Pattern Recognition 42 (2009) 1058 – 1066

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

New models for region of interest reader classification analysis in chest radiographs ☆

M.S. Pattichis[a,*], T. Cacoullos[b], Peter Soliz[c]

[a]Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM 87131-0001, USA
[b]Department of Mathematics, National and Capodistrian University of Athens, Athens Panepistimioupolis, GR-157 84, Athens, Greece
[c]VisionQuest Biomedical, LLC, 4801 Lang NE, Suite 110, Albuquerque, NM 87109, USA

ARTICLE INFO

Article history:
Received 3 December 2007
Received in revised form 6 September 2008
Accepted 29 September 2008

PACS:
87.85.Tu
10.−v
75.Pq

Keywords:
Region of interest classification
Chest radiographs
ROC analysis
Binary classification
Pneumoconiosis
Modeling biomedical systems
Logic, set theory, and algebra
Mathematical procedures and computer techniques

ABSTRACT

In several computer-aided diagnosis (CAD) applications of image processing, there is no sufficiently sensitive and specific method for determining what constitutes a normal versus an abnormal classification of a chest radiograph. In the case of lung nodule detection or in classifying the perfusion of pneumoconiosis, multiple radiograph readers (radiologists) are asked to examine and score specific regions of interest (ROIs). The readers provide size, shape and perfusion grades for the presence of opacities in each region and then use all the ROI grades to classify the lung as normal or abnormal. The combined grades from all readers are then used to arrive at a consensus normal or abnormal classification. In this paper, using area under the ROC curve, we evaluate new mathematical models that are based on mathematical statistics, logic functions, and several statistical classifiers to analyze reader performance in grading chest radiographs for pneumoconiosis as the first step toward applying this technique to early detection of nodules found in lung cancer. In pneumoconiosis, rounded opacities are on the order of 1–10 mm in size, while lung nodules are often not diagnosed until they reach a size on the order of 1 cm.

1. Introduction

Lung cancer has one of the worst survival rates of all the types of cancers. The survival rates have been found to be directly related to its growth rate and size of the pulmonary nodules when detected. Chest radiographs continue to be one of the most common modalities used in detecting the lung nodules. Their early detection in chest radiographs is one of the most challenging tasks performed by radiologists. computer-aided diagnosis (CAD) techniques have been shown to be effective in improving sensitivity and specificity of the radiologists [1–3]. There remains, however, a desire to improve the specificity of current systems in order to reduce the number of false positives without sacrifice of the sensitivity [4].

For early detection, our ultimate goal is to achieve high sensitivity and specificity for small (< 1 cm), isolated lung nodules. To test our approach we have applied and tested our models on small rounded opacities (1.5–10 mm) found in patients who present with pneumoconiosis. In this application of CAD, radiologists are asked to score the degree of interstitial lung disease, such as pneumoconioses, from exposure to dust in work environments.

The pneumoconioses include a number of interstitial lung diseases brought on by the inhalation of dusts in man-made environments, such as coal and other types of mining. Asbestosis, silicosis, and coal workers' pneumoconiosis (CWP) are the primarily lung tissue diseases that are of concern. After years of background exposure or after only a few years of intense exposure, these diseases can progress rapidly and lead to severe lung function impairment. In addition, several studies have shown a positive correlation between CWP or progressive massive fibrosis (PMF), and lung cancer, especially in individuals who smoke. In one form of pneumonconiosis, silicosis, Honma et al. [5] autopsied 764 non-asbestosis cases and found a statistically significant number, 19.1%, with lung cancer. Others have found similar results. For example, Ebihara [6] discovered 48 cases of carcinoma in 450 autopsies (10.7%), while Katabami et al. [7] reported 55 cases in 563 patients (9.7%). The American Thoracic Society (ATS) has described the adverse health effects of exposure to crystalline silica, including lung cancer [8]. Chest radiographs are used for screening and detection of pneumoconiosis and other lung abnormalities, such as lung cancer.

☆Funding for this study was provided through a grant from the National Institute for Occupational Safety and Health (NIOSH), Grant 2R44 OHRGM 03595.

* Corresponding author. Tel.: +1 505 277 0486; fax: +1 505 277 1439.
E-mail addresses: pattichis@ece.unm.edu (M.S. Pattichis), tkakoul@math.uoa.gr (T. Cacoullos), psoliz@visionquest-bio.com (P. Soliz).

0031-3203/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.09.021

For pneumoconiosis, human readers are first asked to provide ratings over six different lung regions, and then use their ratings to provide an overall classification. The procedure is described in the guidelines provided by the International Labor Organization (ILO) [9] and more recently in Ref. [10]. In this paper, we propose several classification techniques for characterizing the performance of the human readers, estimating ground truth and establishing symmetry.

The problem of developing methods for characterizing reader classification is complicated when there is no absolute agreement on what constitutes ground truth. Usually, a region of interest (ROI) is graded by multiple readers, each giving his or her own grade. In this case, it is often found that there is strong variability between the grades. Both the inter-observer and intra-observer variability must be considered [11–15]. In the case of significant differences in the grades between two readers, both readers will meet and agree on a common grade [9]. A similar approach can be followed for more than two readers.

In our study, we develop mathematical models for characterizing and approximating human reader performance. We investigate the use of five statistical classifiers [16–20], ranging from the Logistic classifier, a Bayesian classifier, the *K*-means classifier, to sum and weighted sum classifiers that add up the positive grades in the ROIs. We provide a common, logic function interpretation for all classifiers, allowing us to investigate several classification issues (see Ref. [21], and more recently Ref. [22]). For establishing right–left symmetry, we use weighted kappa statistics [23]. For all classifiers, we compute receiver operating characteristic (ROC) curves, and use the estimated area under each curve as a performance measure [24].

Our approach is related to the development of CAD systems, where we are interested in establishing and understanding how the different ROIs relate to the overall chest radiograph classification (see Ref. [25] for a recent example on the use of symmetry). Research for developing CAD for pneumoconiosis dates back to the 1970s [26–28], with a recent resurgence of interest in the late 1990s [29–35]. More generally, a review of CAD systems for chest radiographs can be found in Ref. [36], while texture analysis methods are given in Ref. [37].

We next provide background information in Section 2. The method is then presented in Section 4. After establishing notation, we provide further motivation for our approach in Section 3. Results are given in Section 5 and a careful discussion is provided in Section 6. Concluding remarks are given in Section 7.

2. Background

2.1. A mathematical model

Given an image, let I denote the set of pixels that are in the spatial support of the image. Define a partition of I by $I = R_1 \cup R_2 \cup \cdots \cup R_M \cup B$, where R_i denotes a non-overlapping ROI, M denotes the number of regions, and B denotes the background region that is not of diagnostic interest. A grade over the ROIs is defined as a random variable X_i of each ROI into a set of binary values $X_i : R_i \to \{0,1\}$. Here, the grade refers to the radiologists' perceptual determination whether lung disease is present or not.

Using the collection of all the grades found for all ROIs $Y = (X_1, X_2, \ldots, X_M)$, a classification decision is made for the entire image. We use the symbol C_i to denote the ith classifier (which could represent a human reader or a computer classification system), and note that $C_i : \{0,1\}^M \to \{0,1,\ldots,P-1\}$, where $\{0,1\}^M$ denotes the M-fold tensor product of the integer set $\{0,1\}$, and P denotes the maximum number of possible classifications.

2.2. A model for chest radiographs

For chest radiographs of the lung, the mathematical model is simplified, since we only need to describe the presence of a mark ($\times$) or its absence in each region. When a mark is present in the ith lung region, we have that $X_i = 1$. When a mark is absent, we thus set the corresponding random variable to zero. In this case, the random variables become binary $X_i : R_i \to \{0,1\}$, and the joint distribution for all ROIs $Y = (X_1, X_2, \ldots, X_M)$ becomes a multivariate Bernoulli.

The international labor organization medical protocol for pneumoconiosis divides the lung into six ROIs [9]. We are naturally mostly interested in small opacities, where opacity presence may not be so obvious. Our focus is also limited to regular opacities of type q, for which we could establish the largest population for our study.

We use R_1, R_2, R_3 for the right lung and R_4, R_5, R_6 for the left lung. We use R_1, R_2 for the two upper regions, R_2, R_5 for the middle ones, and we divide the lung into six ROIs and use a six-dimensional Bernoulli random vector $\mathbf{x} = (x_1, \ldots, x_6)$ to denote the collection of all of the ROIs. Here, the presence of lung opacities is signified by assigning a value of 1 to the corresponding random variable. Thus, $x_1 = 1$ indicates that the reader has marked the presence of opacities in the upper-right lung. For the classification of the entire chest radiograph, we will only consider two cases: normal (0) and abnormal (1).

Right lung	Left lung
x_1 (UR)	x_4 (UL)
x_2 (MR)	x_5 (ML)
x_3 (LR)	x_6 (LL)

Thus, both the ROIs and the classification of the entire chest radiograph can be thought of as binary random variables. To model the classification by each reader in terms of the ratings of the individual ROIs, we want to consider binary logic functions $C(x_1, \ldots, x_6)$. Generally, this formulation is not possible, since it is possible to have the same binary vectors to map to both normal and abnormal classifications. To understand how this situation arises, consider the case that we have two different patients with the same ROI binary vectors. If a human reader classifies one of the lungs as normal and the other one as abnormal, we have that the same ROI binary vector will map to both normal and abnormal. To avoid this problem, for each human reader, for each unique ROI binary vector, we count the number of times that he/she classifies a given binary vector as normal and the number of times that he/she classifies it as abnormal. Then, if the ROI pattern is most often classified as being normal, we assign a 1 to the classifier function. Else, we assign a 0. We use the term *maximum-likelihood* classification to describe this binary classification approach.

3. Motivation

The use of logic functions will provide us with an intuitive approach for summarizing binary classifiers. We will also use logic functions to address issues associated with classifier symmetry, disease growth and generalizability.

In logic function theory, optimal sums of products provide us with an efficient way of describing the data set in terms of elementary prototypes; the products. This approach provides us with an efficient way for describing arbitrary binary classifiers. For example, the classifier represented by the optimal sum $C(x_1, \ldots, x_6) = x_1 x_4 + x_5'$, summarizes the classifier in terms of two simple lung-prototypes $x_1 x_4$ and x_5. To determine the output for any input lung pattern, we simply need to examine if both the upper lung regions are marked $(x_1, x_4 = 1)$ or whether the middle-left lung region is not marked

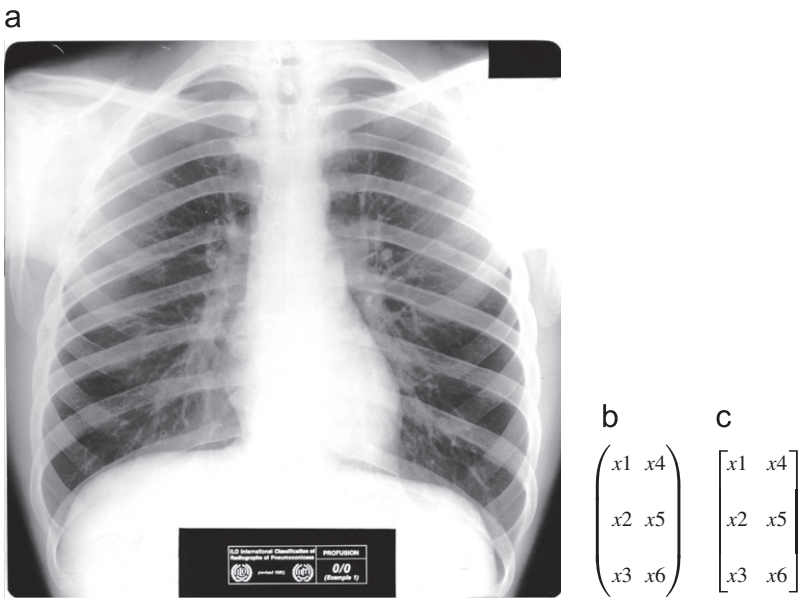

Fig. 1. Lung opacity representation using logic expressions. The lung is divided into six-subregions: upper right (x_1), middle right (x_2), lower right (x_3), upper left (x_4), middle left (x_5) and lower left (x_6). We show: (a) An international labor organization (ILO) standard chest radiograph, (b) the corresponding product term $x_1x_2 \ldots x_6$, and (c) the corresponding sum term $x_1 + \cdots + x_6$. Here, we note that product terms require the presence of lung opacities in each region. For example, x_1x_2 implies opacity presence in both the first and the second region. In contrast, sum terms only require the presence of lung opacities in at least one region.

($x_5 = 0$). If either condition is satisfied, the lung should be classified as positive. Else, we should classify it as negative. To visualize the lung prototypes, we group them together in the same shape as the chest radiograph. This is illustrated in Fig. 1(b). Thus, in our example, we would write

$$C(x_1, \ldots, x_6) = x_1x_4 + x_5' = \begin{pmatrix} 1 & 1 \\ - & - \\ - & - \end{pmatrix} + \begin{pmatrix} - & - \\ - & - \\ - & 0 \end{pmatrix},$$

where we place a 1 in regions that must be marked, a 0 is placed for regions that must not be marked and a $-$ is placed for regions that are not considered.

Similarly, the optimal product of sums representation provide us with an alternative way of describing the classifiers in terms of elementary prototypes. For example, the classifier represented by the optimal product of sums $C(x_1, \ldots, x_6) = (x_1 + x_4)(x_2 + x_5)$, requires that both lung prototypes $x_1 + x_4$ and $x_2 + x_5$ are satisfied. Thus, in this case, a lung would need to be marked as positive in at least one of the upper regions ($x_1 = 1$ or $x_4 = 1$) and also in at least one of the middle regions ($x_2 = 1$ or $x_5 = 1$). For visualizing sum-terms we use square brackets as illustrated in Fig. 1(c). In this case, the classifier would be summarized as

$$C(x_1, \ldots, x_6) = (x_1 + x_4)(x_2 + x_5) = \begin{bmatrix} 1 & 1 \\ - & - \\ - & - \end{bmatrix} \cdot \begin{bmatrix} - & - \\ 1 & 1 \\ - & - \end{bmatrix}.$$

We also note that these representations are optimal in the sense that we cannot find another representation with fewer terms and/or fewer variables. However, it maybe possible to find an equivalent representation with the same number of terms and the same number of variables [21].

Given the relative symmetry between the right and the left lungs, we want to establish whether the classified lung examples follow this symmetry. We will discuss a statistical method for establishing this symmetry in Section 5.1. Here, in terms of logic functions, we note that a symmetric classifier would remain invariant when we exchange the left variables with the right variables:

$$C_{LR}(x_1, x_2, x_3, x_4, x_5, x_6) = C_{LR}(x_4, x_5, x_6, x_1, x_2, x_3). \quad (1)$$

Now, if the location of the marked disease is irrelevant, the classifier function generalizes Eq. (1) to

$$C_{sym}(x_1, x_2, x_3, x_4, x_5, x_6) = C_{sym} \text{ (any permutation of } x_1, \ldots, x_6). \quad (2)$$

To establish disease growth, let us introduce set notation for the binary variables. Let the set S denote the list of variables that are marked positive. For example, if $x_1 = x_4 = 1$, we would have that $S = \{x_1, x_4\}$. Now, suppose that a classifier marks S as positive: $C(S) = C(x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 0, x_6 = 0) = 1$. Then clearly, if the disease grows, we would expect that more variables would be marked as positive. Here, we refer to this disease growth criterion as it applies to different radiographs from either the same person or from different persons, from the same group of people. As our medical team suggested, in pneumoconiosis, the disease progresses from the top to the bottom of the lung. Thus, individuals with disease in the middle and lower parts of the lung would be expected to exhibit the disease in the top parts of the lung. In our example, please recall that $x_1 = x_4 = 1$ refers to the presence of disease in the top regions. Thus, progression to lower regions implies that any new pattern A will include x_1, x_4 in addition to more, lower regions. Clearly, this model can be extended to any growth and it is not limited to our example. Mathematically, we write

$$C(A) = 1 \quad \text{for all } A \text{ that satisfy } S \subseteq A. \quad (3)$$

It is important to establish that the requirements (1)–(3) are not mutually exclusive. We clearly have that Eq. (1) implies Eq. (2). As we shall show next, a simple summation classifier satisfies all requirements. For the summation classifier, we first count the number of positive ratings using (for ordinary addition)

$$s(\mathbf{x}) = x_1 + \cdots + x_6$$

and then allocate the radiograph as normal if $s(\mathbf{x}) < N$ (some positive integer N). Else, the radiograph is classified as abnormal.

A simple generalization of the summation classifier is a weighted classifier $s(\mathbf{w}^T\mathbf{x})$ for some weight vector $\mathbf{w}$. It is easy to see that this classifier can satisfy Eqs. (1) and (3) for symmetric weights.

A related, powerful binary classifier is the Logit classifier (see Section 4.4.2). We will consider the summation classifier, a weighted sum classifier and the Logit classifier in Section 4.4.

It is interesting to note that the requirements of Eqs. (1)–(3) also affect the generalizability of the classifier. Clearly, both symmetry and disease growth requirements provide us with a recipe of required classifier performance for inputs that may not even exist in the training data set. For example, if the training set requires that $x_1 = x_2 = x_3 = 1$ should be classified as positive, the disease growth requirement requires that $x_1 = \cdots = x_4 = 1, \ldots, x_1 = \cdots = x_6 = 1$ should be classified as positive, whether they belong to the training set or not. Furthermore, these requirements add to the robustness of the approach.

4. Method

4.1. Material

Since 1989, the Miners' Colfax Medical Center (MCMC) Outreach Program, in collaboration with pulmonary physicians from the University of New Mexico Health Sciences Center, has offered free screening for mining-related diseases to active and retired miners in the southwestern United States. Persons who have worked in mines of any type are eligible for this program. Participants are examined at the MCMC in Raton, NM or in a mobile clinic that periodically visits mining communities in New Mexico and south central Colorado. Archived posterior–anterior chest radiographs from this program were used to investigate the nature and presentation of pneumoconioses in approximately 200 miners. Radiographs from the cohort were analyzed by two NIOSH-certified B-readers and served as a benchmark for assessing the feasibility of a partially or fully automated radiograph image analysis system. Institutional review of the protocol for use of the radiograph set and protection of patient confidentiality was done by the University of New Mexico Health Sciences Center Human Research Review Committee (HRRC).

The Miners' Colfax Medical Center Outreach Clinic database consists of thousands of chest radiographs. The database was sufficient to find the necessary distribution of the disease progression to test the prototype system. UNM Health Sciences Center has established a population-based registry of patients with ILD in Bernalillo County, New Mexico. Chest X-rays for these patients are all read by two NIOSH-certified B readers. A database of chest radiographs was collected as part of the pre-employment and ongoing health surveillance physicals for Grants Mineral Belt uranium miners. For our study in this paper, the X-ray images were recorded from 186 and 33 subjects with profusion of category 0 and 1, respectively. For each image, the grading of the six chest regions (right upper, right middle, and right lower, and their corresponding left regions) was recorded. For the results presented in Section 5.2, we used 157 chest radiographs (32 abnormal and 125 normal cases) that were graded by two human readers.

4.2. Classifier model agreement validation

In this subsection we provide a summary of how we use ROC curves to evaluate agreement between classifiers. We also provide details on how we train and test this agreement.

As input features to our classifiers, we generate the entire list of all (64) possible binary input combinations. In other words, we generate the input feature values from $(x_1, x_2, x_3, x_4, x_5, x_6) = (0, 0, 0, 0, 0, 0)$ to $(x_1, x_2, x_3, x_4, x_5, x_6) = (1, 1, 1, 1, 1, 1)$. We then run the classifiers to classify each possible input. For computing the human reader classification we use our training set to compute the most frequent reader's classification to the presented feature vector (pattern). In the case when the generated input pattern did not appear as an actual lung in the training set, we use a k-nearest neighbor classification using Mahalanobis distances (with $k = 3$) to other patterns that appeared in the training set. Here, for the ROC curves, we remind the reader that our data set consisted of 157 chest radiographs graded by two human readers (see Section 4.1).

To interpret points on the ROC curve, we provide a quick summary as it relates to our application. In our approach of measuring agreement between two classifiers, we designate one of them as the *ground truth classifier* and the second one as the *proposed classifier*. Here, the *ground truth classifier* refers to the classifier that we are trying to approximate using the *proposed classifier*. In Section 4.3 we discuss how to estimate the actual physical ground truth based on the certified reader classifications.

It is thus important to recognize that for training purposes, we use the designated *ground truth classifier* as the ground truth. For example, suppose that we are using the *K*-means classifier to approximate Reader #1. For training, we average all the six-dimensional lung input patterns that Reader #1 gave a normal classification to find the center for the normal cluster. For the abnormal case, we average all the six-dimensional patterns that Reader #1 gave an abnormal classification to determine the center for the abnormal cluster.

Based on this approach, for a test pattern, we have: (i) a true positive (TP) when both classifiers agree on an abnormal classification, (ii) a false positive (FP) when the ground truth classifier classifies an input as normal while the proposed classifier classifies it as abnormal, (iii) a true negative (TN) when both classifiers agree on a normal classification and (iv) a false negative (FN) when the ground truth classifier classifies an input as abnormal while the proposed classifier classifies it as normal.

We then define sensitivity and specificity in the usual way, $Sensitivity = TP/(TP + FN)$, $Specificity = TN/(TN + FP)$ in terms of the numbers of TPs, FPs, TNs and FNs. As usual, the ROC curves are generated by plotting *Sensitivity* versus $1 - Specificity$. In our case, we have that $TP + FN = 32$, $TN + FP = 125$ which gives the classification accuracy (in %) in terms of sensitivity and specificity using $CC = (32 * Sensitivity + 125 * Specificity)/157 * 100$.

For each point of every ROC curve that we display, we perform validation using the leave one out method. Thus, we test each pattern by training over the rest of them (63). We then compute the *Specificity* and *Sensitivity* for all 64 possible binary input patterns.

The ROC curve itself provides us with a standard visual method for evaluating the agreement. We also compute the ROC area for each case. We define the best operating point on the ROC curve as the point that is closest to $Sensitivity = Specificity = 1$ (using Euclidean distance, see Table 2).

4.3. Ground truth estimation

To provide estimates of the ground truth, suppose that we have ratings from two readers: reader *A* and reader *B*. Now, let assume that $x_i = a_i$ (0 or 1) represents the ratings of reader *A*, while $x_i = b_i$ represents the ratings of reader *B*. We then consider the minimum and maximum estimators of the ground truth using

$$\mathbf{x}_{min,i} = \min(a_i, b_i), \quad i = 1, \ldots, 6, \quad \text{and} \tag{4}$$

$$\mathbf{x}_{max,i} = \max(a_i, b_i), \quad i = 1, \ldots, 6. \tag{5}$$

The definitions given in Eqs. (4)–(5) can be easily extended to any number of ROIs. Clearly, we would expect that random vector $\mathbf{x}_{min}$ will underestimate the number of positive ratings, while $\mathbf{x}_{max}$ will overestimate the number of positive ratings. In general, it is important to note that the readers may classify a pattern $\mathbf{x}$ as either normal (0) or abnormal (1). To compute the maximum-likelihood rating of each reader, we simply compare the number of times that a particular pattern has been classified as normal versus abnormal. If the

pattern is classified more often as normal, then the maximum likelihood rating for the pattern is the normal classification. Similarly for the abnormal case.

Assuming that r_i represents the maximum-likelihood rating of the ith reader, we define the minimum and maximum classifications as

$$C_{min} = \min(r_1, r_2) \quad \text{and} \tag{6}$$

$$C_{max} = \max(r_1, r_2). \tag{7}$$

Here, we see that the minimum classifier will tend to underestimate the ground truth, while the maximum classifier will tend to overestimate.

4.4. Classifier models

4.4.1. Bayes classification

The Bayes discriminatory rule, with respect to prior probabilities π_0 for normal (0 or negative) and π_1 for abnormal (1 or positive) ($\pi_0 + \pi_1 = 1$) is:

If $\pi_0 p_0(\mathbf{x}) \geqslant \pi_1 p_1(\mathbf{x})$
then allocate sample point $\mathbf{x}$ to normal (0)
else allocate sample point to abnormal (1)

Our classification rule is based on the well-known Bayes formula $P[i|\mathbf{x}] = \pi_i p_i(\mathbf{x})$ relating the posterior probability $P[i|\mathbf{x}]$ in terms of the prior π_i and the probability function of $\mathbf{x}$ under $p_j(\mathbf{x})$, $(j = 0, 1)$. In this case, the probability functions are six-dimensional Bernoulli with $2^6 = 64$ unknown probabilities $p_j(\mathbf{x}) = P_j[X_i = x_i, i = 1, \ldots, 6]$, ith $x_i = 0$ or 1, $i = 1, \ldots, 6$. As an example, to estimate $p_0(\mathbf{x})$ for each of the 64 probabilities, we simply count the number of times that each pattern is classified as normal and divide by the total number of patterns. It is important to note that complete classification using the Bayes classifier is not possible due to an insufficient number of samples. Please note that our comment on an insufficient number of samples is made with respect to having a sufficient number of cases for every possible combination of the 64 possible input patterns. Here, please note that advanced cases were not allowed to continue. Also, since the disease growth is symmetrical, highly asymmetric cases do not occur. Of course, the same comment applies to all statistical classifiers that require knowledge of $p_0(\mathbf{x})$ and $p_1(\mathbf{x})$ for every possible value of $\mathbf{x}$. In the case of a missing input, we are thus led to use a nearest neighbor classifier as detailed in Section 4.2.

4.4.2. Logistic discrimination

In the logistic form for the posterior probabilities for two categories (normal or abnormal) in the well-known Cox–Day–Kerridge approach (see, e.g., Ref. [19]), the fitted $y = 0$ or 1 takes the form

$$\text{Logit } \hat{y} = 1/(1 + \exp[\alpha_0 + \alpha_1 x_1 + \cdots + \alpha_n x_n]). \tag{8}$$

In the present application, $\mathbf{x} = (x_1, \ldots, x_6)$, $x_i = 0$ or 1, and $n = 6$.

4.4.3. K-means classification

In K-means classification, we first compute means for each category. For each training sample we compute the Mahalanobis distance for correlated random variables to each of the means.

4.4.4. Summation classifier

For the summation classifier, we simply add all the binary ratings:

$$s(\mathbf{x}) = x_1 + x_2 + \cdots + x_6. \tag{9}$$

The basic classifier would then need to classify the lung as normal if $s(\mathbf{x})$ is small. Else, for larger values of $s(\mathbf{x})$, we would classify the lung as abnormal. To accomplish this and also evaluate the performance using a ROC curve, we need to consider an adaptive threshold for $s(\mathbf{x})$. This is accomplished by allocating the radiograph as normal if $s(\mathbf{x}) < 1 + V$ where V is a free parameter. Else, the radiograph is classified as abnormal. Here, V is varied to generate different operating points.

For normal classification, the use of $s(\mathbf{x}) < 1 + V$ indicates that a small number of positive ratings have been counted. For example, for $V = 0$, we would classify a radiograph as normal if no positive ratings have been marked. On the other hand, for $V = 6$, all radiographs would be classified as normal while for $V = -1$, all radiographs would be classified as abnormal.

4.4.5. Weighted-sum classifier

For the weighted-sum classifier, we consider the weighted averages

$$\xi_1(x_1, x_2, x_3) = 1 + x_1 + 2x_2 + 4x_3, \tag{10}$$

$$\xi_2(x_4, x_5, x_6) = 1 + x_4 + 2x_5 + 4x_6. \tag{11}$$

For this classifier, we allocate $\mathbf{x}$ as abnormal if either $\xi_1 + \xi_2 > 4 + 4V$, $\xi_1 > 2 + 2V$ or $\xi_2 > 2 + 2V$. Else, the radiograph is classified as normal. Here, as for the summation classifier, we vary V to generate different points on the ROC curves.

The weighted classifier is an extension of basic summation classifier. Here, as recommended by our medical team, more emphasis is placed on the middle and lower parts of the lung. Thus, we would expect that disease would progress from the top (x_1, x_4) to the middle and lower portions of the lung. Thus, in defining ξ_1, ξ_2, we use a weight of 2 for x_2, x_5 and a weight of 4 for x_3, x_6. Then, to account for asymmetries, we define a classification rule based on the sum of both lungs: $\xi_1 + \xi_2$ as well as the individual lung sums ξ_1, ξ_2. As for the summation classifier, the use of different values for V allows us to classify based on different numbers of positive ratings.

5. Results

5.1. Spatial symmetry

In order to evaluate the validity of a specific classifier, it is important for us to establish that it satisfies clinical requirements on disease growth. In this case, we are interested to know if disease appears equally on the right and left lungs.

To investigate right–left symmetry, we form the right lung sum $S_{123} = x_1 + x_2 + x_3$, and compare it against the left lung sum $S_{456} = x_4 + x_5 + x_6$. The results are shown in Table 1, where we have restricted the data set to all pairs for which both of the readers provided a q-rating.

In perfect left-right symmetry, all the off-diagonal entries would be zero. The agreement between the two lung grades is evaluated

Table 1
Left-right agreement for pooled data

		Right sum				Total
		0	1	2	3	
	0	6	8	0	0	14
Left	1	9	116	13	1	139
Sum	2	0	18	78	5	101
	3	0	2	9	29	40
Total		15	144	100	35	294

Left-right lung agreement is based on pooled data where both raters selected q (non-q normal ratings were not considered). The weighted kappa statistic between right and left lungs was 0.72 with a 95% confidence interval of 0.65–0.78. The simple kappa statistic was 0.65 with a 95% confidence interval of 0.58–0.72.

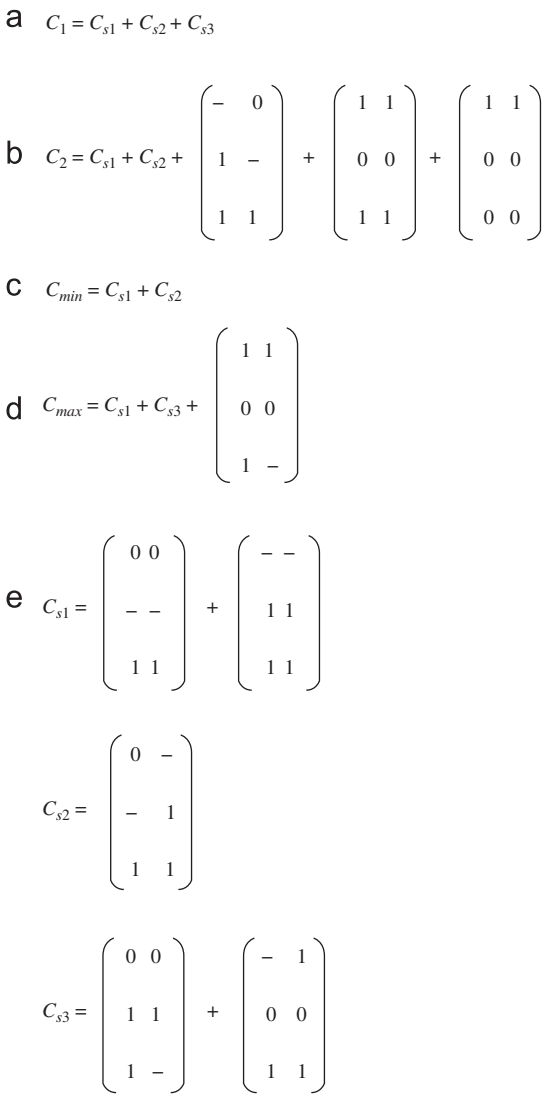

Fig. 2. Optimal lung decompositions as a sum of prototypes. Here, the sum should be interpreted as a logical `or`. A positive score for the whole lung requires that the given lung opacity scores match at least one of the lung prototypes. The logic description provides an optimal list of possible abnormal lung prototypes for: (a) the first reader, (b) the second reader, (c) the Min classifier (lower bound on ground truth), (d) the Max classifier (upper bound on the ground truth) and (e) the list of common lung prototypes. For each lung prototype, '1' denotes that opacity presence is required, '0' denotes that opacities should not be present and '–' implies that there is no requirement for opacity presence or absence.

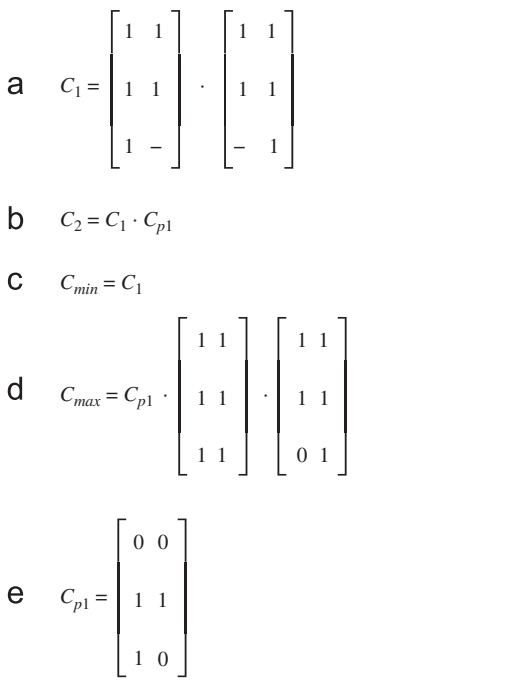

Fig. 3. Optimal Lung decompositions as a product of prototypes. Here, the product should be interpreted as a logical `and`. A positive score for the whole lung requires that the given lung opacity scores should satisfy the requirements for all prototype lungs. In particular, the input scores should match at least one of the marked regions in the prototype lungs. The logic description provides an optimal product list of possible abnormal lung prototypes for: (a) the first reader, (b) the second reader, (c) the Min classifier (lower bound on ground truth), (d) the Max classifier (upper bound on the ground truth) and (e) the list of common lung prototypes.

using both simple and weighted kappa statistics, as shown in Table 1 [23]. For this example, the overall inter-rater agreement was found to be very good [23]. Similarly, if examining other types of symmetries, such as up-down, or rotational symmetries, we simply form an appropriate sum of variables that we require to remain fixed.

5.2. Classification modeling results

In this section, we present logic-function representations for the two readers, the minimum and maximum classifiers (see Section 4), and ROC curves for the five classifiers given in Section 4. We use the term *classification modeling* to refer to the logic-function representations for the two readers. We also use the term *classification analysis* to refer to the analysis of the logic-function representations. We provide an interpretation of the results in Section 6.

We present optimal sum of product decompositions in Fig. 2. Alternatively, we provide optimal product of sum decompositions in Fig. 3. For the decompositions, we use the lung prototype notation that is presented in Fig. 1. In each case, we present optimal decompositions for each classifier using elementary lung prototypes. We provide more detailed descriptions of each decomposition in the captions of Figs. 2 and 3.

For the optimal sum of product decompositions, we present comparative results using the statistical and summation classifiers in Figs. 4 and 5. Recall that we use the term *best operating point* for the ROC curve point that is closest to the point represented by *Specificity* = *Sensitivity* = 1. This point is represented as the upper-left point in the ROC curves of Fig. 4. We summarize the best operating points for the summation classifier in Table 2.

6. Discussion

In what follows, we will provide a critical interpretation of the results. First, we note that the results in Section 5.1 clearly indicate that the lung symmetry criterion applies (see Section 3). In other words, we expect disease to appear symmetrically on the right and left lung regions. In what follows, we also discuss the disease growth criterion and the strong agreement among the classifiers (see Figs. 4 and 5).

From Fig. 2, we see that C_{s1} is shared by all classifiers: C_1, C_2, C_{min} and C_{max}. In both lung prototypes of C_{s1}, we can see that the presence of lung opacities in the lower regions implies an abnormal classification. Furthermore, in all common terms: C_{s1}, C_{s2}, C_{s3}, we can see that we have opacities in at least one of the two lower lung regions. In the lowest estimate of the abnormal lung: C_{min}, we can see that all lung prototypes require the presence of opacities in the lower regions (see Fig. 3(c)). Thus, when both

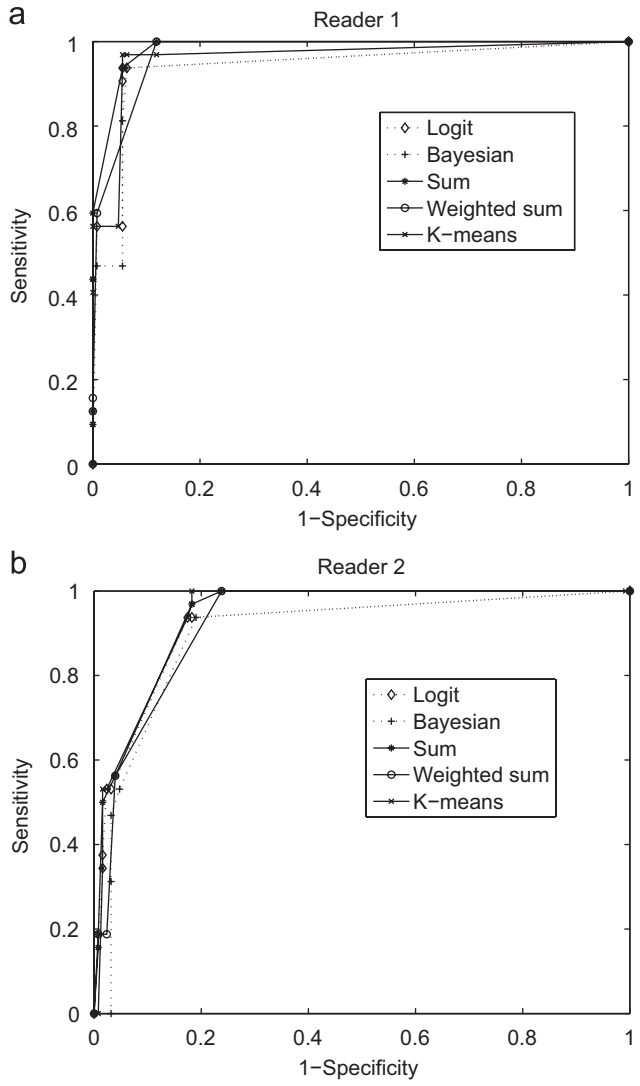

Fig. 4. Receiver operating characteristic (ROC) curves as measures of agreement between the proposed methods and (a) reader # 1, (b) reader # 2.

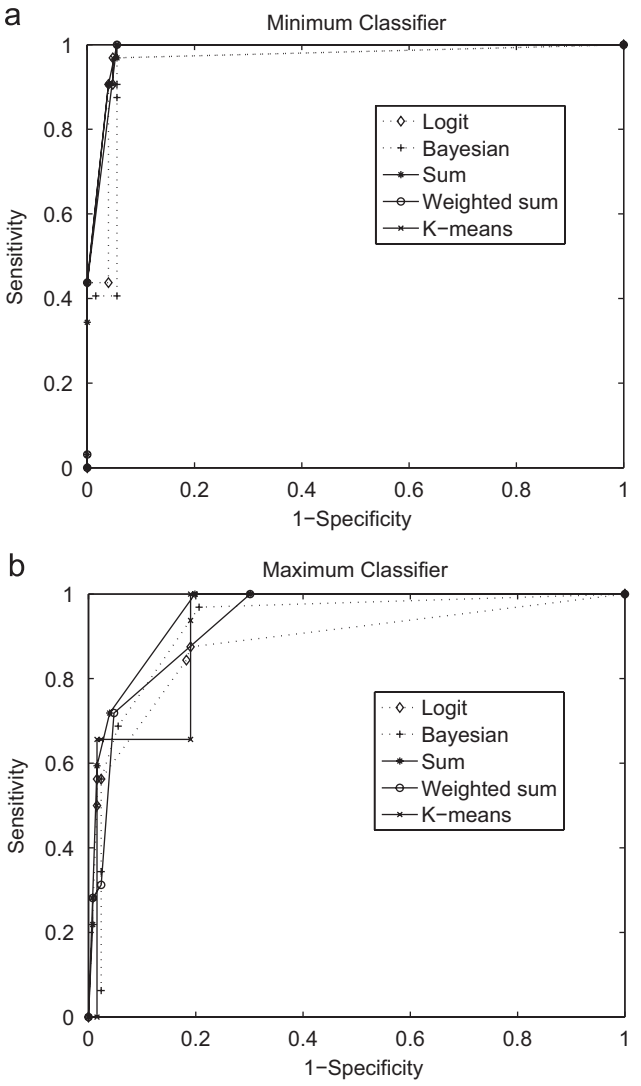

Fig. 5. Receiver operating characteristic (ROC) curves as measures of agreement between the proposed methods and (a) the minimum classifier and (b) the maximum classifier. In (a), the minimum classifier selects the minimum between the reader classifications. In (b), the maximum classifier selects the maximum between the reader classifications.

lower lung regions are marked, we expect the lung to be marked as abnormal. This is in agreement with clinical expectations that the disease progresses from the top to the bottom of the lung. Thus, if abnormalities appear in the lower regions, we are confident that the case is abnormal.

For the optimal product of sums representation, it is interesting to examine the C_{max} classifier shown in Fig. 2(c). Here, we note that both C_{max} and C_2 share the common term of C_{p1}. Now, to satisfy C_{p1}, C_2, C_{max}, we require that at least one of the middle or lower regions be marked as positive. On the other hand, the expressions for C_{min} and C_1 are satisfied even for the simple case where opacities appear in the upper parts of the lung. As we shall explain next, we believe that this is an inevitable artifact that significantly limits the applicability of the product of sums approach.

To understand the limitation of the product of sums approach, we first note that the optimal product is computed by computing an optimal sum of products on the complementary classifier $C'(x_1, \ldots, x_6)$. In other words, the optimal product of sums provides a summary of the negative-ratings (zeros) on the lungs while the optimal sum of products summarizes the positive ratings. However, there is a disproportionate number of negative-ratings as compared to positive ratings. The reason for this is simple. When pneumoconiosis was established beyond reasonable doubt, the miner was removed from exposure to reduce the spread of the disease. As a result, the database included a limited number of advanced cases. And as an artifact, the product of sums approach wrongly assumes the missing data sets to be negative. Note that this violates our requirement for covering the case for disease progression. We are thus led to only consider the sum of products results as being valid.

Table 2
Summation classifier approximations to all readers

Reader	ROC area	Best param.	Best sens.	Best spec.
Reader # 1	0.99	1	0.94	0.94
Reader # 2	0.94	1	0.97	0.82
Min class.	0.97	0	1.00	0.94
Max class.	0.96	1	1.00	0.80

Please note that the measurements reflect the agreement between the model and the readers. For example, an ROC area of 1 for Reader # 1 indicates that the summation classifier can be used to perfectly model the first reader. This is not to be confused with the ground truth based on the physical lung.

From Figs. 4 and 5, it is interesting to note that we obtained similar classification results from all classifiers. However, please note that the Logit, Bayesian and *K*-means classifiers are more complex than the summation and weighted-sum classifiers. In particular, unlike the statistical classifiers: Logit, Bayesian and *K*-means, both the summation and weighted-sum classifier only require a single parameter (see discussion on V in Sections 4.4.4 and 4.4.5). Furthermore, it is important to note that the summation classifier satisfies the symmetry and disease growth criteria that we discussed in Section 3. In general, we cannot guarantee that the Logit, Bayesian and *K*-means classifiers would satisfy these criteria. Thus, we chose the summation classifier of Eq. (3) as the simplest and most robust classifier. As we demonstrate in Table 2, the summation classifier can also closely approximate both readers, the minimum and maximum classifiers.

In terms of estimating the physical lung ground truth, it is clear from Table 2 that the maximum classifier appears to be the best choice. At the best operating point, the same summation classifier (with the same parameter $V = 1$) can accurately approximate both readers and the maximum classifier. At the best operating point, the summation classifier will classify a chest radiograph as being abnormal if two or more ROI regions were marked for disease presence.

7. Conclusion

In this paper, we have presented a number of new models for characterizing region of interest classifiers. In Section 3, we discussed the use of logic functions as a unifying model for characterizing all binary classifiers and also discussed classifier requirements in terms of symmetry and disease growth. This discussion led to our study of summation, weighted sum, and the Logit classifiers. Furthermore, we have introduced the Bayes classifier and the *K*-means classifier with Mahalanobis distance for correlated random variables.

We have found that the sum-of-products decomposition is an effective method for characterizing region of interest classifications. This approach provides for a very efficient way to summarize binary classifiers in terms of elementary lung prototypes that can also be shared among classifiers (see Fig. 2). It is also interesting to note the excellent performance of the simple summation classifier, allowing us to approximate all other classifiers while also satisfying our requirements for symmetry and disease growth.

References

[1] H. Suzuki, N. Inaoka, H. Takabatake, M. Mori, H. Natori, A. Suzuki, An experiment system for detecting lung nodules by chest X-ray image processing, IEEE SPIE Biomedical Image Processing II, vol. 1450, 1991, pp. 99–107.

[2] J. Lin, S.B. Lo, A. Hasegawa, M.T. Freedman, S.K. Mun, Reduction of false positives in lung nodule detection using a two-level neural classification, IEEE Trans. Med. Imaging 15 (1996) 206–217.

[3] X.-W. Xu, K. Doi, T. Kobayashi, H. MacMahon, M.L. Giger, Development of an improved cad scheme for automated detection of lung nodules in digital chest images, Med. Phys. 24 (9) (1997) 1395–1403.

[4] H. Yoshida, K. Doi, Computerized detection of pulmonary nodules in chest radiographs: reduction of false positives based on symmetry between left and right lungs, in: SPIE Medical Imaging, 2000, pp. 97–102.

[5] K. Honma, K. Chiyotani, K. Kimura, Silicosis, mixed dust pneumoconiosis, and lung cancer, Am. J. Ind. Med. 32 (6) (1997) 595–599.

[6] I. Ebihara, A pathological study of carcinoma of the lung and pneumoconiosis, Nihon Kyobu Shikkan Gakkai Zasshi (japanese journal with english abstract available).

[7] M. Katabami, H. Dosaka-Akita, K. Honma, Y. Saitoh, K. Kimura, Y. Uchida, H. Mikami, Y. Ohsaki, Y. Kawakami, K. Kikuchi, Pneumoconiosis-related lung cancers: preferential occurrence from diffuse interstitial fibrosis-type pneumoconiosis, Am. J. Respir. Crit. Care Med. 162 (1) (2000) 295–300.

[8] A.T. Society, Adverse effects of crystalline silica exposure, Am. J. Respir. Crit. Care Med. 155 (1997) 761–768.

[9] I.L. Office, Guidelines for the Use of ILO International Classification of Radiographs of Pneumoconiosis, Geneva, 1980.

[10] K.G. Hering, M. Jacobsen, E. Bosch-Galetke, H.J. Elliehausen, H.G. Hieckel, K. Hofmann-Preiss, W. Jacques, U. Jeremie, N. Kotschy-Lang, B.M.T. Kraus, W. Raab, H.J. Raithel, W.D. Schneider, K. Strassburger, S. Tuengerthal, H.J. Woitowitz, Further development of the international pneumoconiosis classification—from ilo 1980 to ilo 2000 and to ilo 2000/german federal republic version, Pneumologie 57 (10) (2003) 576–584.

[11] H. Amandus, E. Pendergrass, W. Morgan, Pneumoconiosis: inter reader variability in the classification of a type of small opacities in the chest radiograms, Am. J. Respir. 12 (1974) 740–743.

[12] J. Bourbeau, P. Enst, Between- and within-reader variability in the assessment of pleural abnormality using the ilo 1980 international classification of pneumoconiotic chest films, Occup. Med. 14 (1988) 537–543.

[13] N.H. Klerk, A.W. Musk, A. James, J.J. Glancy, W.O. Cookson, Comparison of chest radiograph reading methods for assessing progress of pneumoconiosis over 10 years in wittenoom crocidolite workers, Br. J. Ind. Med. 47 (2) (1990) 127–131.

[14] D.C. Muir, C.D. Bernholz, W.K. Morgan, J.O. Roos, J. Chan, W. Maehle, J.A. Julian, A. Sebestyen, Classification of chest radiographs for pneumoconiosis: a comparison of two methods of reading, Br. J. Ind. Med. 49 (12) (1992) 869–871.

[15] L.S. Welch, K.L. Hunting, J. Balmes, E.A. Bresnitz, T.L. Guidotti, J.E. Lockey, T. Myo-Lwin, Variability in the classification of radiographs using the 1980 international labor organization classification for pneumoconioses, Chest 114 (6) (1998) 1740–1748.

[16] J.A. Anderson, Separate sample logistic discrimination, Biometrika 59 (1) (1972) 19–35.

[17] J.A. Anderson, Ch. Logistic Discrimination With Medical Applications, Discriminant Analysis and Applications, Academic Press, New York, 1973 pp. 1–16.

[18] D. Hosmer, S. Lemershow, Applied Logistic Regression, second ed., Wiley, New York, 2000.

[19] P. McCullagh, J.A. Nelder, Generalized Linear Models, second ed., Chapman & Hall/CRC, New York, 1999.

[20] J.L. Teugels, Some representations of the multivariate bernoulli and binomial distributions, J. Multivariate Anal. 32 (2) (1990) 256–268.

[21] E.J. McCluskey, Introduction to the Theory of Switching Circuits, McGraw-Hill, New York, 1965.

[22] J.F. Wakerly, Digital Design Principles and Practices, fourth ed., Prentice-Hall, New York, 2005.

[23] J.L. Fleiss, Statistical Methods for Rates and Proportions, second ed., Wiley, New York, 1981.

[24] C.E. Metz, Fundamental ROC analysis, in: J. Beutel, L. Kundel, R.L.V. Metter (Eds.), Handbook of Medical Imaging, vol. 2, SPIE Press, 2000, pp. 751–770.

[25] H. Yoshida, Local contralateral subtraction based on bilateral symmetry of lung for reduction of false positives in computerized detection of pulmonary nodules, IEEE Trans. Biomed. Eng. 51 (5).

[26] R.P. Kruger, W.B. Thompson, A.F. Turner, Computer diagnosis of pneumoconiosis, IEEE Trans. Systems Man Cybern. 4 (1) (1974) 40–49.

[27] R.S. Ledley, H.K. Huang, L.S. Rotolo, A texture analysis method in classification of coal workers' pneumoconiosis, Comput. Biol. Med. 5 (1–2) (1974) 53–67.

[28] A.M. Savol, C.C. Li, R.J. Hoy, Computer-aided recognition of small rounded pneumoconiosis opacities in chest X-rays, IEEE Trans. Pattern Anal. Mach. Intell. 2 (5) (1980) 479–482.

[29] H. Kobatake, K. Oh'ishi, J. Miyamichi, Automatic diagnosis of pneumoconiosis by texture analysis of chest X-ray images, in: IEEE ICASSP, 1987, pp. 610–613.

[30] Y. Ugurlu, K. Ohkura, T. Obi, A. Hasegawa, M. Yamaguchi, N. Ohyama, Detection of increasing profusion of opacities from a sequence of personal chest radiographs, IEEE International Conference on Image Processing, vol. 3, 1999, pp. 402–406.

[31] Z. Huang, D. Yu, J. Zhao, Application of neural networks with linear and nonlinear weights in occupational disease incidence forecast, in: IEEE Asia-Pacific Conference on Circuits and Systems, 2000, pp. 383–386.

[32] H. Kondo, T. Kouda, Detection of pneumoconiosis rounded opacities using neural network, Joint 9th IFSA World Congress and 20th NAFIPS International Conference, vol. 3, 2001, pp. 1581–1585.

[33] H. Kondo, T. Kouda, Computer-aided diagnosis for pneumoconiosis using neural network, in: IEEE Symposium on Computer-Based Systems, 2001, pp. 467–472.

[34] P. Soliz, M.S. Pattichis, J. Ramachandran, D.S. James, Computer-assisted diagnosis of chest radiographs for pneumoconiosis, SPIE Medical Imaging Conference, vol. 1, San Diego, California, 2001, pp. 667–675.

[35] M.S. Pattichis, C.S. Pattichis, C.I. Christodoulou, D.S. James, L. Ketai, P. Soliz, A screening system for the assessment of opacity profusion in chest radiographs of miners with pneumoconiosis, in: 5th IEEE Southwest Symposium on Image Analysis and Interpretation, Santa Fe, New Mexico, 2002, pp. 130–133.

[36] B. van Ginneken, B.M.T.H. Romeny, M.A. Viergever, Computer-aided diagnosis in chest radiography: a survey, IEEE Trans. Med. Imaging 20 (12) (2001) 1228–1241.

[37] B. van Ginneken, S. Katsuragawa, B.M.T.H. Romeny, K. Doi, M.A. Viergever, Automatic detection of abnormalities in chest radiographs, using local texture analysis, IEEE Trans. Med. Imaging 21 (2) (2002) 139–149.

About the Author—MARIOS S. PATTICHIS received a B.Sc. (high honors and special honors) in computer sciences in 1991, a B.A. (high honors) in mathematics in 1991, an M.S. in engineering in 1993, and a Ph.D. in computer engineering in 1998, all from the University of Texas at Austin. Currently, he is an Associate Professor with the Department of Electrical and Computer Engineering and An Associate Professor with the Department of Radiology at the University of New Mexico (UNM), Albuquerque, New Mexico. His research interests are in the areas of medical image and video processing, digital image and video models, radar image processing, SIMD, and reconfigurable

computer architecture applications. He is an associate editor for Pattern Recognition, a guest editor of the special issue in Computational Intelligence in Medical Systems to be published by the IEEE Transactions in Information Technology in Biomedicine, and the general chair of the 2008 IEEE Southwest Symposium on Image Analysis and Interpretation, to be held in Santa Fe, New Mexico. At UNM, he received the 2004 ECE distinguished teaching award and the 2006 School of Engineering Harrison Faculty Recognition Award. He is a Senior Member of IEEE.

About the Author—PETE SOLIZ received his Ph.D. from the University of Oklahoma in 1966. From 1966 to 1986, he was with the Air Force Research Laboratories. In 1997, he became Vice President and Director of Biomedical Imaging for Kestrel Corporation. He is currently the President and CEO of VisionQuest Biomedical, a company that he founded in 2006. He is also adjunct Assistant Research Professor with the University of Iowa Department of Ophthalmology and an Adjunct Research Professor with the University of New Mexico (UNM) School of Medicine. At UNM, Dr. Soliz is collaborating with the Department of Radiology and Radiation Physics in research into radiographic image quality and lung disease classification. He has published 48 conference and journal articles and has received over $6.9 million in grants from the National Institutes of Health, the National Medical Testbed, and Pfizer since 2000.

About the Author—THEOPHILOS CACOULLOS received his Ph.D. in Mathematical Statistics from Columbia University in 1962. In 1963–1965, he was an Assistant Professor with the Department of Statistics at the University of Minnesota. In 1965–1968, he was an Assistant Professor and then Associate Professor in the Department of Industrial Engineering and Operations Research at the School of Engineering at New York University. From 1968 to 1999, he has been a Professor of Probability and Statistics in the Department of Mathematics at the University of Athens. He is currently Professor Emeritus at the University of Athens, Greece.
He has held visiting positions at MIT, Stanford, McGill, Columbia, and the University of New Mexico. He founded the Greek Statistical Institute in 1980 and was President of the Institute from 1981 to 1985. He has been a fellow of IMS since 1986, a fellow of the Royal statistical Society since 1963, and an elected ordinary member of ISI (1971).