent computed tomography exams for further evaluation (97 regions). All of the original images were taken between 1991 and 1996. For the database, a “truth” file was prepared by two board certified radiologists for the digitized (Lumiscan model 75; Lumisys, Sunnyvale, CA) 2048 pixel by 2048 pixel images based on the PA radiograph, CT results when applicable, the full radiology report, and the pathology report when applicable. Overall, the database consists of 94 negative ROIs and 145 positive ROIs. Since under initial examination, all of these regions were deemed “suspicious” enough for follow-up, we can consider them all false positives. Considering that 145 of these ROIs were actually positive gives an original positive predictive value of 60.7% for the radiologists.

Please note that for this database, all of the negative regions were deemed suspicious for a nodule upon initial examination by the radiologist, making this a very difficult database. The average of all of the (a) positive and (b) negative ROIs are shown in Fig. 1. The difficulty of this database can be seen by observing the obvious signal in the center of the negative ROIs.

In addition to this database, three radiologists previously performed a ROC study over all of the images by selecting a probability of the region being a nodule for 237 of the regions in this database (two regions were used for training). Analysis of the radiologists’ ROC ratings yielded areas which ranged from 0.72 to 0.83. This level of radiologist performance for area under the ROC curve corresponds well to other studies of lung nodule databases for sets of cases which were deemed to be between very subtle (0.753) and subtle (0.876) in difficulty.<sup>30</sup>

## C. Description of three layer classifier

Figure 2 presents a flow chart of the three layer classifier and a detailed description of each layer is now presented.

### 1. Layer 1: Subregion Hotelling observers

The Hotelling observer is the optimal linear detector for a known signal given certain information about the imaging

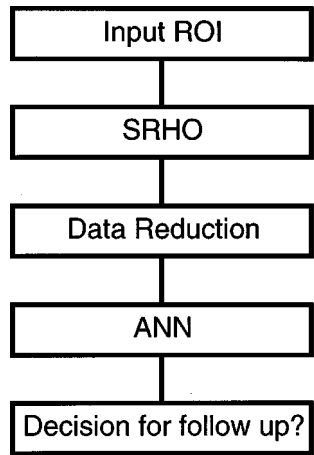


FIG. 2. Flow chart of three layer classifier.

system and the object to be detected. The HO uses estimates of the signal to be detected, the general background, and the image covariance matrix to calculate a set of linear weights. The HO results in an output value by summing up the values of the weights multiplied by the image data. The higher this output weight, the more likely it is to have the signal present. Application of the HO to a large ROI is prohibitive, as too many image samples would be needed to properly estimate the covariance matrix. For example, in the database we have developed, the 128 by 128 pixel ROIs would require a covariance matrix of 16 384 ( $128 \times 128$ ) by 16 384 ( $128 \times 128$ ) elements. Collecting a database of real images large enough to obtain a stable estimate of a covariance matrix this size would prove to be exceedingly difficult.

To combat this size difficulty, many researchers have investigated using a channelized HO model, where radially symmetric vision channels are used to reduce the dimensionality of the problem. Initially we tried this approach, only to find that it did not work well for lung nodule detection. We believe that this failure was due to neither the normal anatomy nor the nodule signal in the lungs being radially symmetric. The need to relax this radial symmetry constraint lead us to develop new methods to use the pixel-wise HO.

Instead of using the pixel-wise HO over an entire ROI, we chose to use a subregion Hotelling observer over a small region of each ROI, because a small region observer would require significantly fewer samples to properly estimate the necessary covariance matrix. To cover the whole region of the ROI, we then chose to tile a matrix of SRHOs over the full region being examined (see Fig. 3). This tiling process results in many SRHOs being used to reduce the complexity of the image data down to the number of SRHOs used. Since each small observer will be “observing” or analyzing every pixel in a small region, it will be sensitive to changes in high frequency noise power spectra and structured noise, including anatomy. The result of each SRHO on a single ROI is a scalar. If a database of  $M$  ROIs is examined, then the output is a vector of length  $M$  for that specific small observer. If  $N$  SRHOs are used, then  $N$  vectors or “features” will be the output. We consider the output of each SRHO as a “feature” as it examines different regions of the ROI, similar to applying different texture measurements to a ROI.

## 2. Layer 2: Rank ordering and data reduction

The result of layer 1 of the classifier is a number of image features, where each feature is the result of the application of

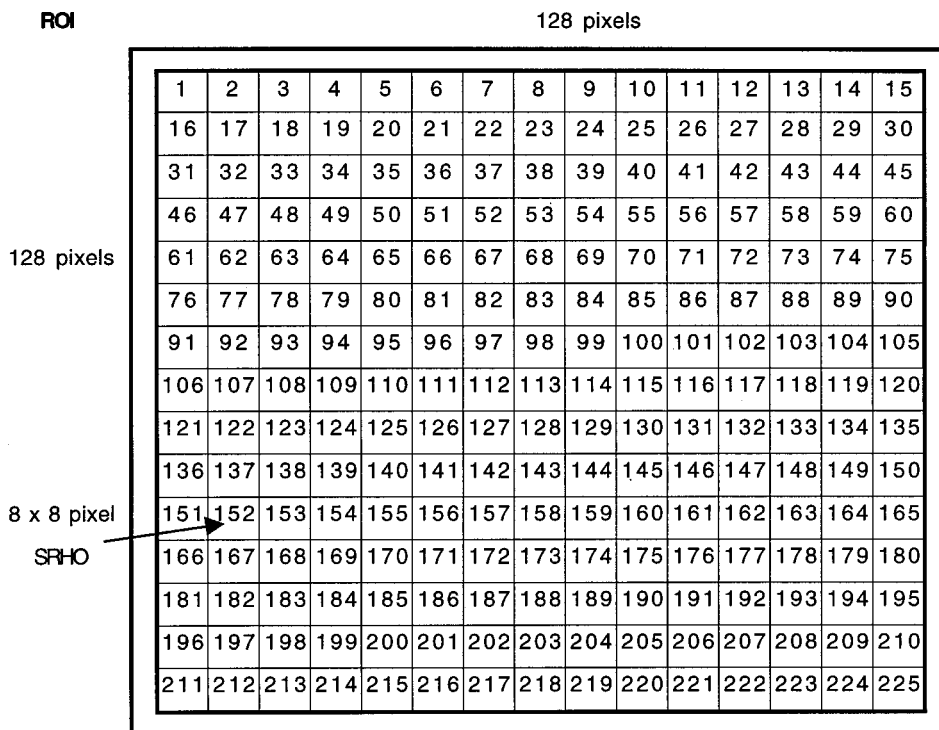


FIG. 3. The 15 by 15 grid of SRHO is shown as placed in each ROI. Each of the SRHOs covers 8 pixels by 8 pixels. The overall ROI is 128 pixels by 128 pixels, where 120 by 120 of them are covered by the different observers.

a SRHO to one particular subregion of the full ROI. To further reduce the dimensionality and to simplify the classifier, a rank ordering of the important features was needed. Once the features are rank ordered, different subsets can be examined to find which one provides the highest classification performance. For the study presented here, the SRHO features were rank ordered using the absolute value of the weights of a perceptron (P1).

To obtain the rank ordering, the perceptron P1 was trained on the output of all the SRHOs for all the ROIs using a leave one out methodology until the mean squared error was minimized. Each of the inputs was normalized from 0 to 1. The overall weights of the perceptron were examined and then rank ordered by the absolute values with the assumption that a large positive or large negative weight would correspond to features that had a larger effect on the outcome.

Once all of the features were rank ordered, an additional perceptron (P2) was used to determine the optimal subset of features. To expedite this feature elimination process and ensure fair comparisons across different subsets of features, P2 used fixed training parameters such as training rates and the number of training iterations. In a manner analogous to stepwise discriminant analysis, the rank ordered features were added in groups of five, starting from the most significant features, and the perceptron P2 was re-optimized. The process was repeated, until the round robin  $A_Z$  started to decline, indicating the P2 perceptron was becoming overtrained. This reduced subset of features which yielded the highest  $A_Z$  was then forwarded to the next layer.

### 3. Layer 3: Combination and classification

The reduced feature data set was then used as the input to another perceptron (P3). Unlike P2 from the previous layer, P3's training parameters were carefully varied to optimize  $A_Z$ . With this final optimized perceptron, the final decision to classify a region as a pulmonary nodule or not was made and the overall system analysis was performed. The SRHO features each examine the high frequency content of a certain region in the ROI. By combining the SRHO features with a perceptron, low frequency content is also being incorporated into the classification procedure, as using many of the tiled SRHO will span portions of the entire ROI.

### D. Procedure

For the study presented here, 225 separate SRHOs were arranged in a 15 by 15 grid across the 128 by 128 pixel ROIs. Each of the SRHOs was designed to "observe" an 8 by 8 pixel subregion within the ROI. Hence, the 15 by 15 SRHOs covered the center 120 by 120 pixel region of the ROI (see Fig. 3). A leave-one-out training and testing methodology was used to generate 225 ( $15 \times 15$ ) features, where each feature is the output of each individual SRHO. Signal and background were modeled as the average of the positives and negatives (minus the testing case) and the overall covariance matrix was formed by an appropriately weighted combination of the positive and negative covariance matrices. All of the test values were then collected and organized as fea-

### Comparison of SRHO/perceptron System to Radiologists

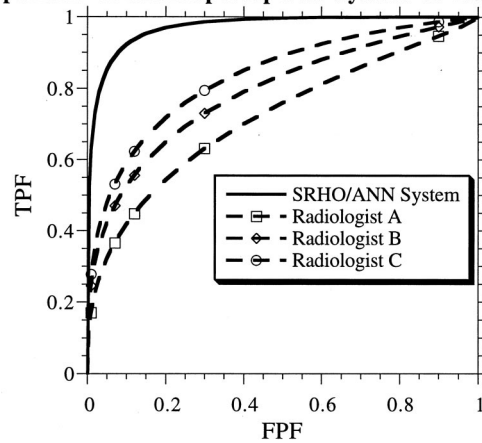


Fig. 4. Comparison of ROC curves for the SRHO/perceptron system to three radiologists. The  $A_Z$  for the SRHO/perceptron system is 0.972, while the  $A_Z$  for the three radiologists were 0.718, 0.789, and 0.834.

tures. The result of this first step was a data reduction from a 128 by 128 pixels region to 225 values or features per ROI. Each SRHO covers approximately a 3 by 3 mm area.

For layer 2, these 225 features were used as the inputs to a perceptron (P1) which was trained to minimize the mean squared error based on the truth file. The features were then rank ordered by examining the absolute value of the weights of the trained perceptron. Another perceptron (P2) was then trained on a reduced set of the features. Starting from the most significant, features were added five at a time and the P2 perceptron was trained and tested. This procedure continued until the value of the area under the ROC curve started to decline, at which point the optimal subset of features was chosen.

For the last layer, another perceptron (P3) was then optimized on the specific reduced set of features and additional metrics were calculated and tests were performed. Calculations of ROC area and partial area, as well as statistical comparisons of those metrics, were performed using the ROCKIT program (Charles Metz, University of Chicago).

## III. RESULTS

After application of the first layer, the highest ROC  $A_Z$  for any individual SRHO was 0.65. For layer 2, the 225 features were then rank ordered and examined in groups with increments of five features. The maximal ROC  $A_Z$  was recorded when a subset containing the top 80 features was used. This subset of features was next used as the input to perceptron P3, which was then optimized. The area under the curve for the final classification was  $0.972 \pm 0.009$ , which corresponded to a partial area under the curve (the normalized area above 90% sensitivity on the ROC curve) of  $0.813 \pm 0.0524$ . For comparison, the  $A_Z$  for the three radiologists who performed the same ROC study was  $0.718 \pm 0.033$ ,  $0.789 \pm 0.029$ , and  $0.834 \pm 0.026$ . Figure 4 shows a comparison of the four different ROC curves for this study. The statistical test to compare the SRHO/perceptron system to

each radiologist was performed and a  $p$  value of less than 0.001 was calculated. Additionally, at 98.6% sensitivity, the overall classifier had a specificity of 71.3% and a positive predictive value of 84.1%.

#### IV. DISCUSSION

The purpose of this study was to investigate the use of subregion Hotelling observers in conjunction with simplified artificial neural networks for the automated classification of regions suspicious of being solitary pulmonary nodules from a database of regions from chest radiographs. The exact classification task was to determine if a region should be sent to follow-up for being suspicious of being a nodule or if no follow-up is needed. Rib crossings, vessels, and calcified nodules need no further follow-up. This type of classifier is being developed to mimic the way in which a radiologist views a chest x ray. We are choosing to classify a positive region as any region that the radiologist feels should go to follow-up for additional information. We will not be attempting to classify nodules as benign or malignant. In essence, we will be training our classifier as a decision on "suspicious nodule, needs follow-up" or not. For this database, all of the negative regions were deemed suspicious for being a nodule upon initial examination by the radiologist. Because of this fact, all of our negatives can be considered false positives by the radiologist and this gives an original positive predictive value of 60.7% for the radiologists.

A three layer classifier was developed and tested on the above database. The first layer is based on subregion Hotelling observers, the second layer performs rank ordering and data reduction, and the third layer does the final combination and classification of the remaining features. This three layer classifier was then trained and tested on the database and area under the ROC curve was calculated as  $0.972 \pm 0.009$ , which corresponds to a partial area of  $0.813 \pm 0.0524$ . Comparing the performance of this system to that of the three radiologists who had previously performed the ROC study on this database ( $0.718 \pm 0.033$ ,  $0.789 \pm 0.029$ , and  $0.834 \pm 0.026$ ) gave a  $p$  value of less than 0.001. Additionally, we calculated the specificity of the system at 98.6% sensitivity to be 71.3%, with a corresponding positive predictive value of 84.1%. At this threshold setting, two positive cases would be missed, while 67 of the 94 suspicious regions would be correctly identified as negative.

For comparison, Wu *et al.* showed that their system gave  $A_c$  values of 0.93 for obvious positive nodules and 0.66 for subtle nodules.<sup>11</sup> Their database was created from a set of false positives generated from the front end of a CAD system, while ours were generated from false positives from radiologists. The performance of our radiologists on the ROC study suggest that our database consists of nodules that are very subtle (0.753) to subtle (0.876),<sup>30</sup> and thus our initial classification results are very impressive and show much promise.

While the system developed here is not a CAD system in itself, we envision two ways in which it can be used. First, it could be used by radiologists as a classification tool. The

radiologist would point to a region and the classifier would determine if the region was a nodule or not. This could help the radiologists to be more confident in their decisions. While we have been told by radiologists that this type of tool would be useful, we feel there are more strides to be gained by incorporating this into a CAD system, which could potentially help radiologists improve upon the detection task. To incorporate this system into CAD, we would need a front end linear image processing tool which would have high sensitivity and be able to select regions that are potentially nodules. The SRHO/perceptron system could then be used to reduce the overall number of false positives and improve the overall performance of the CAD system. Incorporating this system into CAD is one of the future studies we wish to examine.

In conclusion, our preliminary results suggest that using subregion Hotelling observers in combination with perceptrons can provide a successful classification scheme for subtle and difficult pulmonary nodules. Further research will allow this approach to be incorporated into a larger computer aided detection system to aid radiologists in the detection of pulmonary nodules.

#### ACKNOWLEDGMENTS

The authors would like to thank Neal Vittitoe, Ph.D., Rene Vargas-Voracek, Ph.D., and Jeff Drayer, M.D. for collecting the lung nodule database. Additionally, we would like to thank Edward Patz, M.D., Paige McAdams, M.D., and Carl Ravin, M.D. for their help in determining the truth files for the database. We would also like to give thanks to Jeremy Erasmus, M.D., William Foster, M.D., and Phyllis Kornguth, M.D., Ph.D. for participating in the observer study.

<sup>a</sup>)Electronic mail: alan.baydush@duke.edu

<sup>b</sup>)Also at: Department of Medical Physics and Imaging, Cedars-Sinai Medical Center.

<sup>1</sup>S. H. Landis, T. Murray, S. Bolden, and P. A. Wingo, "Cancer statistics, 1999 [see comments]," *Ca—Cancer J. Clin.* **49**, 8–31 (1999).

<sup>2</sup>K. Shaffer, "Role of radiology for imaging and biopsy of solitary pulmonary nodules," *Chest* **116**, 519S–522S (1999).

<sup>3</sup>H. MacMahon, K. Doi, C. H., M. L. Giger, S. Katsuragawa, and N. Nakamori, "Computer-aided diagnosis in chest radiology," *J. Thorac. Imaging* **5**, 67–76 (1990).

<sup>4</sup>T. Kobayashi, X. W. Xu, H. MacMahon, C. E. Metz, and K. Doi, "Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs," *Radiology* **199**, 843–848 (1996) (96220553).

<sup>5</sup>M. L. Giger, K. Doi, H. MacMahon, C. E. Metz, and F. Yin, "Pulmonary nodules: Computer-aided detection in digital chest images," *RadioGraphics* **10**, 41–51 (1990).

<sup>6</sup>J. S. Lin, A. Hasegawa, M. T. Freedman, and S. K. Mun, "Differentiation between nodules and end-on vessels using a convolution neural network architecture," *J. Digit. Imaging* **8**, 132–141 (1995).

<sup>7</sup>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* **8**, 1201–1214 (1995).

<sup>8</sup>S. C. B. Lo, S. L. A. Lou, J. S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE Trans. Med. Imaging* **14**, 711–718 (1995).

<sup>9</sup>K. Nakamura, H. Yoshida, R. Engelmann, H. MacMahon, S. Katsuragawa, T. Ishida, K. Ashizawa, and K. Doi, "Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks," *Radiology* **214**, 823–830 (2000).

- <sup>10</sup>Y. C. Wu, K. Doi, M. L. Giger, C. E. Metz, and W. Zhang, "Reduction of false positives in computerized detection of lung nodules in chest radiographs using artificial neural networks, discriminant analysis, and a rule-based scheme," *J. Digit Imaging* **7**, 196–207 (1994).
- <sup>11</sup>Y. C. Wu, K. Doi, and M. L. Giger, "Detection of lung nodules in digital chest radiographs using artificial neural networks: a pilot study," *J. Digit Imaging* **8**, 88–94 (1995).
- <sup>12</sup>X. W. Xu, K. Doi, T. Kobayashi, H. MacMahon, and M. L. Giger, "Development of an improved CAD scheme for automated detection of lung nodules in digital chest images," *Med. Phys.* **24**, 1395–403 (1997) (97449551).
- <sup>13</sup>S. Garg and C. E. Floyd, Jr., "Artificial neural network analysis for nodule detection in digital chest radiographs," *Radiology* **181**(P), 144 (1991) (abstract).
- <sup>14</sup>M. G. Penedo, M. J. Carreira, A. Mosquera, and D. Cabello, "Computer-aided diagnosis: A neural-network-based approach to lung nodule detection," *IEEE Trans. Med. Imaging* **17**, 872–880 (1998) (99156631).
- <sup>15</sup>S. B. Lo, M. T. Freedman, J. S. Lin, and S. K. Mun, "Automatic lung nodule detection using profile matching and back-propagation neural network techniques," *J. Digit Imaging* **6**, 48–54 (1993).
- <sup>16</sup>H. MacMahon, R. Engelmann, F. M. Behlen, K. R. Hoffmann, T. Ishida, C. Roe, C. E. Metz, and K. Doi, "Computer-aided diagnosis of pulmonary nodules: Results of a large-scale observer test," *Radiology* **213**, 723–726 (1999) (20056719).
- <sup>17</sup>R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, "Psychophysical study to test the ability of the Hotelling trace criterion to predict human-performance," *J. Opt. Soc. Am. A Opt. Image Sci. Vis* **3**, P126–P126 (1986).
- <sup>18</sup>R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, "Hotelling trace criterion and its correlation with human-observer performance," *J. Opt. Soc. Am. A Opt. Image Sci. Vis* **4**, 945–953 (1987).
- <sup>19</sup>H. C. Gifford, M. A. King, D. J. de Vries, and E. J. Soares, "Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J. Nucl. Med.* **39**, 771 (1998).
- <sup>20</sup>H. C. Gifford, R. G. Wells, and M. A. King, "A comparison of human observer LROC and numerical observer ROC for tumor detection in SPECT images," *IEEE Trans. Nucl. Sci.* **46**, 1032–1037 (1999).
- <sup>21</sup>H. C. Gifford, M. A. King, D. J. de Vries, and E. J. Soares, "Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J. Nucl. Med.* **41**, 514–521 (2000).
- <sup>22</sup>S. D. Wollenweber, B. M. W. Tsui, D. S. Lalush, E. C. Frey, K. J. LaCroix, and G. T. Gullberg, "Comparison of Hotelling observer models and human observers in defect detection from myocardial SPECT imaging," *IEEE Trans. Nucl. Sci.* **46**, 2098–2103 (1999).
- <sup>23</sup>H. H. Barrett, T. Gooley, K. Girodias, J. Rolland, T. White, and J. Yao, "Linear discriminants and image quality," *Image Vis. Comput.* **10**, 451–460 (1992).
- <sup>24</sup>H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for assessment of image quality," *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9758–9765 (1993).
- <sup>25</sup>H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. 2. Fisher information, Fourier crosstalk, and figures of merit for task-performance," *J. Opt. Soc. Am. A Opt. Image Sci. Vis* **12**, 834–852 (1995).
- <sup>26</sup>W. E. Smith and H. H. Barrett, "Hotelling trace criterion as a figure of merit for the optimization of imaging-systems," *J. Opt. Soc. Am. A Opt. Image Sci. Vis* **3**, 717–725 (1986).
- <sup>27</sup>H. H. Barrett, C. K. Abbey, and B. G. Gallas, "Stabilized estimates of Hotelling-observer detection performance in patient-structured noise," *Proc. SPIE* **3340**, 27–43 (1998).
- <sup>28</sup>M. Eckstein, C. Abbey, and J. Whiting, "Human vs model observers in anatomic backgrounds," *Proc. SPIE* **3340**, 16–26 (1998).
- <sup>29</sup>J. A. Drayer, N. Vittitoe, R. Vargas-Voracek, A. H. Baydush, C. E. Floyd, Jr., and C. E. Ravin, "Characteristics of regions suspicious for pulmonary nodules on chest radiographs," *Acad. Radiol.* **5**, 613–619 (1998).
- <sup>30</sup>J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *Am. J. Roentgenol.* **174**, 71–74 (2000).