USCEE REPORT #433

# A Study to Determine the Feasibility of Computer Diagnosis of the Pneumoconioses

by

Richard P. Kruger

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
Powell Hall of Engineering
University Park/MC-0272
Los Angeles, CA 90089 U.S.A.

## RESEARCH ABSTRACT

A study was undertaken to determine the feasibility for the possible automated mass diagnostic screening of pneumoconiosis radiographs. Two distinct textural feature extraction methods involving digital and coherent optical approaches were undertaken. The performance of the two automated diagnostic systems is described in detail in tables presented in the main body of this report. Analagous results are presented for diagnosis obtained from experienced radiologists asked to analyze the same films given to the automated systems.

As this was a feasibility study the available data base was necessarily limited. As the data base is expanded, the statistics of the measured features will become better known. Thus, one may conjecture that performance will tend to improve as the data base is enlarged. Nevertheless, performance to date is most encouraging.

Normal/Abnormal classification accuracy (on a testing basis) for the optical system was 90.8%. The comparable percentages for the physicians ranged from 83.0% to 97.9%. It should be noted that the physician who's accuracy was 97.9% is a radiologist with extensive experience at diagnosing pneumoconiosis. Likewise, the missed abnormal testing rate for the optical system was 2.9%. The comparable rate for the physicians ranged from 1.0% to 6.9%

The digital system also compared well with the physician results, though it is important to note that only a subset of the films was available for digital processing. For example, on a four class problem (a normal class and three progressively more severe abnormal classes), the accuracy of the digital results was essentially equivalent to the combined accuracy of the physicians.

TABLE OF CONTENTS

## INTRODUCTION

Public Law 91173 has specified that each coal worker has the right to the obtaining of his chest roentgenogram at regular intervals to detect the possible presence and progression of coal workers pneumoconiosis. There are approximately 130,000 such miners who will be covered by this act. Examination intervals will not exceed 5 years, and it is likely that in cases showing progressive disease this examination interval will be as short as two years.

Each roentgenogram will be read by at least two readers according to grading standards stipulated in the U. I. C. C. /Cincinnati Classification scheme [1,2] (see Appendix 1). If there is significant disagreement, between the first 2 readers a third reader will decide a final category of disease. The first reader is paid by the mine operator and the second and third readers are paid through taxes under the Congressional appropriations for this act. It seems likely that such procedures will result in a progressive overburdening of the current radiographic diagnostic system as it currently exists, since it is estimated that 300,000 roentgenograms may eventually be examined per year. (It is assumed that this projected number will include current as well as several earlier chest roentgenograms.) Thus, some automated procedure for diagnostic mass screening of such films would be desirable and most probably necessary. It has been estimated that at least 80 percent of all films read are diagnosed as not showing radiographic evidence of pneumoconiosis. Thus, a

cost effective and rapid automatic film reader which consistently screens out all definite normals would significantly relieve the projected overburdening previously referred to. In addition, it is also conceivable that this automatic system could produce quantitative indices pertinent to the extend and time progression of losses of normal pulmonary vascular patterns by small and large regular and irregular radiopacities which characterize the pneumoconioses. Thus, quantitative indices extracted from previous examinations could be compared with current data to help determine on-set and progression of disease. These quantitative measurements would be consistent in time and as such there would be no intra-reader variation. Also, diagnosis would be based on fixed quantitative measurements that would be minimally subject to illogical variations between serial examinations. This would be of substantial benefit considering the intra and inter reader variation referred to by Garland [3] and Yerushalmy [4].

It is fortunate that there presently exists a semi-quantitative grading procedure for radiographic diagnoses of the pneumoconioses. This type of descriptive procedure for diagnosis is usually not available. It has provided a framework within which the physicians and technologists can communicate in undertaking a study of possible means by which this diagnosis can be automated.

Recent investigations by Liddell [5] seem to support the contention that for simple pneumoconiosis the twelve point UICC profusion categories from 0/- to 3/4 exhibit on approximate linear relationship to the amount of dust deposition in the lungs.

Reger and Morgan [6, 7] conducted studies concerning the

effects of inter observer variation and film quality on assignment of profusion category. In only 56.7 percent of the cases was there agreement for a four diagnostic class assignment based on increasing profusion category among four highly experienced readers of pneumoconiosis films. The percentage was surprising 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. They also concluded that film quality had an appreciable and significant effect on diagnostic categorization. A marked tendency to place underpenetrated (under exposed) films in a higher profusion category and overpenetrated (over exposed) films in a lower profusion category than the same films properly exposed was noted. These tendencies were least exhibited by the most experienced readers. It was concluded that standardized exposure suggested by Jacobson [8] and refusal to interpret over or underpenetrated films would be advantageous. Finally, intra observer variations on several examinations of the same films with experienced readers was less significant than the other previously mentioned effects.

## IMAGE TEXTURAL DISCRIMINATION

The problem of manually detecting and grading simple pneumoconioses opacities from radiographs appears to be largely one of discrimination between normal pulmonary vascularity (lung markings) pattern and partial or complete obliteration of this normal tree like pattern by opacities of various sizes and profusions which themselves exhibit a more or less textural nature. Complicated disease, on the other hand, is characterized by larger radiopacities which perhaps are not characterized nearly so well by a textural description. Rosenfeld and Troy [9] described image texture ideally as the "repetitive arrangement of a unit pattern over a given area". They also stated, however, that in naturally occurring, as opposed to synthesized imagery, it would be difficult to identify such unit patterns or determine their repititive arrangement. Therefore the previous description is not literal and should be used only as a guide in the analysis of natural imagery.

Pickett [10] refers to two kinds of human subjective textural analyses. These might be described as deliberate and impressionistic. In the latter we receive an immediate impression of say the coarseness or fineness of a visual texture. In a sense, we get an immediate answer to a set of questions not consciously asked. Deliberate analysis, on the other hand, implies that the observer is looking more closely at a pattern or texture which would perhaps be the case in trained observation. In empirical studies Gibson [11] has shown that regular textures seem to convey stronger impressions than irregular ones and also concluded that his observers relied on impres-

sionistic analysis. It has also been concluded that many textures are discriminated by assessments of size, shape, orientation, and repetition rate. Julesz [12] demonstrated that first and second order probabilities of the image field gray level were the highest that the human observer could perceive. Moreover, differences in such probabilities appear to be discriminated on the basis of differences in the size, shape, or orientation of clusters formed by contiguous like-shaded elements. Changes in arrangement that come from varying third or higher order probabilities, i.e., changing relationships among non-adjacent picture elements, are not visible in an immediate impression. While this study used dot clusters as experimental data, this suggests the use of higher order approximate probability density functions as a possible approach to computer texture discrimination. Therefore, it is a psychological fact that the human does perceive and discriminate textural properties of a visual scene. However, the importance of textural properties seen in human perception is less well understood or explored than that of contour identification. Correspondly, techniques for computer textural discrimination and location are not as well developed as those pertaining to contour or boundary analysis. Indeed, these two image characteristics are not mutually exclusive in their effect on the appearance of an image.

In computer analysis of natural image texture Hawkins [13] made several observations. First the notion that natural texture is strictly locally repetitive is only approximately true. That is, what is being replicated is a pattern class within which all examples are

regarded as equivalent. Furthermore, replication may be subject to spatial nonlinear phase shifts and, therefore, only be approximate. Second, non-random arrangement of pattern parts appears to be associated with meaningful textures. He concludes with the observation that texture classification may well be one of the difficult tasks in the field of image processing.

Since our approach to textural feature extraction of pneumoconioses opacities will of necessity be experimental, some probable statistical indices will be explored. With statistical analysis it is hoped that useful measures of the organization and arrangement of the textures in question will be derived without having to focus attention on the specific structural properties. For instance, these statistical measurements will not attempt to recognize or trace the normal vascular tree or the pneumoconiotic lessons per se, but merely compute quantitative impressions which will characterize them.

Textural discrimination of images will be attempted in both the spatial and spatial frequency domain. The specific approach, therefore, is to consider pneumoconiotic opacities or normal lung markings as distinct image textures. The task is to experimentally apply several textural discriminators and subsequently apply feature selection and pattern recognition techniques to obtain a limited diagnostic classification.

## AUTOMATIC LOCATION OF ANATOMICAL POINTS OF REFERENCE WITHIN THE POSTERIOR-ANTERIOR RADIOGRAPH

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 [14, 15]. In addition to its diagnostic mass screening potential, the algorithm located the larger anatomical features present in the P-A film. Figure 1 is an illustration of the anatomical features located and the anatomy measured by the current version of this algorithm.

1. THr - lateral thoracic distance at the diaphragm level.
2. MDL - the chest midline.
3. MLE - maximum left heart extend normalized with respect to THr.
4. MRE - maximum right heart extent normalized with respect to THr.
5. CTR - MRE+MLE (cardio-thoracic ratio).
6. DIAG - a diagonal cardiac measurement normalized with respect to THr.
7. DIAG 1 - a diagonal cardiac measurement normalized with respect to THr.
8. TA - the estimated thoracic cavity area shown as a dotted line.
9. LA - left cardiac area normalized with respect to TA.

10. RA - right cardiac area normalized with respect to TA.

11. RP - a 4th order polynomial which models the right superior mediastinum.

12. LP - a 4th order polynomial which models the left superior mediastinum and left atrial appendage.

13. R 1 through R 6 - these rectangles divide both the left and right lung fields into 3 zones or regions in a manner similar to that extent zones of the current UICC pneumonconiosis grading procedure. Figure 2 shows several P-A films with their cardiac borders and lung zones detected using the algorithm which executes in less than 10 seconds/film.

While the application of the previous algorithm will not be of prime importance during the feasibility period, it does serve to demonstrate several things. 1. It will diagnose an abnormality of cardiac size and shape (CO) and suspect cor pulmonale (CP) both of which are obligatory symbols in the UICC classification scheme. 2. It will indicate when the cardiac outline is shaggy or ill-defined. 3. It does accurately locate the 6 rectangular lung zones within which the primary diagnosis of pneumoconiosis manually made. It was the major thrust of this study to determine if it was possible to reliably differentiate between normal lung vascularity and the progressive obliteration of those normal lung markings by radiopacities caused by pneumoconiosis.

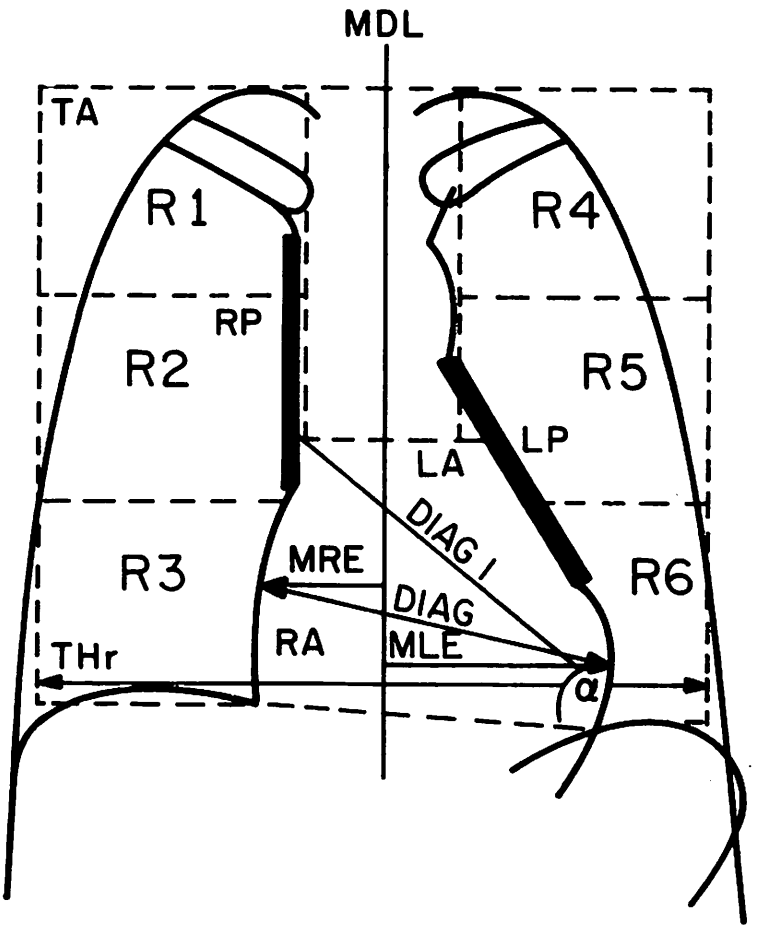



FIGURE 1

ANATOMICAL FEATURES LOCATED AND/OR MEASURED

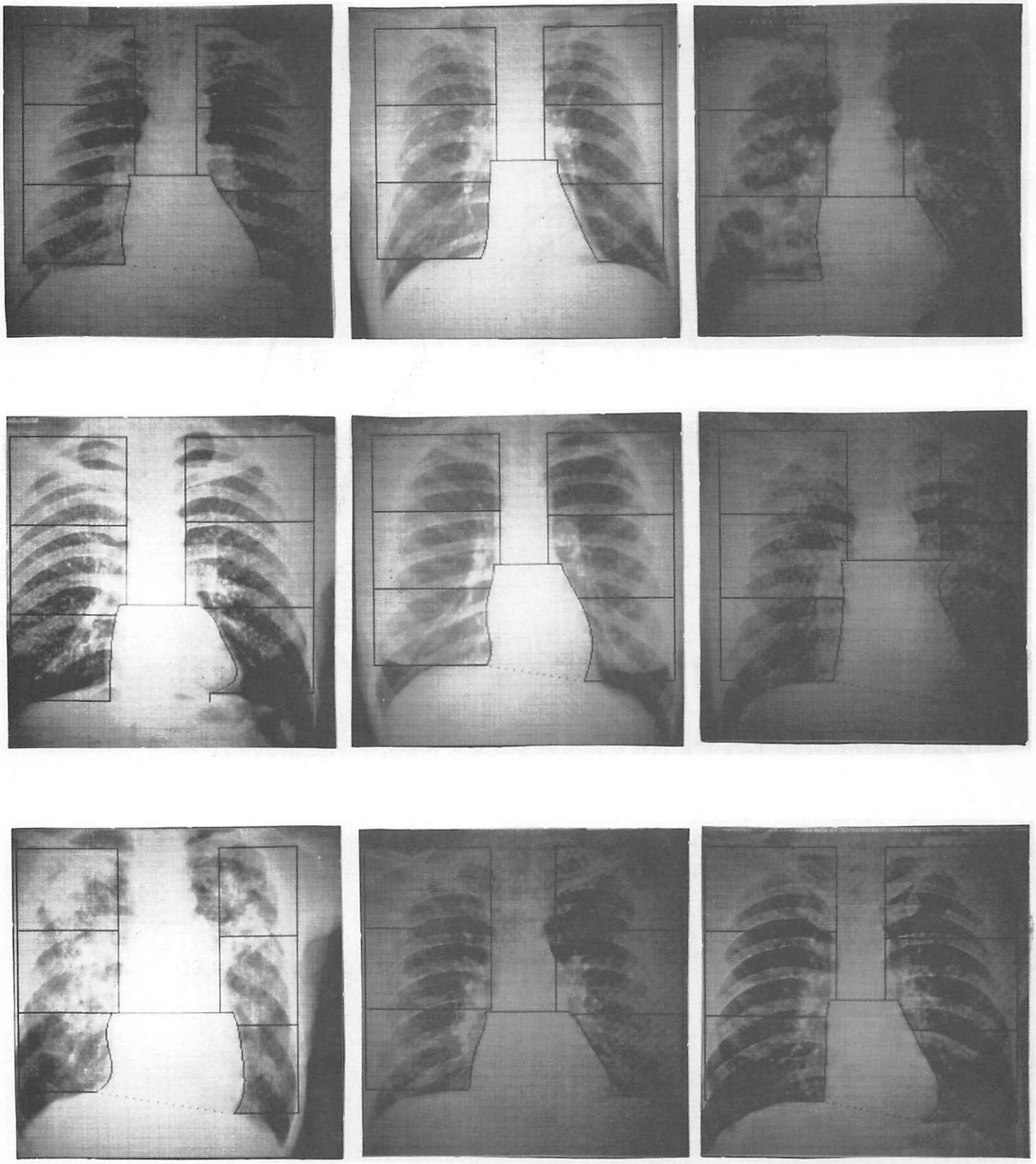


FIGURE 2: SEVERAL PNEUMOCONIOTIC AND NORMAL P-A CHEST FILMS WITH CARDIAC PROJECTIONS OUTLINED AND DENOTED MEASURES EXTRACTED

## DATA BASE

The film base for the feasibility study was selected by Dr. George Jacobson with the aid of Dr. A. Franklin Turner of the LA-County-USC Medical Center Department of Radiology. Of the 102 abnormal films approximately 40% were 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 Drs. Jacobson and Turner as a committee. 33 of the normal films were selected from USC-LA County employee films of individuals with no known work history in a dusty environment and no clinical or radiographic evidence of lung dysfunction. 6 ACR standard films from profusion category 0/1 were also included. All films were either judged of (+) or (+ -) quality according to the UICC classification criteria.

It was decided to produce 3.6 by 4.5 inch Logetronic reproductions of specific lung zones as a primary input film data for this study. These reproductions were to roughly correspond to the 6 lung zones extractable by the previously mentioned algorithm. The following specifications were followed on all reproductions. 1. The reproduced lung region was to be one to one in size with the equivalent area on the full P-A film. 2. The regional areas selected for reproduction should contain an ample mix of right and left lung fields from as many of the 6 lung zones as possible in order that regional bias in the automated feature extraction and measurement could be minimized. 3. The lung regions selected in the abnormal cases for reproduction were to as

accurately as possible characterize the severity and type of lesion present in the field. 4. 100% dodging was used on all reproductions. The list of abnormal diagnoses for the films selected for this study are shown below in Table 1.

| Proposed Computer Classes | | | Proposed Computer Classes | | |
|---|---|---|---|---|---|
| 4 | # 1 | $p^3/3$ | 4 | #20 | $q^3/4$ |
| 3 | # 2 | $u^2/3$ | 4 | #21 | $p^3/4$ |
| 4 | # 3 | $q^1/1$ $t^2/3$ $t^3/3$ | 2 | #22 | $t^1/1$ |
| 4 | # 4 | $p^1/1$ $s^2/2$ $T^2/3$ | 3 | #23 | $t^2/3$ |
| 2 | # 5 | $t^1/2$ | 4 | #24 | $p^3/4$ $s^3/3$ |
| 4 | # 6 | $r^1/1$ C WD | 4 | #25 | $p^3/4$ |
| 4 | # 7 | $r^1/1$ C 1D (CP) | 3 | #26 | $p^2/2$ $s^2/2$ |
| 2 | # 8 | $s^1/1$ | 4 | #27 | $t^3/3$ |
| 2 | # 9 | $q^1/1$ | 4 | #28 | $r^3/4$ |
| 4 | #10 | $p^3/3$ A WD | 4 | #29 | $r^3/3$ |
| 2 | #11 | $t^1/2$ | 3 | #30 | $r^1/2$ $u^2/2$ |
| 2 | #12 | $s^1/1$ | 1 | #31 | $q^2/2$ |
| 3 | #13 | $r^2/3$ | 2 | #32 | $t^1/1$ |
| 3 | #14 | $r^2/2$ C WD (D1) | 3 | #33 | $s^2/3$ |
| 3 | #15 | $q^2/2$ C 1D | 2 | #34 | $s^1/2$ |
| 3 | #16 | $q^2/2$ | 2 | #35 | $q^1/1$ |
| 4 | #17 | $q^2/2$ A 1D | 2 | #36 | $t^1/1$ |
| 2 | #18 | $q^1/2$ | 4 | #37 | $p^3/3$ |
| 4 | #19 | $q^2/3$ B Cn | 4 | #38 | $r^2/2$ B WD |

TABLE 1

TABLE 1 (continued)

| Proposed Computer Classes | | |
|---|---|---|
| 4 | #39 | $t^3/3$ |
| 4 | #40 | $q^3/3$ |
| 4 | #41 | $u^3/3$ |
| 3 | #42 | $p^2/2$ $s^2/2$ |
| 3 | #43 | $q^2/2$ $s^1/1$ |
| 4 | #44 | $s^3/3$ |
| 3 | #45 | $u^2/2$ |
| 2 | #46 | $r^1/1$ $t^1/1$ |
| 3 | #47 | $r^2/1$ $u^2/1$ |
| 4 | #48 | $q^2/2$ C |
| 3 | #49 | $r^2/2$ $s^1/1$ |
| 3 | #50 | $q^2/2$ $t^1/1$ |
| 3 | #51 | $q^2/2$ |
| 3 | #52 | $q^2/3$ |
| 4 | #53 | $q^3/4$ |
| 4 | #54 | $q^3/4$ |
| 4 | #55 | $q^3/4$ |
| 2 | #56 | $r^1/2$ |
| 2 | #57 | $r^1/2$ |
| 3 | #58 | $r^2/1$ |
| 3 | #59 | $r^2/3$ |
| 4 | #60 | $r^3/3$ |
| 3 | #61 | $p^2/3$ |
| 3 | #62 | $p^2/3$ |
| 4 | #63 | $p^3/3$ |
| 4 | #64 | A w d |

| Proposed Computer Classes | | |
|---|---|---|
| 1 | #65 | normals |
| 1 | #66 | normals |
| | • | • |
| | • | • |
| | • | • |
| 1 | #97 | normals |
| 3 | #98 | s 1/1 p 2/2 |
| 3 | #99 | u 2/2 |
| 2 | #100 | t 1/2 |
| 3 | #101 | s 1/1 u 2/3 |
| 2 | #102 | r 1/1 |
| 2 | #103 | r 1/1 |
| 1 | #104 | q 0/1 |
| 1 | #105 | q 0/1 |
| 1 | #106 | q 0/1 |
| 2 | #107 | q 1/1 |
| 2 | #108 | q 1/1 |
| 1 | #109 | s o/1 |
| 1 | #110 | 1 0/1 |
| 1 | #111 | p 0/1 |
| 2 | #112 | s 1/0 |
| 2 | #113 | s 1/0 |
| 2 | #114 | t 1/1 |
| 2 | #115 | t 1/1 |
| 2 | #116 | r 1/1 |
| 2 | #117 | r 1/1 |
| 3 | #118 | s 2/2 |

TABLE 1 (continued)

| Proposed Computer Classes | | |
|---|---|---|
| 4 | #119 | q 3/3 |
| 4 | #120 | q 3/3 |
| 3 | #121 | q 2/2 |
| 3 | #122 | q 2/2 |
| 2 | #123 | q 1/1 |
| 2 | #124 | q 1/1 |
| 4 | #125 | t 3/3 |
| 4 | #126 | t 3/3 |
| 3 | #127 | p 2/2 |
| 3 | #128 | p 2/2 |
| 4 | #129 | r 3/3 |
| 4 | #130 | r 3/3 |
| 2 | #131 | p 1/1 |
| 2 | #132 | p 1/1 |
| 4 | #133 | p 3/3 |
| 4 | #134 | p 3/3 |
| 2 | #135 | s 1/1 |
| 2 | #136 | s 1/1 |
| 4 | #137 | t 3/3 |
| 4 | #138 | t 3/3 |
| 2 | #139 | u 1/1 |
| 4 | #140 | u 3/3 |
| 2 | #141 | u 1/1 |

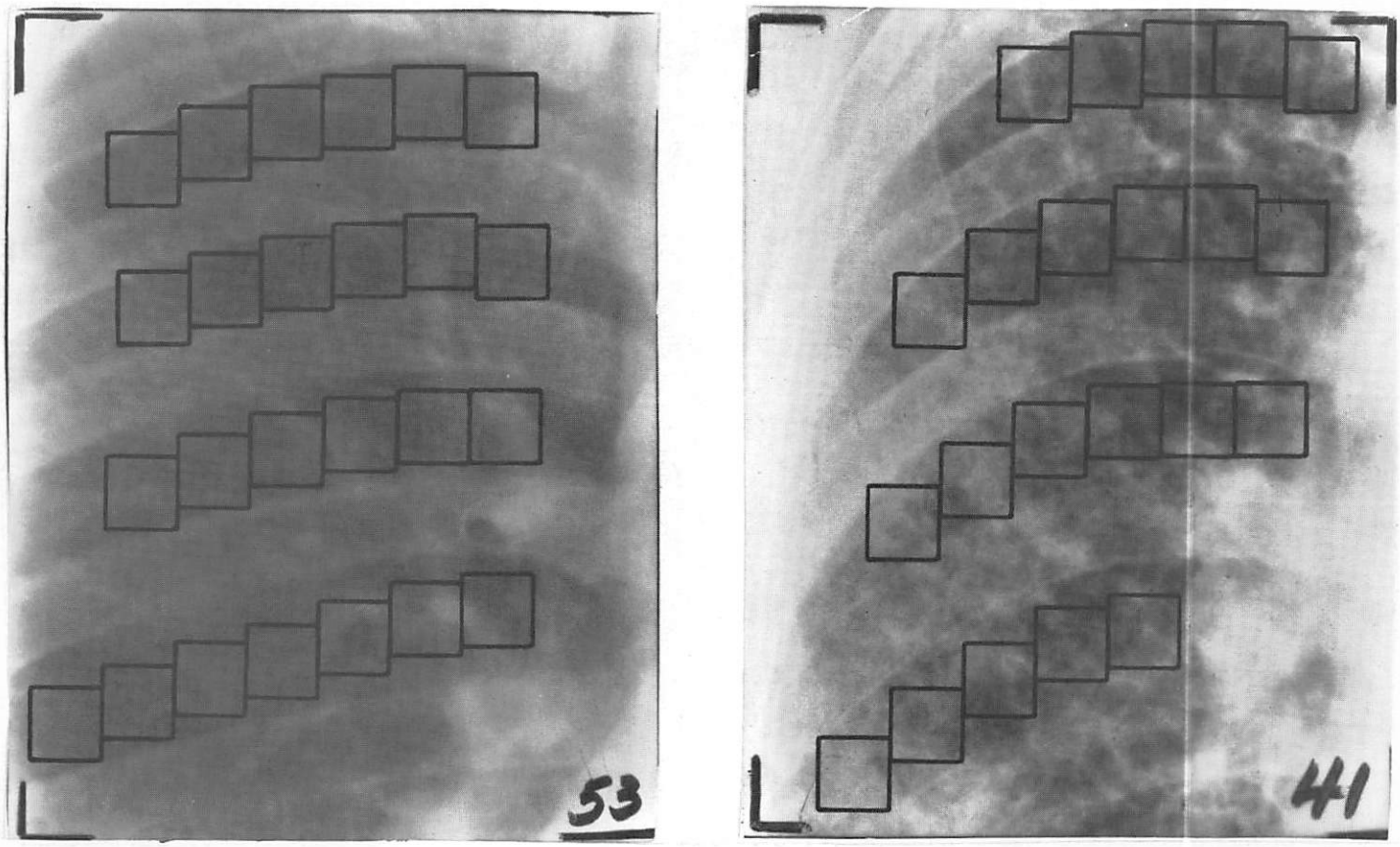
## INITIAL DIGITAL DATA COLLECTION

Ninety seven Logetronics reproductions were sent to the Los Alamos Scientific Laboratory for digitization on a Photometric Data Systems scanning microdensitometer. This device is sensitive over a 0-4 film density range with 1 of 256 possible density values recordable for each image picture elements (pixel). Each film was digitized to a resolution of 9.8 pixels/mm. This represented a 900 by 900 matrix of pixels over a 3.6 inch square. While this would seem to be an overspecification on density sensitivity and resolution, the risk to data base validity is much more severe if the data is less accurate than necessary than if it is more accurate than necessary. This purposeful overspecification was also undertaken tc explore the possibility of developing more lenient specifications at a later time.

## DATA MANAGEMENT AND EXTRACTION FROM THE DIGITAL IMAGES

Figure 3 is an example of a normal and abnormal film from the data base. It was decided after consultation with Dr. Turner, 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 more density radiographic structures. These posterior inter-rib spaces were therefore manually extracted in the following manner. First a computer generated grid was superimposed upon each of the digitized images. This grid allowed visual extraction of from [4 to 10] 100 by 100 pixel squares from each inter-rib space. These squares are delineated for the examples in Figure 3. When extracting squares an effort was made to avoid the most radiographically dense portion of the hilar regions, however no effort was made to avoid the less prominent anterior ribs. Since there were 3 to 4 such inter-rib spaces/film, the complete digital data base consisted of over 1800 such squares and 298 inter-rib spaces.

It was decided to consider each posterior inter-rib space as a separate input to the diagnostic classifier for the following reasons. 1. Since the regional films in the data base were from all six lung zones, any eventual automated classification technique would not be sensitive to any single square nor any specific lung zone. 2. Feature extraction on several 100 by 100 pixel squares within an inter-rib




(a) (b)

FIGURE 3

NORMAL (a) AND ABNORMAL (b) LUNG ZONE FILMS
WITH INTER-RIB SQUARES SELECTED

space should be more indicative of diagnostic classification than any single square. 3. A data base of 298 inter-rib spaces expands the original data base by over 3 times permitting a more mathematically significant automated classification. 4. These local digital examinations will complement the more global coherent optical techniques to be discussed later.

## TEXTURAL FEATURE NORMALIZATION AND EXTRACTION IN THE INTER-RIB SPACES

A system diagram of an automated pattern recognition system is based on four main, but not necessarily distinct sequential subsystems. The input sensor has been discussed in the digital case. The input data consisted of 298 inter-rib spaces from the lung zone films where each inter-rib space contained 4 to 10 non-overlapping 100 by 100 pixel squares. Thus the input data base consists of 298 samples. The data from all the squares in an individual inter-rib space was combined to form a gray level histogram for that complete inter-rib space. A monotonic transformation of gray levels was performed which produced an inter-rib space with 8 equally likely gray levels [16, 17, 18]. This preprocessing step is designed to negate the effect of additive and multiplicative constants introduced due to inconsistency in photographic 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 occurrences. The textural feature extraction measures were all based on spatial gray level dependence matrices [19, 20] under the assumption that visual texture-context information in each inter-rib space is contained in the spatial relationship between image picture elements at several fixed distances and angular orientations as illustrated in Figure 4. More specifically, it shall be 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

DIAGRAM OF DISTANCE AND ANGULAR DEPENDENCE FROM CENTRAL REFERENCE

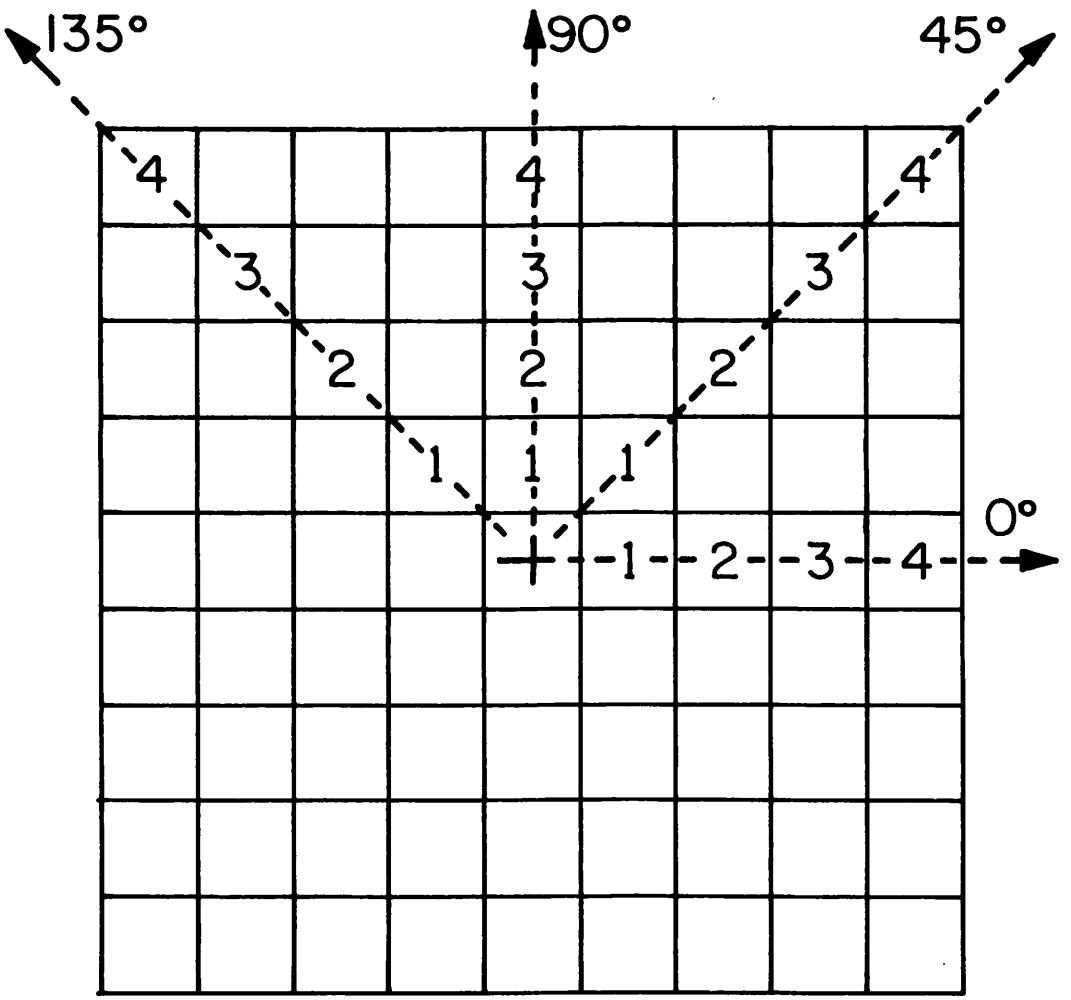



FIGURE 4

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. Therefore an 8 by 8 symmetric count matrix is formed within each 100 pixel square and then all the count matrices per inter-rib space are first summed to form one total count matrix and finally normalized to create a matrix of relative frequencies for each inter-rib space as a function of a and d.

$$p(i, j, a, d) \qquad \begin{aligned} i &= 0, 1, \cdots, 7 \\ j &= 0, 1, \cdots, 7 \\ a &= 0^\circ, 45^\circ, 90^\circ, 135^\circ \\ d &= 1, 3, 7, 11 \end{aligned} \tag{1}$$

Therefore for each of the 298 spaces there were 4 x 4 or 16 such 8 by 8 matrices per space. The following 5 textural measurements $T_k(a, d)$ $k = 1, \cdots, 5$ were computed for each matrix.

$$T_1(a, d) = \sum_{i=0}^{7} \sum_{j=0}^{7} i \cdot j \; p(i, j, a, d) \tag{2}$$

$$T_2(a, d) = \sum_{i=0}^{7} \sum_{j=0}^{7} (i-j)^2 p(i, j, a, d) \tag{3}$$

$$T_3(a, d) = \sum_{i=0}^{7} \sum_{j=0}^{7} \frac{p(i, j, a, d)}{1+(i-j)^2} \tag{4}$$

$$T_4(a,d) = -\sum_{i=0}^{7}\sum_{j=0}^{7} p(i,j,a,d)\log p(i,j,a,d) \tag{5}$$

$$T_5(a,d) = \sum_{i=0}^{7}\sum_{j=0}^{7} |i-j|\, p(i,j,a,d) \tag{6}$$

for $a = 0°, 45°, 90°, 135°$ $d = 1, 3, 7, 11$

$T_1$ is an autocorrelation measure designed to measure image coarseness since the coarser the texture, the greater the (i, j) difference before the match becomes poor. $T_2$ is a dissimilarity measure often called the moment of inertia and is another coarseness measure. $T_3$ measures the extent to which the same or similar gray levels tend to be neighbors. This measure and measure $T_2$ will be very highly correlated in a negative sense. $T_4$ is a conditional entropy measure and measures image homogeneity. The more homogeneous the image, the smaller $T_4$ will be. A maximum will occur when all p(i, j) are equal or at maximum inhomogeneity. $T_5$ is another dissimilarity measure which is similar to $T_2$ and highly correlated with it in a positive manner. A total of $16 \times 5 = 80$ 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).

$$\overline{M}_k(d) = \frac{1}{4}\sum_{a=1}^{4} T_k(a,d) \tag{7}$$

$$R_k(d) = \max T_k(a) - \min T_k(a) \qquad a = 0°, 45°, 90°, 135° \tag{8}$$

$$V_k(d) = \frac{1}{4} \sum_{a=1}^{4} (T_k(a,d) - \overline{M}_k(d))^2 \tag{9}$$

for $k = 1, \cdots, 5$ $\qquad d = 1, 3, 7, 11$ .

There are 3 measures time k = 5 measurements time 4 distances or 60 final measures. The final $\overline{M}$, R and V measures did not possess a strict directional bias and were therefore explicitely a function of (d) and only implicitely a function of (a). These measures have reduced from 40,000 to 100,000 pixels in each inter-rib space down to 60 textural measures.

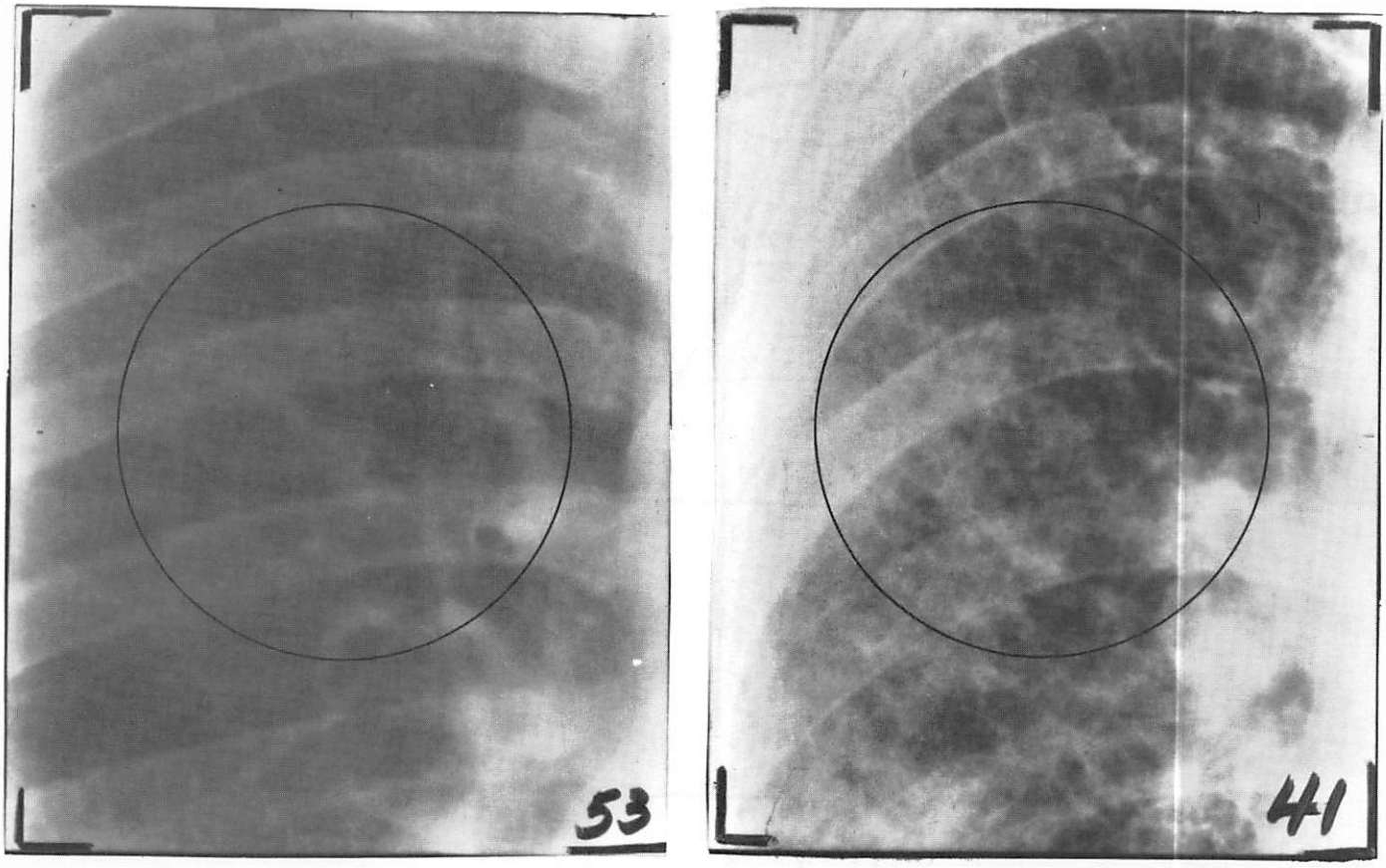
## FOURIER TRANSFORM DOMAIN FEATURE EXTRACTION USING A COHERENT OPTICAL APPROACH

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 to be presented now treat each of the 141 lung region films as an entity and as such will treat more global aspects of visual texture to include the influence of both posterior and anterior rib projections as well as the lung field. Figure 5 describes in simplest form, the apparatus used to extract the spatial frequency measures of visual texture. A helium-neon laser emits a light which passes through a collimating lens and then through the input film image. The transmitted light from the film is next passed through a positive thin lens which performs the Fourier transformation. The transformed image is then projected onto a detector in the transform plane and appropriate measurements are obtained. For the images in this experiment a circular 2.5 inch diameter aperture illuminated a circular area in the center of the rectangular film of the appropriate lung zone. No attempt was made to select a specific circular area other than to illuminate a region of the lung near the center of the film. Figure 6 is an example of an illuminated circular area within 2 typical regional films. The 2.5 inch aperture was selected since it was the largest available in the apparatus used.

The Fourier transform equation is

$$\hat{F}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, e^{j2\pi(xu+yv)} dxdy \qquad (10)$$

(^) denotes complex




(A) (B)

FIGURE 6

CIRCULAR APERTURE SUPERIMPOSED ON
SAME FILMS AS FIGURE 3

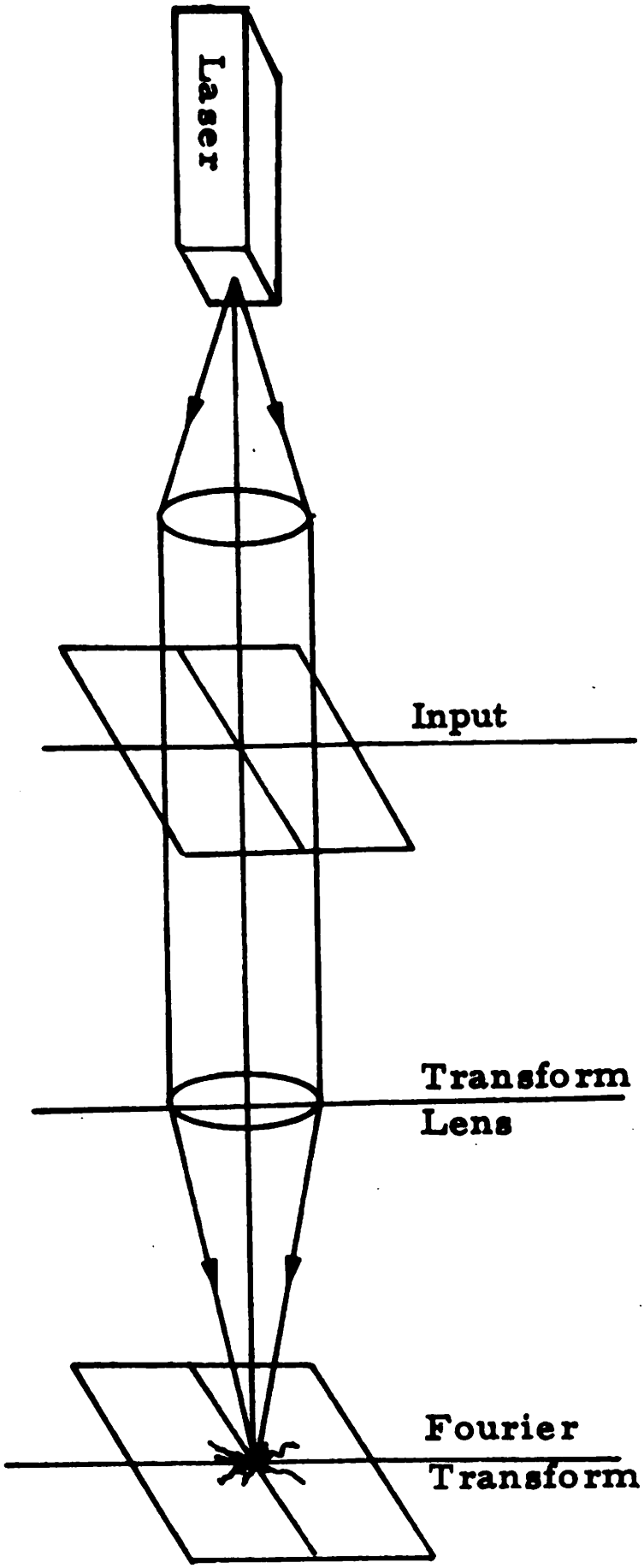



FIGURE 5: IDEALIZED OPTICAL APPARATUS

$f(x, y)$ is the input image with two spatial variables $x$ and $y$ which may be measured in inches. The Fourier transform $\hat{F}(u, v)$ is complex valued with two spatial frequency variables $u$ and $v$ usually measured in cycles/inch. The chosen transform textural measures will be insensitive to phase and as such will measure aspects of the magnitude of the transform denoted as $|\hat{F}(u, v)|$ where $|\hat{F}(u, v)|$ measures the image energy that is contained at a frequency $(u, v)$ or in any band of frequencies. It is well known that high frequency information pertains to the amount of sharpness of edge information in an image. It can be hypothesized for instance that lung regions with pneumoconiosis opacities will generate more of this higher frequency edge information than a normal film. It is also known that images with a marked directional bias will produce transforms with a related directional bias. Several investigators [21, 22, 23] have used these transform properties for terrain cell and lung vascularity classification studies, with varying degrees of success. A detector in the transform plane which has N annular rings was used to obtain the total contribution of N radial frequencies which would accurately reflect the amount of edge by the relative strength of the higher energy annulli. In a like manner a detector with N angular wedges could measure directional bias which is independent of frequency.

Figure 7 is a block diagram of the Recognition Systems, Inc. Recording Optical Spectrum Analyzer ROSA-3. The device consists of an optical apparatus which forms the $|\hat{F}(u, v)|$ at the transform plane where a 64 element photodetector shown in Figure 8 senses the energy in 32 annular frequency ranges and 32 angular directions.

The $20^{\circ}$ dead zones serve as a read out path for the 64 detectors which have a self-ranging amplifier with a 6 decade dynamic range. Figure 9 depicts the physical configuration of the device. The 64 textural measures are obtained in approximately 100 μseconds and punched onto computer cards. The 64 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 has a similar purpose to histogram equalization in the digital data and in this case negates such film dependent parameters as dynamic range differences. The normalized energies in the rings were then logarithmically transformed such that their distributions were more nearly Gaussian. An example of an annular ring signature is shown in Figure 10.

At this point there exist 60 spatial and 64 coherent optical textural measures. The next section will deal with criteria for selecting a subset of these features as a last preprocessing step before diagnostic classification.

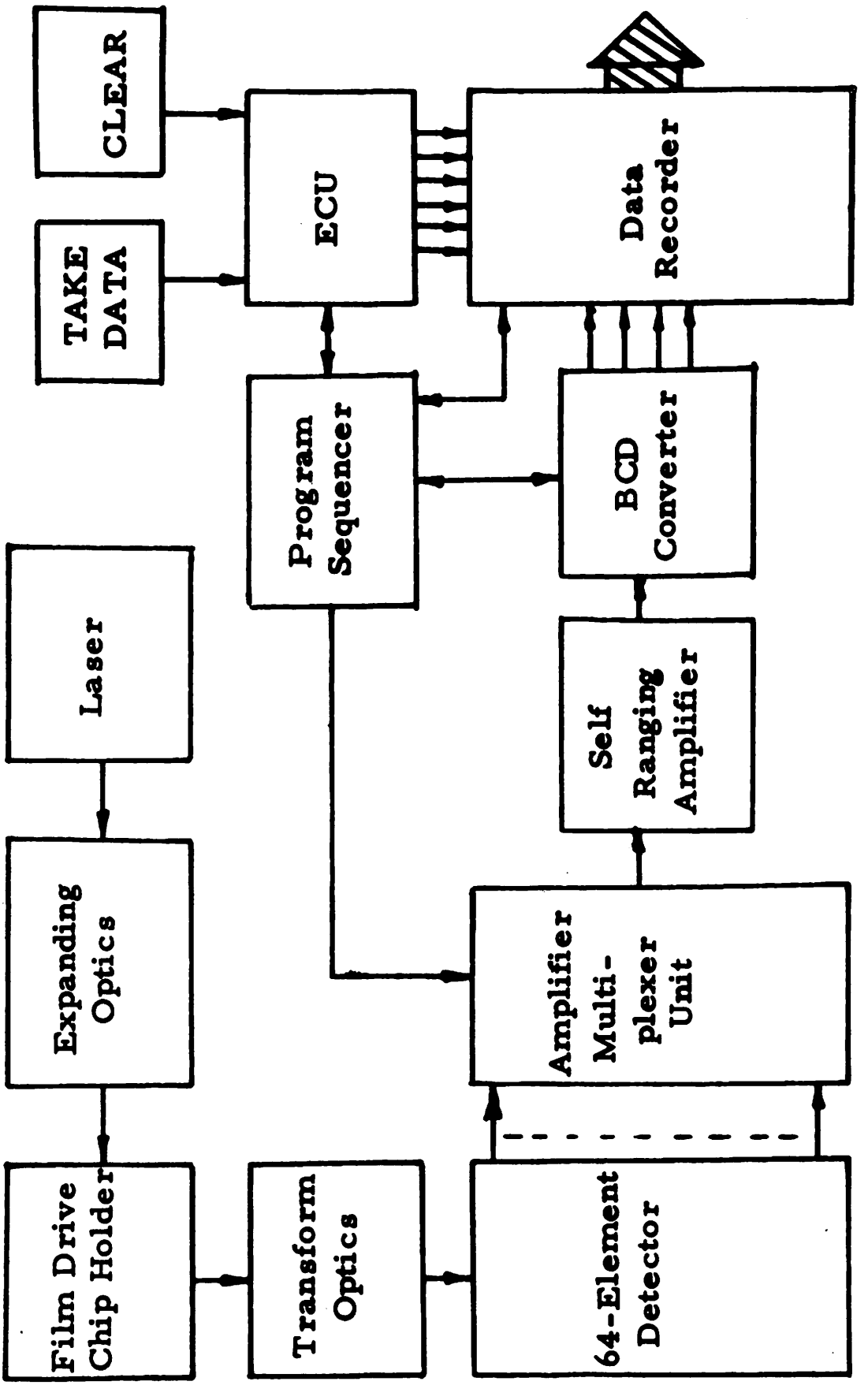



FIGURE 7: BLOCK DIAGRAM OF ROSA-3 OPTICAL SPECTRUM ANALYZER

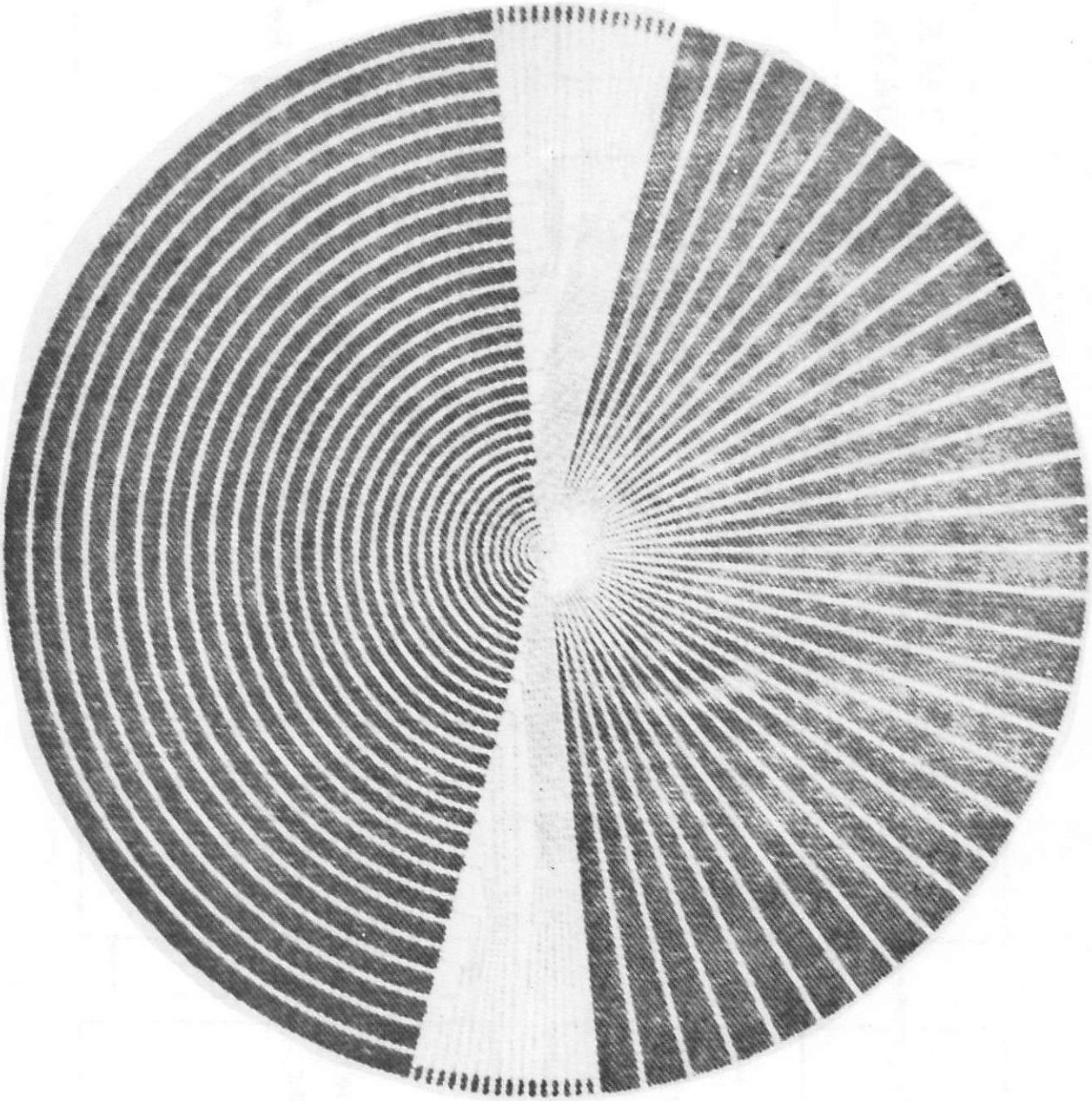


FIGURE 8: 64 ELEMENT PHOTODETECTOR

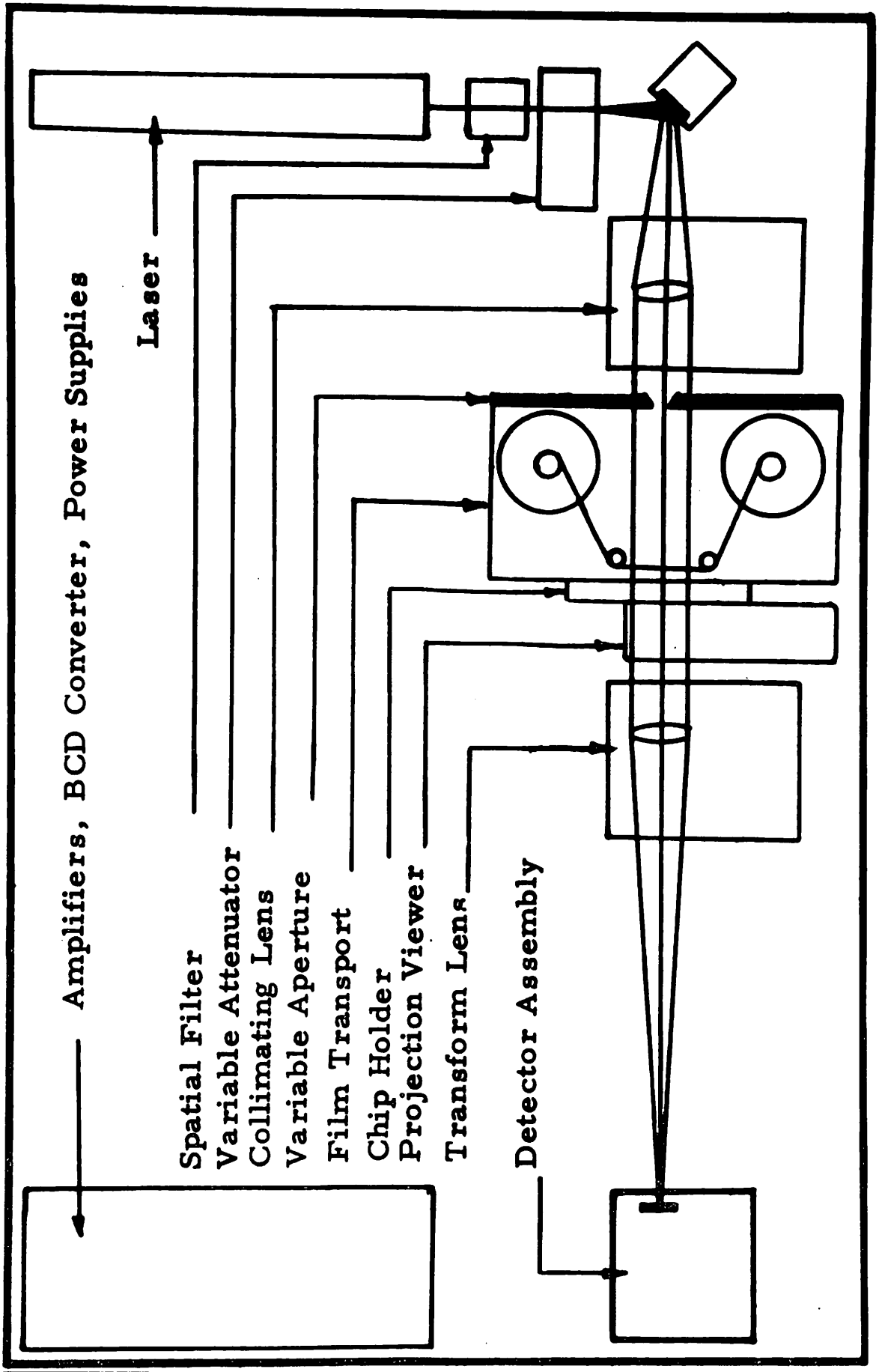



FIGURE 9: PHYSICAL CONFIGURATION OF ROSA-3

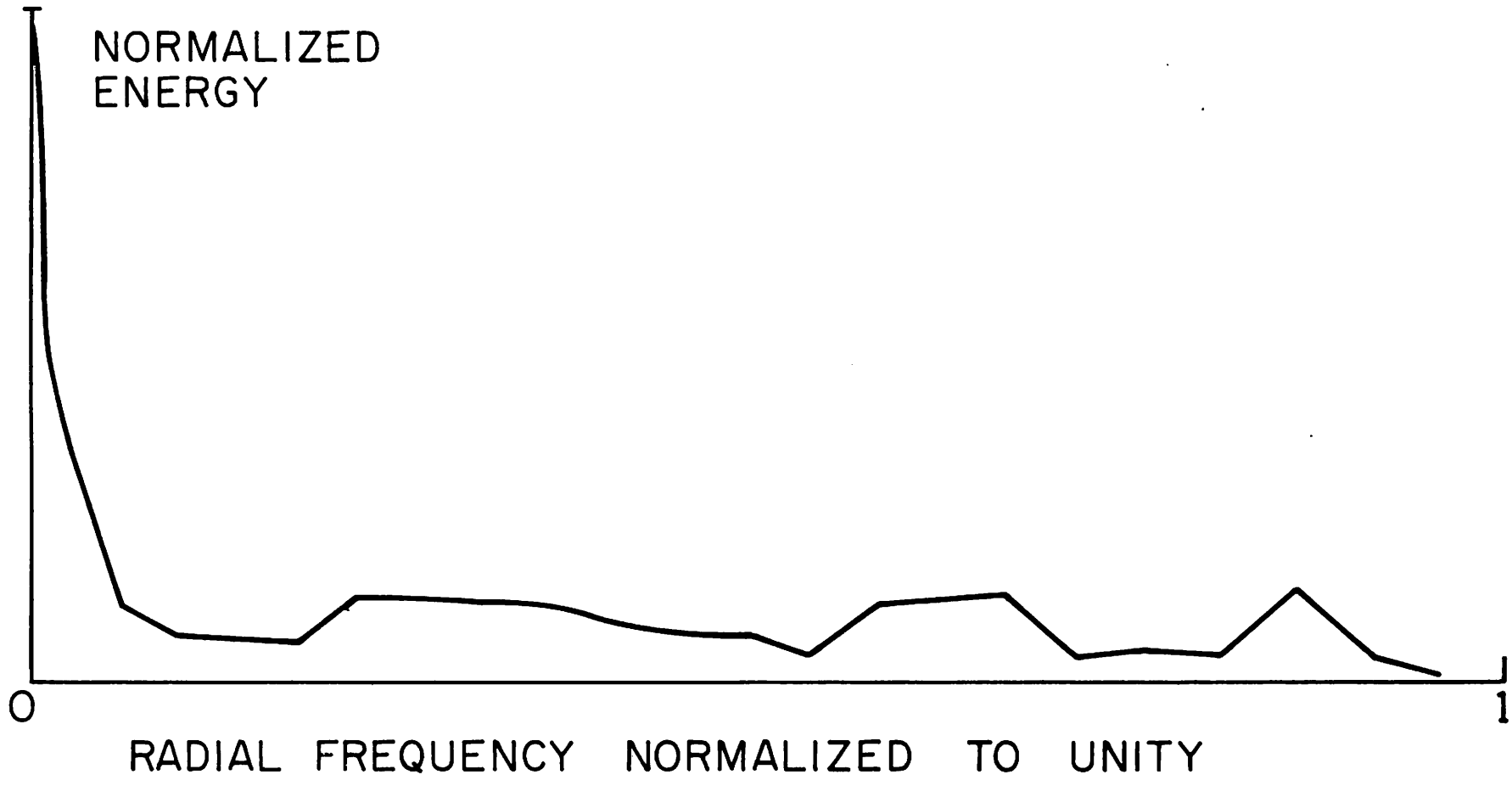

NORMALIZED
ENERGY
0
1
RADIAL FREQUENCY NORMALIZED TO UNITY

FIGURE 10

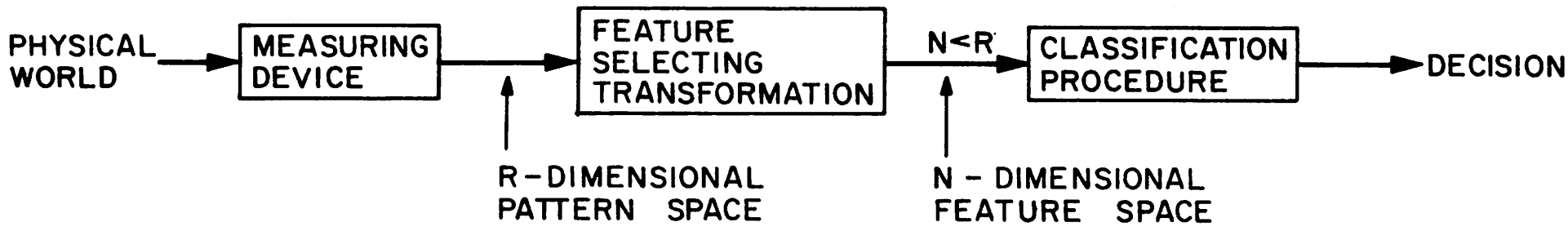
## PATTERN RECOGNITION

The general pattern recognition problem may be described as follows: a phenomenon in the physical world is to be assigned to one of a fixed number of "classes". Measurements are taken of this physical situation. Those measurements providing the most information for classification are selected. Finally, a decision procedure chooses which class to assign to the pattern.

More formally, it is assumed that measurements of the physical world are represented by scalar values. If there are R measurements taken, then the value of these measurements may be represented as a point in an R-dimensional vector space referred to as the "Pattern Space". In man-made systems it is usually assumed that the more data, the better. Thus, the value of R is usually quite large. For effective classification to take place, this pattern space must be transformed into a "Feature Space" such that only that information useful for classification is preserved. Let N be the dimensionality of the feature space. For a computationally effective classifier, N must in general be much smaller than R. A general system diagram is shown in Figure 11.

For the purposes of this work, it is assumed that the set of classes to which a pattern is to be assigned has been previously determined. Furthermore, it is also assumed that a set of patterns

("Prototypes") of known classification is available to us. These prototypes may then be used to "train" the pattern recognition system.




GENERAL PATTERN RECOGNITION SYSTEM

FIGURE 11

## TEXTURAL FEATURE SELECTION

The set of textural feature measurements will contain both redundancy and features which are of little value in separating the classes. For a classifier to work successfully, these features must be removed. In addition, the larger the set of measurements the classifier must deal with, the greater the numerical inaccuracy in computing discriminant functions will be. This is due to round off errors in digital computers. Thus, it is advantageous to make the feature space as low dimensional as possible, determining which of the original measurements contain the most useful information for the classifier.

In order to select meaningful features, a measure of the value of a feature must be defined. For a two class problem where the classifier assumes a Gaussian distribution of features, the "Divergence" [24] is such a measure. The divergence of a set of features is defined as

$$J(S_1, S_2) = \int_{-\infty}^{\infty} [p(\underline{x}|S_1) - p(\underline{x}|S_2)] \ln\left[\frac{p(\underline{x}|S_1)}{p(\underline{x}|S_2)}\right] d\underline{x} \tag{11}$$

where $S_1 \Rightarrow$ class 1, $S_2 \Rightarrow$ class 2. It can be shown that maximizing the divergence measure minimizes the bound on the probability of error $P_e$ such that

$$P_e \leq (P(S_1)P(S_2))\left[\frac{J(S_1, S_2)}{4}\right]^{-\frac{1}{4}} \tag{12}$$

when the classifier is a maximum likelihood estimator and the underlying distributions are multivariate Gaussian.

If classes $S_1$ and $S_2$ are assumed multi-variate Gaussian distributed the R dimensional $\overline{x}$ vector is distributed

$$p(\overline{x}/S_k) = \frac{\exp\left[-\frac{1}{2}(\overline{x}-\mu_k)^t[\Phi_k]^{-1}(\overline{x}-\mu_k)\right]}{(2\pi)^{R/2}|[\Phi_k]|^{\frac{1}{2}}} \tag{13}$$

$$p(\overline{x}/S_k) \sim N(\mu_k, [\Phi_k]) \qquad k = 1,2 .$$

Where $\mu_k$ is a mean vector and $[\Phi_k]$ is the covariance matrix for each class. The divergence distance is optimized over all linear transforms [T], of dimension N by R, which implies Gaussian distributions in the transformed space as well. Thus

$$p(\overline{x}/S_k[T]) \sim N(\mu_k^T, [\Phi_k^T]) \qquad k = 1,2 \tag{14}$$

and

$$\mu_k^T = [T]\mu_k \qquad [\Phi_k^T] = [T]^t[\Phi_k][T]$$

Therefore the divergence measure becomes a function of the linear transformation and will be denoted $J(S_1, S_2, T)$. The divergence can therefore be expressed as [25]

$$J(S_1, S_2, T) = \tfrac{1}{2}\mathrm{tr}([\Phi_2^T]^{-1}[\Phi_1^T] + [\Phi_1^T]^{-1}[\Phi_2^T] - 2[I]) + \tag{15}$$

$$+\tfrac{1}{2}\mathrm{tr}\left(([\Phi_1^T]^{-1} + [\Phi_2^T]^{-1})((\mu_1^T - \mu_2^T)(\mu_1^T - \mu_2^T)^t)\right)$$

where tr is the trace of the matrix.

It would be desirable to find which set of features, taken together, would be optimal. However, in practice this is not possible. A compromise is to calculate the value of features one at a time and then to choose those with the highest value. If Gaussian statistics are assumed and if features are analyzed one at a time, the divergence measure of the ith feature becomes

$$J(S_1, S_2, i) = \frac{(\sigma_i^{(1)} - \sigma_i^{(2)})^2 + 2(\sigma_i^{(1)} + \sigma_i^{(2)})(\mu_i^{(1)} - \mu_i^{(2)})^2}{2\sigma_i^{(1)}\sigma_i^{(2)}} \tag{16}$$

where $\sigma_i^{(k)}$ and $\mu_i^{(k)}$ are the variance and mean of feature i for class k and [T] has one non-zero term of unity value at location (1, i)

$$[T]_{1xR} = [00 \cdots \underset{i}{1} \cdots 0]_{1xR} \quad . \tag{17}$$

The divergence measure is defined for only the two class situation. In order to use the measure in a $k > 2$ class problem the sum of the paired divergences will be used as an optimization criteria. Therefore

$$J_k = \sum_{i=1}^{k} \sum_{j=i+1}^{k} J(S_i, S_j) \qquad k \geq 2 \tag{18}$$

and the features will be ranked by total summed divergence.

An additional problem must be dealt with. 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 is defined where

$$C = [c_{ij}], \tag{19}$$

$$c_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}}\sqrt{S_{jj}}}, \quad S_{ij} \text{ the } (i, j)\text{th element of the covariance matrix.}$$

The correlation of the i and jth feature is given by $c_{ij}$ with $c_{ij} = 1$ implying perferct correlation.

The procedure for choosing a set of features then becomes to choose the single feature with the highest divergence measure. Next, discard those features which are correlated with the chosen feature to a degree greater than some predefined bound. Continue by choosing from the remaining features the one with the highest divergence measure. Discard features correlated with any of the chosen set, and so on.

## CLASSIFIERS

A very large number of approaches to classification appear in the Pattern Recognition literature. Only a few however, have seen application in a practical situation. The classifiers discussed here will be referred to in terms of discriminant functions. That is, for a K diagnostic class problem, we will create K discriminant functions $g_1(\overline{x}), \cdots, g_k(\overline{x})$ where the $g_i$'s are functions of the textural feature measurements. These discriminant functions may be interpreted as assigning each possible set of feature measurements a value corresponding to its "closeness" to a given class, the decision rule is then to choose the diagnostic class whose corresponding discriminant function has the highest value.

Let $\overline{x}$ be a vector representing a point in the N-dimensional feature space. Then assign $\overline{x}$ to class i if

$$g_i(\overline{x}) > g_j(\overline{x}) \text{ for all } i \neq j \qquad 1 \leqq i, j \leqq K . \tag{20}$$

Classifier may be divided into two broad categories — "statistical" and "distribution free". Distribution free classifiers make no assumptions about the underlying statistical distribution of features. This can be an advantage, particularly if the number of training prototypes is small as the distribution would not be well defined. A common classifier of this type is the minimum distance with respect to class means approach.

Let $\overline{X} = \{x_1, \cdots, x_N\}$ be a set of feature measurements.

Then $g_k(\overline{x}) = (\sum_{i=1}^{N} x_i \mu_i^{(k)}) - \frac{1}{2} \sum_{i=1}^{N} \mu_i^{(k)^2}$ $k = 1, \cdots, K$ when (21)

$\mu_i^{(k)}$ is the mean value of the ith feature for the kth class. For the problem at hand, we know that textural feature measures for normal prototypes are quite similar while those for abnormal prototypes have a large variation. For optimal results, a classifier must take this into consideration.

Statistical classifiers make assumptions about the underlying distributions of features. In general, this assumption will not be completely valid. If, however, the distributions can be modeled sufficiently well, then diagnostic classification accuracy will be good. The most common assumption made about feature statistics is that they are multi-variate Gaussian as stated previously. If a classifier is desired which maximizes the "likelihood" of correct classification, then the discriminant functions become [26]

$$g_k(\overline{x}) = -\frac{1}{2}\overline{x}^T[\Phi_k]^{-1}\overline{x} + \overline{x}^T[\Phi_k]^{-1}\overline{\mu}^{(k)} - \frac{1}{2}\overline{\mu}^{(k)^T}[\Phi_k]^{-1}\overline{\mu}^{(k)}$$

$$+ \ln P(S_k) - \frac{1}{2}\ln(\det[\Phi_k]) \qquad k = 1, \cdots, K \qquad (22)$$

where $\Phi_k$ is the $N \times N$ covariance matrix of the feature distributions and $P(S_k)$ is the a priori probability of a sample being from diagnostic class K. In a biomedical application, it is usually desired that there be a very low probability of assigning a normal classification to an abnormal subject. The discriminant function may easily be adjusted to take this into consideration.

While this is a very powerful classifier, it has some strict limitation. If the underlying distributions are sufficiently dissimilar from a Gaussian distribution, then classification accuracy will degrade severely. Also, computation of the discriminant function requires the inversion of a covariance matrix. If the features are highly correlated, then computation of this inverse may be difficult or impossible. Thus, if this type of classifier is used the feature selection process must produce a feature space in which measurements are sufficiently decorrelated.

## FEATURE SELECTION OF SPATIAL AND SPATIAL FREQUENCY MEASURES

There were 60 features extracted in each of 298 inter-rib spaces from 95 films. The divergence measure was applied separately to both a two class and four class situation. The two class problem was a simple normal-abnormal discrimination. The four class divergence measure and computer diagnosis will be based on a normal and 3 abnormal class situation with increasing UICC profusion category. This is shown in Table 2a

| Computer Class | | UICC Class | | | |
|---|---|---|---|---|---|
| Normal | 1 | 0/- | 0/0 | 0/1 | |
| | 2 | 1/0 | 1/1 | 1/2 | |
| Abnormal | 3 | 2/1 | 2/2 | 2/3 | |
| | 4 | 3/2 | 3/3 | 3/4 | (A B and C lesions) |

TABLE 2a

No attempt was made to separate lesion types only profusions. While these classes are somewhat arbitrary, they will yield some useful indications of the ability to automate the severity of disease determination. The energy normalized annular ring features from 141 films were also subjected to a divergence measure criteria with diagnostic classes as in Table. 2a. The features selected in order of preference for 2 class and 4 class frequency and spatial domain textural measures are shown in Table 2b.

| | | | |
|---|---|---|---|
| Selected Spatial Measures | 2 Class | $R_4(3)$, $\overline{M}_3(1)$, $V_4(3)$, $\overline{M}_5(3)$, $V_5(3)$, $\overline{M}_5(1)$, | $\overline{M}_1(3)$, $\overline{M}_1(7)$ |
| | 4 Class | $R_4(3)$, $\overline{M}_5(1)$, $R_5(7)$, $V_4(3)$, $\overline{M}_5(3)$, $V_5(3)$, $\overline{M}_1(3)$, | $\overline{M}_1(7)$ |
| Selected Frequency Domain Measures | 2 Class | Annular Rings 5, 22, 7, 14, | 17, 29 |
| | 4 Class | Annular Rings 5, 8, 22, 14, 17, 29 | |

TABLE 2b. SELECTED MEASURES

Some of the initially selected measures were not subsequentially used for diagnostic classification. The justification for this will be discussed later. However, it is interesting to note that the spatial measures selected were mostly from distance 3 indicating a possible overspecification in resolution in the digital scans. The selected spatial frequency measures were roughly divided into low, medium and high frequency ranges.

It should be noted that the digital and optical data bases contained 95 and 141 films respectively. Thus the digital data base is a subset of the optical one. It was felt necessary to expand the optical data base in order to include enough data to make a realistic statistical diagnostic classification possible. The expanded data base also included 8 0/1 and 1/0 films which were not submitted for digital analysis. In other respects the data bases were quite similar. However when comparisons are made between digital, optical, and physician rates this difference in data bases will be accounted for. The respective data base compositions are shown in Table 3.

| | Digital | Optical | Physician |
|---|---|---|---|
| Class 1 | 32 | 39 | 39 |
| Class 2 | 14 | 33 | 33 |
| Class 3 | 20 | 28 | 28 |
| Class 4 | 29 | 41 | 41 |
| TOTAL | 95 | 141 | 141 |

TABLE 3

To convert the film number per class to number of inter-rib spaces per class a multiple of 3 will yield an accurate approximation.

## COMPUTER CLASSIFICATION RESULTS

The spatial and frequency domain measures shown in Table 2 b were subjected to one further test in order to possibly further reduce the dimensionality of the N dimensional feature vector $\overline{x}$. The Gaussian discriminant function algorithm was trained successively for N = 1, 2, ••• in each of the four cases shown in Table 2b in an effort to determine when and if the training classification percent ceased to significantly increase when new features were sequentially added. The results are shown in Figures 12, 13, 14, and 15. The following conclusions were drawn. The two class frequency domain measures beyond N = 4 were not significant. However, for the corresponding 4 class diagnosis, all N = 6 measures were deemed beneficial. The spatial domain textural measures for the 2 class and 4 class diagnosis indicated that N = 6 and N = 7 were sufficient respectively. These selected measures were used in all the subsequent test results.

The computer or physician diagnostic results are all discussed in terms of confusion matrices [27] the general form of which are shown in Figures 16 and 17. Within this context, a true delineation indicates that true class and assigned class were identical where as a false delineation indicates a disagreement between true and assigned classification. Within this context a true negative implies a normal that was correctly classified as such. On the other hand a, false positive denotes a normal which was mistakenly classified in an abnormal category or categories. A false negative implies an abnormal which was incorrectly assigned to a normal category.

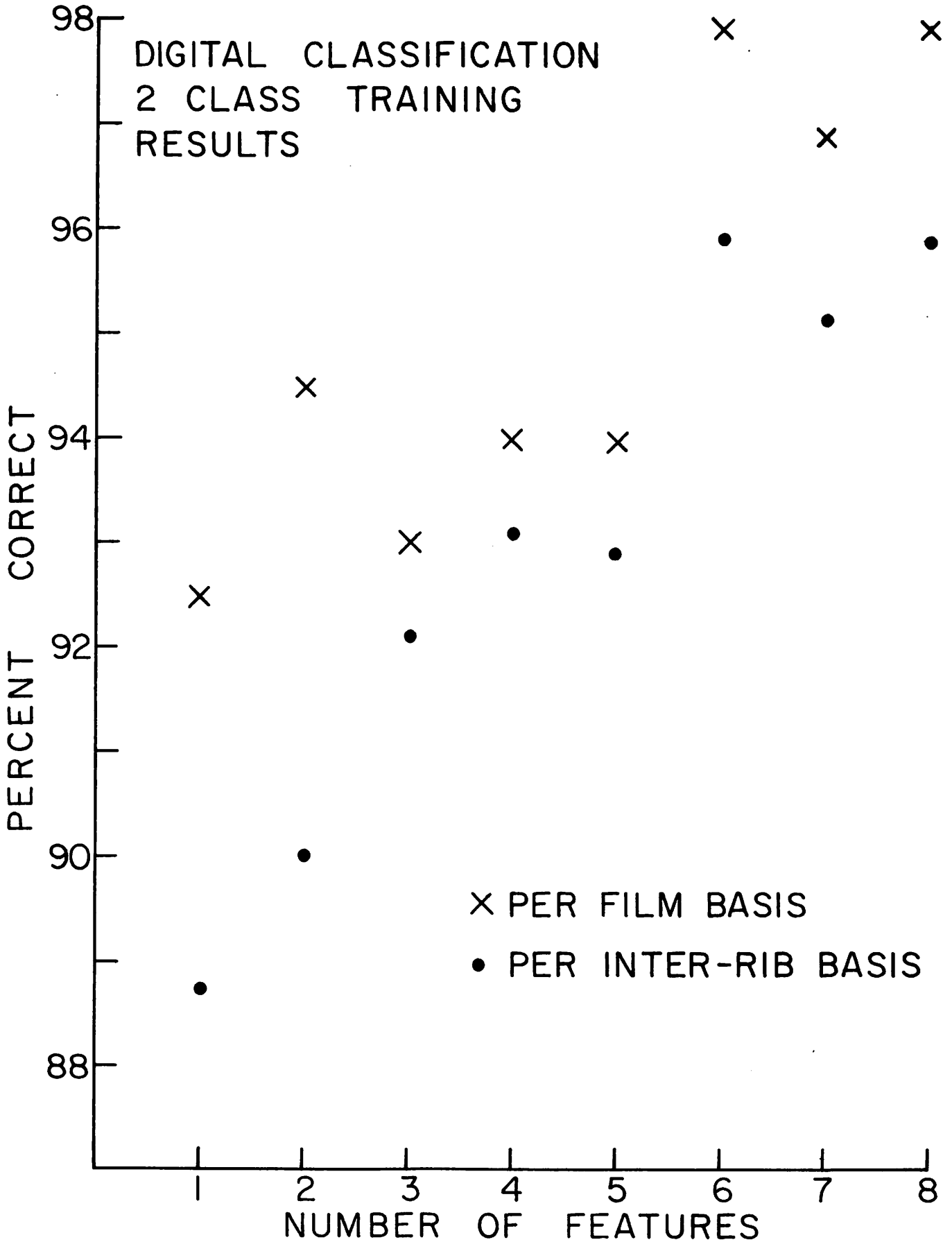

DIGITAL CLASSIFICATION
2 CLASS TRAINING
RESULTS
98
96
94
92
90
88
PERCENT CORRECT
1
2
3
4
5
6
7
8
NUMBER OF FEATURES
× PER FILM BASIS
• PER INTER-RIB BASIS

FIGURE 12

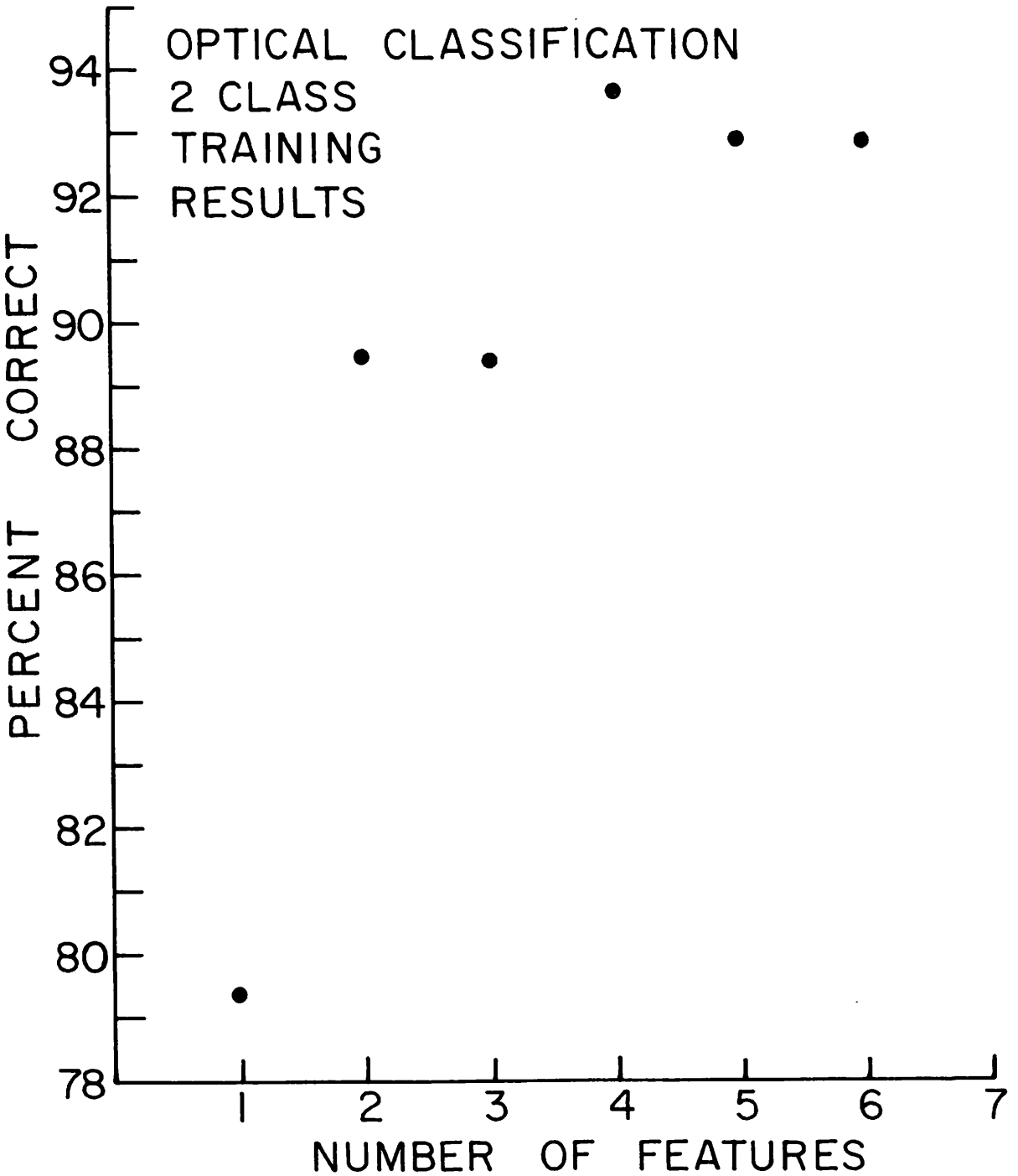

OPTICAL CLASSIFICATION
2 CLASS
TRAINING
RESULTS
PERCENT CORRECT
94
92
90
88
86
84
82
80
78
1
2
3
4
5
6
7
NUMBER OF FEATURES

FIGURE 13

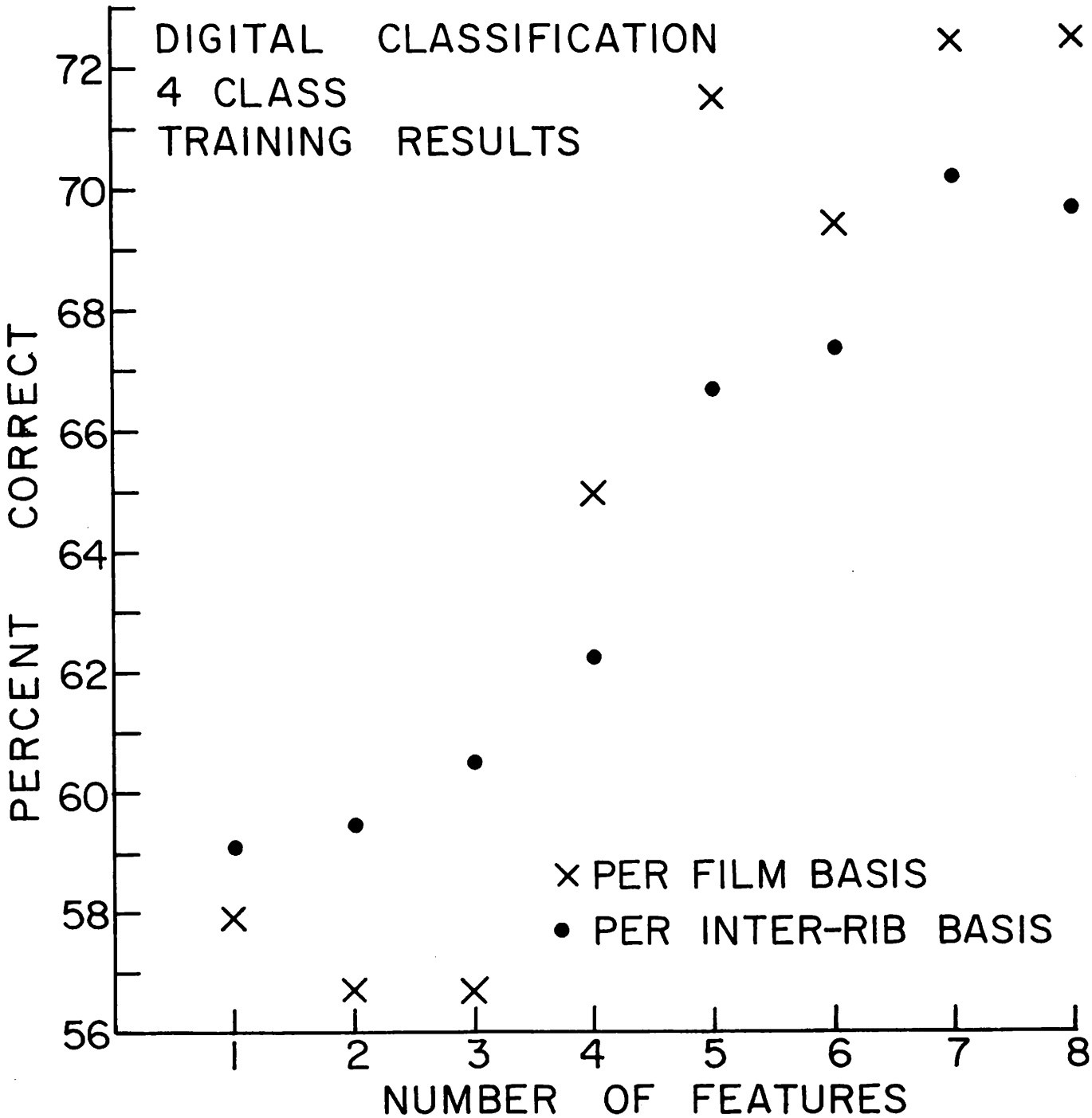

DIGITAL CLASSIFICATION
4 CLASS
TRAINING RESULTS
PERCENT CORRECT
72
70
68
66
64
62
60
58
56
1
2
3
4
5
6
7
8
NUMBER OF FEATURES
× PER FILM BASIS
• PER INTER-RIB BASIS

FIGURE 14

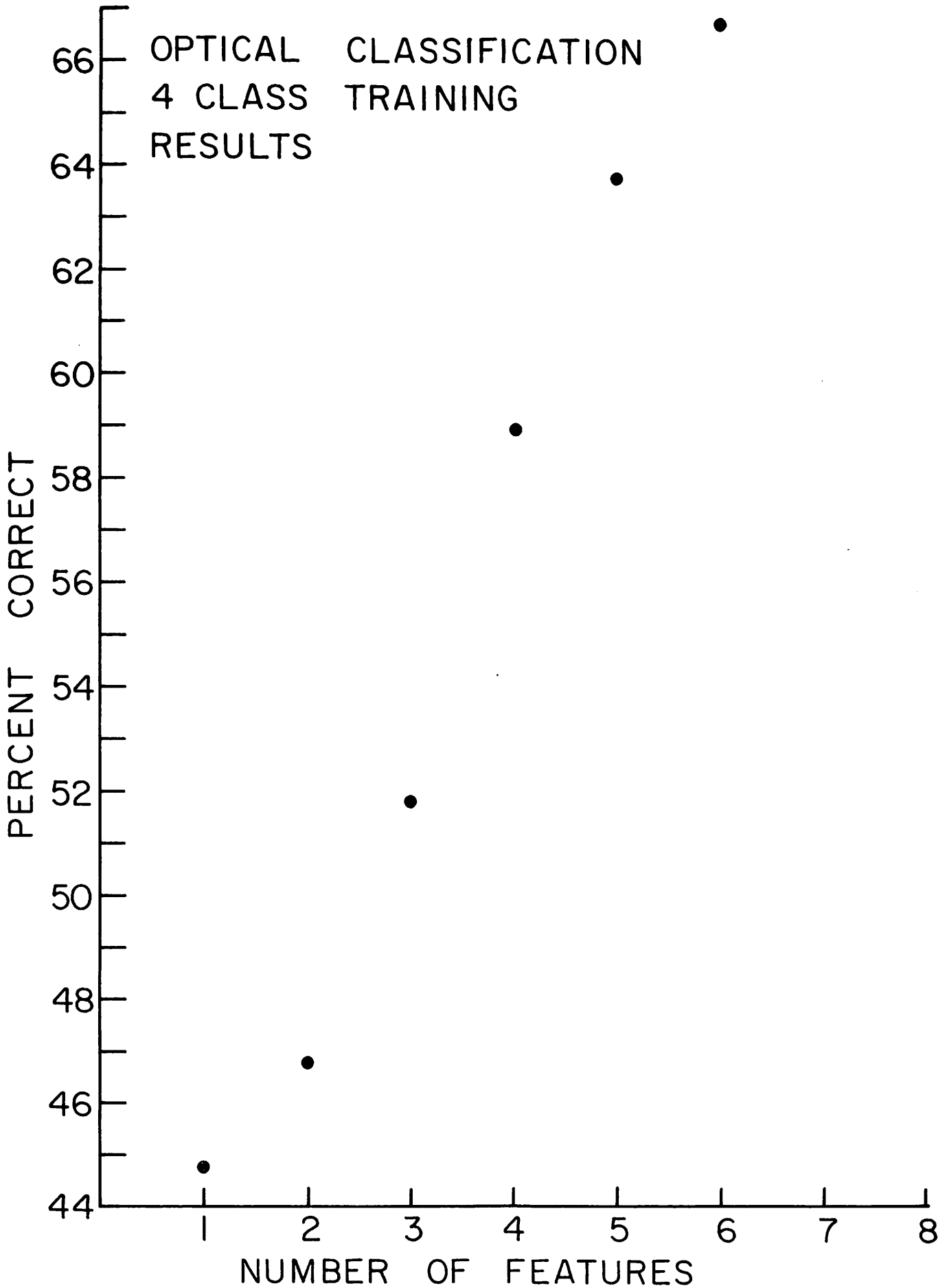
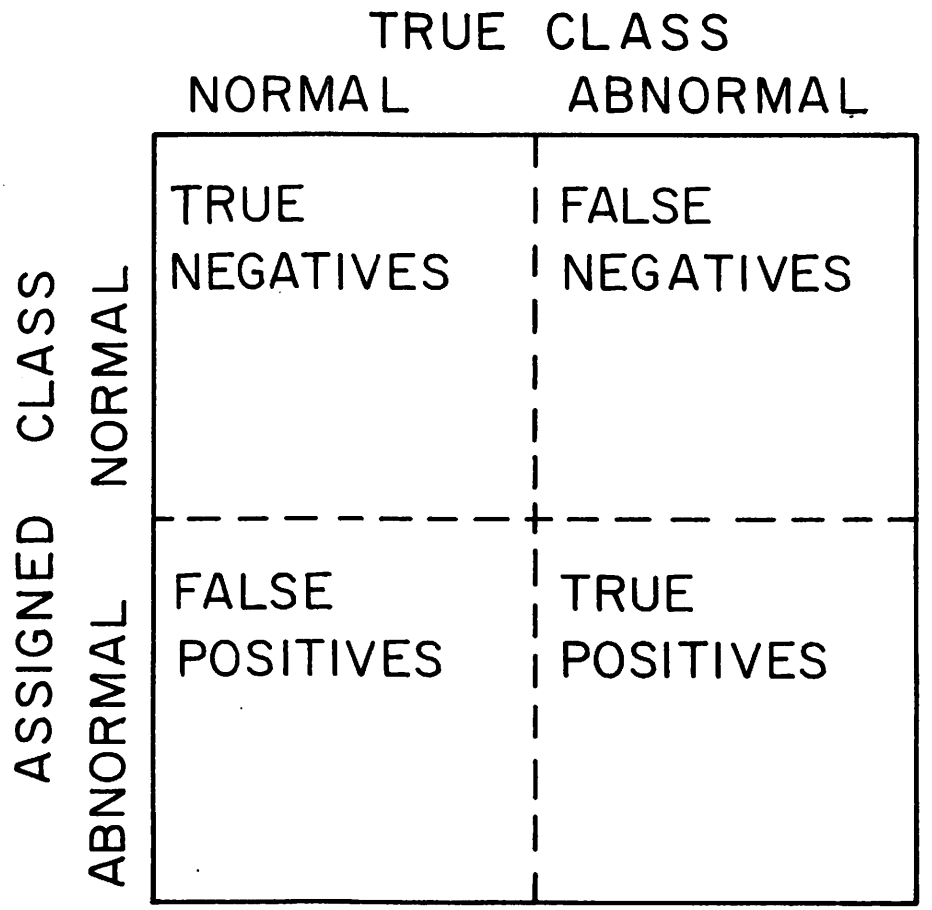
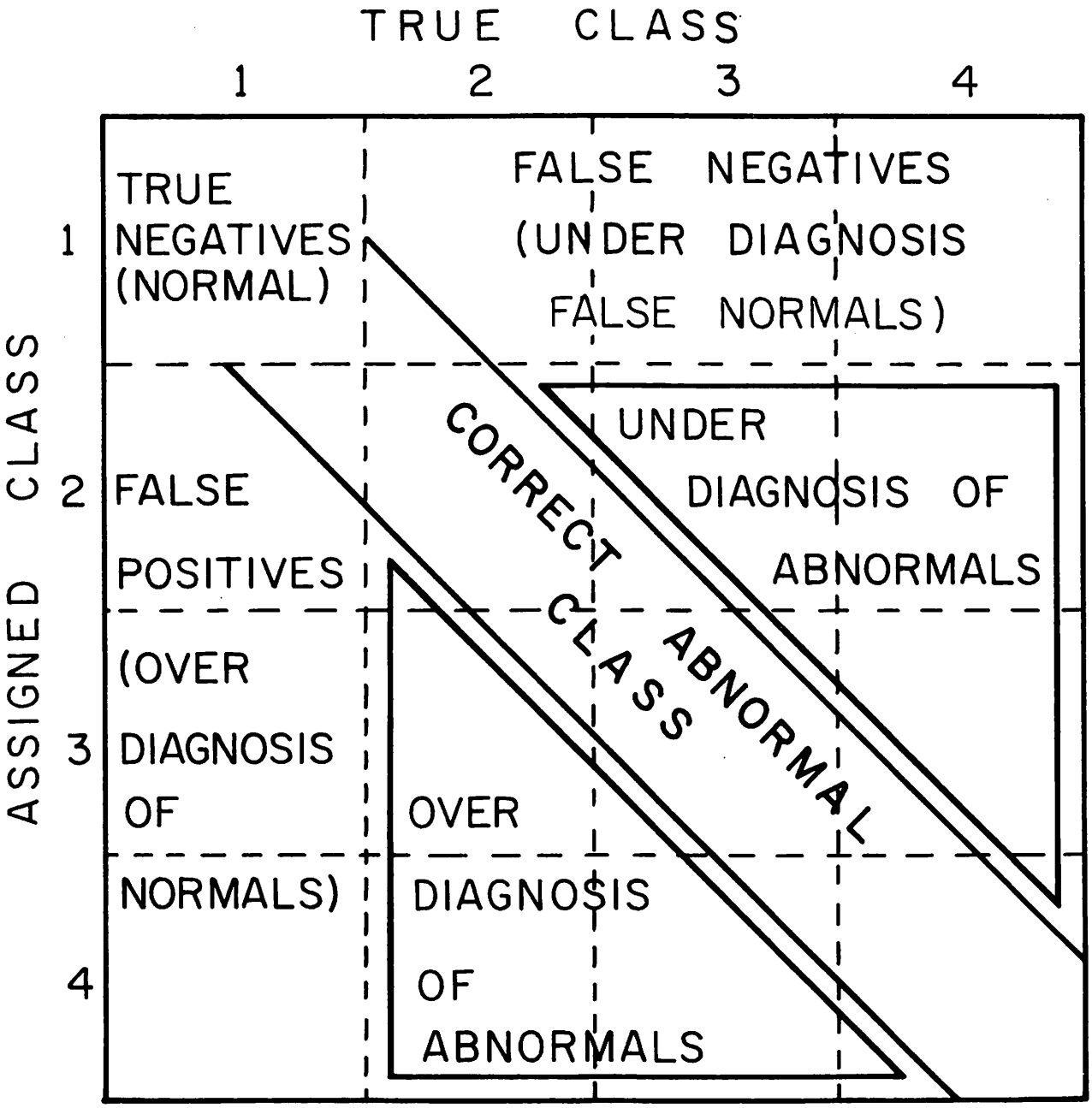

OPTICAL CLASSIFICATION
4 CLASS TRAINING
RESULTS
PERCENT CORRECT
66
64
62
60
58
56
54
52
50
48
46
44
1
2
3
4
5
6
7
8
NUMBER OF FEATURES

FIGURE 15




GENERAL BLOCK DIAGRAM OF A 2 CLASS CONFUSION MATRIX

FIGURE 16

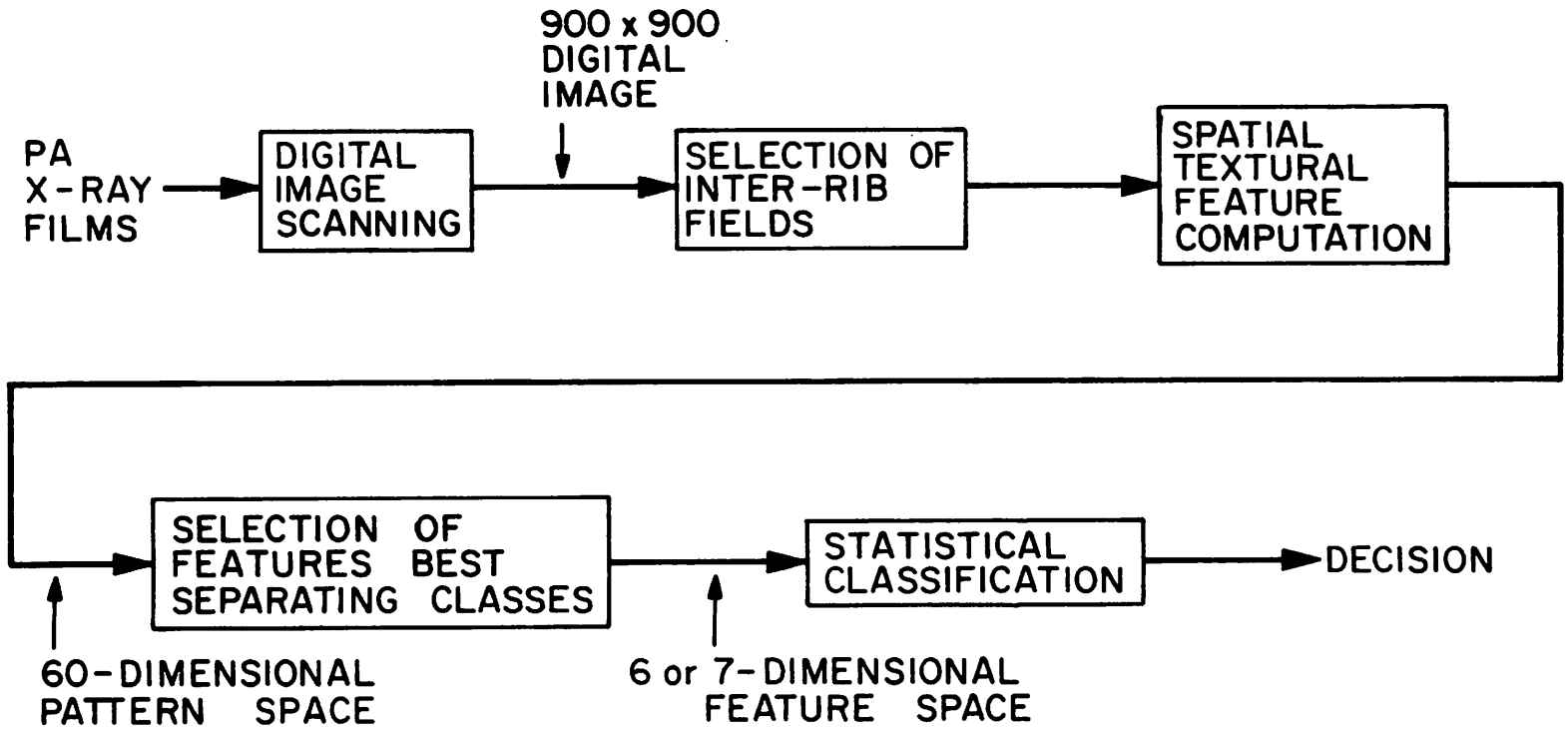



GENERAL BLOCK DIAGRAM OF A 4 CLASS CONFUSION MATRIX

FIGURE 17

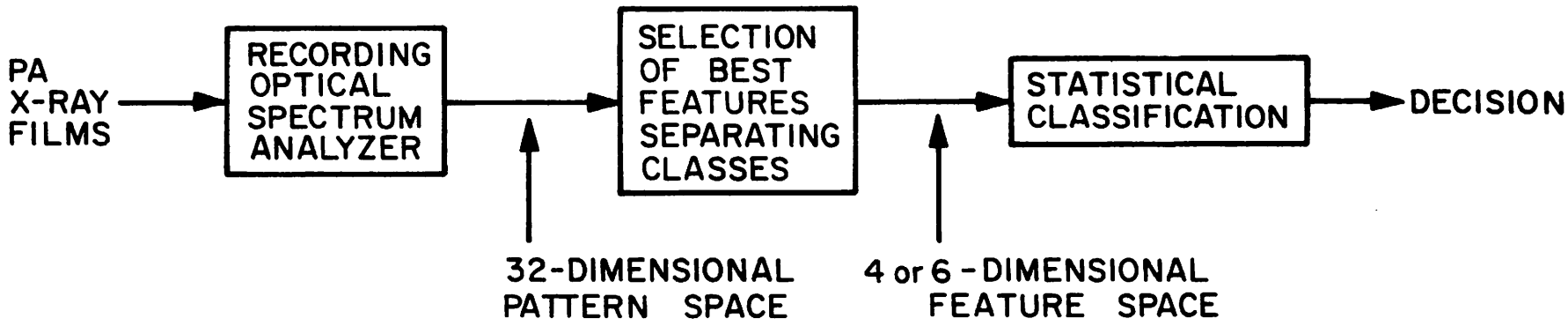
The computer diagnostic testing procedure for both the 2 and 4 class case consisted of a removing one sample from the data base, training on the remaining samples and resubmitting the withdrawn sample for reclassification. This is a fair test since the classifier does not "see" the withdrawn sample until it is asked to diagnostically assign it to a class. A second more severe test was also performed upon the digital and optical data bases for the two class case. 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 are logical extremes. In the first test only one sample is withdrawn and as such the test must be repeated 141 times for the optical and 298 times for the digital data bases, respectively. The second test is only repeated twice for each of the data bases. Insufficient data prevented such a second test for the four class problem.

When discussing digital accuracy rates on a per film rather than a per inter-rib space basis, the following rules of correspondence were applied. If a film contained 2 inter-rib spaces, the assigned diagnostic class for the film was chosen at the highest of the two classes present on the film. If a film contained 3 inter-rib spaces, and two were spaces assigned to the same class, then the film was placed in that class. If all three spaces were classed differently, the highest numbered diagnostic class present was chosen for the film. If a film contained 4 inter-rib spaces, a majority of 3 in any class would assign the film to that class. If this majority did not exist, the film was




UNIVERSITY OF SOUTHERN CALIFORNIA AUTOMATED DIGITAL TEXTURAL DISCRIMINATION PROCEDURE

FIGURE 18




FOURIER OPTICAL SYSTEM FOR TEXTURAL DISCRIMINATION

FIGURE 19

assigned to the highest numbered diagnostic class present in the film. Therefore an effort was made to over diagnose the film when a consensus was not possible. There were 7 films with 2 inter-rib spaces, 75 with 3 spaces and 18 with 4 spaces. Irrespective of specific tests performed a System Diagram for the Optical and Digital approaches are shown in Figures 18 and 19 respectively.

The confusion matrix for the 2 class normal-abnormal digital training results is shown in Table 4, with table values expressed as percentages.

| | | true class 1 | true class 2 |
|---|---|---|---|
| assigned class | 1 | 94.7 | 3.6 |
| | 2 | 5.3 | 96.4 |

TABLE 4. TWO CLASS TRAINING RESULTS

The normal-abnormal diagnostic rate was computer to be 95.9% It should also be noted that the false positive rate 5.3% versus a false negative rate is only 3.6%. This of course is the conservative medical diagnosis for a mass screening situation. On a per film basis only 2 films in 95 were missed. The abnormal film was of profusion category 1/1. The one at a time removal test procedure yielded the confusion shown in Table 5.

| | | true class 1 | true class 2 |
|---|---|---|---|
| assigned class | 1 | 92.6 | 3.6 |
| | 2 | 7.4 | 96.4 |

TABLE 5

The computed normal-abnormal diagnostic rate was 95.2%. On a per-film basis 3 films were missed for a corresponding diagnostic rate of 96.9%. When the more severe second test was performed the following confusion matrix was found in Table 6.

| | true class 1 | true class 2 |
|---|---|---|
| assigned class | 90.0 | 5.2 |
| | 10.0 | 94.8 |

TABLE 6

The normal-abnormal rate was 92.9% On a per-film basis this was 96.8%.

The 1 missed abnormal was of profusion category 1/1.

One quite obvious conclusion is that the normal-abnormal diagnostic rate is quite stable, using digitally derived textural features. The two class transform domain results will now be presented. A corresponding digital rate will be given for the 95 films common

between the two film bases. The two class training results yielded the confusion matrix shown in Table 7. The normal-abnormal diagnostic

| | | true class 1 | true class 2 |
|---|---|---|---|
| assigned class | 1 | 79.5 | 1.0 |
| | 2 | 20.5 | 99.0 |

TABLE 7

rate was 93.6%. The true negative rate was only 79.5%. This contributed to a rather high 20.5% false positive rate. However upon examination of the 8 missed normal it was found that 6 of them were from profusion category 0/1 which is a doubtful normal. Therefore once again the diagnosis is a conservative one. The missed abnormal was a 1/1. One more result is of interest. When only films common between the digital and optical data bases are examined, the normal-abnormal diagnostic rate is 97% representing 3 misses in 95 films. The one at a time removal test procedure yielded the confusion matrix in Table 8. The normal-abnormal

| | | true class 1 | true class 2 |
|---|---|---|---|
| assigned class | 1 | 74.4 | 2.9 |
| | 2 | 25.6 | 97.1 |

TABLE 8

rate was 90.8%. Of the 10 missed normal films the same 6 were again from profusion category 0/1. The three missed abnormals were from profusion category 1/1. When a common data base with digital data is compared, 5 films were missed for a diagnostic normal-abnormal rate of 95%. The more severe test procedure yielded the following matrix table 9. The

| | | true class | |
|---|---|---|---|
| | | 1 | 2 |
| assigned class | 1 | 69.7 | 4.0 |
| | 2 | 30.3 | 96.0 |

TABLE 9

normal-abnormal diagnostic rate was 88.7%. Of the 12 missed normals 6 were again of category 0/1 in profusion. The 4 missed abnormals were all of profusion category 1/1. Using the data base of the digital measurements indicated 5 misses out of 95 for a normal-abnormal rate of 94.8%.

The four class diagnosis designed to grade severity as well as normality versus abnormality yielded lower over-all results. This was to be expected considering the rather arbitrary class assignment. The four class results will also be presented using a confusion matrix the general form of which is shown in Figure 17. The seven selected digital textural measures listed in table 2b yielded the confusion matrix shown in Table 10 for the four class training situation. The normal-abnormal diagnostic rate was 95.5%. The false positive rate was 4% and the false negative rate was 4.6%. There

| assigned class \ true class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 96.0 | 9.3 | 1.4 | 4.4 |
| 2 | 4.0 | 62.8 | 14.3 | 18.9 |
| 3 | 0.0 | 23.3 | 81.0 | 37.8 |
| 4 | 0.0 | 4.6 | 3.1 | 38.9 |

TABLE 10. Digital Four Class Training Results

was an overall 70.1% correct classification in the four class situation when the sum of the major diagonal inter-rib spaces was divided by the total number of inter-rib spaces. On a per film basis this percentage rose to 72.6%. There was significant over diagnosis of class two and under diagnosis of class four. The four class testing results computed by the one at a time removal method yielded the results shown in table 11.

| assigned class \ true class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 93.7 | 13.0 | 1.8 | 4.1 |
| 2 | 6.3 | 47.8 | 23.2 | 22.7 |
| 3 | 0 | 30.4 | 69.6 | 40.2 |
| 4 | 0 | 8.7 | 5.4 | 33.0 |

TABLE 11. Digital Four Class Testing Results

The overall correct classification rate dropped to 61.9%. On a per film basis this was 63.8%. The normal-abnormal diagnostic rate is 94.2%. The false positive rate is 6.3% as opposed to a false negative rate of 5.6%. Once again there was a significant over diagnosis of class 2 and under diagnosis of class 4. The four class digital training and testing results on a per film basis are shown in Tables 12 and 13. The normal-abnormal training rate

| | | true class | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| assigned class | 1 | 93.8 | 7.1 | 0 | 0 |
| | 2 | 6.2 | 71.5 | 10.0 | 3.4 |
| | 3 | 0 | 7.1 | 85.0 | 41.4 |
| | 4 | 0 | 14.3 | 5.0 | 55.2 |

TABLE 12. Digital Training Rates on a Per Film Basis

was 98%. The false positive rate was 6.2% and the false negative rate was 1.6%. The overall 4 class correct diagnostic rate was 72.6% as mentioned earlier.

The corresponding confusion matrix for the one at a time removal test procedure yielded the following results:

| | | true class | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| assigned class | 1 | 97.0 | 7.1 | 0 | 3.2 |
| | 2 | 3.0 | 35.7 | 15.0 | 9.7 |
| | 3 | 0 | 42.9 | 80.0 | 45.2 |
| | 4 | 0 | 14.3 | 5.0 | 41.9 |

TABLE 13. Digital Testing Rates on a per-film basis.

The overall 4 class correct rate was 65.5%. The false positive rate was 3.0% and the false negative rate was 3.2%. The normal-abnormal diagnostic rate was 96.8%. As was seen in corresponding tables for the digital data on a per inter-rib space basis, there was significant under diagnosis of class 2 and under diagnosis of class 4.

The four class selected optical textural features yielded the following training results

| | | true class | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| assigned class | 1 | 76.9 | 6.1 | 4.7 | 2.0 |
| | 2 | 18.0 | 63.6 | 17.8 | 12.4 |
| | 3 | 0.0 | 6.1 | 46.4 | 12.4 |
| | 4 | 5.1 | 24.2 | 32.1 | 73.2 |

TABLE 14a. Optical Four Class Training Confusion Matrix

The overall correct diagnostic rate was 66.7% on 141 films. There was a significant over diagnosis in classes 2 and 3. The normal-abnormal diagnostic rate was 92.3%. Of the nine missed normals, six were from profusion category 0/1 as in the two class case previously discussed. This elevated the false positive rate to 23.1%. The false negative rate was 2.1%. If the confusion matrix is reconstructed for just the 95 films common between the digital and optical data bases, table 1 b is formed.

true class

| assigned class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 90.7 | 7.1 | 0 | 0 |
| 2 | 6.2 | 42.9 | 13.5 | 7.0 |
| 3 | 0 | 7.1 | 45.5 | 18.5 |
| 4 | 3.1 | 42.9 | 41.0 | 74.0 |

TABLE 14b. Optical Four Class Training Confusion Matrix on Digital Subset

It is seen that the true negative rate was increased significantly when the 0/1 profusion category was removed. The only other class which changed significantly was class two. However since it now only consists of 14 films, a wide variability can be expected. The normal-abnormal rate increased to 95.8% reflecting the removal of the 0/1 profusion films. The overall four class diagnostic rate increased to 68.5%. The false positive rate is 9.3% and the false negative rate is 1.6%.

The four class diagnostic testing results based on a one at a time removal are given in Table 15a. The overall four class accuracy was

true category

| assigned class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 74.4 | 15.2 | 3.6 | 2.4 |
| 2 | 20.5 | 42.4 | 21.4 | 26.8 |
| 3 | 0.0 | 9.1 | 32.1 | 26.8 |
| 4 | 5.1 | 33.3 | 42.9 | 43.9 |

TABLE 15a. Optical Four Class Testing Confusion Matrix.

reduced to 49.6%. The false positive rate was 25.6% and the false negative rate was 6.8%. The normal-abnormal rate was 88%. When the equivalent 95 film digital subset was considered, the following confusion matrix was formed.

| | | true class | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| assigned class | 1 | 90.7 | 7.1 | 0 | 0 |
| | 2 | 6.2 | 35.7 | 10.5 | 19.4 |
| | 3 | 0 | 21.4 | 42.1 | 29.0 |
| | 4 | 3.1 | 35.7 | 47.4 | 51.6 |

TABLE 15b. Optical Four Class Testing Confusion Matrix on Digital Subset.

The overall 4 class diagnostic rate was 61.0% with a false positive rate of 9.3% and a false negative rate of 1.6%. The normal-abnormal rate was 96.0%.

## PHYSICIAN DIAGNOSIS

There presently exists a tri-level grading procedure consisting of certified A, B, and C readers for pneumoconiosis film interpretation. The five C level readers are considered the leading diagnosticians in this country in pneumoconiosis film interpretation and were instrumental in developing the current UICC classification scheme. The approximately 100 B readers are more numerous and less experienced than the C readers but are often staff radiologists at one of several pnemoconiosis grading centers such as the Los Angeles County-University of Southern California Medical Center. The A readers who perform the initial diagnostic screening are physicians who are not necessarily radiologists but who have been certified to read pneumoconiosis films as a result one of several two-day training sessions sponsored by the American College of Radiology. In a typical situation the A reader interprets the film and forwards it to the Public Health Service for transmission to a B reader. If the A and B readers do not agree on an interpretation of the film it is sent to one of the C readers for a decision. In rare instances the five C readers acting as a committee constitute a D reader.

Six radiologists were requested to diagnosis the 141 films submitted for automatic analysis. The instruction sheet given them is shown in Appendix II. Two of the six readers originally selected and diagnosed approximately 67 percent of the films in this study. The remaining 33 percent were standard films used in the current physician training and certification program. One of these two initial readers also participated in the selection of these 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 P-A chest films/day for the County of Los Angeles. The other four readers consisted of one C reader, and three B readers ranging from highly experienced to lesser experienced. As a group these six radiologists represent over 130 man years of radiological reading experience.

The physicians used the entire P-A radiograph with rectangular lung zones within which they were to make their diagnosis labelled and numbered. This allowed the readers to grade the films within an anatomical context. The results of their individual and average performance for the 141 films is shows in the following tables

| | | true class 1 | true class 2 |
|---|---|---|---|
| assigned class | 1 | 94.9 | 1.0 |
| | 2 | 5.1 | 9.9 |

the normal-abnormal diagnostic rate 97.9%

TABLE 16 (a): READER 1 (Original B Reader)

| | | true class 1 | true class 2 |
|---|---|---|---|
| assigned class | 1 | 84.6 | 3.9 |
| | 2 | 15.4 | 96.1 |

The normal-abnormal rate 92.9%

TABLE 16 (b): READER II (C Reader)

| | | true class 1 | true class 2 |
|---|---|---|---|
| assigned class | 1 | 41.0 | 1.0 |
| | 2 | 59.0 | 99.0 |

The normal-abnormal rate 83.0%

TABLE 16 (c): READER III (less experienced B reader)

| | | true class 1 | true class 2 |
|---|---|---|---|
| assigned class | 1 | 97.4 | 2.0 |
| | 2 | 2.6 | 98.0 |

The normal-abnormal rate 97.9%

TABLE 16 (d): Reader IV (experienced B reader)

true class

| assigned class | 1 | 2 |
|---|---|---|
| 1 | 79.5 | 1.0 |
| 2 | 20.5 | 99.0 |

The normal-abnormal rate 93.6%

Table 16(e). Reader V

(experienced B reader)

true class

| assigned class | 1 | 2 |
|---|---|---|
| 1 | 97.4 | 6.9 |
| 2 | 2.6 | 93.1 |

The normal abnormal rate 94.3%

Table 16(f). Reader VI

(Original C reader)

When all the 6 x 141 = 846 physician observations are averaged the following confusion matrix is formed

true class

| assigned class | 1 | 2 |
|---|---|---|
| 1 | 82.5 | 2.6 |
| 2 | 17.5 | 97.4 |

The normal-abnormal rate 93.4%

Table 17(a). Average Results of all Readers.

If the original two readers are removed in order to negative their possible bias the following table results:

true class

| assigned class | 1 | 2 |
|---|---|---|
| 1 | 75.6 | 2.0 |
| 2 | 24.4 | 98.0 |

The normal-abnormal rate 91.8%

Table 17(b). Averaged results without original committee.

If Tables 17a and b are repeated for just the 95 films common between the optical and digital film bases. The following matrices are formed.

true class

| assigned class | 1 | 2 |
|---|---|---|
| 1 | 87.0 | 2.9 |
| 2 | 13.0 | 97.1 |

Normal-abnormal rate 93.7%

Table 18(a). Averaged results of all readers on digital subset.

true class

| assigned class | 1 | 2 |
|---|---|---|
| 1 | 80.0 | 2.0 |
| 2 | 19.5 | 98.0 |

Normal-Abnormal rate 92.1%

Table 18(b). Averaged results on digital subset without the original committee.

The four class averaged physician diagnostic confusion matrix for all 141 films is shown in Table 19a.

true class

| assigned class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 82.5 | 7.1 | .6 | .4 |
| 2 | 17.1 | 57.1 | 11.9 | 4.5 |
| 3 | .4 | 24.7 | 53.0 | 22.0 |
| 4 | 0 | 11.1 | 34.5 | 73.2 |

Table 19a. Average four class confusion matrix for all physicians.

The over-all correct classification was 68.0%.

The same confusion matrix for the four physicians not on the original committee is shown in Table 19b.

true class

| assigned class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 75.6 | 6.1 | 0 | 0 |
| 2 | 23.7 | 50.0 | 9.8 | 3.7 |
| 3 | .6 | 30.3 | 48.2 | 17.7 |
| 4 | 0 | 13.6 | 42.0 | 78.7 |

Table 19b. Averaged four class matrix for four readers not on the original committee.

The overall correct classification was 65.2%.

The corresponding tables for the 95 films common between the optical and digital film sets yield the following confusion matrices.

true class

| assigned class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 87.0 | 10.7 | .9 | .5 |
| 2 | 13.0 | 50.0 | 14.0 | 4.3 |
| 3 | 0 | 26.6 | 45.6 | 24.7 |
| 4 | 0 | 13.1 | 39.5 | 70.4 |

Table 20a. Averaged result of all readers on digital film subset

The overall correct classification rate was 68.8%.

If the two readers on the original committee are excluded.

true class

| assigned class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 80.5 | 8.9 | 0 | 0 |
| 2 | 19.5 | 42.9 | 10.5 | 2.4 |
| 3 | 0 | 32.1 | 42.1 | 20.2 |
| 4 | 0 | 16.1 | 47.4 | 77.4 |

Table 20b. Average result with original committee excluded.

The overall correct classification rate was 67%.

One other measure may perhaps be of use. A strict measure of intraobserver variation was not undertaken during this study. However a comparison of the original committee of two readers with the averaged result of their separate reading at a later time may provide some indication of an upper bound on inconsistency. This confusion matrix is shown in Table 21. The false positive and false negative rates were 3.8% and 3.9%

respectively. The normal-abnormal rate was 96% and the overall four class correct classification was 77%.

| | | true class | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| assigned class | 1 | 96.2 | 9.1 | 1.8 | 1.2 |
| | 2 | 3.8 | 71.2 | 16.1 | 6.1 |
| | 3 | 0 | 13.6 | 62.5 | 30.5 |
| | 4 | 0 | 6.1 | 19.6 | 62.2 |

Table 21. Averaged result of original two readers upon rediagnosis.

## DISCUSSION OF RESULTS

The results of this study have been presented in a series of two and four class confusion matrices. The testing confusion matrices for the digital methods shown in tables 5, 6, and 11 indicate that reliable indications of normality versus abnormality and to a lesser degree of severity can be made on an inter-rib space basis independent of anatomical location within the P-A radiograph. A comparison of the overall correct classification rate,false positive and false negative rates of the digital testing results on a per film basis as shown in Table 13 with the averaged physician diagnostic rates of Tables 18 and 20 indicate that the digital methods are competitive. When comparing these results with those for the optical methods shown in Table 15b, once again they are competitive. In neither the optical or digital cases are the physician's overall accuracy rate exceeded. However, both the optical and digital results yield a lower false positive and approximately equal false negative rates with those exhibited by the physicians.

A comparison for all 141 films between the optical methods as shown in Tables 8, 9, and 10 and the physician rates shown in Tables 17 and 19 indicate approximately equal false negative rates with higher false positive rates using the optical methods. The overall correct classification rate in a four class situation is significantly higher for the averaged physician results.

It would seem that within the scope of this study it is possible to conclude that comparable indications with experienced diagnosticians of normality or abnormality of vascular patterns within the lung can be ascertained through automated methods using such indices as normal-abnormal rate and, false positive and negative rates. It also must be said that at this time manual interpretation by experienced readers is still a somewhat superior judge of severity of disease. A partial reason for this may be that the relatively small number of films utilized in this study yielded too few films per class for adequate computer training in the four class situation. However a device to screen the 80 to 90 percent negative coal worker population does seem entirely plausible.

## CONCLUSIONS

Classification results obtained during the feasibility study are most promising. Two entirely distinct systems both gave normal-abnormal classification rates comparable to those of a board of experienced radiologists. The automated systems were also reasonably successful at providing additional information as to the severity of disease in abnormal cases.

For an automated diagnostic screening system to be practical a number of criteria must be met. The system should be able to process radiographs relatively quickly. The normal-abnormal classification rate should be comparable to the accuracy rate of the physicians currently performing the screening manually. The rate of abnormal cases classified as normal should be held particularly low. It is felt that the success of the two systems described in this report indicates the feasibility of meeting the above criteria.

Each of the systems discussed has its own special advantages. The optical system is particularly fast. As a result, the per film expense of diagnostic screening should be relatively low. The purely digital system, while considerably slower, also has advantages. The system is considerably more flexible and able to compute and analyze data in addition to textural measures over the lung field. In particular, such features as anatomical structure may be dealt with in a unified manner. There is also evidence that digital measures could

yield a better severity diagnosis.

This suggests the possibility of a hybrid system. Initial screening could be done optically with the system adjusted to give a missed abnormal rate essentially as low as desired. A second stage would analyze the initially selected abnormal radiographs in greater detail. The UICC classification system includes anatomical structure as a significant feature and these could easily be incorporated at this step. Final determination of normal-abnormal classification would then be made, and an estimate of the severity of the disease in abnormal cases given.

Acknowledgement

The author would like to thank the staff of the Image Processing Laboratory and especially Mr. Mark Sanders and Mr. Carl Wedberg for their excellent cooperation. The author also wishes to acknowledge the advice and contributions of the staff of the LA County USC Medical Center Department of Radiology and in particular Dr. A. Franklin Turner and Dr. George Jacobson. I would also 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. I finally would like to acknowledge the significant research contribution of Mr. William B. Thompson of the Computer Sciences Program to this effort.

## REFERENCES


1. Bohlig, H. et al, "UICC/Cincinnati Classification of the Radiographic Appearances of the Pneumoconioses," Chest, Vol. 58, no. 1, July, 1970.

2. Key, M.M., Kerr, L.E. (editors) Pulmonary Reactions to Coal Dust Academic Press, 1971.

3. Garland, L.H., "Studies on the Accuracy of Diagnostic Procedures," American Journal of Roentgenology Radiation Therapy and Nuclear Medicine, Vol. 82, No. 1, pp. 25-38, July 1959.

4. Yerushalmy, J., "The Statistical Assessment of the Variability in Observer Perception and Description of Roentgenographic Pulmonary Shadows," Radiologic Clinics of North America, Vol. VII, No. 3, December, 1969.

5. Liddell, F.D.K., "Validation of Classification of Pneumoconiosis," Presented at the International Symposium on Coal Worker's Pneumoconiosis New York Academy of Sciences, Sept. 1971.

6. Reger, R.B., and Morgan, W.K.C., "On the Factors Influencing Consistency in Radiologic Diagnosis of Pneumoconiosis," American Review of Respiratory Disease, Vol. 102, 1970.

7. Reger, R.B., Smith, J., Kibelstis, A., and Morgan, W.K.C., "The Effect of Film Quality and Other Factors on Roentgenographic Categorization of Coal Workers Pneumoconiosis," Amer. J. Roentger. Rad. Ther. Nuc. Med., Vol. 115, No. 3, July, 1972.

8. Jacobson, G., Bohlig, H., and Kivi luoto, R., "Essential of Chest Radiography," Radiology, Vol. 95, May 1970.

9. Rosenfeld, A., Troy, E., "Visual Texture Analysis," Proceedings of the UMR-Kelly Communications Conference, Rolla, Mo., Oct. 1970.

10. Pickett, R. H., "The Perception of a Visual Texture," Journal of Experimental Psychology, Vol. 68, pp. 13-20, 1964.

11. Gibson, J.J., "The Perception of Visual Surfaces," American Journal of Psychology, Vol. 63, pp. 367-384, 1950.

12. Julesz, B., "Visual Pattern Discrimination," IRE Transactions IT-8, pp. 84-92, 1962.

13. Hawkin, J.K., "Parallel Electro-Optical Picture Processing," in Pictorial Pattern Recognition, G. C. Cheng, R. S. Ledly, et.al. Editors, Thompson Publishing Co., Washington, D. C., 1968.

14. Kruger, R. P., "Computer Processing of Radiographic Images," Ph.D. Dissertation, University of Missouri, June, 1971.

15. Kruger, R.P., Townes, J.R., et.al., "Radiographic Diagnosis Via Feature Extraction and Classification of Cardiac Size and Shape Descriptors," IEEE Biomedical Transactions, May 1972.

16. Andrews, H.C., Tescher, A.G., and Kruger, R.P., "Image Processing by Digital Computer," IEEE Spectrum, Vol. 9, No. 7, July, 1972.

17. Rosenfeld, A., Troy, E., "Visual Texture Analysis," Technical Report 70-116, June, 1970, University of Maryland.

18. Haralick, R., Anderson, D., "Textural-Tone Study with Applications To Digitized Imagery," Technical Report 182-2, University of Kansas, November, 1971.

19. Rosenfeld, op. cit., 17.

20. Haralick, op. cit., 18.

21. Lendaris, G.G. and Stanley, G.L., "Diffraction Pattern Sampling for Automatic Pattern Recognition," Proceedings of IEEE, vol. 58, February, 1970.

22. Hall, E.L., Kruger, R.P., Dwyer, S.J., et.al., "A Survey of Preprocessing and Feature Extraction Techniques for Radiographic Images," IEEE Transactions on Computers, vol. C-20, no. 9, September, 1971.

23. Sutton, R.N., and Hall, E.L., "Texture Measures for Automatic Classification of Pulmonary Disease," IEEE Transactions on Computers, Vol. C-21, No. 7, July, 1972.

24. Andrews, H.C., Introduction to Mathematical Techniques in Pattern Recognition, John Wiley & Sons, Inc., New York, 1962.

25. ibid, 24

26. ibid, 24.

27. Lusted, L.B., "Perception of the Roentgen Image: Applications of Signal Detectability Theory," Radiologic Clinics of North America, Vol. 7, No. 3, December, 1969.

APPENDIX I

# UICC/CINCINNATI CLASSIFICATION OF RADIOGRAPHIC APPEARANCES OF PNEUMOCONIOSES

| | | Codes | Definitions |
|---|---|---|---|
| SMALL OPACITIES | Rounded | | |
| | Profusion | | The category of profusion is based on assessment of the concentration of opacities in the affected zones. The standard films define the mid categories |
| | | 0/- 0/0 0/1 | Category 0 - small rounded opacities absent or less profuse than in category 1 |
| | | 1/0 1/1 1/2 | Category 1 - small rounded opacities definitely present but relatively few in number |
| | | 2/1 2/2 2/3 | Category 2 - small rounded opacities numerous. The normal lung markings are usually still visible |
| | | 3/2 3/3 3/4 | Category 3 - small rounded opacities very numerous. The normal lung markings are partly or totally obscured. |
| | Type | p q r | The nodules are classified according to the approximate diameter of the predominant opacities. p - rounded opacities up to about 1.5mm diameter. q - rounded opacities exceeding about 1.5mm and up to about 3mm diameter r - rounded opacities exceeding about 3mm and up to about 10mm diameter |
| | Extent | Lung zones | The zones in which the opacities are seen are recorded. Each lung is divided into thirds - upper, middle, lower zones. Thus a maximum of 6 zones can be affected |
| | Irregular | | |
| | Profusion | | The category of profusion is based on assessment of the concentration of opacities in the affected zones. The standard films define the mid categories. |
| | | 0/- 0/0 0/1 | Category 0 - small irregular opacities absent or less profuse than in category 1 |
| | | 1/0 1/1 1/2 | Category 1 - small irregular opacities definitely present but relatively few in number The normal lung markings are usually visible |
| | | 2/1 2/2 2/3 | Category 2 - small irregular opacities numerous. The normal lung markings are usually partly obscured. |
| | | 3/2 3/3 3/4 | Category 3 - small irregular opacities very numerous The normal lung markings are usually totally obscured. |
| | Type | s t u | As the opacities are irregular, the dimensions used for rounded opacities cannot be used, but they can be roughly divided into three types s - fine irregular or linear opacities t - medium irregular opacities. u - coarse (blotchy) irregular opacities |
| | Extent | Lung zones | The zones in which the opacities are seen are recorded. Each lung is divided into thirds - upper, middle, lower zones - as for rounded opacities |

| | | | |
|---|---|---|---|
| **LARGE OPACITIES** | Size | A B C | Category A - an opacity with greatest diameter between 1cm and 5cm, or several such opacities the sum of whose greatest diameters does not exceed 5cm<br>Category B - one or more opacities larger or more numerous than those in category A whose combined area does not exceed one third of the area of the right lung<br>Category C - one or more large opacities whose combined area exceeds one third of the area of the right lung. |
| | Type | wd id | As well as the letter 'A', 'B' or 'C', the abbreviation 'wd' or 'id' should be used to indicate whether the opacities are well defined or ill defined. |
| **OTHER FEATURES** | **Pleural thickening** | | |
| | Costophrenic angle | Right Left | Obliteration of the costophrenic angle is recorded separately from thickening over other sites. A lower limit standard film is provided. |
| | Other sites | 1 2 3 | Grade 0 - not present or less than grade 1.<br>Grade 1 - up to 5mm thick and not exceeding one half of the projection of one lateral chest wall. A lower limit standard film is provided.<br>Grade 2 - more than 5mm thick and up to one half of the projection of one lateral chest wall or up to 5mm thick and exceeding one half of the projection of one lateral chest wall<br>Grade 3 - more than 5mm thick and extending more than one half of the projection of one lateral chest wall. |
| | **Diaphragm** | | |
| | Ill defined | Right Left | The lower limit is one third of the affected hemidiaphragm. A lower limit standard film is provided. |
| | **Cardiac outline** | | |
| | Ill defined (shagginess) | 1 2 3 | Grade 0 - up to one third of the length of the left cardiac border or equivalent.<br>Grade 1 - above one third and up to two thirds of the length of the left cardiac border or equivalent.<br>Grade 2 - above two thirds and up to the whole length of the left cardiac border or equivalent.<br>Grade 3 - more than the whole length of the left cardiac border or equivalent. |
| | **Pleural calcification** | | |
| | Diaphragm<br>Walls<br>Other sites | 1 2 3 | Grade 0 - no pleural calcification seen.<br>Grade 1 - one or more areas of pleural calcification, the sum of whose greatest diameters does not exceed 2cm.<br>Grade 2 - one or more areas of pleural calcification, the sum of whose greatest diameters exceeds 2cm but does not exceed 10cm.<br>Grade 3 - one or more areas of pleural calcification, the sum of whose greatest diameters exceeds 10cm. |

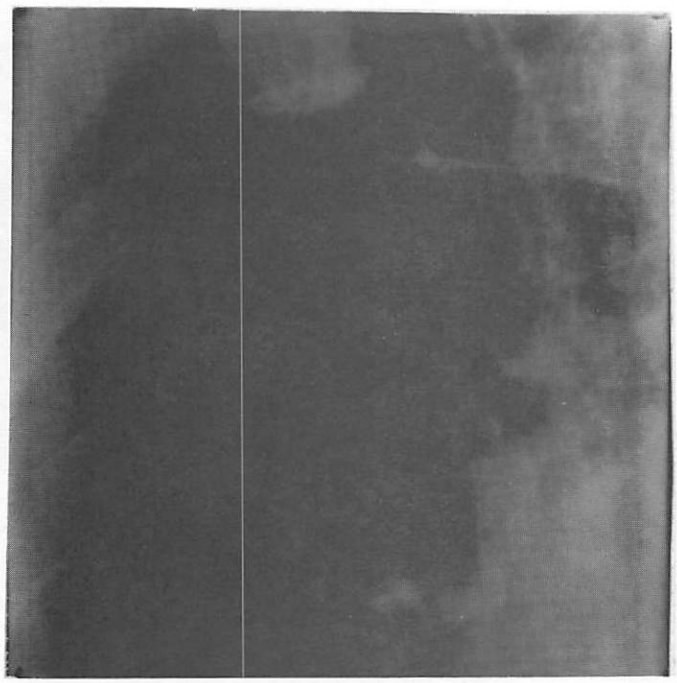
## OTHER SYMBOLS

**OBLIGATORY**

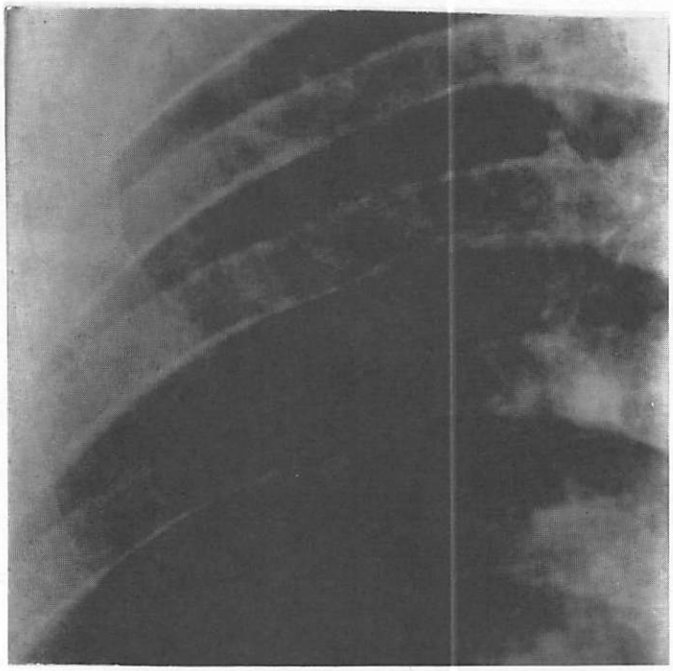
- ca - suspect cancer of lung or pleura.
- co - abnormality of cardiac size or shape.
- cp - suspect cor pulmonale.
- es - eggshell calcification of hilar or mediastinal lymph nodes
- tba - opacities suggestive of active clinically significant tuberculosis
- od - other significant disease. This includes disease not related to dust exposure, e.g. surgical or traumatic damage to chest walls, bronchiectasis, etc.

**OPTIONAL**

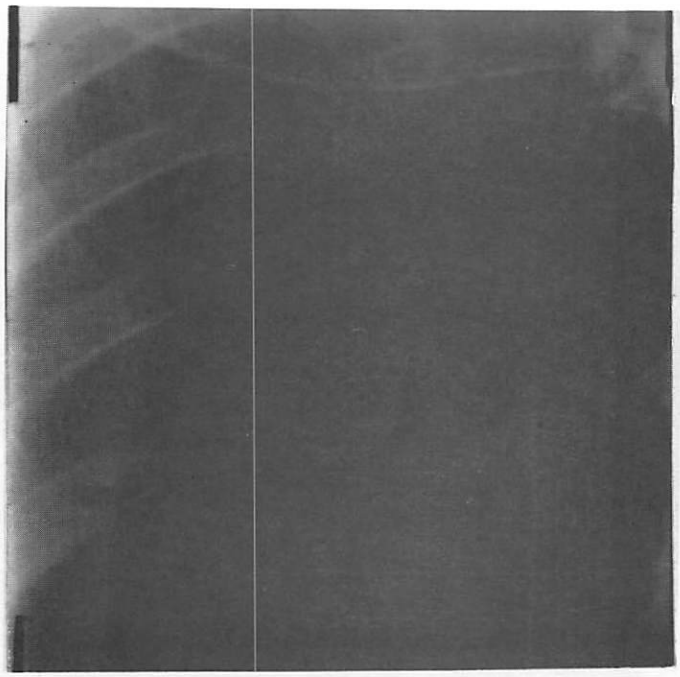
- ax - coalescence of small rounded pneumoconiotic opacities
- bu - bullae.
- cn - calcification in small parenchymal opacities.
- cv - cavity.
- di - marked distortion of the intra-thoracic organs.
- em - marked emphysema.
- hi - marked enlargement of hilar shadows.
- ho - honeycomb lung.
- k - Kerley (septal) lines.
- px - pneumothorax.
- rl - pneumoconiosis modified by rheumatoid process.
- tb - inactive tuberculosis.



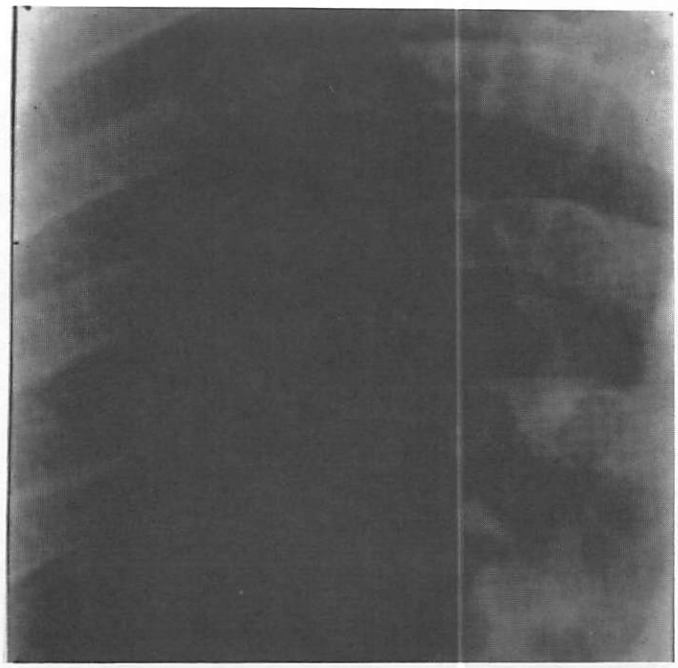
p 3/3



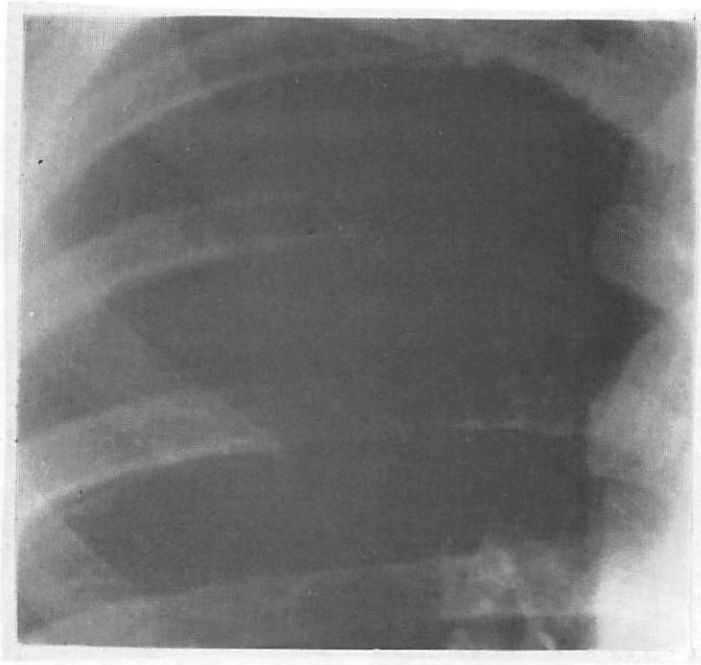
u 2/2



normal

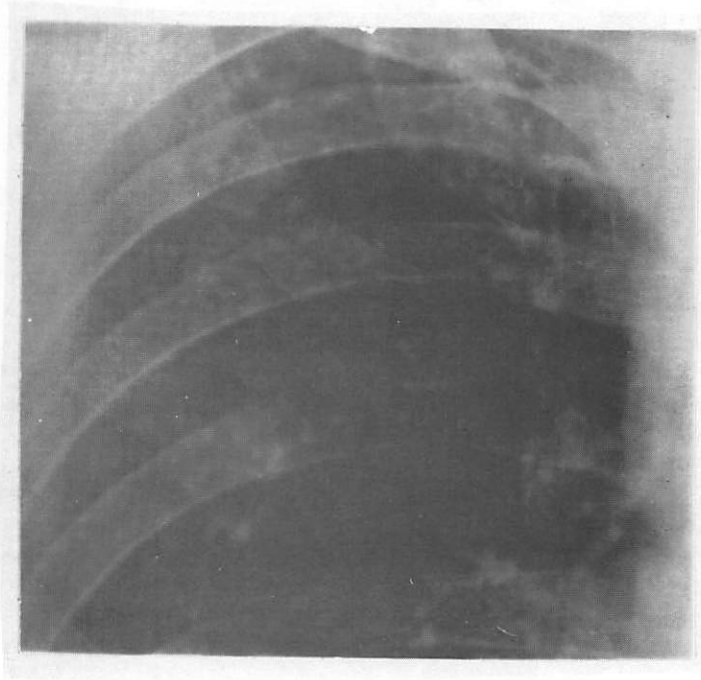


normal

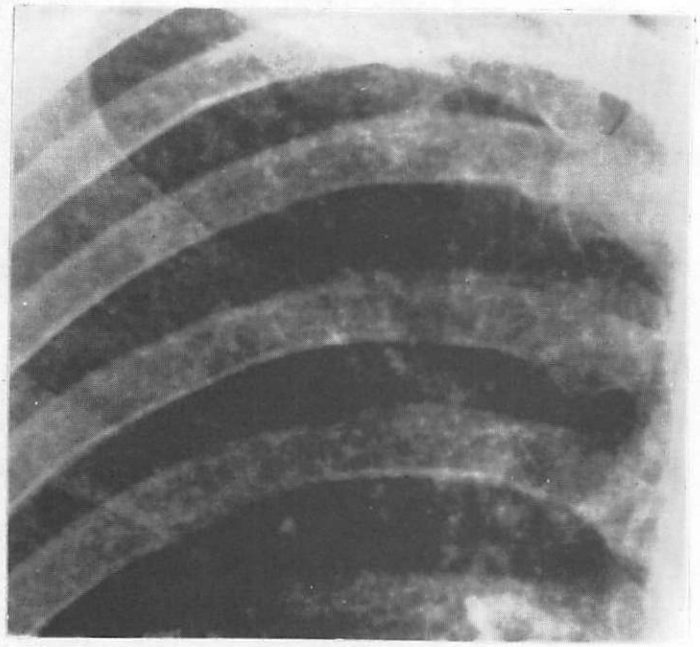


normal

u 2/2 B



t 1/2



p 2/2

## APPENDIX II

## INSTRUCTIONS

Before you is a group of P-A films taken from the Los Angeles County-USC Medical Center files in which 141 lung zones have been selected. Approximately 33 % of these films are standard films used in various physician training sets. The rest were selected and read by a local committee. All films were judged either to be normal or pneumoconiotic using the UICC classification scheme when originally diagnosed, and all were of (+) or (+ -) quality. The films, considered as a whole, depict all 6 lung zones. In addition, there are more right lung zones than left lung zones. Also, more right upper and middle zones were included than lower zones. Every attempt was made to include the lung region which was most indicative of diagnostic classification. All profusion categories are represented as well as all lesion types and sizes.

Please make a film reading using the format established by the UICC grading procedure on each of the numbered films. We wish you to grade only small and large opacities. Therefore, do not indicate "Other Features" such as pleural thickening, shaggy heart, etc. or "Other Symbols" either obligatory or optional. In addition, when lesions of more than one profusion and/or type are present, indicate the most representative variety only on your grading form. When you have completed all readings, please indicate the total time spent on this task. While this is for internal use only, please sign the finished form. Films should be read sequentially as the numbers

appear on the folders. All folders with a black edge have one zone selected. A gray edge marked "2" has two zones, and a brown edge marked "3" has three. Do not read the second zone until its number appears in sequence. When reading the second or subsequent zone, do not refer to the diagnosis on any previous zone on that film. It is advised that the folders marked "2" and "3" be separated at the first pass through the stack of films.