Reprinted from INVESTIGATIVE RADIOLOGY, Vol. 11, No. 4, July-August, 1976


# Automated Computer Screening of Chest Radiographs for Pneumoconiosis

A. FRANKLIN TURNER, MD, RICHARD P. KRUGER, PHD, AND WILLIAM B. THOMPSON, PHD

**The results of two complementary approaches for performing diagnostic screening for the presence of coal workers' pneumoconiosis (CWP) from the routine posterior-anterior chest radiograph are presented. The first is a digital approach utilizing the measurement of image texture, while the second uses hybrid optical-digital methods involving the optical Fourier transform. Both approaches yield classification results comparable to experienced radiologists.**

**Key words: computer, chest, pneumoconiosis.**

THE DIAGNOSIS OF COAL WORKERS' PNEUMOconiosis (CWP) is dependent on a history of exposure to coal dust and the presence of certain distinctive features on a chest radiograph. The severity of the condition is indicated by the extend, profusion, and character of opacities observable on the chest radiograph. The chest film is not only the most reliable method for diagnosing the disease during life, but also the only means of assessing progression.

In 1969, the Federal Coal Mine Health and Safety Act was enacted. Public Law 91173 has specified that each coal worker has the right to obtain his chest roentgenogram at regular intervals not to exceed 5 years to detect the possible presence and progression of coal workers' pneumoconiosis. There are approximately 200,000 claims for disability due to CWP from active or retired coal workers or their survivors. Each roentgenogram is read by at least two certified readers according to the ILO U/C 1971 international classification of radiographs of the pneumoconioses.[4, 10, 11]

Current manual screening utilizing the forementioned international standards consists of an hierarchy of readers with increasing diagnostic skills relative to this disease. The initial reader, designated an A reader, is either a radiologist or a general practitioner who is located in the region where the radiograph was obtained. The second, or B reader is often geographically remote and is a board certified radiologist at one of the small number of pneumoconiosis screening centers established nation wide. If there is more than one minor profusion category difference between the A and B readings, a significant disagreement exists and one of the C readers decides a final category. In rare instances, a panel of C readers constitute a D reader. There are approximately 150 A readers, 30 B readers and 5 C readers actively screening for this disease; however, the 5 C readers are largely responsible for the ILO U/C standards. Recently published NIOSH epidemiological data from this reader hierarchy is also shown in Table 1.[16] It indicates that approximately 13% of active miners are diseased. It was also reported that perhaps 28% of retired miners are diseased.

From the LAC-USC Medical Center, Los Angeles, California.

Presented at the 21st Annual Meeting of the Association of University Radiologists, University of British Columbia, May 10-12, 1973.

The authors wish to acknowledge the cooperation of Mr. Niels Jensen and the staff of Recognition Systems, Inc. for the use of the ROSA-3 device and Dr. George Jacobson, Chairman of the Department of Radiology, and his staff, Drs. Sargent, Halpern, and Gordonson, for their assistance. We would also like to acknowledge Professors George Bekey and William Pratt of the Biomedical Engineering and Image Processing Institutes for their cooperation and facilities, as well as Professor Harry C. Andrews for his advice on pattern recognition techniques.

The reported significant difference in radiological category between A and B readers which necessitated a C reading was 27%. This inter-reader variation in pneumoconiosis screening is not surprising when viewed within the context of earlier studies of intra- and inter-reader variation in chest radiographic diagnosis of similar appearing tuberculosis lesions conducted by Garland and Yerushalmy.[5, 25]

Reger and Morgan[19, 20] also conducted studies concerning the effects of inter-reader variation and film quality on assignment of profusion category. In only 56.7% of the cases was there agreement on a four diagnostic class assignment based on increasing profusion category among four experienced readers of pneumoconiosis films. The percentage was surprisingly low, since approximately one half of the films selected showed no evidence of pneumoconiosis, although many were suffering from other respiratory ailments such as tuberculosis or showed evidence of other radiographic abnormalities.

Based on present projections of future reading loads, it is likely that present manual reading procedures will result in a progressive overburdening of the radiographic diagnostic system as it currently exists. Thus, some automated procedure for diagnostic mass screening of such films will be desirable and most probably necessary. Since it has been demonstrated that a high percentage of all films read are diagnosed as not showing radiographic evidence of pneumonconiosis, a cost effective and rapid automatic film reader to consistently screen out all definite normals would significantly relieve the projected overburdening previously referred to.

Previous research has indicated that automatic anatomical feature location, measurement, and diagnostic classification of the superior mediastinal and cardiac projections in large numbers of standard posterior-anterior chest radiographs is feasible with accuracy rates comparable to manual diagnosis.[13] A modified version of that algorithm has also divided the right and left lung fields into six rectangular zones as stipulated in the ILO U/C grading procedure. It would therefore seem possible to automatically locate each lung zone for subsequent machine diagnosis. A feasibility study was conducted and has been reported in detail[12] using both digital and optical methods for feature extraction and subsequent computer categorization. The results of these experiments will be briefly outlined.

TABLE 1. Profusion Categories and Probability of Occurrence

| | Major Perfusion Category | Number of Cases | A Priori Class Probabilities |
|---|---|---|---|
| Simple disease | 0 | 51,491 | .878 |
| | 1 | 4,630 | .079 |
| | 2 | 1,772 | .030 |
| | 3 | 138 | .002 |
| Complicated disease | 4 | 688 | .011 |

*Image textural discrimination.* The problem of manually detecting and grading simple pneumoconioses from radiographs appears to be largely one of discrimination between normal pulmonary vascularity (lung marking) patterns and partial or complete obliteration of this normal tree-like structure by opacities of various sizes and profusions which themselves exhibit a more or less textural nature. Rosenfeld and Troy[21] described image texture ideally as "the repetitive arrangement of a unit pattern over a given area." It was also stated that, in natural imagery, it is often difficult to identify such unit patterns or determine their repetitive arrangement. Therefore, the previous description should be used only as a guide in the analysis of natural imagery.

Pickett[18] referred to two kinds of human subjective textural analyses. In empirical studies Gibson[6] has shown that regular textures seem to convey stronger impressions than irregular ones, and he also concluded that his observers relied on impressionistic analysis. In computer analysis of natural image texture, Hawkins[9] made several observations. First, the notion that natural texture is strictly locally repetitive is only approximately true. Furthermore, replication may be subject to spatial nonlinear phase shifts and, therefore,

only be approximate. Second, non-random arrangement of pattern parts appear to be associated with meaningful textures.

Textural discrimination of images was attempted in both the spatial and frequency domains. Therefore, the specific approach was to consider pneumoconiotic opacities or normal lung markings as distinct image textures. The task was to experimentally compute several textural features and subsequently apply a feature selection and supervised statistical pattern recognition technique to obtain a diagnostic classification for discrimination between normal and abnormal lung zones.

*The film data base.* It was decided to create a data base consisting of 4 by 5 inch zonal reproductions from 141 chest radiographs. There were several reasons for this choice, not the least of which was the fact that neither the digital or optical device was capable of processing the complete 14 by 17 inch film for input to the computer. There was also a conscious effort to establish accurately the profusion category and lesion type for purposes of computer training. It was felt that zone selection would be an aid toward this goal. The zonal selections provided an ample mix of all lung regions to minimize bias during computer training.

It was hypothesized that, having demonstrated the computer detectability of lung zones, a machine could render a diagnosis within those zones competitive with manual diagnosis.

The film base for the feasibility study was selected by a committee consisting of a B and C reader at the LAC-USC Medical Center Department of Radiology. There were 141 radiographs, 102 were abnormal, 33 were normal and 6 were of profusion category 0/1 from the American College of Radiology standard films. Of the 102 abnormal zonal films, approximately 40% were extracted from standard films used by the American College of Radiology to train physicians to detect and grade pneumoconiotic lesions. The remaining abnormal films were selected and graded by this committee. Thirty-three of the normal films were selected from chest radiographs of LAC-USC employees with no known pulmonary disease and no history of exposure to dust, and who had no clinical or radiological evidence of lung disease.

A representative sample which consisted of 95 zonal reproductions were digitized by a flatbed scanning microdensitometer at a resolution of 9.8 pixels per mm. The optical studies used all 141 reproductions.

*Data management and textural feature extraction from the digital images.* It was decided that visual diagnosis of simple pneumoconiosis lesions is often arrived at by inspecting the lung regions between the more visually prominent posterior ribs. This seemed logical since it is in these inter-rib spaces that normal vascularity is least obstructed by visual interference from denser radiographic structures. Therefore, posterior inter-rib space regions were extracted manually (Fig. 1). No effort was made to avoid the less prominent anterior ribs or inter-rib spaces not involved with disease. Since there were 3 or 4 such inter-rib spaces on each zonal film, the complete digital data base consisted of 298 inter-rib spaces.

It was decided to consider each posterior inter-rib space as a separate input to the diagnostic classifier in order to create an automated classification technique not sensitive to any inter-rib space or any specific lung zones. This would constitute a local region for textural analysis. Thus the digital input data base consisted of 298 samples.

The data from an individual inter-rib space were combined to form a gray level histogram for that complete inter-rib space. A transformation of gray levels was performed which produced an inter-rib space with 8 equally likely gray levels.[2] This preprocessing step was designed to negate the effect of monotonic distortions introduced due to inconsistencies in photography and/or digitization of the original images by constraining all inputs to the feature extractor to be identical with respect to first order probability of gray level occurrence. The spatial textural measures were all based on spatial gray level dependence matrices[3, 8, 22] under the assumption that visual texture-context information in an inter-

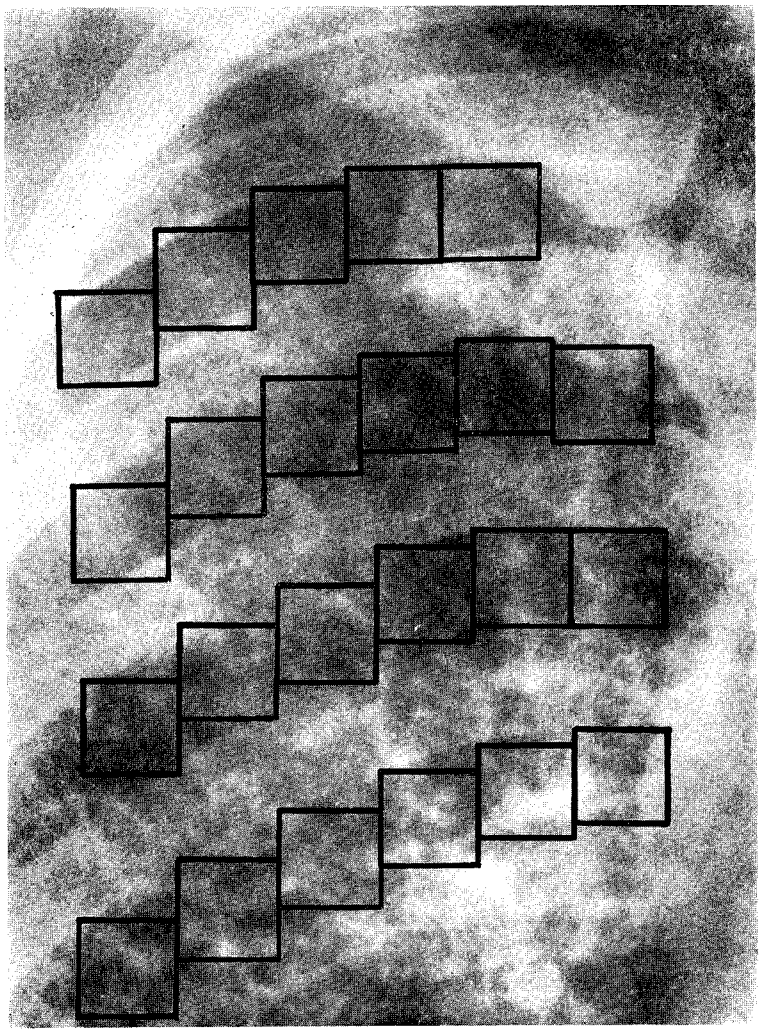

FIG. 1. A 4 by 5 inch zonal reproduction from the right upper lobe region of a 14 by 17 inch radiograph. The square demonstrates the manual extraction of the posterior inter-rib spaces. There are four inter-rib spaces on this zonal reproduction and a total of 23 manually extracted squares. Each inter-rib space was used as a separate input to the diagnostic classifier.

rib space is contained in the spatial relationship between image picture elements at several fixed distances and angular orientations. More specifically, it was assumed that this texture-context information is adequately specified by the symmetric matrix of relative frequencies p (i,j) with which two neighboring pixels are separated by a distance (d) and an angle (a) for each (i,j) gray level pair in the space. For this application, (d) is the number of image lines separating the two pixels of interest. An 8 by 8 symmetric count matrix was formed for each inter-rib space as a function of (a) and (d). For each of the 298 spaces, there were 16 such 8 by 8 matrices per space. Five textural measurements $T_k$ (a,d) k=1, . . . 5 were computed for each matrix. $T_1$ was an auto correlation measure designed to measure image coarseness. $T_2$ was a dissimilarity measure often called the moment of inertia. $T_3$ measured the extent which the same or similar gray levels tended to be neighbors. $T_4$ was a conditional entropy measure and measured image homogeneity . $T_5$ was another dissimilarity measure which was similar to $T_2$. A total of 80 textural measures were extracted from each inter-rib space with each textural feature a function of angle (a) and distance (d). The number of textural features was reduced from 80 to 60 by calculating the mean $\overline{(M)}$ variance (V) and range (R) at a given distance (d) for each of 4 angles (a). The

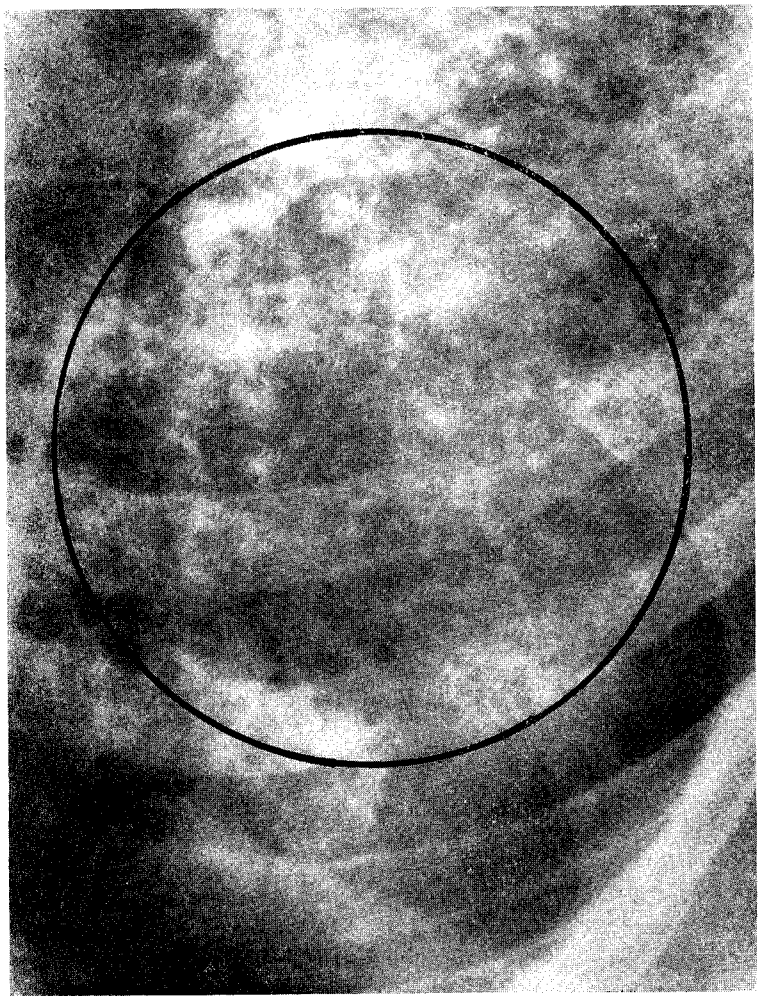

FIG. 2. A 4 by 5 inch zonal reproduction from the same chest radiograph as Fig. 1. This demonstrates a circular 2.5 inch diameter aperature illuminated by the laser light in the center of the appropriate lung zone.

final M, R and V measures did not possess a strict directional bias and were therefore explicitly a function of (d) and only implicitly a function of (a).

*Fourier transform domain feature extraction using a coherent optical approach.* The textural features extracted in the previous section were derived from digital spatial domain data localized to individual posterior inter-rib spaces. The Fourier domain measures treated each of the 141 lung region films as an entity for measured aspects of visual texture. A helium-neon laser emitted a light which passed through a collimating lens and then through the input film image. The transmitted light from the film was next passed through a positive thin lens which performed the Fourier transformation. The transformed image was then projected onto a detector and appropriate energy measurements were obtained. For the images in this experiment, a circular 2.5 inch diameter aperture illuminated a circular area in the center of the appropriate lung zone film. Figure 2 is an example of an illuminated circular area with a typical zonal film.

It is well known that high frequency information pertains to the amount of edge information in an image. It can be hypothesized that lung regions with pneumoconiosis opacities will also generate more of this higher frequency edge information than a normal film. Several investigators have previously used this transform property for terrain, cell, and lung vascularity classification studies with

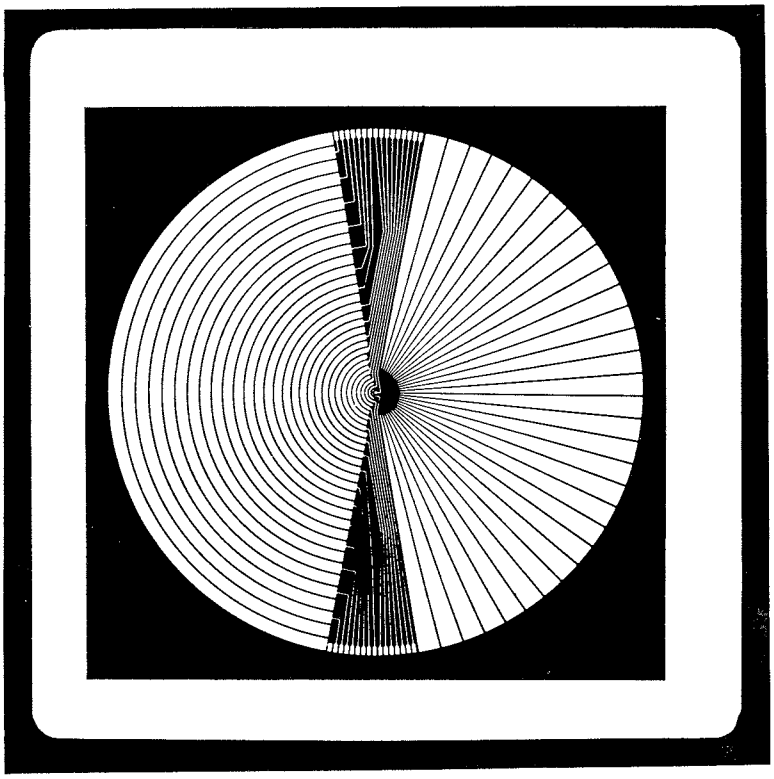

FIG. 3. Detector. There are 32 annular rings and 32 wedges. The 20° vertically oriented sector acts as a filter. The energies which result from the posterior rib edges fall onto this region of the detector, reducing the effect of the posterior ribs in the energy measurements.

some degree of success.[7, 14, 24] Therefore, a detector in the transform plane consisting of 32 annular rings was used to obtain the total contribution of 32 radial frequencies to accurately reflect the amount of edge by the relative strength of the annulli. The maximum spatial frequency measurable in this experiment corresponded to 8.8 line pairs/mm. This exceeds observable resolution for this class of films as measured by Morgan and Rossmann.[17, 23]

The construction of the detector (Fig. 3) is such, that the 20° vertical sector used to carry wires for readout of the 32 annular rings, acts as a filter for certain structures in the image. Since the output in the transform plane is orthogonal to the input plane, the energies which result from the posterior rib edges falls on the 20° vertical wedge regions thus reducing the effect of the posterior ribs in the energy measurements. The spectral measurements were normalized to unit energy by dividing each measurement by the total energy in the transform, which in this case was the sum of the energy in all the annular rings. This normalization compensated for linear inter-film differences in dynamic range and average density. The normalized energies in the rings were then logarithmically transformed to create distributions which were more nearly Gaussian.

*Textural feature selection and classification approach.* Quite often a group of feature measurements contain both redundant features and those which are of little value in separating the classes. For a classifier to work successfully, these unwanted features must be removed. In addition, the larger the set of measurements the classifier must deal with, the greater the numerical inaccuracy in computing the needed discriminant functions. Thus, it is advantageous to reduce the dimensionality of the feature space by determining which of the original measurements contain the most useful information for the classifier.

In order to select this subset, a measure of the value of a feature must be defined. For a two class problem, where the classifier assumes

TABLE 2. Digital Test Results

| | | Assigned Class | |
|---|---|---|---|
| | | 1 | 2 |
| True class | 1 | 90 | 10 |
| | 2 | 5.2 | 94.8 |

TABLE 3. Optical Test Results

| | | Assigned Class | |
|---|---|---|---|
| | | 1 | 2 |
| True class | 1 | 69.7 | 30.3 |
| | 2 | 4.0 | 96.0 |

TABLE 4. Combined Physician Results

| | | Assigned Class | |
|---|---|---|---|
| | | 1 | 2 |
| True class | 1 | 82.5 | 17.5 |
| | 2 | 2.6 | 97.4 |

a Gaussian distribution of features, the "divergence" measure is often used.[1]

It would be desirable to find which combination of N features, taken together, would be optimal. However, in practice, this is often not feasible. A compromise is to calculate the divergence of R features one at a time and then to choose N of those with the highest divergence value.

The divergence measure is defined for only a two class situation. In order to use this measure in a $k > 2$ class problem, the sum of the paired divergences is often used as an optimization criteria.[1] The features are ranked one at a time by the total summed divergence. If two features are highly correlated, they not only contain redundant information, but they make it extremely difficult to carry out statistical classification. This is particularly true of classifiers which must invert a covariance matrix. To identify the degree of correlation between features, a normalized correlation matrix C of dimension R by R was defined, and was computed for all R features. The final N features were chosen by first ordering the R original features by the divergence measure and accepting only features whose absolute value of correlation with all previously accepted features did not exceed 0.8. Features were added to the final set in this manner until the classification accuracy during training no longer improved.

The $k=2$ class situation consisted of a normal-abnormal determination where any film above major profusion category 0 was deemed abnormal. In this manner, an R dimensional feature space was reduced to an N dimensional feature space. The selected features for the two class diagnostic classification using the digital or optical measures never exceeded six in number.

*Computer classification results.* The computer or physician diagnostic results are all discussed in terms of confusion matrices.[15]

The computer diagnostic testing procedure consisted of removing one sample from the data base, training on the remaining samples and resubmitting the withdrawn sample for reclassification. This was a fair test, since the classifier did not "see" the withdrawn sample until it was asked to assign it diagnostically to a class. A second more severe test was also performed. This test consisted of removing one-half of the data from each class and training on the remaining data. The removed half was then submitted to the classifier for diagnosis. This was repeated twice so that all data was classified in a test situation. In many respects, these two testing procedures were logical extremes.

When discussing digital accuracy rates on a per zonal film rather than a per inter-rib space basis, the following rule of correspondence was applied. If a film contained any abnormally classified inter-rib space, the assigned diagnostic class for the film was abnormal.

The one-at-a-time removal test procedure on digitally derived features yielded the following results. The computed normal-abnormal diagnostic rate* was 94.7%. On a per-film basis, 3 films were missed for a corresponding diagnostic rate of 96.8%. When the more severe second test was performed, the confusion ma-

* The normal-abnormal diagnostic rate is the total number diagnosed correctly expressed as a percentage.

trix in Table 2 was found. The normal-abnormal rate was 92.4%. On a per-film basis, this was 96.8%. The one missed abnormal was of profusion category 1/1.

One quite obvious conclusion is that the normal-abnormal diagnostic rates were quite stable, using digitally derived textural features. The two class testing rates using transform domain features indicated the following. For the one-at-a-time test procedure, the normal-abnormal rate was 89.4%. Of the 10 missed normal films, 6 were from profusion category 0/1. The 3 missed abnormals were from profusion category 1/1. When a common film base with digitally derived features is compared, 5 out of 95 films were missed for a diagnostic normal-abnormal rate of 94.7%. The more severe test procedure yielded the matrix shown in Table 3. The normal-abnormal diagnostic rate was 82.9%. Of the 12 missed normals, 6 were again from profusion category 0/1. A common film base of the digital film base once again indicated 5 misses out of 95 for a normal-abnormal rate of 94.7%.

*Physician diagnosis.* Six radiologists were requested to diagnose the identical 141 lung regions submitted for automatic analysis. Two of the six readers originally selected the films in this study. One of these two initial readers also participated in the selection of American College of Radiology standard films and is a C reader. The other initial reader is an experienced B reader who has published extensively in various aspects of chest radiographic interpretation and screens approximately 100 chest films per day for the County of Los Angeles. The other four readers consisted on one C reader and three experienced B readers. As a group, these six radiologists represented over 130 man years of radiological reading experience.

The physicians viewed the entire radiograph with the 141 lung zones within which they were to make their diagnosis labeled and numbered. They were asked to disregard any information not included in the delineated zone. This procedure allowed the readers to grade individual zones within an anatomical context.

When all the $6 \times 141 = 846$ physician observations were averaged, the confusion matrix in Table 4 was formed. In this instance, the normal-abnormal rate was 89.95%. It was also discovered that the original two readers were in agreement with themselves 97% of the time in a normal-abnormal decision.

In summary, the false positive rates for the six readers ranged from 59.0% to 2.9% with an average of 17.5%. Correspondingly the false negative rate ranged from 1.0% to 6.9% with an average of 2.6%. The averaged physician rates showed no significant changes when computed on the basis of the 95 films submitted for digital analysis.

As this was a feasibility study, the available data base was necessarily limited. As the data base is expanded and statistics of the measured features become better known, it may be conjectured that performance will tend to improve. Nevertheless, performance was most encouraging.

*Address for correspondence: A. Franklin Turner, MD, Professor of Radiology & Medicine, LAC-USC Medical Center, Box 778, 1200 North State Street, Los Angeles, California 90033.*

REFERENCES

1. Andrews, H. C.: Introduction to Mathematical Techniques in Pattern Recognition. New York, John Wiley & Sons, Inc., 1972.

2. Andrews, H. C., Tescher, A. G., and Kruger, R. P.: Image processing by digital computer. *IEEE Spectrum* **9**:20, 1972.

3. Ausherman, D. A.: Texture Discrimination within Digital Imagery. Ph.D. Dissertation, Univ. of Mo., December 1972.

4. Bohlig, H., *et al.*: UICC/Cincinnati Classification of the radiographic appearances of the pneumoconioses. *Chest* **58**:57, 1970.

5. Garland, L. H.: Studies on the accuracy of diagnostic procedures. *Am. J. Roentgenol. Rad. Ther. Nucl. Med.* **82**:25, 1959.

6. Gibson, J. J.: The perception of visual surfaces. *Am. J. Psychol.* **63**:367, 1950.

7. Hall, E. L., Kruger, R. P., Dwyer, S. J., *et al.*: A survey of preprocessing and feature extraction techniques for radiographic images. *IEEE Trans. Computers* **C-20**:1032, 1971.

8. Haralick, R., and Anderson, D.: Textural-Tone Study with Applications to Digitized Imagery. TR 182-2, Univ. of Kansas, November 1971.

9. Hawkin, J. K.: Textural Properties for Pattern Recognition. *In* Picture Proc. and Psychopictorics, New York, Academic Press, 1970; pp. 347-370.

10. Jacobson, G., and Lainhart, W. S. (eds.): ILO U/C 1971 international classification of radiographs of the pneumoconioses. *Med. Radiogr. Photogr.* **48**:66, 1972.

11. Key, M. M., and Kerr, L. E. (eds.): Pulmonary Reactions to Coal Dust, New York, Academic Press, 1971.

12. Kruger, R. P., Thompson, W. B., and Turner, A. F.: Computer diagnosis of pneumoconiosis. *IEEE Trans. Systems, Man, and Cybernetics,* **SMC-4**:40, 1974.

13. Kruger, R. P., Townes, J. R., Hall, D., *et al.*: Radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors. *IEEE Biomed. Trans.,* **BME-19**:174, 1972.

14. Lendaris, G. G., and Stanley, G.L.: Diffraction pattern sampling for automatic pattern recognition. *Proc. IEEE,* **58**:198, 1970.

15. Lusted, L. B.: Perception of the roentgen image: applications of signal detectability theory. *Rad. Clinics North Am.* **7**:435, 1969.

16. Moore, R. T.: NIOSH Report. Presented Sept. 1972 Meeting, Rad. description of the Pneumoconioses, Cleveland, Ohio.

17. Morgan, R. H., Bates, L. M., Gopalarao, U. V., *et al.*: The frequency response characteristics of x-ray films and screens. *Am. J. Roentgenol. Rad. Ther. Nucl. Med.* **92**:426, 1964.

18. Pickett, R. M.: Visual analysis of texture in the detection and recognition of objects. *In* Picture Proc. Psychopictorics. New York, Academic Press, 1970; pp. 289-308.

19. Reger, R. B., and Morgan, W. K. C.: On the factors influencing consistency in radiologic diagnosis of pneumoconiosis. *Am. Rev. Resp. Dis.* **102**:905, 1970.

20. Reger, R. B., Smith, J., Kibelstis, A., *et al.*: The effect of film quality and other factors on roentgenographic categorization of coal workers pneumoconiosis. *Am. J. Roentgenol. Rad. Ther. Nucl. Med.* **115**:462, 1972.

21. Rosenfeld, A., and Troy, E.: Visual texture analysis. *Proc. UMR-Kelly Commun. Conf.,* Rolla, Mo., October 1970.

22. Rosenfeld, A., and Troy, E.: Visual Texture Analysis. Tech. Report 70-116, Univ. of Md., June 1970.

23. Rossman, K.: Measurement of the modulation transfer function of radiographic systems containing fluorescent screens. *Phys. Med. Biol.* **9**:551, 1964.

24. Sutton, R. N., and Hall, E. L.: Texture measures for automatic classification of pulmonary disease. *IEEE Trans. Computers* **C-21**:667, 1972.

25. Yerushalmy, J.: The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. *Radiol. Clin. North Am.* **7**:381, 1969.