INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

IMAGE ANALYSIS AND MATHEMATICAL METHODS FOR DIAGNOSIS IN PATHOLOGY: APPLICATIONS TO NEOPLASTIC DISEASES OF THE BREAST, OVARY, AND PARATHYROID

by

ANDREW JEFFREY EINSTEIN

A dissertation submitted to the Graduate Faculty in Biomedical Sciences,
in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

1998

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

November 11, 1997
Date

Craig J. Benham
Chair of Examining Committee
Craig J. Benham, Ph.D.

November 13, 1997
Date

Executive Officer
Terry Ann Krulwich, Ph.D.

Supervisory Committee

Joan Gil
Joan Gil, Dr. med., Habil.

Craig J. Benham
Craig J. Benham, Ph.D.

Sylvan Wallenstein
Sylvan Wallenstein, Ph.D.

The City University of New York

Abstract

IMAGE ANALYSIS AND MATHEMATICAL METHODS FOR DIAGNOSIS IN PATHOLOGY: APPLICATIONS TO NEOPLASTIC DISEASES OF THE BREAST, OVARY, AND PARATHYROID

by

Andrew Jeffrey Einstein

Advisor: Professor Joan Gil

Image analysis and mathematical tools are developed for the resolution of difficult differential diagnoses in pathology. Methodological developments center on three main areas: quality assurance, the quantitative characterization of chromatin appearance, and classification. The impact of a new cytologic staining method, the ultrafast Papanicolaou procedure, on nuclear morphometry is evaluated. It is found to improve on conventional Papanicolaou staining in terms of speed, with no important quantifiable differences in nuclear morphology. Variance component models are used to analyze the reproducibility of several interactive methods used for the segmentation of nuclear images, demonstrating that thresholding-based methods similar to those incorporated into commercial instruments are unacceptably variable, while an arc-fitting method is most reproducible. Two new approaches to chromatin texture characterization are introduced, a statistical method quantitating the amount of local gray level variation, and an approach based on fractal geometry. Chromatin appearance in breast epithelial cell nuclei is shown to be fractal. Nuclei are characterized by Minkowski-Bouligand and spectral fractal dimensions, and by lacunarity, a measure of a fractal's "gappiness." These properties are shown to differ

between benign and malignant breast epithelial cell nuclei. Classification methods are compared, and logistic regression and artificial neural network models are developed and optimized.

These methods are applied to the cytologic diagnosis of breast epithelial cell lesions, and histopathologic diagnoses of parathyroid lesions and ovarian dysplasia. Using fractal measures of chromatin appearance, and logistic regression for classification, 39 of 41 breast cytology specimens are correctly diagnosed as benign or malignant. An optimized neural network is able to correctly classify all 41 cases; however, questions of overtraining are raised. Using nuclear diffuseness, the measure of gray level variability, to characterize chromatin appearance, and a neural network for classification, all 16 parathyroid biopsies considered are correctly diagnosed as normal, adenoma, or carcinoma. Statistical and neural network analyses show ovarian dysplasia in epithelial inclusion cysts adjacent to cancer and incidentally diagnosed to be indistinguishable. This suggests that the associated changes in tissue morphology reflect a single underlying pathologic process, supporting the hypothesis that ovarian dysplasia seen in epithelial inclusion cysts represents a preinvasive malignant change.

Preface

This dissertation uses the "compilation format." It is organized primarily around eight manuscripts, each published or submitted for publication, which constitute chapters one through eight. While each manuscript touches on various aspects of image analysis in pathology, the first two chapters highlight issues of quality control, the middle five chapters focus on the numerical characterization of chromatin texture, and the final chapter surveys classification methodologies. Each chapter is introduced by a Preface which places the subject of the manuscript in context, and lists work performed by my coauthors. The dissertation also includes a general introduction and literature review, a summary, and a comprehensive bibliography.

Acknowledgments

This dissertation would not have been possible without the help of many people. I am most fortunate to have Dr. Joan Gil as my advisor. He has imparted innumerable lessons to me as a teacher, editor, sounding board, and friend. My chairmen, Drs. Craig Benham and Alan Schiller, have been generous both in their counsel and their support of my travels to scientific conferences. The Department of Biomathematical Sciences has provided a congenial home for my research. The faculty and my fellow students create a unique, interdisciplinary environment from which I have greatly benefited, while the administrator, Ms. Toni Hill, and the computer and office staffs have made the logistics of research, publication, and travel manageable. I also thank Dr. Terry Ann Krulwich and the graduate school office at Mount Sinai for making my graduate experience such a pleasant one.

We are fortunate to have numerous scientific collaborators who, through their diverse perspectives, have broadened my education. My sincere appreciation to (in alphabetical order) Drs. Joe Barba, Carol Bodian, Liane Deligdisch, Sylvan Wallenstein, Hai-Shan Wu, and Grace Yang, and to Zhiyuan Liu. These studies also would have been impossible were it not for the materials and diagnoses provided by Drs. Ira Bleiweiss, David Burstein, Miguel Sanchez, and Pam Unger. My life was made considerably easier by the technical assistance provided by Joe Silberfarb and Cindy Sheppard.

For financial support, I am grateful to the National Institutes of Health. Additionally, I am grateful for travel funds provided by the International Society for Stereology, by Dr. Gabriele Losa, Istituto Cantonale di Patalogia, and by the Centre for Nonlinear Dynamics in Physiology and Medicine, McGill University.

On a personal note, this milestone would have been unattainable without the constant love and support of my family. My profound appreciation to my parents, Irene and Dr. Fred Einstein, and to my sister Jacqueline. I am also sure that my two omas would have been so proud to see me complete this dissertation. Lastly, my thanks to God for giving me life, sustaining me, and enabling me to reach this day.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Despite the introduction of promising molecular biomarkers, the diagnoses of malignancies and related conditions are still made by a pathologist on a subjective basis. Diagnosis is based on tissue properties which are amenable to quantification by image analysis and computerized techniques. Although trained pathologists generally experience little difficulty in diagnosing the bulk of biopsies, fine-needle aspiration (FNA) cytologies, or desquamative cytologies submitted to the pathology service, there are a number of diagnoses which are extremely difficult to resolve, due either to the morphological similarity between different conditions or to a lack of familiarity with the morphology of rare-occurring lesions. For other diagnoses, performance varies widely between pathologists and institutions. This introduces an undesirable element of uncertainty into patient care. A number of ancillary tools, some of them instrumental, have already gained acceptance as auxiliary means of clarifying certain problems. We aim at introducing computer-based image analysis as a routine methodology in anatomic pathology. In this dissertation, we develop image analysis and mathematical tools for the resolution of problematic differential diagnoses in pathology. The methodological developments discussed focus on three main areas: quality assurance, the quantitative characterization of chromatin appearance, and classification. We apply the methods developed to provide highly accurate, image analysis based approaches to the cytologic diagnosis of breast epithelial cell lesions and the histopathologic diagnosis of parathyroid lesions, as well as to demonstrate the morphologic similarity between incidentally diagnosed ovarian dysplasia and ovarian dysplasia in epithelial inclusion cysts adjacent to cancer.

Significance & Background

Cancer is the second leading cause of death in Americans. One in every three people will develop cancer in his lifetime, and one in five will die from it.[1] In 1997 alone, an estimated 1,382,400 Americans will be diagnosed with cancer while 560,000 will die.[2] The human and economic costs are staggering.

The definitive role in the diagnosis of cancer is borne by the surgical pathologist. Even with a high degree of clinical suspicion, a patient's condition is neither definitively established nor safely assumed without a tissue diagnosis.[3] When biopsies have not been obtained, tragic errors have been committed, such as radical mastectomies being performed for fat necrosis.[4] Adherence to this practice is routinely monitored by accrediting and regulatory agencies, and policies ensuring its observance are found in the bylaws of most hospitals. The responsibility of the surgical pathologist is to provide the clinician with an accurate and comprehensive diagnosis, enabling him to determine the optimal regimen of treatment and, insofar as possible, gauge prognosis.[3] Thus, the appropriate selection of cancer treatment, be it curative or palliative, by surgery, chemotherapy, or radiotherapy, hinges on the pathologist's diagnosis. Considering the significant morbidity and mortality associated with each of these modalities, and the even more devastating consequences of withholding treatment when it is necessary, it is essential for the diagnosis to be correct. In many cases, diagnosis is straightforward for an experienced pathologist. In some cases, it is not. Rare diagnoses and subtle morphologic changes such as those found in premalignant lesions can present a particular problem. Sometimes, objective diagnostic criteria may be lacking. In some venues, pathologists may be less well trained. In these situations, adjuvant techniques could improve diagnosis.

An example of such a challenging diagnosis is the identification of parathyroid neoplasia. A tumor of the parathyroid glands may take the benign form of adenoma, or the malignant form of carcinoma. For parathyroid carcinoma, radical resection of the gland

and surrounding tissues at the initial operation is the definitively preferred treatment. However, carcinomas lacking the classic macroscopic features of a thick fibrous capsule and local invasion may be misinterpreted as nonmalignant. Lacking these gross morphological signs, the diagnosis of carcinoma is dependent solely upon the histopathologic examination.[5]

In many cases the diagnosis of carcinoma is both difficult[6] and controversial. The most recent review of parathyroid carcinoma by Bondeson *et al.*[7] takes issue with an earlier landmark study by Schantz and Castleman,[8] which had ranked the presence of mitoses as the single most useful criterion of malignancy. Bondeson *et al.* cite Snover & Foucar's study which identifies mitoses with nonmalignant cases;[9] they suggest that different inclusion criteria could account for this discrepancy. Certainly, objective criteria would be beneficial.

Moreover, parathyroid carcinoma is exceedingly rare. Blichert-Toft *et al.*[10] report a prevalence of primary hyperparathyroidism of 0.4-1.0%, while Christensson *et al.*[11] note that their value for prevalence, evaluated at 0.36%, is higher than those reported in a handful of studies dating back to the late 1960s and early 1970s. Of patients in Western countries with primary hyperparathyroidism, carcinoma has been variously reported as the cause in 0.34%[12] to approximately 5.3% of cases.[13] Obara and Fujimoto[14] note that only 271 cases of functioning parathyroid carcinoma have been reported in the literature since 1933. It is quite possible that a pathologist called upon to diagnose a case of parathyroid carcinoma will have never seen one previously. This rare and controversial diagnosis stands to benefit from supplementary techniques such as image analysis.

Another diagnostic dilemma illustrating the potential benefit of these methods is the analysis of breast fine-needle aspiration (FNA) biopsies. Eighty percent of the over 560,000 open breast biopsies performed annually are benign.[15] FNA has been reintroduced in the past two decades as a means to obviate the need for many of these biopsies, which have considerable morbidity and cost. FNA is inexpensive, easy to

perform, and virtually painless. It requires no advance preparations, can be quickly interpreted, and results in minimal breast scarring.[3] In an era of cost containment, it is notable that FNA may achieve a savings of $700 over two step excisional biopsy at a cost of a 0.1% decrease in 10-year survival.[16] While the precise role of FNA in the management of mammary lesions has not yet crystallized, proponents of the "triple-test" of palpation, mammography, and FNA dispense with the need for open biopsy in patients with three positive or negative findings.[17] However, although well trained pathologists can determine the benign or malignant status of breast lesions from FNA with high accuracy, this diagnostic performance is not ubiquitous. Rather, diagnostic performance is highly dependent on how experienced the aspirator and cytopathologist are. A review[17] of 83 published series on fine-needle aspiration (FNA) biopsy of the breast reported false-negative rates ranging from 0 to 35%, indicating that breast FNA diagnostic performance varies widely between institutions. Thus, ancillary image analysis techniques could serve an important role for quality assurance in this setting as well.

An Overview of Image Analysis in Pathology

Introduction

The methodology we apply to the diagnosis of cancer goes by a variety of names: image analysis, image cytometry, and morphometry. While the terms reflect different emphases, in this overview we will not carefully distinguish between them. The general paradigm with which we are presented is as follows: A collection of biopsies or cytologies of a tissue is gathered from a number of patients representing each of the various diagnoses under consideration. These are viewed by oil immersion light microscopy. In our image analysis laboratory we have microscopes connected via video cameras to microcomputers. By viewing regions of tissue on a touch-screen monitor attached to the computer, one is able to select the images of nuclei. Several biopsies from each diagnostic outcome are

considered; from each biopsy a number of digitized images of nuclei are collected. The images are normalized to adjust for variation in lighting and staining conditions. Morphometric descriptors, or features, characterizing quantities such as nuclear size, shape, regularity, and chromatin texture, are computed. Using a classification algorithm, a schema is proscribed for determining the diagnosis of unclassified biopsies based upon their descriptors. Thus, the flow of morphometric information begins with the collection of nuclear images, proceeds to the computation and selection of descriptors, and ultimately is applied to a classification. In the paragraphs that follow, we will discuss each step in more detail.

Selection and Preparation of Specimens

Initially, we need to acquire a collection of biopsies (or cytology specimens) representative of each diagnosis in the narrow range under consideration. Ideally, the sample size for each diagnostic group should be the same, although in practice this does not always hold due to difficulty in obtaining specimens from certain groups. In tissue processing, involving embedding, fixation, sectioning, and staining, the emphasis should be on using as reproducible and standard a procedure as possible. While completely uniform conditions are desirable, some variability in processing conditions, such as fixation pH and duration, can be tolerated without significantly affecting quantitative histopathologic variables.[18] Materials are typically acquired from the archival collection of Mount Sinai Hospital. They are prepared by technicians in the Department of Pathology using procedures common in commercial processing laboratories. For histologic specimens, this involves the following procedure. After grossing, the specimens are immersed in 10% phosphate buffered formaldehyde and placed inside an automatic tissue processor for one additional hour of fixation with 10% formaldehyde. This is followed by a second hour of fixation in the automatic tissue processor with 63% ethanol - 10% phosphate buffered formaldehyde, and then by treatment with the usual ascending ethanol

chain, xylene, and paraffin embedding. Sectioning is at a nominal thickness of 5 μm. Hematoxylin and eosin (H&E) staining is done in an automatic instrument, with 30 seconds of eosin staining and 7 minutes of hematoxylin staining. This system is routinely used for the majority of specimens at Mount Sinai, and in our experience shows reproducible results. Cytologic materials are generally prepared in one of three ways: air drying with a Diff-Quik stain, the wet-fixed Papanicolau technique, and the ultrafast Papanicolaou procedure. We use ultrafast Papanicolau staining of our cytologic specimens. This recent advance combines the speed and larger nuclear profile size of Diff-Quik with the crisp nuclear detail and clean background of conventional Pap staining,[19] making it an ideal preparatory method for cytomorphometry.

Image Acquisition and Sampling

Our present image acquisition system is a self-assembled system, based on a Gateway 2000 486DX2/50E microcomputer augmented with Sprynt i860 image processing boards and the Semper 6 Plus graphics program (Synoptics Ltd., Cambridge, UK), connected through a Sony DXC-M2 camera to a Nikon Optiphot microscope equipped with a 100× Nikon Plan objective with a numerical aperture of 1.25. Systematic sampling of all the nuclei on a slide is not feasible with current technology. Sampling is most typically performed by randomly positioning a slide on the microscope's stage and consecutively selecting a designated number of nuclei meeting specified inclusion criteria. Generally, only nonoverlapping, complete nuclei with sufficient contrast will be included. An alternate sampling scheme, requiring the use of a trained pathologist or cytotechnologist, involves selection of the "worst" nuclei visible on screening. As these two sampling rules will chose different nuclei, it is not surprising that they result in different values for average and variance of nuclear size.[20] Probably more important than the specific sampling scheme is the consistency with which it is applied. Studies typically include 40 to 100 nuclei per specimen. Sample size requirements are studied in detail in Chapter 2.

Segmentation

Castleman[21] defines the image segmentation process as "one that partitions a digital image into disjoint (nonoverlapping) regions." For our purposes, these regions of interest are nuclei, or more properly, nuclear profiles. Approaches to segmentation may be manual, semi-automatic, or automatic. The literature is replete with many suggestions for automatic segmentation algorithms, but few of them were developed for biologic materials. For some purposes, especially applications in engineering and materials science, the extant automatic segmentation algorithms are quite effective. Unfortunately there is no such algorithm for the majority of biologic applications such as the morphometry-based diagnoses we are attempting. As Marchevsky and Erler[22] have pointed out,

> "...a basic impediment to the widespread application of interactive and automated image analysis lies in the inherent characteristics of biological images. Many tissues and preparations lack sufficient contrast to easily separate or segment objects and areas from background areas. If the boundaries of features to be measured (*e.g*, nuclei) have pixel gray values similar or identical to those of adjacent areas, automated edge detection becomes difficult and time-consuming and often fails, when compared with manual boundary tracing methods. Another problem with automated image segmentation arises when the boundaries of objects are successfully defined but unwanted image components are segmented along with them. Therefore, automated image-segmentation algorithms must accurately and selectively detect the boundaries of objects to be measured."

Segmentation remains a research problem under active investigation, and a host of automatic segmentation algorithms and their hybrids have been proposed. As a general strategy, an automatic segmentation algorithm may seek to identify either regions of interest or the boundaries between regions.[23]

Automatic Segmentation: Methods and Problems

An important class of automatic segmentation algorithms are the thresholding methods. In global thresholding, a single gray level is specified. Pixels with intensity values below this cutoff are included in the segmented image, and those with values above it are considered background. This method works only for very simple images. More sophisticated adaptive or dynamic thresholding algorithms vary this cutoff depending on the position in the image, considering factors such as neighboring gray levels and the location of a pixel in the composite image . One such method has been used with some success for karyotype analyses in cytogenetics.[21] Since thresholding often yields noisy segmented images, it is often advantageous to apply image-smoothing operations.[23] For example, Barba *et al.*[24] identified edges in cytologic images by preprocessing with a two dimensional median filter before thresholding local entropy measures.

Entropy, in fact, is just one quantity that has been employed to identify edge pixels. Hall[25] compares 10 local edge detectors, including Roberts, Laplacian, Sobel, Frei-Chen, and Davies operators. These methods use two by two or three by three masks; statistics based upon these masks are computed and can subsequently be thresholded. Boundary tracking (contour following) methods are another group of edge detection procedures. They are only effective for very low-noise images, as the tracking algorithms are prone to getting lost and crossing the borders of the image to be segmented.[21]

Numerous methods approach segmentation as a clustering or classification problem, using more complicated inclusion rules than thresholding. Each pixel may be characterized by some feature vector of its properties, including quantities such as gray level, spectral values, and contrast or texture in a local region. In the region growing methods, pixels are added to existing regions, or alternatively used to start new regions if their properties are sufficiently dissimilar from the extant regions.[26] Pixels may be identified with regions by using a number of different classification methods. A pixel may be associated with the region or point to which its feature vector is closest; closeness may

be in terms of Euclidean distance or some other metric, *e.g*, Mahalanobis distance.[25] Another approach employs a neural network to classify pixels.[27] These methods tend to be computationally intensive.

A final class of automatic segmentation algorithms uses the methods of mathematical morphology, an school of image processing using set theoretic concepts of shape developed by the French mathematicians Matheron[28] and Serra.[29] A variety of morphological filters, based upon small images called structuring elements, are employed, often in conjunction with each other. For example, a shrink-and-blow algorithm may be used in some cases to "separate" two attached nuclei. Shrinking is achieved by repeatedly employing skeletonization and erosion, two mathematical morphology operations.[23] Mathematical morphology is incorporated into the segmentation routines of some commercial instruments such as the Becton-Dickinson CAS 100, and has been combined with other methods, such as multiple thresholding, with some encouraging results.[30]

While they have no interactive component and thereby are at least nominally "objective," our experience, and that of other working in the field, indicates that no automatic segmentation algorithm is entirely satisfactory. Typical are the comments of Irinopoulou *et al.*[31]: "[In] earlier studies ... full automation was achieved only at the cost of skipping all pale nuclei ... Since the presence of such nuclei might be of prognostic relevance, it was decided that the present work would take into account all complete, nonoverlapping and well-focused nuclei using semiautomated image analysis." Despite the wealth of attempts, no single method is sufficiently flexible to account for all the variability present in histologic and cytologic images. Problems include noise, voids within nuclei omitted from the segmented image, unwanted debris included in the image, attached images, interference from texture, and overlapping structures.

Interactive Segmentation: Manual and Semi-Automatic Approaches

One group of interactive segmentation methods are the fully manual procedures. Using a tracing device such as a mouse, trackball, or touch screen,[32] one specifies the outline of a nucleus or the pixels contained within the nucleus. Our older image analysis system used manual segmentation; nuclear profiles were traced on a touch-sensitive screen using a stylus. Since tracing may be corrupted by jitter, and noise in the form of jitter has been show to result in overestimation of profile size,[33] it is often desirable to specify a step size such that pixels separated by less than this distance are not added to the contour.[26] Manual segmentation is usually tedious and more time-consuming than other methods. Moreover, it may introduce an element of subjectivity into segmentation. This potential subjectivity is a weakness it shares with all interactive methods.

A second class of interactive segmentation methods involve the selection of points on the periphery of a nucleus and the computation and display of a contour from these points. In our experience, the best of these semi-automatic procedures for the type of images with which we work is the arc-fitting method developed by Z. Liu in our laboratory. Three points are selected on the margin of a nuclear profile, and the arc connecting these points is computed and displayed. The curve is extended by adding more points to define more arcs, until the whole nuclear profile is outlined. This procedure enables the rapid and accurate identification of nuclear profiles. It is incorporated into our 486-based image analysis system. Other similar approaches include connecting the points to form a polygon (which is not realistic for images of nuclei) and an ellipse-fitting method.[34]

An additional type of interactive segmentation method is based upon automatic algorithms but provides for some user input, either in the form of incorporating editing capabilities or interactive selection of grey-level thresholds. This approach is popular in commercial instrumentation. Editing functions may include provisions for separating, merging, resizing, selecting, and deleting images.[22] Selecting and deleting images,

however, may open a Pandora's box of problems relating to sampling. For example, if large nuclei are more likely to be attached, and attached nuclei are deleted, the average nuclear area descriptor may be biased.

Quality Control in Segmentation

For image analysis procedures to be adopted as adjuvant tests in cancer diagnosis, sources of variability in the diagnostic process must be addressed. One important potential source of variability stems from the influence of segmentation on descriptor values. This involves both the particular segmentation algorithm chosen and the morphometric descriptors computed. Neither a segmentation algorithm nor a descriptor that yields significantly different descriptor values in repeated segmentations of the same nuclei can be relied upon. In Chapter 2 we provide a framework for analyzing the variability of segmentation methods and apply this to several interactive methods. In Chapter 4, we investigate the effect of segmentation on computation of the nuclear diffuseness.

Image Normalization

While data is acquired, in the form of image intensity matrices, using the systems in the morphometry laboratory, these images are often transferred to other systems for further segmentation, normalization, descriptor computation, and classification. Much of this work can be done using a package written for the Silicon Graphics Indigo^2 workstation. The graphics capabilities of this machine permit viewing of nuclei to ensure image quality and allows for additional segmentation, on a pixel by pixel basis, if required. The package also normalizes images and performs the computationally intensive task of calculating descriptors.

Each image acquired is adjusted to normalize for small variations in lighting conditions and staining intensity, using a contrast compensation method developed by Barba & Gil.[35] This is an important step for morphometry using texture analysis, as the

same image acquired under different conditions may yield vastly different textural descriptor values. The particular method used in our lab was designed so that identical images obtained under different optical conditions are transformed into images with similar contrast by compensating for the nonlinear response of the camera.

In a study testing the effect of image normalization on texture descriptors often used in pathology (Haralick descriptors[36]), contrast compensation was shown to be far more effective than traditional normalization approaches based upon histogram equalization. If normalization is 100% effective, the same nucleus acquired under different conditions should have transformed images with identical statistics. While the values of these statistics under contrast compensation were observed to be in close agreement, they were not identical. The discrepancy is attributed to the camera's nonlinearity and the discretization process involved in the analog to digital converter's assignment of 8 bit gray levels. Specific details about the contrast compensation method and about the optical characteristics of our camera, as described by J. Barba, appear in Chapter 4.

Descriptor Computation

One of the most crucial steps in morphometry is the selection of appropriate descriptors. Since images as represented by intensity matrices contain too much information to present to a classification algorithm, salient features need to be compactly and quantitatively described. In practice, we compute measures of nuclear size, shape, optical density, color, texture, and/or tissue architecture.

Size Descriptors

Simple as it may be, an indispensable feature in the characterization of nuclei is nuclear profile area. With the exception of the prostate, it is observed in most organs that carcinomas have larger average nuclear profile areas than benign tissue. Biological, this may reflect higher DNA content. Area may be computed by counting pixels or by a more

sophisticated method based upon a discrete analog of Green's theorem. Besides from average nuclear profile size, other statistics such as mean, standard deviation, or higher moments of area may be appropriate descriptors. A recent paper[37] classified prostatic neoplasms based upon the size distribution plots of nuclear profile areas. Perimeter is another descriptor of nuclear size, albeit of questionable utility since it is especially dependent upon segmentation. Moreover, work on fractals[38] has called into question the entire meaning of perimeter, as it may be a scale-dependent artifact. Maximal chord length across a cell is still another measure of size.

Shape Descriptors

Shape descriptors are another class of features used in morphometry, and also are very dependent on segmentation.[39] A standard measure of shape is the circularity factor, given by

$$\text{Circularity Factor} = \text{Perimeter}^2 / (4\pi * \text{Area}) \ .$$

Circularity factor takes a value of one for a circle and increases for more irregular shapes. Its utility is also doubtful, for not only is circularity factor dependent on the segmentation, but also shapes appearing completely different may have the same circularity values.[22] Similar measures of circularity have been suggested which are just simple transformations of the circularity factor. For example, form factor, the inverse of circularity factor, has been applied to the classification of erythrocytes in sickle cell anemia.[40] As nuclear profiles are not necessarily circular, but may be more elliptic, ellipticity factors have been suggested[41] based upon best fit ellipses, which can be determined a number of ways.[42] Bending energy is a descriptor characterizing the irregularity of contours and thereby purportedly reflects physical processes in nucleus development.[43] It was used in some early studies to describe nuclear shape in polymorphonuclear leukocytes.[44] Concavity factors have been suggested to quantitate irregularity or "crinkeliness."[41] Fourier

approaches to nuclear shape description have been unsuccessful due to an observed lack of rotational invariance and size independence.

The fractal dimension of a nuclear outline has also been proposed as a descriptor of shape.[45] While fractal dimension does show some promise in characterizing outlines of neurons[46] and electron micrographs,[47,48] the small number of pixels in profiles of the epithelial cell nuclei that we measure would seem to preclude any accurate measurement of this feature. Note that fractal measures of nuclear texture do not share this weakness, as all pixels and their gray values are used for this calculation, as opposed to merely the location of perhaps 200 segmentation-dependent edge pixels. It comes as no surprise then, that texture features have proven to be more useful than those describing shape. While shape analysis has found some uses, *e.g*, as a prognostic discriminant in colorectal carcinomas,[49] in many others it fails to provide diagnostically relevant information.[50]

Densitometric Descriptors

Densitometric features such as the optical density (OD) and integrated optical density (IOD) are determined based on Beer's law, which relates absorbance in OD units to the ratio of transmitted to incident light intensity.[51] In the literature, optical density is frequently confused with pixel gray values. When used with appropriate stains, densitometric measurements can be used to determine DNA ploidy. In fact, ploidy by image cytometry is the most widely accepted use of image analysis in pathology today, and most major hospitals have a commercial instrument.

Spectral Descriptors

When using a color camera, gray values are determined from a weighted average of red, green, and blue signals. Depending on the camera, this may either be computed from the three signals, or delivered directly as the NTSC video signal,[23] given by

$$\text{Brightness} = 0.30\ \text{Red} + 0.59\ \text{Green} + 0.11\ \text{Blue} \quad .$$

Instead of just using gray levels, it is possible to characterize images based upon spectral information. While spectral descriptors may have particular appeal when used in conjunction with special stains, such as Feulgen for DNA, they are not often used in morphometry.

Architectural Descriptors

Architecture is the second order structure of tissues, describing the spatial relationship of nuclei to each other. As such, it is more useful for histologic than cytologic tissue characterization. Architectural descriptors incorporating information about the spatial relationship between nuclei include factors such as the number of cells per millimeter and the average minimum distance between a nucleus and the basement membrane. As with size descriptors, the individual measurements can be combined by taking averages, standard deviations, etc. An alternative approach represents tissue architecture using graph theoretic methods.[52,53] Our group has used architectural descriptors in conjunction with manual segmentation procedures in the diagnosis of ovarian dysplasia,[54] but their routine application awaits more advanced scene segmentation methodology.

Textural Descriptors for Chromatin Characterization

Consideration of the morphological appearance of chromatin inside the nuclei of interest is one of the main criteria used by pathologists in making a diagnosis from an anatomic pathology specimen. This is particularly true for malignancies. In addition to the size and number of nucleoli, chromatin features of interest include fine or coarse appearance, clumping, margination, and the formation of voids. Chromatin takes two forms in interphase nuclei: heterochromatin, a condensed, presumably inactive form, and euchromatin, which is more diffuse. The relationship between chromatin structure and neoplasia is complex. While electron microscopic studies have shown that some cancerous, hyperchromatic nuclei have larger and/or more numerous masses of

heterochromatin, one also finds pale, euchromatic nuclei in tumors.[55] In terms of morphometry, the hope is that average properties of nuclei differ between diagnostic categories, or alternatively, that their variabilities do.

An image's texture is the spatial distribution of its intensity values. The texture of nuclei reflects factors such as chromatin arrangement and the presence of nucleoli. In subtle ways, texture can reflect morphologically, the molecular changes which lead to neoplasia. Quantitation of nuclear texture has been applied to a number of biological settings as a tool in making difficult diagnostic and prognostic classifications.

Several approaches have been proposed for numerically describing texture in nuclear images. Initial attempts derived from general methods of texture description developed in the field of pattern recognition. While these methods were developed for applications in geology and satellite reconnaissance, they have been applied to pathology as well. They generally yield a large number of features (*e.g*, 44!) which may not have obvious biological interpretation. While methods exist which can be used to select the most valuable textural descriptors, a high ratio of descriptors computed to sample size raises serious issues of overtraining. Nevertheless, these approaches to texture remain popular, to a large degree owing to their implementation in commercial machines such as the SAMBA and Becton Dickinson CAS lines.

The most popular of the general, as opposed to chromatin-specific, approaches to texture description is based on Markovian analysis and identified with the work of Robert Haralick.[36] Following preprocessing to reduce the number of gray levels g in the image, typically to 8 or 10, the so-called co-occurrence matrix is computed. This $g * g$ matrix contains as its (i, j)th entry the conditional probability of gray level i occurring after gray level j. Second order statistics of the co-occurrence matrix, such as angular second moment, contrast, and entropy, are computed and serve as the texture descriptors. More than twenty such statistics have been defined and are routinely computed by commercial instruments. The co-occurrence matrix can be computed with adjacent pixels in the 0°, 45°,

90°, and 135° directions and using a variety of step sizes. Statistics for the four different directions are combined by computing the average and range. The step size is ordinarily taken as small as possible; Pressman[56] has shown that this leads to the highest classification rate. Diegenbach and Baak,[57] in an early application of Markovian descriptors to the diagnosis of uterine cancer, demonstrated that many of the features have high pairwise Spearman rank correlation coefficients; this indicates that not all features are required to characterize the textural information found in the co-occurrence matrix. Although a recent paper[58] using caricatures of nuclei with various qualitative textural features attempts to correlate Markovian descriptors with factors such as nucleoli, granularity, and margination, this connection remains uncertain. Markovian descriptors have now been employed in a variety of studies, including the diagnosis of giant cell tumor of bone,[59] small cell lung carcinoma,[60] hepatoma[61], melanoma,[62,63] cervical intraepithelial neoplasia,[64] and colonic malignancy.[65,66]

Another general approach to texture, also borrowed from terrain classification, uses statistics of gray level run length matrices.[67] These matrices, like the co-occurence matrices, are computed for each of the four directions. Element *(i, j)* represents the number of times that the image contains a run of length *j* in the specified direction, of pixels with gray level *i* (or alternatively gray level range *i*). Statistics emphasizing factors such as short runs, long runs, and gray level nonuniformity are computed. Run-length statistics have been used to describe subvisual changes in cervical squamous metaplastic cells[68] as well as in conjunction with Markovian descriptors in many of the applications cited above.

A third general method is based on the neighboring gray level dependence matrix (NGLDM).[69] Rather than considering in the matrix only pixels adjacent in a certain direction, here its *(i, j)*th entry specifies the number of pixels in the image of gray level *i* that have *j* neighbors within a specified radius and having gray level within a specified range centered around *i*. While statistics of the NGLDM compared favorably to two other methods in classifying smoothness of skin lesions,[70] they have not been described in the

histopathology or cytology literature. Similarly, features based on image decomposition into Gabor elementary functions[71] have compared favorably to other texture description methods,[72] yet have not been applied in pathology. Still another approach to texture description is to consider moments (mean, variance, skewness, kurtosis) of the gray level histogram.[62,73]

One or more of the Markovian or run-length descriptors may roughly correspond with the salient textural characteristics of nuclei. Another class of texture descriptors specifically attempts to characterize these important features. No single nuclear change is pathognomonic for cancer. But since morphologic changes in neoplasia typically may include nuclear pleomorphism (indicative of altered DNA content and/or chromosome number), hyperchromasia, prominent nucleoli, and aggregation of chromatin at the nuclear membrane,[74] features which mathematically correspond to these subjective observations should be particularly appropriate choices of morphometric descriptors. The thought is that such an approach will more closely replicate what a pathologist observes when he makes a diagnosis.

Towards this end, several investigators have proposed measures of chromatin clumping, or conversely, of chromatin diffuseness. We have proposed the nuclear diffuseness,[75] a measure of average local variability of the gray level matrix, as a descriptor of chromatin diffuseness. This method was applied to the diagnosis of parathyroid lesions; more details can be found in Chapter 4 below. Spina *et al.*[76] use an elaborate algorithm to identify the frequency of high-contrast "microareas" in an image; however, it is difficult to evaluate how these microareas correspond to chromatin structure. Giménez-Mas *et al.*[77] have recently introduced a method using the top hat transformation of mathematical morphology to quantify chromatin granularity.

A group of related measures of clumping are based upon the notion of textons,[78] or textural primitives. Rather than working with an image of 255 gray levels, some criterion is used to reduce this number, typically to three.[79] The algorithm used may be equivalent

to the preprocessing before computing co-occurrence matrices in the Markovian approach, or alternatively could involve setting cutoffs a fixed percent above and below the mean intensity. Contiguous regions of the same reduced gray level are called textons and numerous statistics of textons have been devised to characterize nuclear chromatin. These include the number and area of black (perhaps representing nucleoli), gray, and white textons as well as "second order" statistics, such as those of Young *et al.*,[80] which attempt to characterize chromatin clumping and heterochromaticity. For example, heterochromaticity is quantified as the ratio of the number of black and white textons to the total number of textons.

Stolz *et al.*[63] employed image analysis techniques to the problem of discriminating between malignant melanoma and benign nevocytic nevi using electron microscopy. Their set of features included measures of the amount of euchromain and heterochromain, the presence of dark chromatin aggregations, and Markovian descriptors. Chromatin structure features were based on a difference image between a nucleus and its own median-filtered image, using a 21×21 pixel filter. This approach to texture quantitation appears to be an effective one for electron microscopy, but it demands a greater resolution than is available in light microscopy.

In addition to descriptors reflecting chromatin condensation, two authors have suggested features to quantify the margination of densely stained material at the nuclear membrane, a condition which may be associated with malignancy[74] or necrosis.[55] Jagoe et. al.[61] divided a nuclear image into eight concentric bands based upon minimal distance to the boundary. To measure margination, they computed both the ratio of the mean gray level of pixels in the outermost band to the mean gray level of all pixels, and a gray level moment about the original boundary. In a study using a nested random factors analysis of variance to identify features discriminating between non-neoplastic hepatocytes and hepatoma cells, neither of these margination descriptors demonstrated a higher between group variance component than within group variance component. Similarly, Young *et*

al.[80] computed the average optical density per pixel in a series of rings formed by performing mathematical morphology erosions of the original profile. They defined nuclear margination as the second radial moment divided by the radius squared. This measure of margination also did not provide considerable discriminatory information in multivariate analyses to identify features diagnostic of adenocarcinoma[66] and bladder tumors.[81] A number of factors could account for the poor performance of these margination descriptors, including their high dependence on segmentation, high quantization error,[80] image resolution too low to accurately characterize margination, and the univariate nature of the feature selection techniques employed. While these initial studies of margination are discouraging, further investigation into descriptors of this biologically meaningful phenomenon is certainly warranted.

An entirely different approach which we have introduced to characterize chromatin structure is to consider the intensity surface plot as a fractal surface and compute its fractal dimensions and lacunarity. It has been suggested in other contexts in mathematical biology that complexity and adaptability may be characteristic of health, while pathology may be associated with a certain loss of structural richness. For example, aging has been associated with decreased complexity of cardiovascular dynamics.[82] The new field of self-organized criticality proposes that many fractals in nature originate from self-organizing dynamical processes.[83] Thus, the complexity of biological structure — mirrored to some degree by textural complexity — should reflect cellular developmental processes. Since these processes differ between normal and malignant cells, we hypothesized that fractal dimension may be a meaningful descriptor of chromatin changes in neoplasia. In Chapters 5 through 7, we introduce methods for the fractal characterization of nuclear texture, demonstrate that chromatin texture in breast epithelial cell nuclei is fractal, and show how differences in fractal properties between benign and malignant epithelial cells can be exploited in cancer diagnosis.

Classification Methods

Image analysis in pathology has little value without classification. Its clinical applicability is ultimately dependent on prescribing a specific method of classifying new cases. Diagnosis is classification.

The problem of choice and optimization of classification methods is a crucial one, yet a review of the literature shows that this problem has rarely been addressed and never been the subject of systematic study. Only a few papers provide comparisons between methods; at this point, choosing the contents of an "optimal" feature set and selecting a classification algorithm and its associated parameters on the basis of a training set of cases are as much an art as they are a science. While most morphometric classificatory studies to date have employed stepwise feature selection and discriminant analysis to classify cases, there is no reason to assume that they are the most effective of the numerous methods which exist. In Chapter 8 of this dissertation we treat the problem of classification, discussing, in order, the role of computers in diagnosis, feature selection, types of classifiers, the details of some important classification algorithms, classifier overtraining/overfitting, and comparisons between classification methods. We refer the reader interested in an overview of classification to Chapter 8.

Complementary Novel Methods for Cancer Diagnosis

The goal of all these morphometric and classificatory techniques is the accurate, valid, and reproducible diagnosis of cancerous and premalignant lesions. Modern biology has spawned other techniques which approach this problem from different angles, and our goal in this section is to discuss these novel methods. Perhaps the most common question we are asked when explaining our work relates to molecular biomarkers. Hardly a week goes by without the appearance of a new biomarker paper. Won't these tumor markers replace morphology in general and morphometry in particular?

In short, while we believe that molecular markers are gaining an increasingly important role in pathology, it is still more one of prognosis and screening than of diagnosis. The distinction is important enough to warrant formal definitions. A predictive factor is a variable which predicts an outcome with a sufficient accuracy. Risk factors, prognostic factors, and diagnostic factors are three types of predictive factors which differ in the predictive accuracy demanded and in the outcome of interest. For a risk factor, the main outcome of interest is incidence and the predictive accuracy is less than 100%. For a diagnostic factor, the outcome of interest is also incidence but the accuracy is nearly 100%. For a prognostic factor, the outcome is death and the accuracy is less than 100%.[84] Diagnosis and prognosis have overlapping but distinct roles in clinical decision making. For examples, diagnosis from frozen sections determines the course of intraoperative surgical treatment, while prognosis is used to identify patients for chemoprevention and monitor patients already diagnosed with a condition. While many pilot studies have identified putative molecular prognostic factors, the same cannot be said for molecular diagnostic factors. Presently, the only biomarkers with diagnostic utility are ß-hCG for choriocarcinoma and, controversially, PSA for prostatic carcinoma.

Numerous molecular markers have been suggested for cancer. While generally they fail to be specific and sensitive enough to diagnose one type of cancer, they may play a significant prognostic role. This is closely connected to the definitions of diagnosis and prognosis above. Since neoplasia is a complex, multifactorial process, molecular changes leading to cancer in one patient may differ from those in another patient. Molecular markers must[85] reflect a biologic mechanism relating the marker presence to tumorigenesis. The many paths to neoplasia and the many steps in each path results in a plethora of potential markers. Markers reflect factors such as loss of tumor-suppressor ability (p53[86]), metastatic ability (nm23[87]), cell adhesion (α-catenin and e-cadherin[88]), growth factor receptors (EGFR[89]) and their structural homologues (c-erbB2 (HER-2/*neu*),[90] erbB3 (HER-3),[91] and erbB4 (HER-4)[92]), antigens expressed by proliferating cells (Ki-67[93] and

MIB 1[94]), mutator phenotype genes involved in DNA repair,[95] angiogenesis (microvessel count[96]), and apoptosis (telomerase[97]). These markers may benefit from image analysis techniques,[98] both for quantifying immunohistochemically assayed biomarkers[99,100] and for counting microvessels.[101]

The down side to all these markers is that it is less likely that any single one will be common to all patients with a given tumor. Thus, sensitivities of a particular biomarker common for a particular cancer will still fall short of the high sensitivity required for a diagnostic factor. For example, the c-erbB2 proto-oncogene is amplified in about 25% of human primary breast cancers.[102] Mutations of the putative tumor suppressor gene BRCA1, while strongly associated with early onset breast and ovarian cancers, are only found in a small fraction of patients with these neoplasms.[103] Moreover, since the same pathways are involved in the etiology of numerous malignancies, and some markers may also be associated with hyperplasia, specificity of a single marker will generally be too low for diagnosis. This, however, does not prevent a biomarker from serving as a prognostic factor. Indeed, immunohistochemical assays for estrogen and progesterone receptors, p53, c-erbB2, Ki-67, and Cathepsin D are now being marketed for breast cancer prognosis.

Another modern method with use in cancer prognosis is flow cytometry. Like molecular biomarkers, flow cytometric ploidy studies are mainly used to provide prognostic information on which to base therapeutic decisions. Their diagnostic role is presently limited to only a few applications, namely the differentiation of partial vs. complete hydatidiform moles, osteosarcoma vs. osteoblastoma, mesothelioma vs. reactive pleuritis, and rhabdomyosarcoma vs. Ewing's sarcoma.[104] Since DNA content can be assessed by both flow and image cytometry, one might view flow cytometry as a competitor to image analysis. However, although in many cases the two methods will perform identically in the evaluation of DNA content,[105] each has its technical limitations. Image cytometry is more sensitive in the detection of tetraploidy and aneuploidy since there is less dilution of neoplastic cells by normal cells, while flow cytometry is more sensitive in

identifying near-diploid aneuploidy since the larger number of cells it can sample results in smaller coefficients of variation.[104] To avoid these technical limitations, Mayall has suggested using both methods.[106] Others have suggested guidelines for when to use one method and when to use the other. Thus, flow and image cytometry really complement each other more than they compete with each other.

What are the prospects for molecular markers in diagnosis? Two approaches that may potentially increase the sensitivity and specificity of biomarkers are the use of longitudinal tumor markers[107] and the aggregation of factors.[84] Rather than comparing a patient's biomarker values with those of the population, they can be compared to his own values at an earlier time. While this approach has been introduced for prognostic purposes,[108] it may have diagnostic prospects. Additionally, since tumor heterogeneity results in low sensitivities of markers, it may be beneficial to integrate a number of markers representing the various etiologic paths to neoplasia. The methods to do this are in fact precisely the classification algorithms discussed in Chapter 8. For this approach to be fully effective, we would require an understanding of all etiologic paths, the "molecular taxonomy" of cancer. This underscores the importance of basic research in molecular genetics.

One step beyond the integration of biomarkers with classification methodology is the incorporation of additional forms of diagnostic information. This may include results of laboratory tests, physical exam findings, flow-cytometric ploidy study results, morphologic findings, and morphometric "biomarkers." At times, these may provide different ways of measuring the same phenomenon, *e.g*, mitosis counting (morphology), Ki-67 (molecular marker), and S-phase fraction (flow cytometry) all demonstrate cellular proliferation. By now it should be apparent that distinctions between classical, morphometric, flow cytometric, and molecular biological approaches to diagnosis are artificial. The best diagnostic approach will be to include all factors that are informative, although economic considerations will necessitate "feature" selection. Such efforts[109,110]

are just beginning and integrative, automated diagnostic approaches remain far from clinical practice, but the prospects are exciting.

References

1. Li FP, Schneider JA, Kantor AF. Cancer epidemiology. In: Holland JF, Frei E, Bast RC, Kufe DW, Morton DL, Weichselbaum RR, eds. Cancer Medicine. Philadelphia: Lea & Febiger, 1993.

2. Parker SL, Tong T, Bolden S, Wingo PA. Cancer statistics, 1997. *CA Cancer J Clin* 1997; **47:** 5-27.

3. Fisher B, Osborne CK, Margolese R, Bloomer W. Neoplasms of the breast. In: Holland JF, Frei E, Bast RC, Kufe DW, Morton DL, Weichselbaum RR, eds. Cancer Medicine. Philadelphia: Lea & Febiger, 1993.

4. Morton DL. Principles of surgical oncology. In: Holland JF, Frei E, Bast RC, Kufe DW, Morton DL, Weichselbaum RR, eds. Cancer Medicine. Philadelphia: Lea & Febiger, 1993.

5. Fujimoto Y, Obara T. How to recognize and treat parathyroid carcinoma. *Surg Clin North Am* 1987; **67:** 343-357.

6. McKeown PP, McGarity WC, Sewell CW. Carcinoma of the parathyroid gland: is it overdiagnosed? A report of three cases. *Am J Surg* 1984; **147:** 292-298.

7. Bondeson L, Sandelin K, Grimelius L. Histopathological variables and DNA cytometry in parathyroid carcinoma. *Am J Surg Pathol* 1993; **17:** 820-829.

8. Schantz A, Castleman B. Parathyroid carcinoma: a study of 70 cases. *Cancer* 1973; **31:** 600-605.

9. Snover DC, Foucar K. Mitotic activity in benign parathyroid disease. *Am J Clin Pathol* 1981; **75:** 345-347.

10. Blichert-Toft M, Mollerup CL, Feldt-Rasmussen UF, Daugaard H, Engel UH. Primary hyperparathyroidism. An underdiagnosed disease in Denmark? *Ugeskr Læg* 1993; **155:** 765-769.

11. Christensson T, Hellström K, Wengle B, Alveryd A, Wikland B. Prevalence of hypercalcaemia in a health screening in Stockholm. *Acta Med Scand* 1976; **200:** 131-137.

12. McCance DR, Kenny BD, Sloan JM, Russell CFJ, Hadden DR. Parathyroid carcinoma: a review. *J R Soc Med* 1987; **80:** 505-509.

13. Jarman WT, Myers RT, Marshall RB. Carcinoma of the parathyroid. *Arch Surg* 1978; **113:** 123-125.

14. Obara T, Fujimoto Y. Diagnosis and treatment of patients with parathyroid carcinoma: an update and review. *World J Surg* 1991; **15:** 738-744.

15. Winchester DP, Senen S, Immerman S, Blum M. A systematic approach to the evaluation and management of breast masses. *Cancer* 1983; **51:** 2535-2540.

16. Layfield LJ, Chrischilles EA, Cohen MB, Bottles K. The palpable breast nodule: a cost-effectiveness analysis of alternate diagnostic approaches. *Cancer* 1993; **72:** 1642-1651.

17. Layfield LJ, Glasgow BJ, Cramer H. Fine-needle aspiration in the management of breast masses. *Pathol Annu* 1989; **24:** 23-62.

18. Ladekarl M. The influence of tissue processing on quantitative histopathology in breast cancer. *J Microsc* 1994; **174:** 93-100.

19. Yang GC, Alvarez II. Ultrafast Papanicolaou stain. An alternative preparation for fine needle aspiration cytology. *Acta Cytol* 1995; **39:** 55-60.

20. Kronqvist P, Collan Y, Kuopio T, Kujari H. Nuclear morphometry in breast cancer: the influence of sampling rules and freezing of samples. *Mod Pathol* 1995; **8:** 187-192.

21. Castleman KR. Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, 1979.

22. Marchevsky AM, Erler BS. Morphometry in pathology. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994: 125-179.

23. Bartels PH, Thompson D. Scene segmentation. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994.

24. Barba J, Jeanty H, Fenster P, Gil J. The use of local entropy measures in edge detection for cytological image analysis. *J Microsc* 1989; **156:** 125-134.

25. Hall EL. Computer Image Processing and Recognition. New York: Academic Press, 1979.

26. Gil J, Barba J. Principles of stereology: computerized applications in anatomic pathology. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994: 79-124.

27. Reichl U, Treskatis S, Turner RFB. A backpropagation neural network for segmentation of optical microscopy images of cells. *Submitted* .

28. Matheron G. Random Sets and Integral Geometry. New York: John Wiley & Sons, 1975.

29. Serra J. Image Analysis and Mathematical Morphology. London: Academic Press, 1982.

30. Barba J, Yuan L, Gil J. Cell contour extraction on multi-threshold images. *Pathol Res Pract* 1992; **188:** 449-453.

31. Irinopoulou T, Rigaut JP, Benson MC. Toward objective prognostic grading of prostatic carcinoma using image analysis. *Anal Quant Cytol Histol* 1993; **15:** 341-344.

32. Gil J, Silage DA. Touch sensitive screen as interactive peripheral in stereology and image analysis. *Acta Stereol* 1987; **6:** 282-288.

33. Cornelisse JTWA, Berg JTP. Profile boundary length can be overestimated by as much as 41% when using a digitizer tablet. *J Microsc* 1984; **136:** 341-344.

34. Wu H-S, Barba J. An efficient semi-automatic algorithm for cell contour extraction. *J Microsc* 1995; **179:** 270-276.

35. Barba J, Gil J: Image analysis in statistical texture analysis *Eighth International Congress for Stereology* 1991.

36. Haralick RM, Shanumgam K, Dinstein I. Textural features for image classification. *IEEE Trans Systems Man Cybernet* 1973; **3:** 610-621.

37. Unger PD, Hoon V, Stone N, *et al.* Computerized interactive morphometry in the differential diagnosis of irradiated prostates. *Anal Quant Cytol Histol* 1995; **17:** 100-108.

38. Mandelbrot BB. The Fractal Geometry of Nature. New York: W. H. Freeman and Company, 1983.

39. Marchevsky AM, Gil J, Jeanty H. Computerized interactive morphometry in pathology: current instrumentation and methods. *Hum Pathol* 1987; **18:** 320-331.

40. Wheeles LL, Robinson RD, Lapets OP, *et al.* Classification of red blood cells as normal, sickle, or other abnormal, using a single image analysis feature. *Cytometry* 1994; **17:** 159-166.

41. Gschwind R, Umbrich CB, Torhorst J, Oberholzer M. Evaluation of shape descriptors for the morphometric analysis of cell nuclei. *Path Res Pract* 1986; **181:** 213-222.

42. Wali R, Colef M, Barba J. Best fit ellipse for cell shape analysis. *SPIE Visual Communications and Image Processing* 1991; **1606:** 665-674.

43. Bowie JE, Young IT. An analysis technique for biological shape - II. *Acta Cytol* 1977; **21:** 455-464.

44. Teich JM, Young IT, Sher SE, Lee JS. Transformation of nuclear morphology during cellular maturation. *J Histochem Cytochem* 1979; **27:** 193-198.

45. Sanders H, Crocker J. A simple technique for the measurement of fractal dimensions in histopathological specimens. *J Pathol* 1993; **169:** 383-385.

46. Smith TG, Jr., Marks WB, Lange GD, Sheriff WH, Jr., Neale EA. A fractal analysis of cell images. *J Neurosci Methods* 1989; **27:** 173-180.

47. Keough KM, Hyam P, Pink DA, Quinn B. Cell surfaces and fractal dimensions. *J Microsc* 1991; **163:** 95-99.

48. Baumann G, Barth A, Nonnenmacher TF. Measuring fractal dimensions of cell contours: practical approaches and their limitations. In: Nonnenmacher TF, Losa GA, Weibel ER, eds. Fractals in Biology and Medicine. Basel: Birkhäuser Verlag, 1994.

49. Mitmaker B, Begin LR, Gordon PH. Nuclear shape as a prognostic discriminant in colorectal carcinoma. *Dis Colon Rectum* 1991; **34:** 249-259.

50. Mitmaker B, Kyzer S, Begin LR, H GP. The value of nuclear morphometry in the management of patients with colorectal polyps that contain invasive adenocarcinoma. *J Surg Oncol* 1992; **51:** 42-46.

51. Erler BS, Hsu L, Truong HM, *et al.* Image analysis and diagnostic classification of hepatocellular carcinoma using neural networks and multivariate discriminant functions. *Lab Invest* 1994; **71:** 446-451.

52. Coleman K, van Diest PJ, Baak JP, Mullaney J. Syntactic structure analysis in uveal melanomas. *Br J Ophthalmol* 1994; **78:** 871-874.

53. Beltrami CA, Della Mea V, Finato N. Structure analysis of breast lesions using neighborhood graphs. *Anal Quant Cytol Histol* 1995; **17:** 143-150.

54. Deligdisch L, Gil J. Characterization of ovarian dysplasia by interactive morphometry. *Cancer* 1989; **63:** 748-755.

55. Ghadially FN. Ultrastructural Pathology of the Cell and Matrix. London: Butterworths, 1988.

56. Pressman NJ. Markovian analysis of cervical cell images. *J Histochem Cytochem* 1976; **24:** 138-144.

57. Diegenbach PC, Baak JPA. Quantitative nuclear image analysis: differentiation between normal, hyperplastic, and malignant appearing uterine glands in a paraffin section. III. The use of texture features for differentiation. *Europ J Obstet Gynec Reprod Biol* 1978; **8:** 109-116.

58. Dawson AE, Cibas ES, Bacus JW, Weinberg DS. Chromatin texture measurement by Markovian analysis. Use of nuclear models to define and select texture features. *Anal Quant Cytol Histol* 1993; **15:** 227-235.

59. Kuwahara H, Shimazaki M, Morikita I, Chanoki Y, Sakurai M. Texture analysis of histological images of giant cell tumor of bone. *Path Res Pract* 1992; **188:** 565-569.

60. Thunnissen FB, Diegenbach PC, van Hattum AH, *et al.* Further evaluation of quantitative nuclear image features for classification of lung carcinomas. *Pathol Res Pract* 1992; **188:** 531-535.

61. Jagoe R, Sowter C, Slavin G. Shape and texture analysis of liver cell nuclei in hepatomas by computer aided microscopy. *J Clin Pathol* 1984; **37:** 755-762.

62. Fleming MG, Friedman RJ. Multiparametric image cytometry of nevi and melanomas. *Am J Dermatopathol* 1993; **15:** 106-113.

63. Stolz W, Abmayr W, Schmoeckel C, Landthaler M, Massoudy P, Braun-Falco O. Ultrastructural discrimination between malignant melanomas and benign nevocytic nevi using high-resolution image and multivariate analyses. *J Invest Dermatol* 1991; **97:** 903-910.

64. Hanselaar AGJM, Vooijs GP, Van't Hof-Grootenboer AE, Pahlplatz MMM. Cytophotometric analysis of cervical intraepithelial neoplasia grade III, with and without synchronous invasive squamous cell carcinoma. *Cytometry* 1990; **11:** 901-906.

65. Verhest A, Kiss R, d'Olne D, *et al.* Characterization of human colorectal mucosa, polyps, and cancers by means of computerized morphonuclear image analyses. *Cancer* 1990; **65:** 2047-2054.

66. Bibbo M, Michelassi F, Bartels PH, *et al.* Karyometric marker features in normal-appearing glands adjacent to human colonic adenocarcinoma. *Cancer Res* 1990; **50:** 147-151.

67. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975; **4:** 172-179.

68. Hutchinson ML, Isenstein LM, Martin JJ, Zahniser DJ. Measurement of subvisual changes in cervical squamous metaplastic cells for detecting abnormality. *Anal Quant Cytol Histol* 1992; **14:** 330-334.

69. Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. *Comput Vision Graph Image Process* 1982; **23:** 341-352.

70. Stoecker WV, Chiang C-S, Moss RH. Texture in skin images: comparison of three methods to determine smoothness. *Comput Med Imaging Graph* 1992; **16:** 179-190.

71. Porat M, Zeevi YY. Localized texture processing in vision: analysis and synthesis in the Gaborian space. *IEEE Trans Biomed Eng* 1989; **36:** 115-129.

72. DuBuf JMH, Kardan M, Spann M. Texture feature performance for image segmentation. *Pattern Recognition* 1990; **23:** 291-309.

73. Remmelink M, Salmon I, Petein M, Henrion S, Pasteels J-L, Kiss R. Computer-assisted quantitative description of chromatin pattern in soft tissue tumors of the adult. *Am J Clin Pathol* 1994; **102:** 780-787.

74. Listrom MB, Fenoglio-Preiser CM. Cellular aspects of neoplasia. In: Fenoglio-Preiser CM, Weinstein RS, Kaufman N, eds. New Concepts in Neoplasia as Applied to Diagnostic Pathology. Baltimore: Williams & Wilkins, 1986.

75. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

76. Spina D, Disanto A, Luzi P, *et al.* Novel, contrast gradient-oriented, automated chromatin texture analysis: I. Feasibility study on nuclei from benign and malignant breast epithelial cell lines in fine needle aspirates. *Virchows Archiv B Cell Pathol* 1992; **62:** 119-124.

77. Gimenez-Mas JA, Sanz-Moncasi MP, Remon L, Gambo P, Gallego-Calvo MP. Automated textural analysis of nuclear chromatin: a mathematical morphology approach. *Anal Quant Cytol Histol* 1995; **17:** 39-47.

78. Julesz B. Textons, the elements of texture perception, and their interactions. *Nature* 1981; **290:** 91-97.

79. Deligdisch L, Miranda C, Barba J, Gil J. Ovarian dysplasia: nuclear texture analysis. *Cancer* 1993; **72:** 3253-3257.

80. Young IT, Verbeek PW, Mayall BH. Characterization of chromatin distribution in cell nuclei. *Cytometry* 1986; **7:** 467-474.

81. de Meester U, Young IT, Lindeman J, van der Linden HC. Towards a quantitative grading of bladder tumors. *Cytometry* 1991; **12:** 602-613.

82. Kaplan DT, Furman MI, Pincus SM, Ryan SM, Lipsitz LA, Goldberger AL. Aging and the complexity of cardiovascular dynamics. *Biophys J* 1991; **59:** 945-949.

83. Bak P, Chen K. The physics of fractals. *Physica D* 1989; **38:** 5-12.

84. Burke HB. Increasing the power of surrogate endpoint biomarkers: the aggregation of predictive factors. *J Cell Biochem Suppl* 1994; **19:** 278-282.

85. McGuire WL. Breast cancer prognostic factors: evaluation guidelines. *J Natl Cancer Inst* 1991; **83:** 154-155.

86. Ostrowski JL, Sawan A, Henry L, *et al.* p53 expression in human breast cancer related to survival and prognostic factors: an immunohistochemical study. *J Pathol* 1991; **164:** 75-81.

87. Steeg PS, Bevilacqua G, Kopper L, *et al.* Evidence for a novel gene associated with low tumor metastatic potential. *J Natl Cancer Inst* 1988; **80:** 200-204.

88. Rimm DL, Sinard JH, Morrow JS. Reduced α-catenin and e-cadherin expression in breast cancer. *Lab Invest* 1995; **72:** 506-512.

89. Nicholson S, Richard J, Sainsbury C, *et al.* Epidermal growth factor receptor (EGFr); results of a 6 year follow-up study in operable breast cancer with emphasis on the node negative subgroup. *Br J Cancer* 1991; **63:** 146-150.

90. Fontana X, Ferrari P, Namer M, Peysson R, Salanon C, Bussiere F. C-erb-B2 gene amplification and serum level of c-erb-B2 oncoprotein at primary breast cancer diagnosis. *Anticancer Res* 1994; **14:** 2099-2104.

91. Kraus MH, Issing W, Miki T, Popescu NC, Aaronson SA. Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor

receptor family: evidence for overexpression in a subset of human mammary tumors. *Proc Natl Acad Sci USA* 1989; **86:** 9193-9197.

92. Plowman GD, Culouscou J-M, Whitney GS, *et al.* Ligand-specific activation of HER4/p180 erbB4, a fourth member of the epidermal growth factor receptor family. *Proc Natl Acad Sci USA* 1993; **90:** 1746-1750.

93. Abbona GC, Papotti M, Gasparri G, Bussolati G. Proliferative activity in parathyroid tumors as detected by Ki-67 immunostaining. *Hum Pathol* 1995; **26:** 135-138.

94. Cattoretti G, Becker MHG, Key G, *et al.* Monoclonal antibodies against recombinant parts of the Ki-67 antigen (MIB 1 and MIB 3) detect proliferating cells in microwave-processed formalin-fixed paraffin sections. *J Pathol* 1992; **168:** 357-363.

95. Cordon-Cardo C. Tumor suppressor genes. *Cancer* 1995; **75:** 2641.

96. Weidner N, Folkman J, Pozza F, *et al.* Tumor angiogenesis: a new significant and independent prognostic indicator in early-stage breast carcinoma. *J Natl Cancer Inst* 1992; **84:** 1875-1887.

97. Kim NW, Piatyszek MA, Prowse KR, *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* 1994; **266:** 2011-2014.

98. Bacus SS, Ruby SG. Application of image analysis to the evaluation of cellular prognostic factors in breast carcinoma. *Pathol Annu* 1993; **28:** 179-204.

99. Charpin C, DeVictor B, Andrac L, *et al.* p53 quantitative immunocytochemical analysis in breast carcinomas. *Hum Pathol* 1995; **26:** 159-166.

100. Esteban JM, Kandalaft PL, Mehta P, Odom-Maryon TL, Bacus S, Battifora H. Improvement of the quantification of estrogen and progesterone receptors in paraffin-embedded tumors by image analysis. *Am J Clin Pathol* 1993; **99:** 32-38.

101. Merchant FA, Aggarwal SJ, Diller KR, Bovik AC. In-vivo analysis of angiogenesis and revascularization of transplanted pancreatic islets using confocal microscopy. *J Microsc* 1994; **176:** 262-275.

102. Slamon DJ, Godolphin W, Jones LA, *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989; **244:** 707-712.

103. Futreal PA, Liu Q, Shattuck-Eidens D, *et al.* BRCA1 mutations in primary breast and ovarian carcinomas. *Science* 1994; **266:** 120-122.

104. Lee S, Tolmachoff T, Marchevsky AM. DNA content analysis ("ploidy") by image analysis. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994.

105. Esposito MJ, Fuchs A. Computerized image analysis and flow cytometric evaluation of ovarian borderline tumors: a study of 24 cases. *Cytometry* 1994; **18:** 218-222.

106. Mayall BH. Current capabilities in clinical applications in image cytometry. *Cytometry Suppl* 1988; **3:** 78-84.

107. Skates SJ, Xu F-J, Yu Y-H, *et al.* An optimal algorithm for screening for ovarian cancer with longitudinal tumor markers. *Cancer* 1995; **75:** 2637.

108. Hölzel WGE, Beer R, Deschner W, Griesmacher A, Müller MM. Individual reference ranges of CA 15-3, MCA and CEA in recurrence of breast cancer. *Scand J Clin Lab Invest Suppl* 1995; **221:** 93-101.

109. Bostwick DG, Burke HB, Wheeler TM, *et al.* Panel statement: the most promising surrogate endpoint biomarkers for screening candidate chemopreventive compounds for prostatic adenocarcinoma in short-term phase II clinical trials. *J Cell Biochem Suppl* 1994; **19:** 283-289.

110. Ganju V, Jenkins RB, O'Fallon JR, *et al.* Prognostic factors in gliomas: a multivariate analysis of clinical, pathologic, flow cytometric, cytogenetic, and molecular markers. *Cancer* 1994; **74:** 920-927.

CHAPTER 1

The Effect of Ultrafast Papanicolaou Staining on Nuclear and Textural Features in Breast Cancer Cytology

Andrew J. Einstein

Grace C. H. Yang

Joseph B. Silberfarb

Joan Gil

Analytical and Quantitative Cytology and Histology 1997; 19: 361-367.

Preface to Chapter 1

The first two chapters in this dissertation focus on quality control issues. If morphometric techniques are to be clinically viable as adjuvant diagnostic tests in the diagnosis of neoplasia, possible sources of variability in the morphometric descriptors characterizing and representing each diagnosis must be controlled. These include factors such as the segmentation of nuclei, staining of specimens, and sampling. This chapter addresses the effect of staining on nuclear morphometry.

Until recently, most breast fine-needle aspiration (FNA) biopsies in the United States were processed using the "Pap" stain developed by Dr. George Papanicolaou in the early 1940s. Over the past several years, Diff-Quik staining has gained in popularity, and now most breast FNAs diagnosed at the Mount Sinai Hospital are Diff-Quik stained. The great advantage enjoyed by Diff-Quik is its rapidity and simplicity; it requires only 4 steps of processing versus 14 for Pap staining, and takes only 30 seconds as opposed to 20 minutes. However, the opacity of Diff-Quik stained cells obscures nuclear detail and renders a diagnosis of cancer more difficult.

The Ultrafast Papanicolaou (UFP) protocol was developed in 1994 by Drs. Grace Yang and Iliana Alvarez, working in the cytology laboratory named for Dr. Papanicolaou at Cornell. Yang and Alvarez claimed that UFP combines the detailed cytomorphology of Pap staining and the speed of Diff-Quik. The procedure has rapidly gained popularity, and, according to a *New York Times* report, is already being used in over 300 cytopathology laboratories across the country. Our first contact with UFP came when, in looking for a series of patients with which to investigate the utility of image analysis in breast FNA diagnosis, we met with Dr. Miguel Sanchez, director of the Cytodiagnosis and Breast Care Center at Englewood Hospital, who had begun to use UFP. While Englewood promised the series of patients we needed, the UFP processing they used was new and still unpublished, and we were uncertain how applicable the diagnostic procedures we planned

to develop (Chapters 5 through 7) would be to FNAs stained using the more established methods. Moreover, in discussing UFP with several pathologists, we encountered a wide range of reactions, ranging from unqualified enthusiasm to rejection of Yang and Alvarez's claims.

This study was designed to investigate the impact of UFP staining on the morphology of breast epithelial cell nuclei. We compared UFP, Pap, and Diff-Quik stained nuclei from the same patients, and found no important quantifiable differences in nuclear morphology between UFP and Pap, but differences between Diff-Quik and each of the other methods. This indicates that the methods developed below in Chapters 5 through 7 should be useful for FNAs prepared using conventional Pap staining. Additionally, it is supportive of Yang and Alvarez's contention that UFP morphology is comparable to that of Pap. However, the results below appear to refute their original contention that nuclei are larger in UFP than in conventional Pap staining.

The FNA biopsies in this study were provided by Dr. Grace Yang, who also wrote most of the paragraph entitled "Preparation of Specimens," beginning from "Using the corner..." The micrographs in Figure 1 were taken by Dr. Joan Gil. Joe Silberfarb assisted with the segmentation.

Objective: The Ultrafast Papanicolaou staining procedure was recently introduced as a protocol combining speed comparable to that of Diff-Quik with the detailed cytomorphology characteristic of Papanicolaou staining. This study was designed to compare quantifiable cytologic features of mammary lesions prepared using each of the three methods.

Study Design: Thirteen patients with mammary lesions were studied, the majority of whom had histopathologic diagnoses of infiltrating ductal carcinoma. For each patient, three specimens were prepared, using 1) Diff-Quik staining after air drying, 2) Papanicolaou staining after wet fixation, and 3) the Ultrafast Papanicolaou procedure. Descriptors of nuclear size, shape, and texture were computed from an average of 61.9 normalized nuclear images per specimen. Differences between preparation methods were analyzed using two-way analysis of variance.

Results: While differences in nuclear size, shape, and texture existed among the three cytologic staining methods, only form factor varied significantly between conventional and Ultrafast Papanicolaou stains. Nuclear areas were larger in Ultrafast Papanicolaou specimens than conventional Papanicolaou specimens but this difference was not statistically significant.

Conclusion: The Ultrafast Papanicolaou procedure improves on conventional Papanicolaou staining in terms of speed, with no important quantifiable differences in nuclear morphology.

Key Words: morphometry, image analysis, stains and staining

The cytologic assessment of fine needle aspiration biopsies is typically performed using a wet-fixed Papanicolaou (Pap)[1] stain and/or an air-dried Diff-Quik (Wright-Giemsa) stain. Using Pap staining, although cells are "transparent" and chromatin is readily visible, the cells from the organ biopsied are obscured by orange-stained erythrocytes.[2]

Additionally, the length of the Pap procedure makes rapid assessment of specimens difficult, and while rapid Pap protocols[3,4] exist, they are troubled with a loss of cellular detail due to inadequate fixation. Using Diff-Quik, although cells are larger and preparation is simpler (4 steps vs. 14) and quicker (30s vs. 20min), less information can be obtained from nuclei because of the opacity of the stain. Consequently, other cytologic features such as cellular dyscohesion, single cells with retained cytoplasm, intracytoplasmic mucin, anisonucleosis, and nuclear enlargement as compared to the nearby erythrocytes[2] are used as criteria of malignancy.

The Ultrafast Papanicolaou (UFP) protocol was introduced in 1995 by Yang and Alvarez[5] as a hybrid of Pap and Diff-Quik, combining the detailed cytomorphology of the former with the speed of the latter. Representative high-magnification images of cells processed using each of the three methods are shown in Figure 1. While subjective claims have been made about nuclear features in the different protocols, putative differences in size, shape, and chromatin appearance have not been analyzed quantitatively. Using image analysis, quantitative measures of such nuclear characteristics can be determined, enabling an objective comparison between the protocols. In a number of contexts, morphometric analyses have been useful in demonstrating differences in nuclear features under the variation of tissue processing conditions. These include aspects of processing such as smear vs. oese preparation,[6] freezing,[7] fixation,[8-11] Böhm post-fixation,[11] embedding,[9,10] and staining.[12,13] The three protocols compared here differ most notably in fixation (air drying vs. wet fixation) and staining (Pap vs. Wright-Giemsa), and the goal of this study was to evaluate the morphologic consequences of these differences. In particular, we compared quantifiable cytologic features of mammary lesions prepared using each of the three methods.

Materials and Methods

Preparation of Specimens

Thirteen patients with carcinomas of the breast were studied. Histologically, ten of the lesions were infiltrating ductal carcinomas (one with lobular features as well), two were lobular, and one was papillary. Three cytology specimens were prepared for each patient. Using the corner of a glass slide to scratch the surface of a freshly excised breast tumor, a semiliquid drop of tumor cells was obtained and placed on another glass slide. A fine needle aspiration type of oval smear was then made with head, body, and tail.[14] For each specimen, three smears were made: one was immediately fixed in 95% ethanol and processed with conventional Papanicolaou stain, while the other two were allowed to air-dry. One of the air-dried smears was stained with Diff-Quik stain, while the other one was rehydrated in normal saline and stained with ultrafast Papanicolaou stain as previously described.[5]

Image Analysis

Image analysis was performed on a self-assembled system based on a Gateway 2000 486DX2/50E microcomputer augmented with Sprynt i860 image processing boards and the Semper 6 Plus graphics program, connected to a Nikon Optiphot microscope through a Sony DXC-M2 camera. For each of the 39 specimens, an individual segmented randomly selected epithelial cell nuclei using the arc-forming method, excluding damaged and overlapping nuclei as well as those with insufficient contrast. The arc-forming method involves the selection of three points on the nuclear membrane, the computation and display of the arc connecting these points, and the extension of this nascent contour by adding more points to define more arcs, until the whole nucleus is outlined. In another study, we found this approach to be the most reproducible method for the segmentation of cytologic images.[15] Subsequent image processing and feature computation was performed using

customized software written in our laboratory for the Silicon Graphics Indigo^2 graphics workstation. Nuclear images were normalized[16,17] to compensate for variations in lighting conditions and staining intensity and then screened to assure that they were complete, properly segmented, and of sufficient contrast for texture analysis. A minimum of 50 nuclei were collected from each slide, with the exception of one Diff-Quik stained slide for which only 44 nuclei could be collected due to scanty cellularity. A total of 2413 images was collected for the 39 slides.

Several quantitative nuclear features, also referred to as descriptors, were determined from each nuclear image. These included measures of size (nuclear area, perimeter, and maximal chord length), shape (form factor), and texture (nuclear diffuseness,[17] heterogeneity, clumping, and condensation[18]). Nuclear areas were measured using a digital approximation to Green's theorem while perimeters were determined using a Pythagorean approach.[19] Characterization of nuclear texture was done according to the approaches of Einstein *et al.*[17] and Young *et al.*[18] The former approach characterizes texture in nuclei by quantifying the amount of local intensity variation throughout the nucleus. For each pixel in the nucleus, the variance of the intensities of all pixels within a prescribed radius of the given pixel is determined; the nuclear diffuseness is defined as the average of these variances over all pixels in the nucleus. The latter approach reduces the number of gray levels in nuclear images from 255 to 3, from which a measure of heterogeneity is determined based on the number of pixels at each gray level. This feature is defined as

$$\text{Heterogeneity} = \frac{N_B + N_W}{N_B + N_G + N_W}$$

where N_B, N_G, and N_W are the number of black, gray, and white pixels, respectively. Clumping and condensation measures also reflect the homogeneity of these three gray levels in a moving window eight pixels wide.

Statistical Analysis

All statistical calculations were done using the JMP statistical package (SAS Institute, Cary, NC) running on an Apple Power Macintosh 8100/110. Means and standard deviations for each preparation method of the mean descriptor values for each patient were determined; thus, the means reported are in fact means of means while standard deviations are standard deviations of means. Descriptors were compared among staining methods by two-way analysis of variance (ANOVA) models, in which stains were treated as fixed effects and patients as random effects.[20] The models included interaction terms. Comparisons were made both among the three staining methods and between each pair of staining techniques. These pairwise comparisons were made using linear contrasts of the means for the three stains. The entire analysis was repeated using only the 10 infiltrating ductal carcinomas.

Results

Results of the study are summarized in Table 1, and the ANOVA tables are summarized in Table 2. *P* values less than 0.05 indicate a significant difference between protocols for a descriptor. Mean descriptor values for each combination of patient and processing method are graphed in Figure 2.

In sum, while there was little quantifiable difference in nuclear features between Pap and UFP processing, there were significant differences between each of these protocols and Diff-Quik staining. In comparisons among the three processing methods, there was a statistically significant difference for six of the eight descriptors: nuclear area, perimeter, maximum chord length, form factor, heterogeneity, and condensation. With the exception of form factor, when such a difference existed it was very highly significant, with very highly significant pairwise differences between Diff-Quik and Pap, and also between Diff-Quik and UFP, but not between Pap and UFP. For form factor, there was a

significant difference among the three methods, and pairwise differences were significant between Diff-Quik and Pap, and between Pap and UFP. However, we tend to discount this because unlike the other features, for which significance was at p values less than 0.001, significance for form factor was at p values between 0.02 and 0.05 and could be an artifact of multiple testing. When Bonferroni correction[21] of the p values was performed, the overall and pairwise differences in form factor were all insignificant. There was no significant difference among the three methods in either diffuseness or clumping, which attempt to describe similar phenomena. Results were identical when the analysis was limited to the 10 patients with infiltrating ductal lesions, with the single exception that the pairwise difference in form factor was only significant between Pap and UFP.

Discussion

As mentioned above, there was little difference in quantifiable nuclear features between Pap and UFP processing. Form factor, a measure of shape and the only descriptor with a significant difference between Pap and UFP, has repeatedly been shown to be highly variable and therefore unreliable for use in morphometry.[15,22,23] In contrast, there was a significant difference in most features between Diff-Quik stained cells and cells processed using either of the other methods. Nuclear size and chromatin appearance were demonstrably different among the methods but similar for the two protocols using Pap staining. This is generally consistent with previous claims. As opposed to the rapid Papanicolaou protocols of Kline[3] and Tao[4] which suffer from a loss of nuclear detail, UFP processed breast epithelial cells exhibited comparable chromatin appearance to that in conventional Pap. Moreover, the rehydration of air-dried aspirate[24] incorporated into the Ultrafast Pap procedure provides advantages over the wet fixation used in conventional Pap staining. Wet-fixed cells can be lost during processing, presenting a problem especially in specimens with scanty cellularity. Air drying rehydration results in the lysis of red blood

cells, which often obscure cellular detail. One additional benefit of air drying is that cells appear larger.[25] While subjectively, nuclei may seem larger in UFP than in conventional Pap-processed slides, the mean nuclear area for UFP (77.69μm^2) was much closer to the mean area of Pap nuclei (76.92μm^2) than of Diff-Quik nuclei (134.25μm^2). Apparently, rehydration has the effect of rounding up flattened air-dried nuclei, and the seemingly larger area of cells in UFP is probably attributable to enlargement of the cytoplasm rather than the nuclei.

The quantitative description of chromatin texture is a difficult problem, for which a number of approaches have been proposed.[17,18,26-28] The descriptors of chromatin appearance used in this study attempt to mathematically describe some diagnostically important textural characteristics of nuclei. As described above, the method of Einstein *et al.*[17] quantifies chromatin diffuseness in terms of local intensity variation throughout the nucleus, while those of Young *et al.*[18] quantify homogeneity, clumping, and condensation. While the similarity of these descriptors between Pap and UFP processing suggests that chromatin appearance is similar in these protocols, the descriptors do not necessarily reflect all diagnostically relevant aspects of chromatin appearance. Thus, there may be differences in chromatin features between Pap and UFP that are not identified in our approach, and consequently the textural findings should be regarded as suggestive, rather than conclusive.

In summary, the Ultrafast Papanicolaou procedure dramatically improves on conventional Papanicolaou staining in terms of speed and background clarity, with no important quantifiable differences in nuclear morphology. Speed is becoming increasingly important in the current managed healthcare system, where fastest test turnaround time is valued. The ultrafast Papanicolaou procedure is fast enough for a cytopathologist to offer on-site diagnosis to patients or clinicians immediately following fine-needle aspiration biopsy, allowing clinical decisions to be made without delay and in some cases reducing the number of office visits required. Additionally, the rapidity of the ultrafast protocol

allows its use by the surgical pathologist for intraoperative cytology, as an adjunct for frozen section.[14]

Acknowledgments

The authors thank Sylvan Wallenstein, Ph.D. for his statistical advice.

References

1. Papanicolaou GN. A new procedure for staining vaginal smears. *Science* 1942; **95:** 438-439.

2. Yang GCH. Ultrafast Papanicolaou stain: a superior stain for fine-needle aspiration cytology applied in conjunction with the rehydration of air-dried smears by normal saline solution technique. *Adv Anatom Pathol* 1995; **2:** 208-211.

3. Kline TS, Kline IK. Guides to Clinical Aspiration Biopsy: Breast. New York: Igaku-Shoin, 1989.

4. Tao L-C. Transabdominal Fine-Needle Aspiration Biopsy. New York: Igaku-Shoin, 1990.

5. Yang GC, Alvarez II. Ultrafast Papanicolaou stain. An alternative preparation for fine needle aspiration cytology. *Acta Cytol* 1995; **39:** 55-60.

6. van Diest PJ, Smeulders AW, Thunnissen FB, Baak JP. Cytomorphometry. A methodologic study of preparation techniques, selection methods and sample sizes. *Anal Quant Cytol Histol* 1989; **11:** 225-231.

7. Kronqvist P, Collan Y, Kuopio T, Kujari H. Nuclear morphometry in breast cancer: the influence of sampling rules and freezing of samples. *Mod Pathol* 1995; **8:** 187-192.

8. Baak JP, Noteboom E, Koevoets JJ. The influence of fixatives and other variations in tissue processing on nuclear morphometric features. *Anal Quant Cytol Histol* 1989; **11:** 219-224.

9. Boon ME, van der Poel HG, Tan CJA, Kok LP. Effect of embedding methods versus fixative type on karyometric measures. *Anal Quant Cytol Histol* 1994; **16:** 131-136.

10. Ladekarl M. The influence of tissue processing on quantitative histopathology in breast cancer. *J Microsc* 1994; **174:** 93-100.

11. Makkus ACF, van't Hof-Grootenboer AE, Pahlplatz MMM, *et al.* Practical aspects of fixatives in high resolution nuclear image analysis. *Cytometry* 1994; **15:** 302-310.

12. Schulte E, Wittekind C. The influence of the wet-fixed Papanicolaou and the air-dried Giemsa techniques on nuclear parameters in breast cancer cytology: a cytomorphometric study. *Diagn Cytopathol* 1987; **3:** 256-261.

13. Schulte E, Wittekind C. The influence of Romanowsky-Giemsa type stains on nuclear and cytoplasmic features of cytological specimens. *Anal Cell Pathol* 1989; **1:** 83-86.

14. Yang GCH, Hoda SA. Combined use of "scratch and smear" technique and ultrafast Papanicolaou stain enhances intraoperative cytology. *Acta Cytol* in press.

15. Einstein AJ, Gil J, Wallenstein S, *et al.* Reproducibility and accuracy of interactive segmentation procedures for image analysis in cytology. *J Microsc* in press.

16. Barba J, Gil J: Image analysis in statistical texture analysis *Eighth International Congress for Stereology* 1991.

17. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

18. Young IT, Verbeek PW, Mayall BH. Characterization of chromatin distribution in cell nuclei. *Cytometry* 1986; **7:** 467-474.

19. Gil J, Barba J. Principles of stereology: computerized applications in anatomic pathology. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994: 79-124.

20. Searle SR, Casella G, McCulloch CE. Variance Components. New York: John Wiley & Sons, 1992.

21. Miller R. Multiple comparisons. In: Kotz S, Johnson NL, eds. Encyclopedia of Statistical Sciences, vol. 5. New York: John Wiley & Sons, 1985: 679-689.

22. Barry JD, Sharkey FE. Observer reproducibility during computer-assisted planimetric measurements of nuclear features. *Hum Pathol* 1985; **16:** 225-227.

23. Chan KW, Chiu KY, Fu KH, Ling JML. Observer variability in microcomputer-assisted morphometric study of nuclear parameters. *Pathology* 1987; **19:** 407-409.

24. Chan JKC, Kung ITM. Rehydration of air-dried smears with normal saline: application in fine-needle aspiration cytologic examination. *Am J Clin Pathol* 1988; **89:** 30-34.

25. Yang GC. The mathematical basis for the increased sensitivity in cancer detection in air-dried cytopreparations. *Mod Pathol* 1994; **7:** 681-684.

26. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975; **4:** 172-179.

27. Haralick RM, Shanumgam K, Dinstein I. Textural features for image classification. *IEEE Trans Systems Man Cybernet* 1973; **3:** 610-621.

28. Pressman NJ. Markovian analysis of cervical cell images. *J Histochem Cytochem* 1976; **24:** 138-144.

Table 1

Comparison of the Diff-Quik, Papanicolaou, and Ultrafast Papanicolaou Protocols

Descriptor	Units	Type	Diff-Quik	Pap	UFP	*p*	*p (DQ-Pap)*	*p (DQ-UFP)*	*p (Pap-UFP)*
Area	μm^2	Size	134.25 ± 33.66	76.92 ± 26.27	77.69 ± 30.88	<0.0001 *	<0.0001 *	<0.0001 *	0.8954
Perimeter	μm	Size	44.34 ± 5.61	33.38 ± 5.66	33.24 ± 6.72	<0.0001 *	<0.0001 *	<0.0001 *	0.9019
Max Chord	μm	Size	15.07 ± 2.01	11.50 ± 2.01	11.41 ± 2.35	<0.0001 *	<0.0001 *	<0.0001 *	0.7918
Form Factor	—	Shape	0.83 ± 0.02	0.82 ± 0.02	0.83 ± 0.02	0.0440 *	0.0403 *	0.8354	0.0227 *
Diffuseness	—	Texture	179.54 ± 89.78	158.18 ± 34.68	147.66 ± 51.99	0.3802	0.3639	0.1744	0.6445
Heterogeneity	—	Texture	0.54 ± 0.06	0.45 ± 0.06	0.43 ± 0.12	0.0007 *	0.0018 *	0.0003 *	0.5307
Clumping	—	Texture	0.82 ± 0.03	0.82 ± 0.03	0.83 ± 0.04	0.2935	0.7936	0.2350	0.1457
Condensation	—	Texture	0.44 ± 0.04	0.37 ± 0.05	0.36 ± 0.10	0.0007 *	0.0011 *	0.0004 *	0.7437

For each combination of feature and protocol, the mean ± standard deviation of the mean feature value is shown. *P* values were determined by two-way ANOVA. Pairwise comparisons were made using linear contrasts of the means for the three protocols. DQ denotes Diff-Quik. Asterisks denote statistical significance.

Table 2

Summary of Analysis of Variance Tables

		Mean Square							
Source	df	Area	Perimeter	Max_Chord	Form_Factor	Diffuseness	Heterogeneity	Clumping	Condensation
Stain	2	841659	31597	3394.6	0.0186	208615	2.6134	0.0446	1.6218
Patient	12	141952	5665	717.4	0.0869	298609	0.8612	0.1525	0.5079
Patient*Stain	24	13790	483	55.5	0.0052	208214	0.2631	0.0347	0.1631
Error	2374	1179	38	4.8	0.0024	3892	0.0100	0.0029	0.0071

Source refers to source of variation. df denotes degrees of freedom.

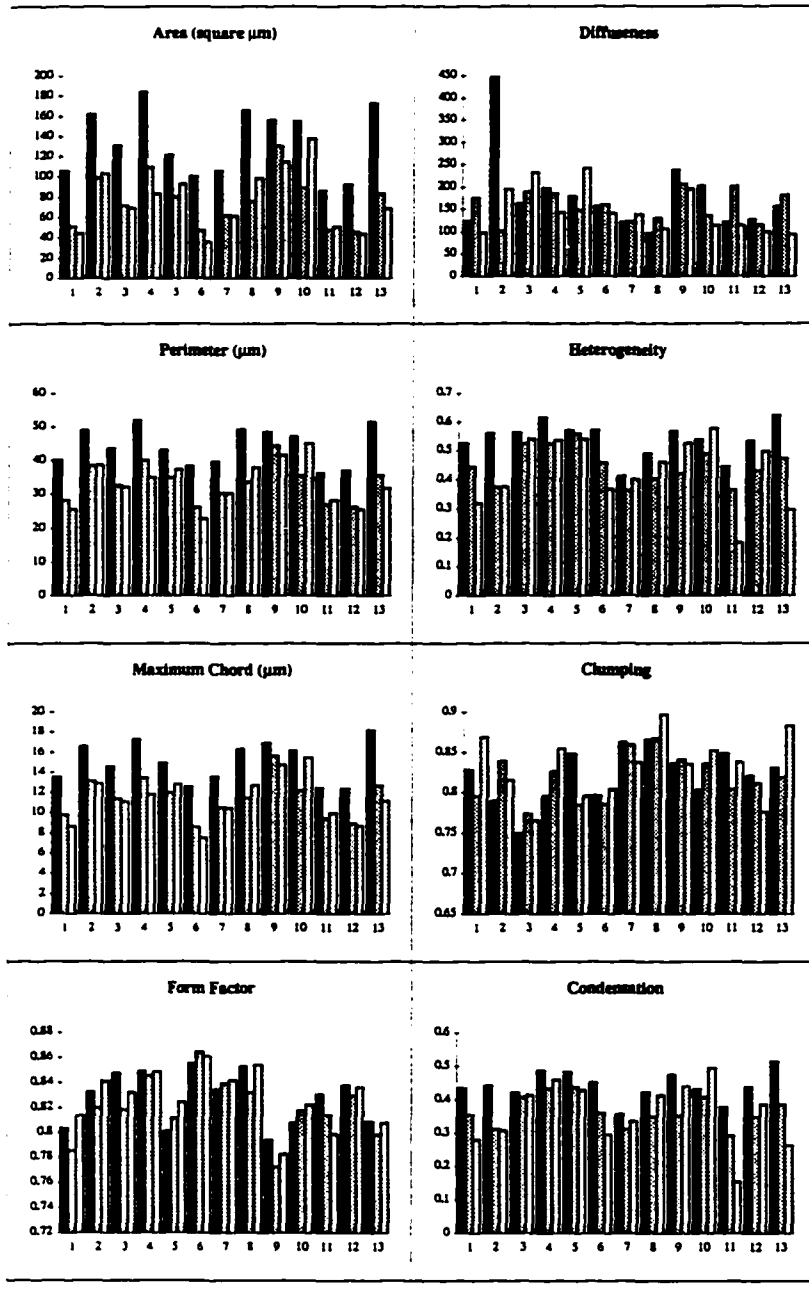
Figure 1

Representative Cells Processed With Each of the Three Protocols

(a) Diff-Quik (b) Papanicolaou (c) Ultrafast Papanicolaou (×2000)

Figure 2

Comparison of the Protocols by Patient



For each patient the three bars represent mean feature values for Diff-Quik, Pap, and Ultrafast Pap processed nuclei, respectively. Histologic types are ductal for patients 1-10, lobular for 11-12, and papillary for 13. Size and shape features are in the left column while texture features are in the right column.

CHAPTER 2

Reproducibility and Accuracy of Interactive Segmentation Procedures for Image Analysis in Cytology

Andrew J. Einstein
Joan Gil
Sylvan Wallenstein
Carol A. Bodian
Miguel Sanchez
David E. Burstein
Hai-Shan Wu
Zhiyuan Liu

Journal of Microscopy 1997; 188: 136-148.

Preface to Chapter 2

Continuing on the theme of quality assurance, this chapter addresses the effect of segmentation on nuclear morphometry. The segmentation of nuclear images is a crucial step in the development of procedures using image analysis for the cytologic diagnosis of cancer, and can serve as a major source of variability. Here we evaluate the reproducibility and accuracy of several interactive segmentation methods, as well as the related topic of sample size requirements for image analysis in cytology.

Drs. Sylvan Wallenstein and Carol Bodian provided statistical guidance in this study. Dr. Miguel Sanchez provided materials and Dr. David Burstein provided diagnoses. As noted in the paper, the ellipse-fit method was developed by Dr. Hai-Shan Wu, and the arc-fit method was developed by Zhiyuan Liu. I wrote section 2.3 based on an earlier version by Hai-Shan Wu. Figures 1, 2, and 3 were produced by Hai-Shan Wu. The five individuals performing the segmentation were Drs. Joan Gil, Hai-Shan Wu, Doina Ciura, and Zhang Libang, and I.

Summary

The segmentation of nuclear images is a crucial step in the development of procedures using image analysis for the cytologic diagnosis of cancer. The purpose of this study is to evaluate the reproducibility and accuracy of several interactive segmentation methods which can be used in this context. Four methods were studied: a thresholding-based method enabling selection of intensity histogram contrast and brightness, manual tracing with a stylus, and arc- and ellipse-fitting routines. Features of nuclear size and shape were derived from nuclei segmented on repeated occasions by several individuals. Variance component models provided a statistical framework for evaluating the intraobserver and interobserver variability of these measurements in terms of their intraclass correlation coefficients. Of the methods tested, the arc-fitting segmentation method gave the most reproducible results, and thresholding the least. Reproducibility was generally very high both between individuals and for repeated segmentations by a single individual. Accuracies of area measurements for the various methods, as determined with respect to point counting, paralleled the reproducibilities of the methods. Sample size requirements were observed to be more dependent on the biologic variability of the tissue sampled than on the particular segmentation method or on the number of individuals performing segmentation.

Key Words: accuracy, cytology, image analysis, intraobserver reproducibility, interobserver reproducibility, Monte Carlo methods, morphometry, segmentation, variance component models

1. Introduction

Many investigators have worked on developing morphometric procedures for the cytologic and histologic diagnosis of malignancies. Recent applications have included the diagnoses

of malignant melanoma,[1] carcinoma of the breast,[2] and parathyroid carcinoma.[3] If such techniques are to be useful as adjuvant diagnostic tests for differentiating between entities of similar morphology, possible sources of variability in the measurement of morphometric features, in particular nuclear profile areas, must be minimized. These sources of variability may include preparation and staining of specimens, sampling, segmentation of nuclei, and methods used for determining feature values from segmented images.[4,5] This study concentrates on segmentation, an important step in the determination of virtually all features.

Castleman[6] defines the image segmentation process as "one that partitions a digital image into disjoint (nonoverlapping) regions." In morphometric studies in pathology, these regions of interest are usually nuclear images, and segmentation consists in accurately delineating the outline of the nuclear envelope. Approaches to segmentation may be fully automatic or may involve some interactive component. An automatic segmentation algorithm of the kind routinely offered by commercial instruments may seek to identify either regions of interest or the boundaries between regions, solely on the basis of pixel grey or color values. The appeal of an effective automatic segmentation algorithm is great. Not only would it eliminate the subjectivity perceived to be inherent to interactive approaches, but also segmentation could be performed much more rapidly. Unfortunately, developing such an algorithm has proved to be an exceedingly difficult problem. While the methods used in commercial instruments are generally not disclosed, the research literature has proposed numerous approaches to automatic segmentation,[7] incorporating thresholding,[8] classification/clustering,[9] region growing,[10] mathematical morphology,[11] and edge detection/boundary tracking methods[12] as well as their hybrids. None has been entirely satisfactory for the real-time segmentation of nuclear images in pathology.[4] Despite the wealth of attempts, no single method is sufficiently flexible to accommodate all the variability present in histologic and cytologic images. The comment of Irinopoulou *et al.*[13] is typical: "full automation was achieved only at the cost of skipping all pale nuclei."

A description of the issues involved from the point of view of image analysis is given by Wu & Barba.[14] Problems may include interference from nuclear and background texture, overlapping structures, attached images, voids within nuclei omitted from the segmented image, unwanted debris included in the image, computational complexity and noise. Notwithstanding these potential pitfalls, commercial instrumentation does not provide calibration methods to check the reliability of automatic segmentation methods incorporated into these devices.

Therefore, it is generally necessary to employ interactive segmentation techniques for image analysis in pathology. One group of interactive segmentation procedures involves using a pointing device such as a stylus on a touch screen,[15] mouse, or trackball to specify the outline of a nucleus. Although such approaches have been used in numerous morphometric and stereologic applications,[16,17] they are generally slower than more automated procedures. Another class of interactive segmentation methods involves the selection of points on the periphery of a nucleus and the computation and display of a contour from these points based on curve fitting. A third group of interactive segmentation methods, popular in commercial instrumentation, is based upon automatic algorithms but provides for some user input. Such input may take the form of restricting the operation to designated cells, of employing editing capabilities, of interactive selection of grey-level thresholds, or of trying alternative methods on the same region of interest and selecting the best segmentation. Editing functions may include provisions for separating, merging, resizing, selecting, and deleting images.[18] Thresholding methods include all pixels with intensities at or above a set level — the threshold — in the segmented image, excluding the remaining pixels. All interactive methods introduce an element of subjectivity, dependent upon the individual performing the segmentation.

The present study was designed to analyze the effect of segmentation method choice on the reproducibility and accuracy of measurements of nuclear features. We begin by describing a number of interactive segmentation methods which can be used for image

analysis in cytology. After introducing statistical methodology which can be used to evaluate reproducibility, we discuss the intraobserver and interobserver reproducibility of these segmentation methods, their time requirements, their accuracies, and their impacts on sample size requirements.

2. Materials and methods

2.1 Cytologic specimens

Cytologic material was obtained from six patients with invasive ductal carcinoma of the breast, and processed using the ultra-fast Papanicolaou method.[19] As opposed to histology, where the investigator works with cut profiles of embedded cells, in cytology, whole cells lie compressed between the slide and the cover slip. This has the effect of presenting an unnaturally enlarged projected area.

2.2 Image analysis

Image analysis was performed on a self-assembled system based on a Gateway 2000 486DX2/50E microcomputer augmented with Sprynt i860 image processing boards and Semper 6 Plus graphics program (Synoptics Ltd., Cambridge, UK), connected through a Sony DXC-M2 color camera to a Nikon Optiphot microscope equipped with a Nikon Plan 100× objective with a numerical aperture of 1.25. Images containing six nuclei from a single patient were digitally stored for easy retrieval (Figure 1). These images were based on the green components of the nuclei's RGB signals; in our experience, this component typically provides the best contrast for cytologic images. Five variably experienced individuals each interactively segmented the six nuclei on five different occasions. We shall refer to the five individuals as readers or observers, and the five repetitions for each reader as observations. The process was repeated using the following interactive segmentation algorithms: 1) a thresholding-based method, 2) manual tracing with a stylus

on a touch-sensitive screen, 3) an ellipse-fitting algorithm taking as input four extremal points of the nuclear outline, and 4) an arc-fitting routine. In the latter method, three points are selected on the margin of a nuclear profile, and the arc connecting these points is computed and displayed; the contour is extended by adding more points to define more arcs, until the whole nuclear profile is outlined. The two curve-fitting methods are discussed in greater detail below. Features describing area, perimeter, maximal chord length, and form factor[20] were derived from each segmented nuclear profile. The accuracy of area measurements was determined for each method, considering area estimated by point counting on a randomly positioned, calibrated, fine (5 pixel × 5 pixel) grid as a gold standard. These areas were estimated according to two observations by each of two readers. Sample size requirements were determined from both the single patient and also from five additional patients, for each of which a single reader segmented at least 50 nuclei using each of the four methods as well as point counting.

2.3 Curve-fitting segmentation methods

Two of the segmentation methods involved the interactive determination and display of nuclear contours. One uses quarter-ellipses, while the other attempts to fit profiles by piecewise arcs.

The ellipse-fit method,[14] illustrated in Figure 2, determines the contour based on four reader-specified points. It exploits two characteristics of nuclear images: first, that nuclear contours often have lower grey intensities than their immediate surroundings, and second, that nuclear profiles are smooth and somewhat elliptical, or at least piecewise elliptical, in shape. The algorithm requires the user to specify the four points corresponding to the extreme left, top, right, and bottom points on the contour. These selected points define a rectangle that encloses the whole nuclear region and partitions the nuclear contour into four segments. Each of the segments is approximated by a quarter ellipse, and it is supposed that nuclear contours lie spatially near these approximation

curves. A cost function is defined on the set of pixels in the image. This cost function is a composite of two factors. One factor reflects the pixel's intensity value, while the other reflects the closeness of the pixel to the approximation curve, assigning large values to pixel locations far from the generated quarter ellipse and small values to pixels close to the ellipse. The algorithm searches for those points which minimize the composite cost function and identifies these extrema as nuclear contour points.

In the arc-fitting method, an unpublished approach developed by Z. Liu in our laboratory and illustrated in Figure 3, a finite number of separated contour points are specified or marked by the user. Points should be selected close enough so that profile convexity is relatively uniform between neighboring marks. The remaining portions of the nuclear contour are to be determined so that the final contour extracted is a closed curve. Suppose that there are $2n$ contour points marked on the cell profile, where n is an integer and the first and last points are identical. The whole contour will be approximated by n pieces of arcs, each of which is determined by three successive marked points. Let the three successive marks be represented by (x_1, y_1), (x_2, y_2), and (x_3, y_3). A circle is determined with the center at (x_0, y_0) and radius of r_0, where

$$x_0 = \frac{\Delta x}{\Delta}, \quad y_0 = \frac{\Delta y}{\Delta}, \text{ and } r_0 = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

for

$$\Delta = \begin{vmatrix} x_1 - x_2 & y_1 - y_2 \\ x_1 - x_3 & y_1 - y_3 \end{vmatrix},$$

$$\Delta x = \begin{vmatrix} \frac{1}{2}(x_1^2 - x_2^2 + y_1^2 - y_2^2) & y_1 - y_2 \\ \frac{1}{2}(x_1^2 - x_3^2 + y_1^2 - y_3^2) & y_1 - y_3 \end{vmatrix},$$

and

$$\Delta y = \begin{vmatrix} x_1 - x_2 & \frac{1}{2}(x_1^2 - x_2^2 + y_1^2 - y_2^2) \\ x_1 - x_3 & \frac{1}{2}(x_1^2 - x_3^2 + y_1^2 - y_3^2) \end{vmatrix}.$$

The arc selected from the circle begins at (x_1, y_1), goes through (x_2, y_2), and ends at (x_3, y_3).

3. Reproducibility of segmentation methods

Various terms, such as reproducibility, variability, unbiasedness, accuracy, precision, and reliability, are used to describe the quality of measured data such as morphometric measurements. These terms are subject to differences in interpretation. In this paper, we use *reproducibility* to refer to agreement among repeat measurements. The first level of reproducibility, called *intraobserver reproducibility*, is the extent to which repeated measurements of the same object (nucleus) by a specific reader agree with each other. *Intraobserver variability* refers to a lack of such agreement. The second level of reproducibility, called *interobserver reproducibility*, is the extent of agreement among different readers measuring the same object; the lack of such agreement is *interobserver variability*. These terms all concern the internal agreement among repeat measurements.

In contrast, *accuracy* refers to the relation between the measurements and some external standard. *Unbiasedness* is used to reflect the equality of the average of the replicated measurements for a particular segmentation method to the "true value." One source of confusion is the fact that the term "unbiased" is used both in a statistical sense, as above, and in a stereologic sense, to refer to an assumption-free method.

While all these concepts are closely linked, their relative importances vary depending on the application. For diagnostic morphometry, where individual measurements are important only insofar as they result in a correct diagnosis, terms describing reproducibility of individual measurements are typically more important than those describing their accuracy. In this section we evaluate the reproducibilities of the four segmentation methods, while in Section 4 we focus on accuracy.

3.1 Statistical terminology

Agreement among repeat measurements of features such as area and perimeter can be statistically characterized by an intraclass correlation coefficient (ICC). ICCs are defined as "measures of the relative similarity of quantities which share the same observational units of a sampling and/or measurement process"[21] and serve a similar role for continuous data as Kappa statistics do for categorical data. They are determined as combinations of variance components in linear models. ICCs typically are expressed as percentages, with an ICC of 100% corresponding to perfect agreement.

Before describing the linear models used to determine the ICCs, a few comments to clarify some terminology used to describe linear models are in order. Our discussion follows Searle *et al.*[22] Linear models are constructed to study the variability found in data, and in particular to attribute variability to various categorizations of data, referred to as *factors*. Factors considered in this study include nucleus, reader, and patient. Each class of a factor is called a *level*; the five people who segmented the nuclei are the five levels of the factor "reader." When we characterize data in terms of factors and their levels, we are interested in the degree to which the different levels of a factor affect the measurement of the variable or feature of interest, for example how a particular reader performing the segmentation affects the measurement of nuclear area. This is referred to as the *effect* of the level of the factor on the feature. Effects are classified as being *fixed effects* or *random effects*. Basically, effects are considered fixed if we are interested in the (finite) levels of a factor for their own sake, and random if we are interested in the observed levels only insofar as they are representative of a (generally infinite) population of levels from which they are drawn. For example, in this study we are not really interested in the readers *per se*, but only as representatives of all the possible people who could be performing the segmentation. Thus, "reader" is considered a random effect. In a different (hypothetical) study, we might want to compare four cytotechnologists in terms of their reliability in

performing segmentation, to determine which of the four should be hired for a new position. In such a study "reader" would be a fixed effect.

Linear models are described in terms of their effects and factors. If all of the effects studied in a linear model are fixed, then it is called a *fixed effects model*; if all of the effects are random, then it is called a *random effects model*; if some are fixed and some are random then it is called a *mixed model*. If a model considers the effects of the levels of a single factor, then it is called a *one-way* model. If the effects of two factors are considered at the same time, then it is a *two-way* model, etc. Factors in two-way and more complicated models can be *crossed* or *nested*. If each level of one factor is observed in combination with each level of the other factor, then the model is crossed. If the levels of one factor are unique to a particular level of the other factor, then the model is nested, or hierarchical.[23] In this study, where the same six nuclei are segmented by each reader, the study design is crossed. Two-way and more complicated models may also include *interaction terms* to quantify the effects of a synergistic effect or interaction among multiple factors, *e.g.*, how differences among readers depend on the nucleus. The calculation of the analysis of variance, and consequently the ICCs, varies depending on whether effects are fixed or random, whether factors are crossed or nested, and whether or not the model includes interaction terms. Thus, it is important that these be properly specified.

3.2 Statistical methodology

The linear models in this study were implemented in JMP (SAS Institute, Cary, NC), while further calculations to determine ICC estimates and sample size requirements were performed using Microsoft EXCEL. Intraobserver variabilities in nuclear features were studied with a family of 80 one-way random effects models, one for each combination of feature, method, and reader. Each of these models is given by

$$y_{ni} = \mu + \alpha_n + e_{ni}$$

where y_{ni} is the observed value of the feature for the ith segmentation of nucleus n, μ is the general mean of the feature value, $\alpha_n \sim N(0, \sigma_\alpha^2)$ (*i.e.*, α_n, which is assumed to be normally distributed with a mean of 0 and a variance of σ_α^2) is the effect due to nucleus n, and $e_{ni} \sim N(0, \sigma_e^2)$ is the residual error, *i.e.*, intraobserver variability. The variance components σ_α^2 and σ_e^2 were estimated using the ANOVA method, which equates mean squares in the analysis of variance (ANOVA) table to their expected values.[22] ICCs in the one-way model measure the agreement among repeat observations by a specified reader of a single feature using a particular method, and are given by

$$\textit{Intraobserver ICC} = s_n^2 / (s_n^2 + s_e^2) \qquad (1)$$

where s_n^2 and s_e^2 are ANOVA estimates of σ_α^2 and σ_e^2, the variance components for nucleus and residual error, respectively. Confidence intervals for intraobserver ICCs were calculated as functions of the test's F statistic according to Searle *et al.*[22]

Segmentation approaches were compared through a family of 16 two-way, random effects, crossed classification, interaction models, with one for each combination of feature and method. Each model is given by

$$y_{rni} = \mu + \alpha_n + \beta_r + \gamma_{nr} + e_{nri}$$

where y_{rni} is the value obtained for the specified feature at the ith segmentation of nucleus n by reader r, μ is the general mean, $\alpha_n \sim N(0, \sigma_\alpha^2)$, $\beta_r \sim N(0, \sigma_\beta^2)$, and $\gamma_{nr} \sim N(0, \sigma_\gamma^2)$ are the effects due to nucleus n, reader r, and nucleus-by-reader interaction, respectively, and $e_{nri} \sim N(0, \sigma_e^2)$ is the residual error. The ANOVA variance component estimates for nucleus, reader, nucleus by reader interaction, and residual error are denoted by s_n^2, s_r^2, s_{nr}^2, and s_e^2, respectively. A similar two-way model, based on two observations from each of two readers on the six nuclei, was used to assess the reproducibility of point counting, the gold-standard segmentation method.

Two types of ICCs, referred to here as the common intraobserver ICC and the interobserver ICC, derive from these two-way models. Their determinations follow an approach similar to that of Shrout & Fleiss.[24] In contrast to the one-way models, where a

separate ICC was postulated for each reader, the two-way models assume a common underlying level of consistency shared by all readers. For a given feature and method, the common intraobserver ICC measures the expected agreement among replicate observations for a reader selected at random. It is given by

$$\textit{Common Intraobserver ICC} = s_n^2 / (s_n^2 + s_r^2 + s_{nr}^2 + s_e^2) \ . \qquad (2)$$

The interobserver ICC quantifies interobserver variability for a particular feature and method, in terms of the expected agreement between observations for two randomly selected readers. It is given by

$$\textit{Interobserver ICC} = (s_n^2 + s_r^2 + s_{nr}^2) / (s_n^2 + s_r^2 + s_{nr}^2 + s_e^2) \ . \qquad (3)$$

3.3 Results

Results concerning agreement for single readers are shown in Table 1, while those involving multiple readers are shown in Table 2. For reasons of brevity, the results for single readers are displayed only for area, while results for other features are summarized below. Area, diagnostically the most important of all morphometric features, is also the most relevant feature for comparing segmentation methods as it reflects most directly the pixels included in the segmented nuclear image. The data suggest a number of results about segmentation and the variation of morphometric measurements, which are discussed in the following sections.

3.4 Intraobserver reproducibility

Intraobserver ICCs are measures of intraobserver reproducibility, gauging the variability of morphometric measurements for a single reader. A high intraobserver ICC reflects high intraobserver reproducibility, *i.e.*, low intraobserver variability, and *vice versa*. Intraobserver reproducibility of area was high for all methods except thresholding, with intraobserver ICCs typically above 95%. Since most intraobserver ICCs were greater than 95%, they exhibited little variability among readers, supporting the assumption of a

common intraobserver ICC (2) among readers. Estimates of these common intraobserver ICCs are shown in Table 2. Intraobserver ICCs behaved similarly for maximal chord length and for perimeter, but were lower for form factor.

3.5 Interobserver reproducibility

Interobserver reproducibility for a single feature and method is expressed both in terms of the interobserver ICC, determined in the two-way model by (3), and the variability of intraobserver ICCs among readers, determined in one-way models by (1). For most combinations of feature and method, interobserver ICCs were greater than 90%, and as noted above, there was little variability of intraobserver ICCs among readers. Thus, in general there was good agreement among readers; most variation in morphometric measurements was due to morphologic differences between nuclei and not a result of intraobserver and interobserver variability in segmentation. There were a few notable exceptions to this low variability characterizing most combinations of feature and method. The implications of these exceptions, in terms of the choice of segmentation methods and features, are discussed below.

3.6 Comparison of segmentation methods

The arc-fitting method was the most reproducible of the four segmentation methods evaluated, reflected by an interobserver ICC for area of 96.7%. The corresponding estimate of common intraobserver ICC was 97.6%, with reader-specific estimates ranging from 95.2% to 99.2%; this range is quite narrow. The reproducibility of the arc-fitting method was followed by that of tracing with a stylus and the ellipse-fitting method. While for area, tracing was more reproducible than the ellipse-fitting method, for other features it was less reproducible. The thresholding method was not reproducible for any feature, with interobserver ICCs between 36.89% and 72.85%. Use of the ultra-fast Papanicolaou procedure, a recent advance in cytology staining,[19] resulted in crisp chromatin detail; while this is a marked advantage for both conventional diagnosis and morphometric texture

analysis, it makes segmentation more difficult due to interference from texture. This is particularly problematic for thresholding methods.

Time requirements are a significant consideration in the choice of segmentation methods. The average time per nucleus required for a single, experienced reader (H.-S.W.) to segment the nuclei was approximately 14 seconds for arc-fitting, 7.3 seconds for ellipse-fitting, 4.5 seconds for thresholding, 26 seconds for tracing with a stylus, and 77 seconds for point counting. These averages do not include time to locate nuclei and reposition the stage, or processing time, which depends on the hardware used and on the complexity of the method. The long time per nucleus required for point counting was caused by the use of a very fine grid; shorter times would be expected were a coarser grid used. In general, the more automated a segmentation procedure was, the less time per nucleus it required. The reproducibility of a segmentation method needs to be balanced against the time required for its implementation. In estimating mean nuclear areas, it may not be worthwhile to use the most reproducible segmentation method if there are many nuclei available, but instead "do more less well".[25] However if there are a limited number of nuclei available, which is often the case in aspiration cytology where many specimens are of scanty cellularity, then a more reproducible but slower segmentation method is appropriate. Similarly, when features more sensitive to segmentation are being determined, *e.g.*, some texture descriptors, the slower, more reproducible method is better. Thus, the choice of a segmentation method should be tailored to the specific requirements of the diagnostic application.

3.7 Reproducibility of features

In general, the high intraobserver and interobserver ICCs suggest that the nuclear features used are sufficiently reproducible to warrant their use in morphometric studies, using any of the segmentation methods except thresholding. However, there were two exceptions:

1) ICCs for form factor, a measure of roundness, were relatively small. Interobserver ICCs were 84% and below, depending on the method, while intraobserver ICCs of less than 80% were observed for each of the methods. This confirms observations of others[26,27] that form factor is not a reproducible feature for use in morphometry.

2) The interobserver ICC for perimeter when tracing with a stylus was low (78.70%). This is not unexpected, since tracing may be corrupted by "jitter," and noise in the form of jitter has been show to result in overestimation of profile boundary lengths.[28,29] Moreover, perimeter measurements are highly dependent on the method used to derive them from segmented profiles,[4] rendering their use in morphometry dubious.

4. Accuracy of Segmentation Methods for Area

Assessment of accuracy is dependent on having a gold standard against which to compare measurements. In measurements of nuclear profiles there are no absolute gold standards; the "true" values of features are unknown or not well defined. Nevertheless, point counting with a fine grid can give an unbiased (assumption-free) estimate of the true area, and is generally thought to be very reproducible. We tested the validity of this assumption using a two-way model as described above, based on two observations from each of two readers. As is shown in Table 2, the estimate of the common intraobserver ICC for point counting was 99.2%, while the estimate of the interobserver ICC was 99.3%. These ICCs are higher than those of the four segmentation methods considered. The high ICCs demonstrate the reproducibility of point counting, a necessary condition for its use as a gold standard.

Using the mean area from the four measurements (two from each of two readers) as the gold standard, errors in area measurements were determined for each combination of

method, reader, and observation as the difference between the measured value and the gold standard. Error measurements for each of the four methods, broken down by reader, are plotted in Figure 4, which is based on a modification of Bland and Altman's[30] approach.

The accuracies of area measurements using the four methods generally reflect the reproducibilities of the methods. Thresholding drastically underestimated the areas of many nuclei. The range of errors in area measurements was somewhat smaller for the arc-fitting method than for the ellipse-fitting method or stylus tracing. The correspondence between the accuracies and reproducibilities is not surprising, for these two factors are interrelated. It is impossible to consistently agree with a gold standard if repeat measurements have little internal agreement.

The significance of the (statistical) bias for a segmentation method can be tested for using a two-way random effects crossed classification model. The model is given by

$$Error\ in\ area_{rni} = \mu + \alpha_n + \beta_r + \gamma_{nr} + e_{nri}$$

where $Error\ in\ area_{rni}$ is the difference between the observed value and the gold standard for the ith segmentation of nucleus n by reader r, μ is the bias, $\alpha_n \sim N(0, \sigma_\alpha^2)$, $\beta_r \sim N(0, \sigma_\beta^2)$, and $\gamma_{nr} \sim N(0, \sigma_\gamma^2)$ are the effects on error due to nucleus n, reader r, and nucleus-by-reader interaction, respectively, and $e_{nri} \sim N(0, \sigma_e^2)$ is the residual error. For each method, testing the hypothesis H_0: $\mu = 0$ yields a p value which characterizes the significance of the observed statistical bias. Estimates of μ were negative for each of the four segmentation methods. These differences were highly significant for each method, and thus all underestimate area with respect to the point counting gold standard. For thresholding, the error in determining area was considerable, on average -17.95 μm^2. For the other three methods, the statistical bias was much smaller: -0.79 μm^2 for arc-fitting, -0.54 μm^2 for ellipse-fitting, and -0.91 μm^2 for tracing. Considering the close agreement among these three methods, it is conceivable that in fact some small systematic error existed in the way point counting was applied.

As can be seen in Figure 4, there is some variability in error among readers. Using the model above, the significance of the effect due to reader was tested. The effect was highly significant for all methods except arc-fitting, where it was of borderline significance ($p = 0.067$). Nevertheless, the magnitudes of the differences in statistical bias between readers were small, typically about 1 μm^2.

As expected, point counting using a fine grid produced extremely reproducible measurements. However, although this approach provides a reproducible and unbiased (assumption-free) estimate of the true area of a nucleus, it does not result in the segmentation of a nuclear image from its background. Without this segmentation, it is impossible to extract other features, *e.g.*, perimeter, shape, texture, and fractal descriptors, from the nuclear image. While it would be possible to specify the points counted within the nuclear image, for example with a mouse, this would be tedious and add to the considerable time per nucleus already required for point counting. Nevertheless, point counting remains the most reproducible method for the determination of nuclear areas and can serve as a gold standard for our purposes.

5. Sample Size Determination

Determining the minimum number of nuclei to segment for a particular application requires the specification of some minimal criterion for the measured feature values to satisfy. One approach typically used to specify sample size requirements is to determine the number of nuclei that need to be segmented such that the measured coefficient of error (CE) of feature values is less than some specified percentage, say *X%*. A second approach required by some authors, *e.g.*, DeHoff,[31] is to calculate the number of nuclei necessary to estimate mean feature values within *Y%* of the true value with *U%* confidence. In fact, these two approaches are closely related. Assuming independence of the features describing the nuclei segmented, the former requires

$$CE = \frac{standard\ deviation}{mean \times \sqrt{sample\ size}} \leq X\%$$

while the latter requires

$$z_{(U\%+1)/2} \times \frac{standard\ deviation}{mean \times \sqrt{sample\ size}} \leq Y\%, \qquad (4)$$

where $z_{u\%}$ is the uth percentile of the standard normal distribution. Thus, the first approach is a special condition of the second, where $X = Y$ and $z_{(U\%+1)/2} = 1$, *i.e.*, $U\% = 68.3\%$. Equating the two sides of (4) and solving for sample size yields the minimum sample size meeting the required condition:

$$sample\ size = variance \times \left(\frac{z_{(U\%+1)/2}}{mean \times Y\%} \right)^2 . \qquad (5)$$

The values of $U\%$ and $Y\%$ in the sample size requirement calculation should depend on the accuracy required by the application. In the stereology literature, it has been suggested that a CE of 5% or less should suffice for most applications,[32] and similar recommendations have been made for cytopathology.[33] In practice, to determine the error tolerance for a diagnostic procedure based on the classification of morphometric or stereologic features, one can perturb mean feature values and determine the effect on classifier performance. In this study, we performed sample size determinations using two particular criteria: 1) $CE \leq 5\%$ (the first approach with $X = 5$), and 2) more stringently, determining the number of nuclei necessary to estimate mean feature values within 5% of the true value with 95% confidence (the second approach with $Y = 5$ and $U = 95$). More precise methods for determining sample sizes to satisfy confidence intervals of specified width have been suggested, for example by Beal.[34]

Sample size requirements were estimated using variance components from the two-way model. In practice, segmentation may be performed either by a single person or by numerous individuals; in the latter case it is expected that there will be added interindividual variability in feature measurements. Thus, for each feature and method, sample size determinations were made for both a single reader system, in which case the variance was

set to the sum of the nucleus and residual variance component estimates, and also a multiple reader system, in which case the variance (s^2) was set to the sum of all variance components. Following (5), the appropriate formulas for sample size requirements are

$$n_{single\ reader} = (s_n^2 + s_e^2) \times (1.96 / 0.05\mu)^2$$

$$n_{multiple\ readers} = (1.96s / 0.05\mu)^2$$

using the first criterion, and

$$n_{single\ reader} = (s_n^2 + s_e^2) / (0.05\mu)^2$$

$$n_{multiple\ readers} = (s / 0.05\mu)^2$$

using the second criterion.

Since all 600 values of each feature in the original evaluation were taken from six nuclei from a single patient, they do not reflect all the morphologic variability that can exist in cytology. For this reason, sample size determinations were repeated for nuclear area using a larger sample of nuclei. A single reader segmented at least 50 randomly sampled nuclei from each of 5 patients, using each of the four methods as well as point counting. Each patient had been diagnosed with invasive carcinoma of the breast, and cytologic specimens were processed and image analysis was performed as above. Sample size requirements were determined for each segmentation method according to (5), both for the 5 patients individually and as a composite, again using both criteria. Variance was estimated for individual patients by the sample variance, and for a composite sample size requirement as ANOVA estimates of the residual error (nucleus) variance component in a one-way random effects nested model. The model is specified by

$$y_{pn} = \mu + \alpha_p + e_{pn} \quad ,$$

where y_{pn} is the measured area of the nth nucleus of patient p, μ is the mean area, $\alpha_p \sim N(0, \sigma_\alpha^2)$ is the effect due to patient p, and $e_{pn} \sim N(0, \sigma_e^2)$ is the residual error attributed to the nucleus. Results of these sample size requirement determinations are summarized in the top part of Table 3. The sample sizes shown are those to determine mean area; sample sizes required to determine mean values of other nuclear features were

lower, with the single exception of mean perimeter using thresholding. Sample size requirements vary systematically between the two criteria compared; the number of nuclei required is always about a factor of four ($z^2_{0.975}$) greater using the second criterion. For a given criterion, most of the variation in estimates of sample size requirements was between patients. For example, using arc-fitting, to estimate the mean area within 5% of the its true value with 95% confidence, one would need to segment 46 nuclei for Patient 2, but 227 nuclei for Patient 5. In contrast, the variation in sample size requirements between segmentation methods was small. For example, sample sizes required to estimate mean area with CE < 5% ranged between 23 and 25 among the five methods. Similarly, sample size requirements differed little between the single and multiple reader systems. Thus, the two factors contributing the most to sample size requirements for image analysis in cytology are the intrinsic biologic variability of the tissue of interest, and the criterion chosen for the maximum variability permitted.

All of the foregoing sample size requirements are based on the assumption that the nuclei sampled are spatially independent. If this is not the case, then a weaker assumption is that CE behaves as a power function of sample size (N), *i.e.*, $CE = kN^{-b}$, rather than $CE = kN^{-1/2}$ as would be implied under independence.[35] As suggested to the authors by T. Mattfeldt, under the weaker assumption, Monte Carlo techniques can be used to estimate sample sizes, even when the stronger assumption of independence does not hold. To estimate the parameters k and b, multiple samples of different sizes can be drawn, and CE estimated from each sample. Here, for each of the 5 patients, 100 samples each of sizes 10, 20, 30, and 40 nuclei were drawn with replacement. The slope of the linear regression line fit to the plot of *log CE vs. log N* estimates $-b$, while its y-intercept estimates *log k*. Having estimated k and b, the number of nuclei required can be estimated by $N \geq \left(\frac{k}{X\%}\right)^{1/b}$ for the first criterion, or by analogy, $N \geq \left(\frac{z_{(U\%+1)/2}\, k}{Y\%}\right)^{1/b}$ for the second criterion. In addition, standard methods[36] can be used to determine 95% confidence intervals on the

regression parameters. All of these calculations were performed in C on a Silicon Graphics workstation, and are summarized in the bottom portion of Table 3.

With spatially independent nuclei, one would expect *CE* to fall off as $1/\sqrt{N}$, *i.e.*, $b = 0.5$. We found the point estimates of b for the 5 methods to be 0.461 to 0.476, while the upper limits of the 95% confidence intervals were from 0.490 to 0.504. Therefore, spatial correlation of the areas of the nuclei seems negligible. Moreover, as can be seen in Table 3, sample size requirement estimates are similar whether the nuclei are assumed to be independent or not. Thus, spatial dependence of randomly sampled nuclei in cytology does not appear to pose a major problem for sample size requirement determination.

In practice, how should one approach the question "How many nuclei should be sampled?" In diagnostic cytology applications, one can get some idea of the allowable tolerance in feature values by determining the effect on classifier performance of randomly perturbing feature values. The loss of classifier performance caused by errors of a given magnitude needs to be weighed against the costs in time and resources of tolerating a smaller magnitude of error. In absolute terms, it has been suggested[32] that a measured CE of 5% is almost always acceptable. Nevertheless, the number of nuclei required to attain a prescribed CE can vary greatly from patient to patient, depending on the biologic variability present. It would be desirable to estimate means and standard deviations of the features of interest from the first several nuclei segmented, in order to provide an initial sample size estimate for the patient. Ideally, running CEs could be maintained, and segmentation of nuclei stopped only when CE is less than the target value for each feature of interest. Based on the results in this study, it appears that 50 or 60 nuclei will be enough in most cases to meet this requirement.

6. Conclusion

General experience in point-counting stereology[25,37] and morphometry[26,27] has been that intraobserver and interobserver variability is not an important source of error when using manual procedures. This study extends this observation to the context of breast cytomorphometry. Contour fitting methods based on interactively selected profile pixels improve on the manual approach both in terms of accuracy and speed. More automated segmentation procedures such as the thresholding and mathematical morphology based approaches used in some commercial instruments are less successful, owing to the complexity and heterogeneity of histologic and cytologic images. Fully automatic segmentation is almost unimaginable in histology, and still exceedingly difficult in cytology, where instruments used for screening identify regions containing nuclei rather than locating the actual borders of nuclear profiles. In the United States, recent Food and Drug Administration approval of two such devices for cervical cytology screening[38,39] has paved the way for other applications of image analysis in pathology, bringing issues of quality control to the forefront. Variance component models provide the proper framework for statistically evaluating the reproducibility of morphometric data, and as such they should play an increasingly important role.

Acknowledgments

This work was presented in part at the Ninth International Congress for Stereology, Copenhagen, August 20-25, 1995. The authors thank Drs. Matt Reed and Torsten Mattfeldt for numerous helpful suggestions, and Drs. Doina Ciurea and Zhang Libang for assistance in performing the study. A. J. Einstein received support from NIH Training Grant GM07280 and from a Hans Elias Bursary.

References

1. Fleming MG, Friedman RJ. Multiparametric image cytometry of nevi and melanomas. *Am J Dermatopathol* 1993; **15:** 106-113.

2. Hutchinson ML, Isenstein LM, Zahniser DJ. High-resolution and contextual analysis for the diagnosis of fine needle aspirates of breast. *Anal Quant Cytol Histol* 1991; **13:** 351-355.

3. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

4. Gil J, Barba J. Principles of stereology: computerized applications in anatomic pathology. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994: 79-124.

5. Smeulders AWM, Dorst L. Measurement issues in morphometry. *Anal Quant Cytol Histol* 1985; **7:** 242-249.

6. Castleman KR. Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, 1979.

7. Pal NR, Pal SK. A review on image segmentation techniques. *Pattern Recognition* 1993; **26:** 1277-1294.

8. MacAulay C, Palcic B. A comparison of some quick and simple threshold selection methods for stained cells. *Anal Quant Cytol Histol* 1988; **10:** 134-138.

9. Foran DJ, Berg RA. A method for quantitative image assessment based on redundant feature measurements and statistical reasoning. *Comput Methods Programs Biomed* 1994; **45:** 291-305.

10. Wu H-S, Barba J, Gil J. Region growing segmentation of textured cell images. *Electr Lett* 1996; **32:** 1084-1085.

11. Wu H-S, Barba J, Gil J. Morphological segmentation of textured cell images. *J Imaging Sci Tech* 1996; **40:** 265-270.

12. Barba J, Jeanty H, Fenster P, Gil J. The use of local entropy measures in edge detection for cytological image analysis. *J Microsc* 1989; **156:** 125-134.

13. Irinopoulou T, Rigaut JP, Benson MC. Toward objective prognostic grading of prostatic carcinoma using image analysis. *Anal Quant Cytol Histol* 1993; **15:** 341-344.

14. Wu H-S, Barba J. An efficient semi-automatic algorithm for cell contour extraction. *J Microsc* 1995; **179:** 270-276.

15. Silage DA, Gil J. The use of a touch-sensitive screen in interactive morphometry. *J Microsc* 1984; **134:** 315-321.

16. Deligdisch L, Miranda C, Barba J, Gil J. Ovarian dysplasia: nuclear texture analysis. *Cancer* 1993; **72:** 3253-3257.

17. Hoffman K, Gil J, Barba J, *et al.* Morphometric analysis of benign and malignant adrenal pheochromocytomas. *Arch Pathol Lab Med* 1993; **117:** 244-247.

18. Marchevsky AM, Erler BS. Morphometry in pathology. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994: 125-179.

19. Yang GC, Alvarez II. Ultrafast Papanicolaou stain. An alternative preparation for fine needle aspiration cytology. *Acta Cytol* 1995; **39:** 55-60.

20. Gschwend JE, Vogel U, Bader C, Mattfeldt T, Hautmann RE. Predictive value of magnetic resonance imaging and computerized tomography for conservative renal surgery in an ex vivo tumor enucleation study followed by step-sectioning. *J Urol* 1996; **155:** 451-454.

21. Koch GG. Intraclass correlation coefficient. In: Kotz S, Johnson NL, eds. Encyclopedia of Statistical Sciences, vol. 4. New York: John Wiley & Sons, 1983: 212-217.

22. Searle SR, Casella G, McCulloch CE. Variance Components. New York: John Wiley & Sons, 1992.

23. Hicks CR. Fundamental Concepts in the Design of Experiments. New York: Holt, Rinehart, and Winston, 1982.

24. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979; **86:** 420-428.

25. Gundersen HJG, Østerby R. Optimizing sampling efficiency of stereological studies in biology: or 'Do more less well!'. *J Microsc* 1981; **121:** 65-73.

26. Chan KW, Chiu KY, Fu KH, Ling JML. Observer variability in microcomputer-assisted morphometric study of nuclear parameters. *Pathology* 1987; **19:** 407-409.

27. Barry JD, Sharkey FE. Observer reproducibility during computer-assisted planimetric measurements of nuclear features. *Hum Pathol* 1985; **16:** 225-227.

28. Gil J, Marchevsky AM, Silage DA. Applications of computerized interactive morphometry in pathology: I. Tracings and generation of graphic standards. *Lab Invest* 1986; **54:** 222-227.

29. Cornelisse JTWA, Berg JTP. Profile boundary length can be overestimated by as much as 41% when using a digitizer tablet. *J Microsc* 1984; **136:** 341-344.

30. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1(8476):** 307-310.

31. DeHoff RT. Sampling of material and statistical analysis in quantitative stereology. In: Elias H, ed. Stereology: Proceedings of the Second International Congress for Stereology. New York: Springer-Verlag, 1967: 119-130.

32. Gundersen HJG, Jensen EB. The efficiency of systematic sampling in stereology and its prediction. *J Microsc* 1987; **147:** 229-263.

33. van Diest PJ, Baak JPA. Morphometry. In: Bibbo M, ed. Comprehensive Cytopathology. Philadelphia: W. B. Saunders Company, 1991: 946-964.

34. Beal SL. Sample size determination for confidence intervals on the population mean and on the difference between two population means. *Biometrics* 1989; **45:** 969-977.

35. Mattfeldt T. The accuracy of one-dimensional systematic sampling. *J Microsc* 1989; **153:** 301-313.

36. Rosner B. Fundamentals of Biostatistics. Belmont, CA: Wadsworth Publishing Company, 1995.

37. Mathieu O, Cruz-Orive LM, Hoppeler H, Weibel ER. Measuring error and sampling variation in stereology: comparison of the efficiency of various methods for planar image analysis. *J Microsc* 1981; **121:** 75-88.

38. Associated Press. Panel backs new tool to find cervical cancer New York Times. New York, 1995; C6.

39. Wied GL. Industrial developments in automated cytology as submitted by their developers. *Anal Quant Cytol Histol* 1993; **15:** 358-370.

Table 1

One-way Analysis of Variance

Feature	Method	Reader	Mean (μm²)	Variance Component Estimates: Nucleus	Residual	Intraobserver ICC (%)	95% CI Lower Limit	95% CI Upper Limit
Area	Arc-Fit	1	73.09	144.17	4.93	96.69	90.13	99.46
Area	Arc-Fit	2	73.86	145.13	7.38	95.16	85.91	99.20
Area	Arc-Fit	3	74.49	171.06	1.65	99.05	97.04	99.85
Area	Arc-Fit	4	72.28	137.14	3.70	97.37	92.06	99.57
Area	Arc-Fit	5	73.93	172.90	1.45	99.17	97.42	99.87
Area	Ellipse-Fit	1	76.20	186.16	5.40	97.18	91.52	99.54
Area	Ellipse-Fit	2	70.22	169.59	11.54	93.63	81.89	98.94
Area	Ellipse-Fit	3	75.44	160.90	7.18	95.73	87.45	99.30
Area	Ellipse-Fit	4	71.80	161.85	4.18	97.48	92.38	99.59
Area	Ellipse-Fit	5	75.19	175.02	1.63	99.08	97.13	99.85
Area	Threshold	1	60.28	101.91	99.30	50.65	15.87	88.23
Area	Threshold	2	61.15	72.61	131.19	35.63	3.74	81.92
Area	Threshold	3	48.27	286.20	55.04	83.87	60.18	97.12
Area	Threshold	4	61.11	118.89	117.94	50.20	15.46	88.07
Area	Threshold	5	51.02	70.36	206.94	25.37	-2.97	76.13
Area	Tracing	1	75.24	157.31	5.80	96.45	89.44	99.42
Area	Tracing	2	72.83	172.32	4.26	97.59	92.70	99.61
Area	Tracing	3	74.73	176.89	5.01	97.25	91.71	99.55
Area	Tracing	4	74.91	214.41	5.36	97.56	92.61	99.60
Area	Tracing	5	69.33	163.74	5.02	97.03	91.08	99.52

Intraobserver ICCs, given by Eq. (1), measure agreement between observations of a single feature for a particular reader and method. Only results for areas are shown.

Table 2

Two-way Analysis of Variance

Feature	Method	Mean Value	Variance Component Estimates: Nucleus	Reader	Interaction	Residual	Common Intraobserver ICC (%)	Interobserver ICC (%)
Area	Arc-Fit	73.53	153.13	0.46	0.95	3.82	97.59	96.70
Area	Ellipse-Fit	73.77	168.60	6.26	2.11	5.99	96.73	92.16
Area	Threshold	56.37	104.17	30.32	25.82	122.08	56.77	36.89
Area	Tracing	73.41	176.10	5.77	0.84	5.09	97.29	93.77
Area	Point Counting	74.32	143.08	0.02	-0.29	1.21	99.16	99.35
Form Factor	Arc-Fit	0.8379	0.0015	0.0000	0.0000	0.0002	87.69	84.00
Form Factor	Ellipse-Fit	0.8631	0.0005	0.0000	0.0001	0.0001	79.90	71.66
Form Factor	Threshold	0.3318	0.0706	0.0026	0.0016	0.0230	76.49	72.21
Form Factor	Tracing	0.7744	0.0018	0.0006	0.0001	0.0011	69.38	51.23
Max Chord	Arc-Fit	10.99	1.08	0.00	0.01	0.04	96.74	95.80
Max Chord	Ellipse-Fit	10.81	1.25	0.01	0.04	0.06	95.83	91.67
Max Chord	Threshold	10.31	0.92	0.04	0.04	0.26	79.08	72.85
Max Chord	Tracing	11.24	1.23	0.04	0.01	0.08	94.29	90.84
Perimeter	Arc-Fit	33.08	5.54	0.02	0.05	0.22	96.18	94.99
Perimeter	Ellipse-Fit	32.66	8.86	0.26	0.08	0.33	96.49	92.91
Perimeter	Threshold	67.39	773.71	33.71	59.39	494.48	63.68	56.84
Perimeter	Tracing	34.39	6.02	0.86	0.06	0.71	90.74	78.70

Mean shows mean feature value for each feature and method. Areas were measured in μm^2, while perimeters and maximal chord lengths were measured in μm. Common intraobserver ICCs, given by Eq. (2), measure the expected agreement between observations for a reader selected at random. Interobserver ICCs, given by Eq. (3), measure the expected agreement between observations for two randomly selected readers.

Table 3
Sample Size Requirement Estimates
(Numbers of nuclei required to satisfy criteria)

		Criterion 1: Measured CE < 5%					Criterion 2: Mean within 5% of True Value with 95% Confidence				
		Arc-Fit	Ellipse-Fit	Threshold	Tracing	Point Counting	Arc-Fit	Ellipse-Fit	Threshold	Tracing	Point Counting
6 Nuclei	Single Reader	11.6	12.8	28.5	13.4	10.0	44.6	49.3	109.4	51.7	40.1
	Multiple Readers	11.7	13.4	35.6	13.9	10.0	45.0	51.7	136.6	53.5	40.1
Assuming Independence	Overall	23.8	23.6	24.2	22.7	24.1	91.4	90.8	93.1	87.0	92.6
	Patient 1	31.7	33.0	36.2	31.1	33.9	121.9	126.6	139.2	119.5	130.1
	Patient 2	11.9	12.9	10.3	10.4	12.0	45.5	49.4	39.4	39.9	46.0
	Patient 3	16.4	16.2	16.0	15.8	15.8	63.2	62.4	61.6	60.6	60.7
	Patient 4	28.9	27.2	30.6	27.7	27.9	111.1	104.3	117.6	106.6	107.2
	Patient 5	59.2	55.4	61.2	59.8	60.3	227.3	212.7	235.1	229.9	231.7
Not Assuming Independence	Overall	23.6	23.6	23.9	22.5	23.6	100.9	99.2	103.1	92.6	99.2
	Patient 1	30.0	30.4	35.0	29.1	31.8	121.1	131.5	148.5	113.3	126.1
	Patient 2	10.9	12.0	8.6	7.4	10.8	42.5	48.0	38.4	37.9	44.2
	Patient 3	14.1	14.8	14.5	14.8	14.5	67.3	62.0	62.4	60.9	60.5
	Patient 4	27.7	26.3	28.7	26.3	26.9	113.0	105.4	123.1	112.4	99.3
	Patient 5	57.0	54.1	57.5	57.1	59.5	253.5	252.4	239.4	256.6	266.5

Figure 1

The Six Nuclei Segmented (Mammary Carcinoma)

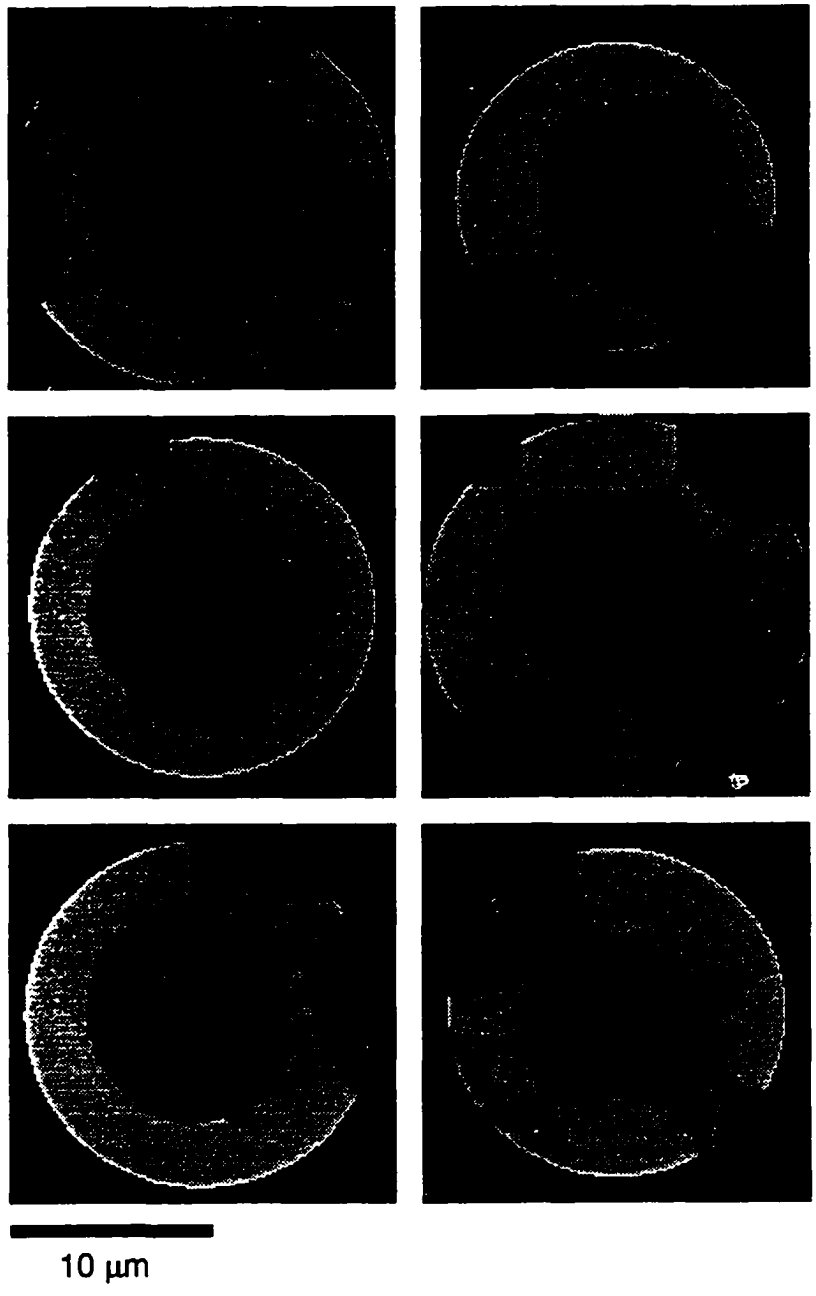


Figure 2

The Ellipse-fitting Method

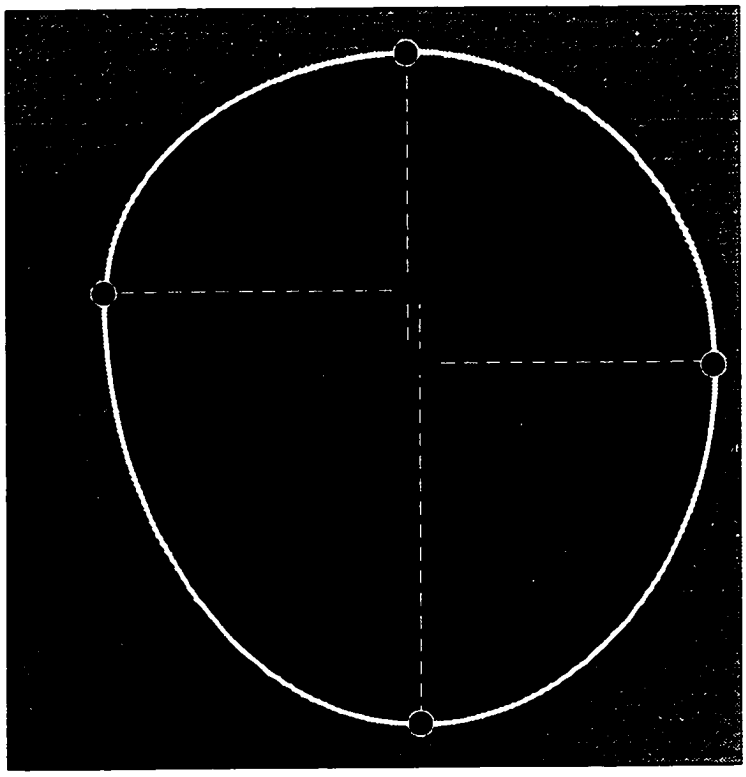


Figure 3

The Arc-fitting Method

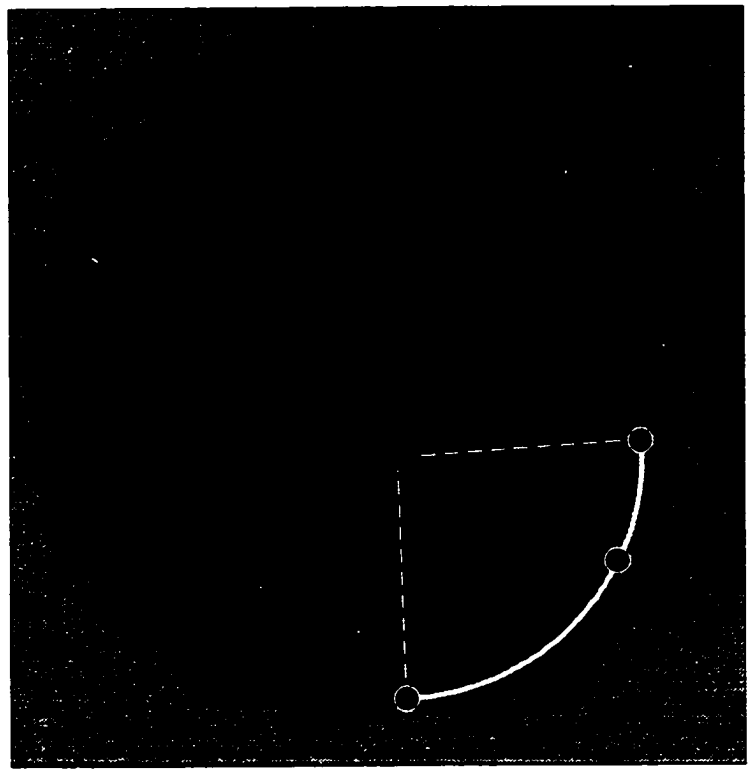
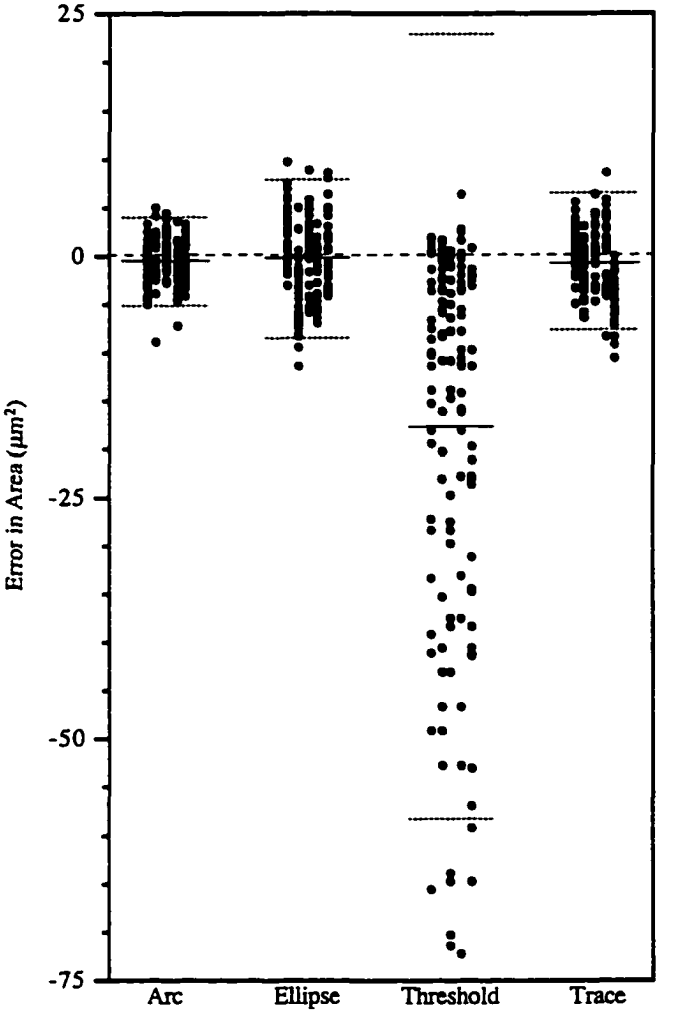


Figure 4

Errors in Area Measurements Using the Four Methods, Broken Down by Reader.



Solid lines represent mean errors while dotted lines represent mean ± 2 × standard deviation. The dashed line represents zero error.

CHAPTER 3

Ovarian Dysplasia in Epithelial Inclusion Cysts: A Morphometric Approach Using Neural Networks

Liane Deligdisch

Andrew J. Einstein

Dora Guera

Joan Gil

Cancer 1995; 76: 1027-1034.

Preface to Chapter 3

Due to the inaccessible location of the ovaries, ovarian cancer is typically diagnosed at a much later stage than the other gynecologic malignancies. Drs. Liane Deligdisch and Saul Gusberg were among the first to recognize the significance of premalignant changes in ovarian epithelium. In 1984 they proposed ovarian dysplasia, now alternatively known as ovarian intraepithelial neoplasia (OIN), as a distinct clinical entity. We conducted this study to quantitatively evaluate the morphologic similarity of two types of OIN appearing in the case literature, *viz.* dysplasia as an incidental finding in ovaries removed for non-neoplastic disorders and dysplasia seen adjacent to overt ovarian carcinoma. Patients were characterized in terms of measures of nuclear size and shape, and texton-based measures of chromatin appearance. By comparison of both the descriptors characterizing patients, using the Kruskal-Wallis test, and their neural network classifications, we found the two types of dysplasia to be morphologically indistinguishable. This suggests that the associated changes in tissue morphology reflect a single underlying pathologic process, in turn lending credence to the hypothesis that ovarian dysplasia seen in epithelial inclusion cysts may represent a preinvasive malignant change.

The biological side of this work is due to Dr. Deligdisch and her visiting fellow, Dr. Dora Guera. Dr. Guera was responsible for the interactive segmentation of nuclei. My contribution, done with the guidance of Dr. Joan Gil, is the technical side, most importantly the statistical and neural network analysis. Portions of this paper were written by Dr. Deligdisch, and portions I wrote. In particular, to the best of my recollection, the first two paragraphs of the Introduction, the section entitled "Biological Materials," and the first, second, fifth, and sixth paragraphs of the Discussion were written by Dr. Deligdisch, while I wrote the third paragraph of the Introduction, all of the Materials and Methods section with the exception of "Biological Materials," all of the Results section, the third and fourth paragraphs of the Discussion, and the Appendix. Figures 1-3 were taken by Dr. Deligdisch, while I produced Figure 4 and the Table. The abstract is a joint effort.

Precis

Ovarian dysplasia in epithelial inclusion cysts adjacent to cancer and incidentally diagnosed are indistinguishable as evaluated by statistical and artificial neural network analysis.

Abstract

Background: Ovarian dysplasia has been described in the ovarian surface epithelium by histologic and morphometric studies. This study evaluates ovarian dysplasia in epithelial inclusion cysts both adjacent to overt carcinoma and also incidentally found in ovaries removed for non-neoplastic diseases including oophorectomies for family history of ovarian cancer, using an artificial neural network.

Methods: Histologic sections from 37 ovaries of which 26 were diagnosed with dysplasia in epithelial inclusion cysts (10 adjacent to carcinoma and 16 incidental) and 11 with benign epithelial inclusion cysts were evaluated by tracing nuclear profiles and assessing measures of nuclear area, shape and texture. These were analyzed using artificial neural networks and also statistically using the Kruskal-Wallis test with the Dunn procedure to compare the morphologic similarity of incidental and adjacent dysplasia.

Results: Neither statistical nor artificial neural network analysis was able to distinguish between incidental and adjacent dysplasia. Both types differed significantly from the control cases.

Conclusions: Neural networks are powerful classification tools when applied to multiple variables extracted from individual cases. In this study, they helped to substantiate the similarity between dysplasia found incidentally and that adjacent to ovarian carcinoma. As dysplasia represents a potential pre-cancerous lesion, its incidental finding may help identify patients at risk for developing ovarian carcinoma.

Key Words: ovarian dysplasia, morphometry, neural networks, ovarian epithelial inclusion cysts

Introduction

Epithelial ovarian cancer is presently the most lethal cancer of the female pelvis, surpassing the lethality of the more common uterine cervix and corpus cancer combined. Its early diagnosis remains elusive despite advances in transvaginal sonography. Biological tumor markers, of which CA 125 is the most utilized, are neither specific nor sensitive enough to be used for screening asymptomatic patients. Clearly, morphologic techniques continue in the forefront of early diagnosis.

With the advent of new surgical laparoscopic techniques, the ovaries become more accessible and laparoscopic biopsies can provide biopsic material that may reveal early precancerous (dysplastic) or preinvasive cancerous lesions. Such precancerous lesions have been described in high risk patients with prophylactic oophorectomies[1] and adjacent to overt carcinoma.[2] We have shown previously the value of computerized image analysis in validating this rather subtle diagnosis. Measurements of cellular stratification and crowding, of nuclear profiles[3,4] and nuclear texture[5] as analyzed by a discriminant classificatory procedure have proven that dysplasia of the ovarian surface epithelium represents a distinct histologic entity, characterized by features intermediate between normal and carcinomatous ovarian epithelium.[3-5] The present study was designed to evaluate differences, if any, in ovarian dysplasia present in the epithelium lining ovarian inclusion cysts in three diagnostic categories: 1) Benign cysts - inclusion cysts lined by benign epithelium seen as incidental findings, considered as controls; 2) "Incidental dysplasia" - dysplasia as an incidental finding in ovaries removed for non-neoplastic disorders (including four cases from patients with a family history of ovarian cancer); 3) "Adjacent dysplasia" - dysplasia seen adjacent to overt ovarian carcinoma.

Our goal in this study is different from that in our previous morphometric studies. Rather than differentiating between patients in various diagnostic categories, we seek to quantitatively explore the relationship between categories. Our approach is twofold: firstly, to compare categories on the basis of the morphometric descriptors characterizing representative cases, and secondly, to determine whether differences between groups, if any, are significant enough to affect the ability to morphometrically distinguish between categories. This analysis is complex because it involves the simultaneous consideration of multiple descriptors. We have applied Kruskal-Wallis tests of standard size, shape, and texture features, and neural network classification of the features.

Materials and Methods

The general paradigm in morphometry is as follows: Specimens are collected from a number of patients representing each of the diagnostic categories under consideration. Each specimen is viewed under a microscope, attached via a video camera to a computer. A large number of digitized images of nuclei are obtained from each patient and stored as gray level matrices. Morphometric descriptors, or features, are computed: these describe characteristics of the nuclear profiles such as size, shape and texture. The average (or other statistics) of these values for each patient are computed. Thus, each patient is characterized by a vector of such features. A classification algorithm is used to define criteria for inclusion in the diagnostic categories, and is then applied to the classification of unknowns, thereby using morphometric information to provide a diagnosis.

Biological Materials

This study includes 37 cases of which 26 were diagnosed as ovarian dysplasia in epithelial inclusion cysts, according to the criteria used for dysplasia of the surface epithelium.[1-5] Specimens in all cases were embedded in paraffin, sectioned at a nominal

thickness of 5 μm, and stained with hematoxylin and eosin. Sixteen cases were incidental findings in ovaries removed for reasons other than neoplasms (Fig. 1); in four of these cases, there was a family history of ovarian cancer. Ten cases showed histologically similar changes seen in the vicinity of overt invasive serous papillary carcinoma (Fig. 2) and eleven cases were histologically benign appearing epithelial inclusion cysts (Fig. 3). Fifty consecutive nuclei of the epithelium lining the inclusion cysts were submitted for morphometric analysis that included nuclear profile area and chromatin texture, as described in a previous publication.[5]

Image Analysis Procedures

We used a self-assembled image analysis system based on a Gateway 2000 (486DX2/50E) microcomputer augmented with Sprynt i860 image processing boards and Seringer 6 Plus graphics program (Synoptics Ltd., Cambridge, UK). Image acquisition was from a Nikon Optiphot microscope with an attached Sony camera. We used a semi-automatic interactive approach, selecting with a mouse three points on the margin of the nuclear profile, interactive computation and display of the arc connecting the points, and extension of the curve by adding points to define more arcs, until the whole nuclear profile is outlined. Subsequently the enclosed area of the nuclear profile was digitized. Perimeter lengths were computed following a Pythagorean approach; area was computed from a digital approximation to Green's theorem.[6] In addition to determining the nuclear profile length, enclosed area, maximal chord length, and circularity factor, we extracted textons from the chromatin.[5,7] Nuclear images were conditioned to normalize contrast for variations in lighting conditions and stain intensity, as previously described.[8,9] The gray level histogram[10] for each adjusted nucleus was divided into three equal ranges and each pixel gray level was replaced by the average gray level of the group to which it belonged. This revealed agglomerates of gray levels (textons) within the chromatin, which were quantitated as to size and number for each case. This information, as well as the size and

shape descriptors, was averaged for each case and used for the statistical and neural network studies.

Statistical Analysis

Possible intergroup differences in morphometric descriptors were analyzed using the Kruskal-Wallis test[11] with the Dunn procedure.[12] The Kruskal-Wallis test may be seen as the non-parametric analogue of the analysis of variance (ANOVA) procedure. Statistically, to demonstrate that a variable differs between k groups of patients, an ANOVA is used. This, however, assumes an underlying normal distribution of the variable for each of the k groups. If this cannot be assumed to hold, the nonparametric analogue, called the Kruskal-Wallis test is used. Since we have repeatedly seen that morphometric descriptors cannot be assumed to have underlying normal distributions, the appropriate statistical test to compare descriptors between our three categories is the Kruskal-Wallis test. When using Kruskal-Wallis, a test statistic H is computed based upon rank sums, corrected for ties. To assess statistical significance (p) and determine if there is an overall difference between the groups with respect to the variable considered, H is compared with a chi-squared distribution. To compare specific groups, the Dunn procedure is used; this involves z statistics based on pairwise differences in rank sums. The values of these statistics are compiled in Table 1 for the features computed in this study.

Neural Network Analysis

The neural networks used in this study were generated using BrainMaker (California Scientific Software, Nevada City, CA). Artificial neural networks are discussed in an Appendix.

The data set was divided into a training set and a test set; the neural network was trained on the former, and evaluated using the feature vectors from the latter. Approximately 75% of the patients in each category were randomly assigned to the training

set, and the remainder were retained to test the trained neural networks. Specifically, the training set comprised 8, 12 and 7 patients from categories one, two and three respectively, while the test set consisted of three, four, and three patients, respectively. Neural networks were generated with 10 input neurons and one layer of 14 hidden neurons. The networks were trained by back propagation using the delta rule with gradient descent, a sigmoidal transfer function, a training tolerance of 0.05, and a testing tolerance of 0.5.

Five neural networks were trained and tested. In the first three networks, the three categories were compared pairwise: benign cysts vs. incidental dysplasia, incidental dysplasia vs. adjacent dysplasia, and incidental dysplasia vs. adjacent dysplasia. Each of these networks contained two output neurons, corresponding to the two diagnoses being compared. All three groups were included in the remaining pair of networks. In one, all dysplastic cases were clumped together, while in the other, each group had its own output neuron.

Results

A Kruskal-Wallis test and Dunn procedure were performed on the average values for each patient, of each of the 10 descriptors. The results are summarized in Table 1. For 8 of the 10 descriptors (the exceptions being circularity factor and the area of the dark textons), the p value was less than 0.05; in each of these cases $p<0.005$. Hence, for these eight features there exists a statistically significant difference between the three categories. However, for none of these 8 features was the z statistic comparing categories 2 and 3 in the significant range. z_{12} was significant for 7 of the 8 features, and z_{13} was significant for all eight features. Thus, even those descriptors shown statistically to be effective in discriminating between the categories, did not distinguish between incidental and adjacent dysplasia. So from the point of view of the morphometric descriptors calculated, incidental ovarian dysplasia is an entity closer to adjacent ovarian dysplasia than it is to benign

ovarian epithelium.

The weights of the output neurons for the test cases in the neural networks run are contained in Table 2. BrainMaker was able to diagnose between benign cases and incidental dysplasias in six of seven cases, and between benign cases and adjacent dysplasias in six of six cases. When considering a combined group of dysplasias from patients with and without cancer, it correctly diagnosed between the dysplasias and controls in nine out of ten cases. Thus, the neural network was able to accurately distinguish between control and dysplastic cases. However, when confronted with the classification of dysplasias into those from cancerous and noncancerous organs, it failed achieving a correct classification of less than 50%. Network 4, comparing all three categories, successfully identified all of the controls but again misclassified four of the seven dysplasias (Fig. 4). Since a random classification scheme would be expected to correctly classify half of the cases, this indicates the inability of the neural network to distinguish between categories two and three based upon their feature vectors. Thus, neither the feature vectors themselves nor the classification based upon them are useful in characterizing the dysplastic cases. On the other hand, both the descriptors and the classification resulting from them successfully distinguish incidental dysplasia from control cases, and adjacent dysplasia from control cases. We conclude that according to the objective criteria of morphometry, dysplasia as an incidental finding in ovaries removed for non-neoplastic disorders, is an entity that seems to be indistinguishable from dysplasia seen adjacent to overt ovarian carcinoma.

Discussion

In the absence of reliable tumor markers, the early diagnosis of ovarian cancer and the identification of precancerous lesions is mostly based on morphologic tissue changes. Recently developed laparoscopic techniques offer an increased accessibility to ovarian

tissue. Ovarian dysplasia could represent a precursor of ovarian epithelial malignancy although, as in dysplasias of other epithelial tissues (cervix, larynx, colon, gallbladder, etc.) its natural history and progression to overt carcinoma are unpredictable. The histologic and morphometric description of ovarian dysplasia,[1-3] as well as studies of tumor markers and oncogenes[13,14] may lead to the identification of patients at risk for ovarian cancer. It should be pointed out that morphologic alterations of tissues have been used for the identification of early malignancy in the vast majority of cancers (cervical, endometrial, breast, lung, colon, etc.).

Epithelial inclusion cysts are known to be the site of origin for benign and malignant epithelial ovarian tumors.[15] They represent invaginations of surface epithelium and are frequently encountered on histologic sections of adult and fetal ovaries;[16,17] in their majority they are devoid of pathologic significance. Descriptions of hyperplastic and atypical changes in the lining epithelium have pointed toward a relationship with ovarian epithelial tumors.[15] As is the case with ovarian surface epithelium,[3,5] the presence of stratification, loss of polarity and nuclear atypia in the inclusion cysts adjacent to overt neoplasm may reflect a "field effect" of the carcinogen and/or an early pre-invasive stage of malignancy, in a manner similar to that encountered in other tissues (colonic polyps and carcinoma, dysplastic nevi and melanoma, CIN and cervical carcinoma, etc.). The nuclear chromatin pattern reflects the proliferative activity of the cell and is used for the histologic diagnosis of most neoplasms. The identification of subtle changes in the nuclear texture by using quantitative morphometric procedures has validated the histologic description of ovarian dysplasia.[5]

The present study was designed to identify potential precursors of ovarian carcinoma by examining morphologic analogies between ovarian dysplasias arising in different settings. Our approach is morphometric. While morphometry studies are typically framed in terms of demonstrating that a particular choice of descriptors and a classificatory algorithm may be used to classify cases as to their diagnostic groups, here we

are interested in exploring the relationship between these groups - a task more akin to hypothesis testing than to classification. Since morphometry is based both upon computation of descriptors and their subsequent classification, groups may be compared based upon both the descriptors of representative cases and our ability to classify cases based upon these descriptors.

Identity of descriptors is characterized by insignificant z statistics in the Dunn procedure, of all descriptors deemed by the Kruskal-Wallis test to differentiate among the groups. In our comparison of incidental and adjacent dysplasia, this was satisfied for ten descriptors which characterize disparate types of nuclear features and had previously been shown to be appropriate choices of descriptors in the diagnosis of ovarian dysplasia.[3,4] For classification, we employed an artificial neural network. While no comprehensive comparative study of morphometric classifiers exists, we believe that the generality and adaptability of neural networks,[18,19] as well as the lack of assumptions they make as to the structure of the data, make them the most powerful classifiers available today. Thus, the inability of our neural network to differentiate between incidental and adjacent dysplasia further supports their morphologic similarity.

The histologic characteristics of ovarian dysplasia adjacent to carcinoma were the same as those seen in ovaries removed for reasons other than malignancy and thus considered incidental findings. It is interesting that four of this group of patients had prophylactic oophorectomies done because of a family history of ovarian carcinoma. This morphometric study demonstrates that both groups of dysplasia were different from the control group, and quite similar to each other. That analysis of both morphometric descriptors and their subsequent classification showed no difference between the two groups of dysplasia suggests that the associated changes in tissue morphology reflect a single underlying pathologic process, lending credence to the hypothesis that ovarian dysplasia seen in epithelial inclusion cysts may represent a pre-invasive malignant change. Indeed, it has been demonstrated that such changes are often noted in the contralateral

ovaries of patients diagnosed with unilateral ovarian carcinoma.[20]

This study confirmed the similarity of the dysplastic lesions seen adjacent to overt cancer and found incidentally at oophorectomy. The diagnosis of "incidental" dysplasia contributes to the identification of patients at risk to develop ovarian carcinoma, and even more so if there is also a positive family history.

Appendix

Artificial neural networks (ANNs) are a form of artificial intelligence.[18,21] A detailed review of their structures, operations, and applications to quantitative pathology can be found in Dytch and Wied.[19] As opposed to specifying rules for assigning values to output variables on the basis of input variables, an ANN is presented with a training set of data from which it effectively "learns" such rules. Neural networks provide the ability to generalize from a constantly growing training set, while making few assumptions as to the nature of the data. In our study, input variables were the 10 morphometric descriptors, and output variables coded for the gold-standard diagnosis by assigning a value of one to the output neuron corresponding to the gold-standard and a value of zero to the other output neurons.

The basic unit of structure for artificial neural networks, as in their biological counterparts, is the neuron. The type of ANN considered here has its neurons arranged in a multilayered architecture. There are layers of input and output neurons, in which each input and output variable has a neuron assigned to it, and additionally there are one or more hidden layers of neurons which enable the networks to capture complicated relationships between input and output variables. To compute the value of a neuron, the signed weighted sum of inputs from the previous layer is computed and then passed through a so-called "transfer function," most typically the sigmoidal function.

As illustrated in Figure 4, each neuron is connected to every neuron in its adjacent

layers, and a weight is assigned to the connection. By repeatedly modifying these weights to reflect the training set, the network is able to learn how to determine output variables based upon inputs. Initially, the weights in a neural network are set at random. Input variables from the training set are used to compute output variables. The difference between these computed output variables and their assigned values in the training set is calculated, and the weights in the network are adjusted to minimized this error, by a process called back-propagation. This process of computing output variables and modifying neuron weights is repeated until the error for each set of variables in the training set is sufficiently small. Then, using these final weights, a new set of input variables can be used to predict output variables.

A possible caveat which exists with the use of neural networks as morphometric classifiers relates to formal sample size considerations. In a recent review of neural network applications in pathology and laboratory medicine, Astion and Wilding,[22] based upon theoretical work by Baum and Haussler,[23] suggest that the total number of training facts should be at least 10 times the number of weights in the neural network. In light of the fact that for morphometric applications, each training fact is a feature vector reflecting perhaps 50 nuclear images, this requirement can be extremely unfeasible to attain. Mathematically, for a network with one layer of hidden neurons, if we let i be the number of input neurons, h be the number of hidden neurons, and o be the number of output neurons, then the total number of connection weights CW is given by the formula:

$$CW = (i+o)*h + h + o\ .$$

Thus according to Baum and Haussler's analysis, the 10-14-2 network used in our study would ideally require 1840 training facts. Even the largest gynecology service will not have records of thousands of cases of this recently-described lesion. This does not, however, preclude a morphometric characterization.

The problem raised by Baum and Haussler's work is not germane only to this study, but applies widely. For example, a recent study by Floyd *et al.*[21] used a neural

network to accurately predict breast cancer from mammographic findings. Their 8-16-1 network would theoretically have required a sample size of 1610 training facts while in actuality, 259 training facts were employed for the jackknife analysis. Before the advent of neural networks, the most popular classificatory algorithm was discriminant analysis, which has an underlying assumption that the input variables are normally distributed. However according to calculations we and others have done, this assumption typically fails to hold in practice. The attitude adopted by the community, however, was a pragmatic one, particularly in view of the possibility of using such schemes for the classification of unknowns. Ultimately, the classification algorithm chosen is the one which performs most effectively. In terms of neural networks, this philosophy is reflected in a recent paper[24] discussing optimization of neural network performance in the diagnosis of lymphocyte-rich effusions; this is done by modifying a number of factors including the network architecture. Despite problems raised by suggested sample size requirements, throughout the paper the emphasis is on making the most accurate network, not the most compact one.

Moreover, Baum and Haussler's analysis may not entirely apply to this study. Firstly, their theoretical analysis is for a training tolerance of 0.1, while we used a stricter tolerance of 0.05. for which the formula for CW need not be the same. Secondly, in contrast to the features in the example of Astion and Wilding (age, gender, and analytes), the morphometric features we used are by no means independent from each other. Nuclear area, perimeter, maximal chord length, and circularity are interdependent descriptors of the nuclear outline, which we perceive as a single variable most easily characterized by a few descriptors. The various texton-based texture descriptors used are similarly related. While it is necessary to use numerous descriptors to capture the salient morphologic features, this does not truly suggest that we are adding an equal number of degrees of freedom. The effects of non-independent descriptors and varying training tolerance on sample size recommendations for neural network generalizability is an issue that requires further theoretical analysis.

Acknowledgments

The authors thank Ms. Zhiyuan Liu, M.E.E. for technical assistance, Ms. Fortune Uy, R.N. for secretarial assistance, and Dr. John Mandeli for a helpful discussion of relevant statistical issues. A.J.E. received support from U.S. Public Health Service Grant 5-T32-GM 7280-16.

References

1. Gusberg SD, Deligdisch L. Ovarian dysplasia: a study of identical twins. *Cancer* 1984; **54:** 1-4.

2. Plaxe SC, Deligdisch L, Dottino P, Cohen CJ. Ovarian intraepithelial neoplasia (OIN) demonstrated in patients with stage I ovarian carcinoma. *Gynecol Oncol* 1990; **38:** 367-372.

3. Deligdisch L, Gil J. Characterization of ovarian dysplasia by interactive morphometry. *Cancer* 1989; **63:** 748-755.

4. Gil J, Deligdisch L. Interactive morphometric procedures and statistical analysis in the diagnosis of ovarian dysplasia and carcinoma. *Pathol Res Pract* 1989; **185:** 680-685.

5. Deligdisch L, Miranda C, Barba J, Gil J. Ovarian dysplasia: nuclear texture analysis. *Cancer* 1993; **72:** 3253-3257.

6. Gil J, Barba J. Principles of stereology: computerized applications in anatomic pathology. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994: 79-124.

7. Gil J, Ciurea D, Lin Z, Barba J: Semi-automatic nuclear texture analysis based on three gray level zones *Second International Congress Histology Laboratory* 1992.

8. Barba J, Gil J: Image analysis in statistical texture analysis *Eighth International Congress for Stereology* 1991.

9. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

10. Castleman KR. Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, 1979.

11. Daniel W. Applied Nonparametric Statistics. Boston: PWS-KENT Publishing Company, 1990.

12. Rosner B. Fundamentals of Biostatistics. Belmont, CA: Wadsworth Publishing Company, 1995.

13. Nouwen EJ, Hendrix PG, Eerdekens MW, DeBroe ME. Tumor markers in the human ovary and its neoplasms: a comparative immunohistochemical study. *Am J Pathol* 1987; **126:** 230-242.

14. Berchuk A, Boente MP, Bast Jr. RC. The use of tumor markers in the management of patients with gynecologic carcinomas. *Clin Obstet Gynecol* 1992; **35:** 45-54.

15. Scully RE. Ovary. In: Henson DE, Albores-Saavedra J, eds. Pathology of Incipient Neoplasia. Philadelphia: W.B. Saunders, 1993: 283-300.

16. Blaustein A, Kantius M, Kaganowicz A, Pervez N, Wells J. Inclusions in ovaries of females aged day 1-30 years. *Int J Gynecol Pathol* 1982; **1:** 145-153.

17. Blaustein A. Surface cells and inclusion cysts in fetal ovaries. *Gynecol Oncol* 1981; **12:** 222-233.

18. Lawrence J. Introduction to Neural Networks: Design, Theory, and Applications. Nevada City, CA: California Scientific Software, 1993.

19. Dytch HE, Wied GL. Artificial neural networks and their use in quantitative pathology. *Anal Quant Cytol Histol* 1990; **12:** 379-393.

20. Mittal KR, Zeleniuch-Jacquotte A, Cooper JL, Demopoulos RI. Contralateral ovary in unilateral ovarian carcinoma: A search for preneoplastic lesions. *Int J Gynecol Pathol* 1993; **12:** 59-63.

21. Floyd CE, Lo JY, Yun AJ, Sullivan DC, Kornguth PJ. Prediction of breast cancer malignancy using an artificial neural network. *Cancer* 1994; **74:** 2944-2948.

22. Astion ML, Wilding P. The application of backpropagation neural networks to problems in pathology and laboratory medicine. *Arch Pathol Lab Med* 1992; **116:** 995-1001.

23. Baum EB, Haussler D. What size net gives valid generalization? In: Touretzky DS, ed. Advances in Neural Information Processing Systems. San Mateo, CA: Morgan Kauffman Publishers, 1989: 81-90.

24. Truong H, Morimoto R, Walts AE, Erler B, Marchevsky A. Neural networks as an aid in the diagnosis of lymphocyte-rich effusions. *Anal Quant Cytol Histol* 1995; **17:** 48-54.

Table 1

Results of Kruskal-Wallis Test with Dunn Procedure for Specific Group Comparison

Descriptor	H	P	z_{12}		z_{23}		z_{13}	
Length	14.6273	0.00067	-3.5716	Sig.	0.1576	N.S.	-3.0563	Sig.
Area	15.7265	0.00038	-3.8798	Sig.	0.7821	N.S.	-2.7564	Sig.
Maximum Chord	13.3556	0.00126	-3.5689	Sig.	0.6904	N.S.	-2.5623	Sig.
Circularity Factor	0.9033	0.63658	-0.8001	N.S.	0.7878	N.S.	-0.0096	N.S.
Area of Dark Textons	2.5684	0.27688	-1.5935	N.S.	0.7878	N.S.	-0.7016	N.S.
Area of Gray Textons	11.2961	0.00352	-3.1534	Sig.	0.1805	N.S.	-2.6603	Sig.
Area of White Textons	12.7281	0.00172	-3.0100	Sig.	-0.5672	N.S.	-3.2216	Sig.
Avg. # of Dark Textons	13.4825	0.00118	-3.3397	Sig.	0.0344	N.S.	-2.9621	Sig.
Avg. # of Gray Textons	13.8601	0.00098	-3.2171	Sig.	-0.3910	N.S.	-3.2446	Sig.
Avg. # of White Textons	9.2162	0.00997	-2.2099	N.S.	-0.9310	N.S.	-2.8400	Sig.

H denotes Kruskal-Wallis statistic. z_{ij} denotes test statistic comparing categories i and j. Sig. denotes statistically significant ($\alpha = 0.05$). N.S. denotes not statistically significant.

Table 2

Neural Network Data

Neural Network 1: Control (1) vs. Incidental Dysplasia (2)

Specimen ID	*Pathologist's Diagnosis*	*Cat. 1 Weight*	*Cat. 2 Weight*	*Computer's Diagnosis*	*Correct/Incorrect*
94-367131	Control	0.996	0.002	Control	Correct
94-393072	Control	0.993	0.005	Control	Correct
94-48771	Control	0.998	0.001	Control	Correct
94-44128r	Incidental Dysplasia	0.009	0.995	Incidental Dysplasia	Correct
93-17858	Incidental Dysplasia	0.971	0.021	Control	Incorrect
94-334516	Incidental Dysplasia	0.005	0.995	Incidental Dysplasia	Correct
94-33325	Incidental Dysplasia	0.000	0.999	Incidental Dysplasia	Correct

Neural Network 2: Incidental Dysplasia (2) vs. Adjacent Dysplasia (3)

Specimen ID	*Pathologist's Diagnosis*	*Cat. 2 Weight*	*Cat. 3 Weight*	*Computer's Diagnosis*	*Correct/Incorrect*
94-44128r	Incidental Dysplasia	0.056	0.949	Adjacent Dysplasia	Incorrect
93-17858	Incidental Dysplasia	0.879	0.122	Incidental Dysplasia	Correct
94-334516	Incidental Dysplasia	0.390	0.558	Adjacent Dysplasia	Incorrect
94-33325	Incidental Dysplasia	0.967	0.027	Incidental Dysplasia	Correct
93-20951	Adjacent Dysplasia	0.147	0.858	Adjacent Dysplasia	Correct
94-359495	Adjacent Dysplasia	0.997	0.004	Incidental Dysplasia	Incorrect
94-324323	Adjacent Dysplasia	0.978	0.024	Incidental Dysplasia	Incorrect

Neural Network 3: Control (1) vs. Adjacent Dysplasia (3)

Specimen ID	*Pathologist's Diagnosis*	*Cat. 1 Weight*	*Cat. 3 Weight*	*Computer's Diagnosis*	*Correct/Incorrect*
94-367131	Control	0.998	0.001	Control	Correct
94-393072	Control	0.510	0.497	Control	Correct
94-48771	Control	0.999	0.001	Control	Correct
93-20951	Adjacent Dysplasia	0.118	0.869	Adjacent Dysplasia	Correct
94-359495	Adjacent Dysplasia	0.000	0.999	Adjacent Dysplasia	Correct
94-324323	Adjacent Dysplasia	0.277	0.692	Adjacent Dysplasia	Correct

Neural Network 4: Control (1) vs. Incidental Dysplasia and Adjacent Dysplasia (2 and 3)

Specimen ID	*Pathologist's Diagnosis*	*Cat. 1 Weight*	*Cat. 2 or 3 Weight*	*Computer's Diagnosis*	*Correct/Incorrect*
94-367131	Control	0.986	0.009	Control	Correct
94-393072	Control	0.762	0.264	Control	Correct
94-48771	Control	0.999	0.000	Control	Correct
94-44128r	Dysplasia	0.515	0.298	Control	Incorrect
93-17858	Dysplasia	0.397	0.610	Dysplasia	Correct
94-334516	Dysplasia	0.434	0.568	Dysplasia	Correct
94-33325	Dysplasia	0.002	0.999	Dysplasia	Correct
93-20951	Dysplasia	0.269	0.770	Dysplasia	Correct
94-359495	Dysplasia	0.000	0.999	Dysplasia	Correct
94-324323	Dysplasia	0.246	0.762	Dysplasia	Correct

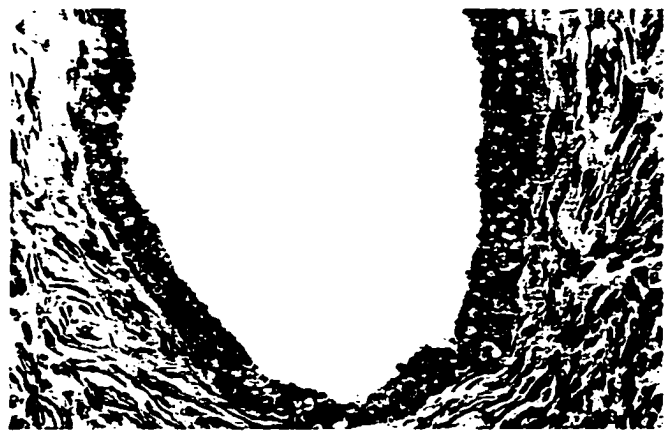
Neural Network 5: Control (1) vs. Incidental Dysplasia (2) vs. Adjacent Dysplasia (3)

Specimen ID	*Pathologist's Diagnosis*	*Cat. 1 Weight*	*Cat. 2 Weight*	*Cat. 3 Weight*	*Computer's Diagnosis*	*Correct/Incorrect*
94-367131	Control	0.996	0.000	0.038	Control	Correct
94-393072	Control	0.821	0.002	0.628	Control	Correct
94-48771	Control	0.999	0.000	0.005	Control	Correct
94-44128r	Incidental Dysplasia	0.520	0.000	0.003	Control	Incorrect
93-17858	Incidental Dysplasia	0.634	0.000	0.762	Adjacent Dysplasia	Incorrect
94-334516	Incidental Dysplasia	0.021	0.966	0.016	Incidental Dysplasia	Correct
94-33325	Incidental Dysplasia	0.002	0.954	0.002	Incidental Dysplasia	Correct
93-20951	Adjacent Dysplasia	0.582	0.004	0.928	Adjacent Dysplasia	Correct
94-359495	Adjacent Dysplasia	0.000	0.999	0.665	Incidental Dysplasia	Incorrect
94-324323	Adjacent Dysplasia	0.385	0.524	0.030	Incidental Dysplasia	Incorrect

Figure 1

Incidental Dysplasia

Patient with prophylactic oophorectomy for family history of ovarian carcinoma



Inclusion cyst shows piling up of epithelium and loss of polarity

H & E, original magnification × 100

Enlarged nuclei, prominent nucleoli, coarse chromatin

H & E, original magnification × 1000

Figure 2

Adjacent Dysplasia

Epithelial inclusion cyst (below) close to serous carcinoma of ovary (above)

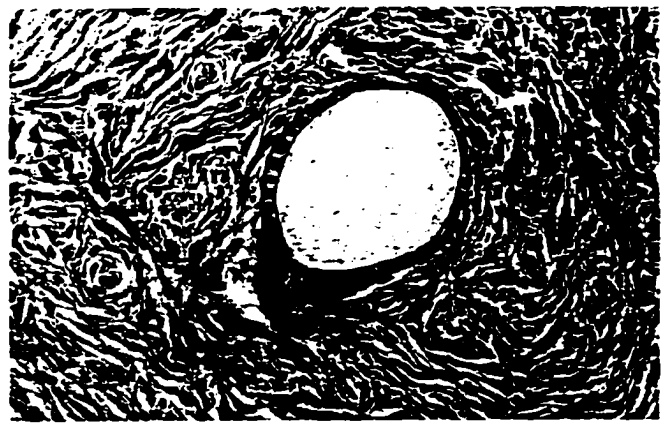
H & E, original magnification × 100

Note similarity to Figure 1

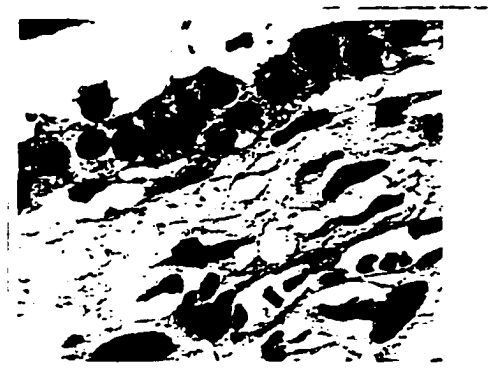
H & E, original magnification × 1000.

Figure 3

Epithelial Inclusion Cyst (Control)



H & E, original magnification × 100

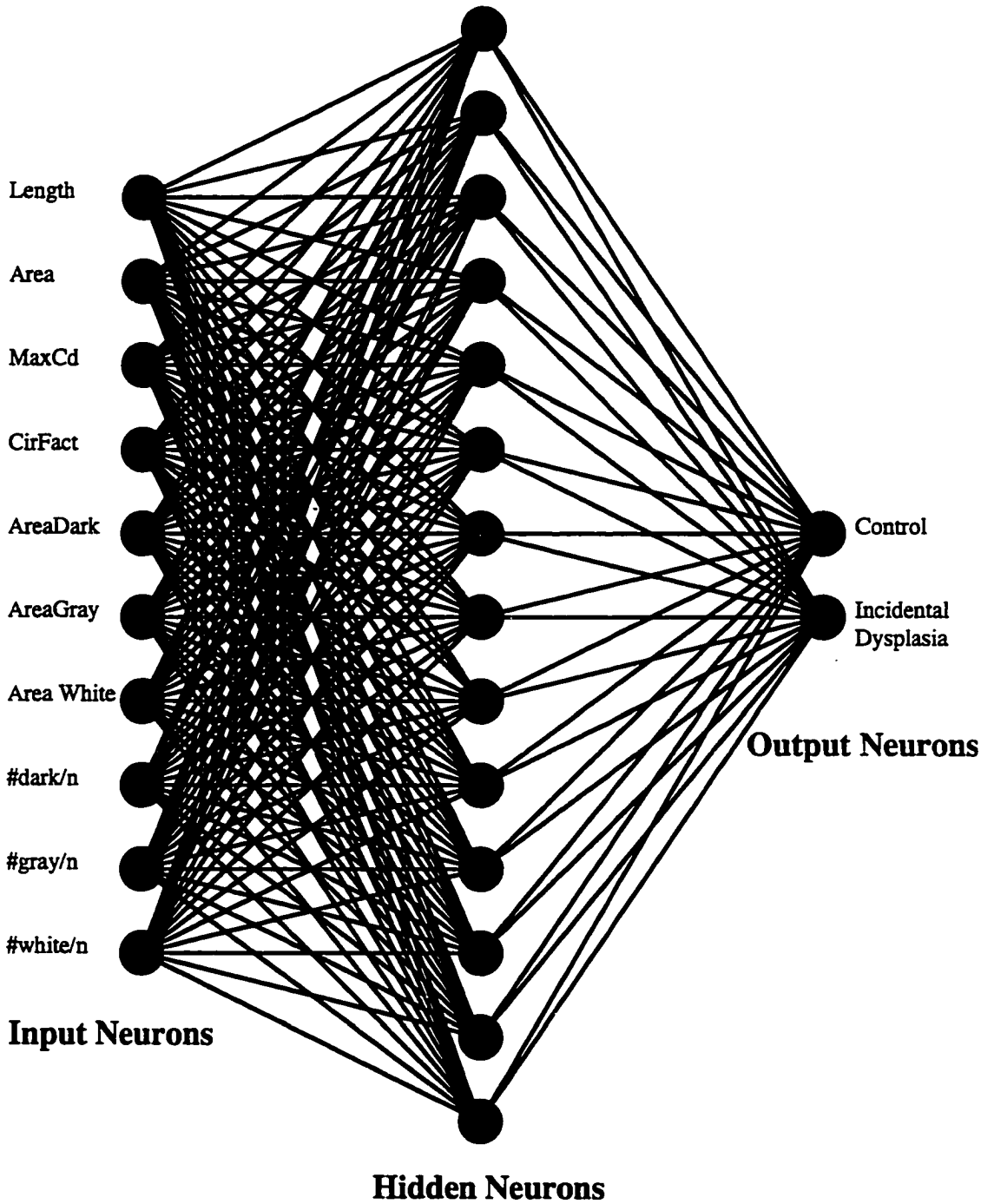


Benign nuclei are smaller, regular, with uniform appearing chromatin

H & E, original magnification × 1000

Figure 4

Neural Network Showing Difference between Control and Incidental Dysplasia



CHAPTER 4

Nuclear Diffuseness as a Measure of Texture: Definition and Application to the Computer-Assisted Diagnosis of Parathyroid Adenoma and Carcinoma

Andrew J. Einstein

Joseph Barba

Pamela D. Unger

Joan Gil

Journal of Microscopy 1994; 176: 158-166.

Preface to Chapter 4

In this chapter, we introduce the nuclear diffuseness, a measure of local gray level variation, to numerically characterize chromatin appearance. A disc is moved over a nuclear image, and at each point the variance of the gray levels of all pixels in the nucleus that are within the disc is computed. Nuclear diffuseness is the mean value of these variances, taken over all pixels in the nuclear image. We study the effects on diffuseness of changing the disc size, of adjusting for edge effects, and of changes in segmentation. Chromatin appearance in normal, adenomatous, and carcinomatous parathyroid biopsies is characterized by diffuseness. Using just diffuseness and nuclear area, and an artificial neural network for classification, all 16 parathyroid biopsies considered are correctly diagnosed.

The moving disc employed in the computation of nuclear diffuseness is a recurring theme in our texture description efforts. It provides a useful tool to measure local textural properties. Under different names, it will appear again in Chapters 5 through 7—in the determination of ε-oscillation used to calculate the Minkowski-Bouligand fractal dimension, and as the "gliding box" in calculations of lacunarity.

The contrast compensation method introduced in this chapter was developed by Dr. Joe Barba, and the section on contrast compensation, including Figure 1, was excerpted from a poster presented by Drs. Joan Gil and Joe Barba at the Eighth International Congress for Stereology. Dr. Pamela Unger provided materials and advice on parathyroid pathology. The second paragraph of the Introduction was inserted at the insistence of an editor, who provided much of the phraseology.

Summary

A measure of texture, the nuclear diffuseness, was formulated for use in biological classification, and specifically to quantitatively characterize chromatin texture. Nuclear diffuseness corresponds to the amount of local intensity variation in the digitized image of a nuclear profile. As a setting in which to test the efficacy of nuclear diffuseness as a diagnostic tool, we considered the identification of parathyroid adenoma and carcinoma. Digitized images of sections of parathyroid chief cell nuclei were obtained from 16 biopsies, and their nuclear diffusenesses, as well as other morphometric descriptors, were computed. With just the average nuclear diffuseness and average nuclear profile area, jackknife (leave-one-out) classification using an artificial neural network was able to correctly and unambiguously diagnose the condition (normal, parathyroid adenoma, or parathyroid carcinoma) in 15 of 16 cases. In one case, the neural network assigned a higher weight to the correct diagnosis, but was unable to conclusively distinguish between normal and adenoma.

Key Words: morphometry, image analysis, nuclear texture, nuclear diffuseness, parathyroid carcinoma, parathyroid adenoma, artificial neural networks

Introduction

Quantitation of nuclear texture has been applied to a number of biological settings as a tool in making difficult diagnostic and prognostic classifications, including the classification of lung carcinoma,[1] colonic adenocarcinoma,[2] hepatoma,[3] pancreatitis and pancreatic adenocarcinoma,[4] giant cell tumor of bone,[5] thyroid carcinoma,[6-8] breast carcinoma,[9,10] borderline ovarian tumor,[11] cervical cancer,[12,13] and malignant melanoma.[14,15] We have developed a measure of nuclear texture, called nuclear

diffuseness, as a tool to quantitatively characterize the texture of nuclear profiles in tissue sections. This texture is variable due to a number of reasons. Most notably, nuclear chromatin can be fine or coarse; in terms of the intensity matrix, fine chromatin should manifest significant local variation, whereas clumping should yield more homogeneity within regions of nuclear matrix. Moreover, nuclei may have within them nucleoli, a nuclear rim, and/or nuclear "voids," all of which effect the distribution of light and dark areas in the digitized gray scale image. The nuclear diffuseness is an attempt to quantify, and thereby objectify, a determination of local intensity variation within the image of a nucleus.

The images we study are nuclear profiles, which arise because they have been hit by a sectional plane — and not nuclei, which are three-dimensional objects that are not observed. Such a plane selects the nuclei with probability proportional to the tangent diameter parallel to the normal to the section. The resulting selection of nuclei is therefore biased, as the largest nuclei have the highest chance for inclusion in the sample. Thereafter, the nuclear profiles which had been sampled with unequal probabilities from the population, are selected interactively as described below.

As a setting in which to test the efficacy of nuclear diffuseness as a diagnostic tool, we considered the particularly difficult identification of parathyroid neoplasia. A tumor of the parathyroid glands may take the benign form of adenoma, or the malignant form of carcinoma.[16] For parathyroid carcinoma, radical resection of the gland and surrounding tissues at the initial operation is the definitively preferred treatment. However, carcinomas lacking the classic macroscopic features of a thick fibrous capsule and local invasion may be misinterpreted as nonmalignant. Lacking these gross morphological signs, the diagnosis of carcinoma is dependent solely upon the histopathologic examination.[17]

For a number of reasons, this diagnosis is an especially relevant setting in which to apply morphometry in addition to conventional techniques:

1) It has precedence in the literature. Jacobi *et al.*,[18] without proposing a classificatory approach, demonstrated higher mean diameter of nuclear profiles in 18 parathyroid carcinomas than in 55 adenomas.
2) In many cases the diagnosis of carcinoma is both difficult[19] and controversial. The most recent review of parathyroid carcinoma by Bondeson *et al.*[20] takes issue with an earlier landmark study by Schantz & Castleman,[21] which had ranked the presence of mitoses as the single most useful criterion of malignancy. Bondeson *et al.* cite Snover & Foucar's 1981 study[22] which identifies mitoses with nonmalignant cases; they suggest that different inclusion criteria could account for this discrepancy. Certainly, objective criteria would be beneficial.
3) Parathyroid carcinoma is exceedingly rare. Blichert-Toft *et al.*,[23] in a Danish-language publication, report a prevalence of primary hyperparathyroidism (pHPT) of 0.4-1.0% in the population, while Christensson *et al.*[24] note that their value for prevalence, evaluated at 0.36%, is higher than those reported in a handful of studies dating back to the late 1960s and early 1970s. Of patients in Western countries with primary hyperparathyroidism, carcinoma has been variously reported as the cause in 0.34%[25] to approximately 5.3% of cases.[26] Shane & Bilezikian[27] reported only 62 cases of carcinoma in a comprehensive twelve-year review of the English-language literature, while Obara & Fujimoto[28] note that only 271 cases of functioning parathyroid carcinoma have been reported since 1933. (This covers only the English language literature for the period 1968-1989; two of the cases included from the 1969 review of Holmes *et al.*[29] were published in French journals.) It is quite possible that a pathologist called upon to diagnose a case of parathyroid carcinoma will have never seen one previously.

Definition of Nuclear Diffuseness

Nuclear diffuseness is defined as an attempt to mathematically describe an observation about chromatin — namely, that fine chromatin displays significant local variation in appearance, while clumping exhibits more homogeneity within small regions of nuclear matrix. To compute the diffuseness of a nuclear profile, begin by computing for each pixel in the profile, the variance of the intensities of all pixels in the profile within a prescribed radius r_{max} of the given pixel. The average of these variances over all pixels in the domain is called the nuclear diffuseness. The average nuclear diffuseness for a specimen is defined as the average of the diffuseness over all nuclei collected.

More rigourously, let $\boldsymbol{M}_2$ be the space of two-dimensional matrices with real entries. Consider a pathology specimen and the set $N = \{N_1, N_2, ..., N_k\}$ of digitized images of nuclei collected from that specimen, $N_x \in \boldsymbol{M}_2$ for all x. Let $N_x(i, j)$ denote the (i, j) entry of N_x.

Now define the function $v:\boldsymbol{M}_2 \rightarrow \boldsymbol{M}_2$ by

$$v(N_x(i, j)) = \underset{N_x}{var} \{N_x(i', j') \neq 0 \mid \sqrt{(i' - i)^2 + (j' - j)^2} \leq r_{max}\}$$

for all (i, j) such that $N_x(i, j) \neq 0$. Let $v(N_x(i, j))$ be 0 everywhere else.

Define the diffuseness function $d:\boldsymbol{M}_2 \rightarrow \Re$ by

$$d(N_x) = \underset{N_x}{mean} \{v(N_x(i, j)) \neq 0\}$$

Then the average diffuseness D for the set of nuclear images N is simply

$$D(N) = \underset{N}{mean} \{d(N_x)\} = (\sum_{N} d(N_x)) / k$$

Materials & Methods

Sixteen parathyroid biopsies were selected from the files of the Department of Pathology of The Mount Sinai Hospital. These cases had been previously diagnosed by a pathologist on the basis of established histological criteria, and were assessed to be non-controversial diagnoses. Five cases were identified as carcinoma, five were identified as adenoma, and six were identified as normal parathyroid tissue. Within each group of biopsies, the cases were chosen at random. Each carcinoma demonstrated fibrous bands, mitoses, local extension, and/or metastases. Of the six normal samples, five were from patients with no demonstrable parathyroid pathology and one was normal tissue from a patient with parathyroid adenoma. The specimens were prepared by a procedure common in commercial processing laboratories in the United States. After grossing, the specimens were immersed in 10% phosphate buffered formaldehyde and placed inside an automatic tissue processor for one additional hour of fixation with 10% formaldehyde. This was followed by a second hour of fixation in the automatic tissue processor with 63% ethanol - 10% phosphate buffered formaldehyde, and then by treatment with the usual ascending ethanol chain, xylene, and paraffin embedding. Sectioning was at a nominal thickness of 5 μm. H&E staining was done in an automatic instrument, with 30 seconds of eosin staining and 7 minutes of hematoxylin staining. This system is routinely used for the majority of specimens at Mount Sinai, and in our experience shows reproducible results.

Two image analysis systems were used in our study, one for data acquisition and the other for nuclear profile selection, gray level normalization, and descriptor computation. Data was collected using an image analysis system built around an expansion box containing an Intel-310/17 multibus microcomputer based on the 16 bit 80286 microprocessor with an Intel 80287 math coprocessor, Imaging Technology AP-512 analog processor, and Imaging Technology FB-512 frame buffers. Images were generated by a Sony DXC-M2 camera attached to a Nikon Optiphot microscope equipped with a

Nikon Plan objective with a numerical aperture of 1.25. The interactive peripheral used to manually select nuclei and their surrounding cytoplasm was a touch-sensitive screen. These images containing nuclei were selected by consecutively tracing around all chief cell nuclear profiles in areas identified as parathyroid tissue, discarding incomplete images and images lacking contrast, until at least fifty nuclei were collected. The images were then transferred to the second image analysis system for further analysis.

The second system used was a Silicon Graphics Indigo^2 and Personal IRIS equipped with custom-written image analysis software. Nuclear profiles were segmented using the system's mouse-based pixel editor. The editor takes an image containing a nuclear profile and enables the user to trace out its border, specifying on a pixel by pixel basis which points are to be included.

Each image acquired is adjusted to normalize contrast for small variations in lighting conditions and staining intensity, using a method developed by Barba & Gil.[30] Camera characteristics are provided by manufacturers and can be generally described by a fractional power transfer function relating camera relative output voltage to the logarithm of relative input luminance.[31] The camera relative output voltage is measured by using the signal level corresponding to apparent black as the zero reference (as indicated by the horizontal drive portion of the video signal) and the voltage level corresponding to maximum brightness as 100 percent. The measurements of luminance are made on a relative basis by arbitrarily specifying the maximum brightness that may be reproduced by the camera as 100 percent luminance. Fig. 1 shows the saticon camera characteristics used in our system, which are obtained by interpolating a set of ten coordinates obtained from the manufacturer's data.

Image contrast models are based on plots of the Weber ratio ($\Delta L/L$) which is obtained from experiments to determine just-noticeable differences in luminance (ΔL) between a small region and its surroundings (L) for varying background luminances.[32] The reciprocal of the Weber ratio may be interpreted as the incremental increase in distinguishable levels for input luminance level. For the luminance range of interest the

Weber ratio is generally assumed constant at 2% leading to the conclusion that 50 levels are needed for the contrast on a scale of 0 to 1. More importantly, this result indicates that equal increments in the logarithm of the luminance should be perceived to be equal changes in contrast. Thus, identical images acquired under different lighting conditions will not be perceived as having similar contrast. If the camera characteristics were truly linear then identical images acquired under different lighting conditions could be transformed to have similar contrast by uniformly shifting the image gray levels so all image histograms have identical mean value. Due to the non-linear characteristics we normalize images with respect to contrast by transforming their pixel gray levels via a fractional power law corresponding to the camera in use and linearly shifting all resulting histograms so they have the same mean.

The input-output transfer function of our saticon camera can be approximated by

$$b = [\,(g - 0.0526) / 0.95\,]^{0.4545}$$

where g and b represent the pixel gray level and transform variable, respectively; both variables are normalized with respect to 255, the number of levels that can be represented by 8 bit pixels. The transformation in some sense approximates the input relative luminance.

As indicated previously, for each patient, at least fifty profiles were collected. Examples of nuclear images from normal, adenomatous, and carcinomatous biopsies are shown in Fig. 2. A number of descriptors were computed for each profile, and these were averaged for each patient. This information was subsequently transferred to BrainMaker, an artificial neural network package from California Scientific Software, for classification.

Results & Discussion

In practice, the problem of how to make a classification based upon morphometry is twofold: 1) *Which descriptors should be computed?* and 2) *How should they be*

combined to make a diagnosis?. Previous investigators have used measures of nuclear profile size, such as average area, average perimeter, and standard deviation of area, as well as a host of texture descriptors, most notably those of Haralick *et al.*[33] The generality of the Haralick descriptors — their applicability to settings as disparate as geology and pathology — is appealing, but the introduction of 28 parameters derived from covariance matrices obfuscates the second part of the diagnosis. Whether classification is done by the more conventional multivariate analysis, or by neural network analysis as we choose, such a large descriptor set is unwieldy. Our approach is to use fewer parameters. Since all descriptors are ultimately derived from the same adjusted intensity matrix, they are not truly independent of each other, and thus using more descriptors does not necessarily provide more usable classificatory information. The two descriptors upon which we choose to make our classification below encapsulate different types of information about the intensity matrix. Perhaps the reason why texture quantitation remains outside the realm of clinical decision-making is because texture has been generally employed with too many parameters and without a readily usable method of combining their values to reach a diagnosis. The former is provided by well-chosen descriptors, and the latter by a trained neural network.

We have seen in previous studies using descriptors of nuclear profiles,[34,35] of argyrophilic nucleolar organizer regions (AgNORs),[36,37] and of nuclear textons,[38] that any single descriptor is insufficient to make a diagnosis. Hence we are faced with the problem of applying classificatory techniques to a group of parameters — in particular, parameters describing textural and other morphometric data. In selecting a textural descriptor, a number of variations of the definition of nuclear diffuseness were computed. The nuclear diffuseness was calculated for 16 values of r_{max} ranging between one and 25. Additionally, a modification of the definition of diffuseness was considered which adjusts for possible edge effects, weighting each pixel by the number of pixels of the nuclear profile (*i.e.*, of non-zero intensity) found within the neighborhood of radius r_{max}; thus pixels at the perimeter of the profile are weighted less than those near the center. This too

was considered for several values of r_{max}. Furthermore, we also considered the average variance of intensity values, a global measure of diffuseness which is equivalent to the value of average diffuseness as r_{max} increases to the length of the longest chord traversing any profile. Figure 3 is a scatter plot of average area versus average diffuseness for four definitions of nuclear diffuseness. For all formulations, high average diffuseness is characteristic of normal cases, high average nuclear profile area of carcinomas, and low diffuseness and area of adenomas. It is seen that the relative values of the diffuseness descriptors change little between the various formulations. The clear separation between the three groups buttresses our conclusion that a narrow differential diagnosis can be safely achieved by the use of physical descriptors. While below we shall justify the choice of $r_{max} = 4$ for our particular data set, in a sense this is "fine-tuning," since there is no statistically significant difference in the diffuseness between the $r_{max} = 4$ unweighted formulation and other definitions. This was determined by performing Wilcoxon signed rank tests comparing the $r_{max} = 4$ unweighted diffusenesses with the other definitions, which were rescaled to account for differences in means between the various formulations.

A possible concern about any texture descriptor relates to its dependence upon the method of segmentation used to capture the nuclear profile images. Since manual segmentation, as well as a number of different algorithms for automated nuclear segmentation, are presently in use, a descriptor of texture should not be overly sensitive to the inclusion or exclusion of several peripheral pixels. To address this concern, we collected images of several nuclei with their surrounding cytoplasm from patients representing each of the diagnostic outcomes. The slides used were the same ones as described in Materials & Methods. Each nuclear profile was segmented three times: within the nuclear membrane, at the nuclear membrane, and outside the membrane, such that the "nuclear" image in fact contained a rim of cytoplasm. The nuclear diffuseness of each image was computed for the three segmentations. These results are displayed in Table 1 for $r_{max} = 4$; similar results were observed for other values of r_{max}. As can be seen, the

diffuseness values for segmentation inside and at the nuclear membrane are similar. A Wilcoxon signed rank test was performed to demonstrate that the difference in diffuseness between the two segmentations is statistically insignificant ($p = 0.349$). However, segmentation including a peripheral rim of cytoplasm surrounding the selected profile greatly increases the diffuseness; the difference in diffuseness between this segmentation and either of the other two is very highly significant. This result is entirely predictable from the definition of diffuseness. Since the lighter stained cytoplasm has intensity values much larger than those of the nucleus, the juxtaposition of a rim of several hundred pixels will result in much larger values of $v(N_x(i, j))$ for peripheral pixels, and consequently a much larger $d(N_x)$. In practice, however, neither effective nuclear segmentation algorithms nor carefully performed manual nuclear segmentation should include substantial amounts of cytoplasm. Review of our nuclear images confirmed the virtual absence of peripheral pixels with intensity values considerably different from those of their neighbors, and thus we feel confident in the diffuseness values computed.

To classify biopsies on the basis of computed values of diffuseness and area, we employed BrainMaker, an artificial neural network package from California Scientific Software. Neural networks are a form of artificial intelligence, wherein a computer can be trained to learn from a set of data and make generalizations. An explanation of the principles and jargon of artificial neural networks is beyond the scope of this paper. The reader unfamiliar with neural networks is referred to Nilsson's classic monograph,[39] and to Dytch & Wied's review of their structures, operations, and applications to quantitative pathology.[40]

BrainMaker was used to generate neural networks with two input neurons, one layer of ten hidden neurons, three output neurons, and a sigmoidal transfer function. The two input neurons correspond to average nuclear diffuseness and average nuclear profile area, while the output neurons represent the three diagnostic outcomes, *viz.* normal, adenoma, and carcinoma. The analysis was repeated for various formulations of the

average nuclear diffuseness. Input was limited to the two descriptors so as to test the diagnostic capabilities of morphometric data; certainly any system used in clinical practice would benefit from the use of a broad range of parameters from which to learn and to use in classification. The network was trained by back propagation using the delta rule with gradient descent, with a training tolerance of 5%. Training was performed 16 times. Each training run was performed on all of the cases less one, and the network was requested to make a diagnosis on the excluded case. Each of the cases was taken in turn to be the excluded case; the outputs of the network for the unweighted $r_{max} = 4$ diffuseness, with outputs between zero and one, are contained in Table 2.

If we associate a diagnosis with the condition assigned the highest output, the computer was able to correctly identify the parathyroid's status in all 16 cases. In 14 of the cases, the neural net assigned a output of greater than 0.94 to the diagnosis. In one case, the diagnosis was associated with an output of 0.873, but the other two diagnoses had very small outputs. In case A91-233, identified as a normal gland, the highest output was associated with a normal diagnosis. However, this was given a value of only 0.363 of a possible 1, and since the output for adenoma was 0.203, the result is inconclusive. In practice, this would signify an ambiguous diagnosis, albeit between benign conditions.

In fact, for values of r_{max} greater than 6, the larger output in A91-233 is associated with adenoma. While for our data set, smaller values of r_{max} do better at diagnosing case A91-233, they have the disadvantage of assigning relatively larger outputs to incorrect diagnoses. For example, when $r_{max} = 1$, the output for the incorrect diagnosis of normal in case 90-17588 is 0.880, while the output for the correct diagnosis of adenoma in case 90-15699 is only 0.759. In contrast, when $r_{max} = 10$, all the correct diagnoses have outputs of at least 0.95. The value of $r_{max} = 4$ was chosen because it both assigns the largest output to normal for A91-233, and also has lesser outputs for the incorrect diagnoses than do formulations with smaller radii. Weighting the diffuseness to adjust for possible edge effects did not improve the outputs, and in fact the neural network using edge adjusted

$r_{max} = 4$ diffuseness misclassified case A91-233. The important result here is not which formulation of diffuseness yields the preferred neural network outputs for our particular set of data, but rather, that the neural network achieved good success in classifying patients based upon the descriptors, and that correct classification is not overly dependent on the choice of r_{max}.

Conclusions

All cases studied were previously diagnosed by a pathologist on the basis of established criteria. However, our study demonstrates that they also could have been separated from each other based upon morphometric descriptors, with classification done by an artificial neural network. In particular, we introduced the nuclear diffuseness and showed that, for our collection of parathyroid biopsies, it contains the necessary textural information to complement average nuclear profile area in making a morphometric classification.

A clinical neural network system for the diagnosis of parathyroid adenoma and carcinoma would benefit from employing more parameters than our experimental one, which was deliberately limited to morphometric information. Any such system, of course, would still require use by a trained physician. In particular, it must be assured that the nuclei selected are from a representative section of tissue. Appropriate factors to include[17] would be serum calcium, immunoassayable parathyroid hormone (iPTH), alkaline phosphatase[27] and chorionic gonadotropin (α- and β-hCG) levels,[41] as well as indicators for the presence of bone disease, a palpable neck mass, renal involvement, and pancreatitis, and results of a ploidy study. Moreover, as the average nuclear diffuseness and average nuclear profile area correspond to physical properties of tissue, they can be used not only in computerized diagnostic systems, but also as added laboratory tests for pathologists and clinicians in diagnosis and decision making.

Acknowledgments

The authors gratefully acknowledge the computer assistance of Zhiyuan "Theresa" Liu, Ragu Nagalingam, Kevin Kelliher, and Janne Ravantti. We appreciate helpful comments from Drs. Torsten Mattfeldt and Karsten Nielsen, and from an anonymous reviewer. We thank Drs. Carol Bodian, Tai-Shing Lau, and Sylvan Wallenstein for discussion of pertinent statistical issues. A. J. E. received support from U. S. Public Health Service Grant 5-T32-GM 7280-16.

References

1. Thunnissen FB, Diegenbach PC, van Hattum AH, *et al.* Further evaluation of quantitative nuclear image features for classification of lung carcinomas. *Pathol Res Pract* 1992; **188:** 531-535.

2. Bibbo M, Michelassi F, Bartels PH, *et al.* Karyometric marker features in normal-appearing glands adjacent to human colonic adenocarcinoma. *Cancer Res* 1990; **50:** 147-151.

3. Jagoe R, Sowter C, Slavin G. Shape and texture analysis of liver cell nuclei in hepatomas by computer aided microscopy. *J Clin Pathol* 1984; **37:** 755-762.

4. Rickaert F, Gelin M, van Gansbeke D, *et al.* Computerized morphonuclear characteristics and DNA content of adenocarcinoma of the pancreas, chronic pancreatitis, and normal tissues: relationship with histopathologic grading. *Hum Pathol* 1992; **23:** 1210-1215.

5. Kuwahara H, Shimazaki M, Morikita I, Chanoki Y, Sakurai M. Texture analysis of histological images of giant cell tumor of bone. *Path Res Pract* 1992; **188:** 565-569.

6. Ferrer-Roca O, Ballester-Guardia E, Martin J. Nuclear chromatin texture to differentiate follicular and papillary carcinoma of the thyroid. *Path Res Pract* 1989; **185:** 561-566.

7. Galera-Davidson H, González-Cámpora R, Mora-Marín JA, *et al.* Cytophotometric DNA measurements in medullary thyroid carcinoma. *Cancer* 1990; **65:** 2255-2260.

8. Salmon I, Kiss R, Franc B, *et al.* Comparison of morphonuclear features in normal, benign and neoplastic thyroid tissue by digital cell image analysis. *Anal Quant Cytol Histol* 1992; **14:** 47-54.

9. Dawson AE, Austin RE, Weinberg DS. Nuclear grading of breast carcinoma by image analysis: classification by multivariate and neural network analysis. *Am J Clin Pathol* 1991; **95:** S29-S37.

10. Spina D, Disanto A, Luzi P, *et al.* Novel, contrast gradient-oriented, automated chromatin texture analysis: I. Feasibility study on nuclei from benign and malignant breast epithelial cell lines in fine needle aspirates. *Virchows Archiv B Cell Pathol* 1992; **62:** 119-124.

11. Drescher CW, Flint A, Hopkins MP, Roberts JA. Prognostic significance of DNA content and nuclear morphology in borderline ovarian tumors. *Gynecol Oncol* 1993; **48:** 242-246.

12. Hanselaar AGJM, Vooijs GP, Van't Hof-Grootenboer AE, Pahlplatz MMM. Cytophotometric analysis of cervical intraepithelial neoplasia grade III, with and without synchronous invasive squamous cell carcinoma. *Cytometry* 1990; **11:** 901-906.

13. Hutchinson ML, Isenstein LM, Martin JJ, Zahniser DJ. Measurement of subvisual changes in cervical squamous metaplastic cells for detecting abnormality. *Anal Quant Cytol Histol* 1992; **14:** 330-334.

14. Fleming MG, Friedman RJ. Multiparametric image cytometry of nevi and melanomas. *Am J Dermatopathol* 1993; **15:** 106-113.

15. Stolz W, Abmayr W, Schmoeckel C, Landthaler M, Massoudy P, Braun-Falco O. Ultrastructural discrimination between malignant melanomas and benign nevocytic nevi using high-resolution image and multivariate analyses. *J Invest Dermatol* 1991; **97:** 903-910.

16. DeLellis R. Atlas of Tumor Pathology: Tumors of the Parathyroid Gland, 3rd series. Washington: Armed Forces Institute of Pathology, 1993.

17. Fujimoto Y, Obara T. How to recognize and treat parathyroid carcinoma. *Surg Clin North Am* 1987; **67:** 343-357.

18. Jacobi JM, Lloyd HM, Smith JF. Nuclear diameter in parathyroid carcinomas. *J Clin Pathol* 1986; **39:** 1353-1354.

19. McKeown PP, McGarity WC, Sewell CW. Carcinoma of the parathyroid gland: is it overdiagnosed? A report of three cases. *Am J Surg* 1984; **147:** 292-298.

20. Bondeson L, Sandelin K, Grimelius L. Histopathological variables and DNA cytometry in parathyroid carcinoma. *Am J Surg Pathol* 1993; **17:** 820-829.

21. Schantz A, Castleman B. Parathyroid carcinoma: a study of 70 cases. *Cancer* 1973; **31:** 600-605.

22. Snover DC, Foucar K. Mitotic activity in benign parathyroid disease. *Am J Clin Pathol* 1981; **75:** 345-347.

23. Blichert-Toft M, Mollerup CL, Feldt-Rasmussen UF, Daugaard H, Engel UH. Primary hyperparathyroidism. An underdiagnosed disease in Denmark? *Ugeskr Læg* 1993; **155:** 765-769.

24. Christensson T, Hellström K, Wengle B, Alveryd A, Wikland B. Prevalence of hypercalcaemia in a health screening in Stockholm. *Acta Med Scand* 1976; **200:** 131-137.

25. McCance DR, Kenny BD, Sloan JM, Russell CFJ, Hadden DR. Parathyroid carcinoma: a review. *J R Soc Med* 1987; **80:** 505-509.

26. Jarman WT, Myers RT, Marshall RB. Carcinoma of the parathyroid. *Arch Surg* 1978; **113:** 123-125.

27. Shane E, Bilezikian JP. Parathyroid carcinoma: a review of 62 patients. *Endocrine Reviews* 1982; **3:** 218-226.

28. Obara T, Fujimoto Y. Diagnosis and treatment of patients with parathyroid carcinoma: an update and review. *World J Surg* 1991; **15:** 738-744.

29. Holmes EC, Morton DL, Ketcham AS. Parathyroid carcinoma: a collective review. *Ann Surg* 1969; **169:** 631-640.

30. Barba J, Gil J: Image analysis in statistical texture analysis *Eighth International Congress for Stereology* 1991.

31. Grob B. Basic Television Principles and Servicing. New York: McGraw-Hill, 1964.

32. Graham CH. Visual space perception Vision and Visual Perception. New York: John Wiley, 1965: 504-547.

33. Haralick RM, Shanumgam K, Dinstein I. Textural features for image classification. *IEEE Trans Systems Man Cybernet* 1973; **3:** 610-621.

34. Unger PD, Watson CW, Liu Z, Gil J. Morphometric analysis of neoplastic renal aspirates and benign renal tissue. *Anal Quant Cytol Histol* 1993; **15:** 61-66.

35. Deligdisch L, Kerner H, Cohen CJ, Dargent D, Gil J. Morphometric differentiation between responsive tumor cells and mesothelial hyperplasia in second-look operations for ovarian cancer. *Hum Pathol* 1993; **24:** 143-147.

36. Hytiroglou P, Harpaz N, Heller DS, Liu ZY, Deligdisch L, Gil J. Differential diagnosis of borderline and invasive serous cystadenocarcinomas of the ovary by computerized interactive morphometric analysis of nuclear features. *Cancer* 1992; **69:** 988-992.

37. Kalir T, Chan KS, Liu Z, Strauchen J, Gil J. Semi-automatic quantitation of nucleolar organizer regions in non-Hodgkin's lymphomas. *Pathol Res Pract* 1994; **190:** 124-128.

38. Deligdisch L, Miranda C, Barba J, Gil J. Ovarian dysplasia: nuclear texture analysis. *Cancer* 1993; **72:** 3253-3257.

39. Nilsson NJ. The Mathematical Foundations of Learning Machines. San Mateo, CA: Morgan Kaufmann Publishers, 1990.

40. Dytch HE, Wied GL. Artificial neural networks and their use in quantitative pathology. *Anal Quant Cytol Histol* 1990; **12:** 379-393.

41. Stock JL, Weintraub BD, Rosen SW, Aurbach GD, Spiegel AM, Marx SJ. Human chorionic gonadotropin subunit measurement in primary hyperparathyroidism. *J Clin Endocrinol Metab* 1982; **54:** 57-63.

Table 1

Nuclear Diffuseness as a Function of Location of Segmentation with Respect to the Nuclear Membrane

Diagnosis	Specimen	Nucleus ID	Inside Membrane	At Membrane	Outside Membrane
Normal *	A91-196	2	92.66	111.04	154.28
Normal	A91-196	3	52.46	55.14	188.12
Normal	A91-196	5	63.04	79.71	193.70
Normal	A91-132	3	158.81	143.72	264.82
Normal	A91-132	4	50.47	54.44	119.93
Normal	A91-132	6	63.45	59.25	176.76
Adenoma	90-18593	2	140.07	136.35	261.47
Adenoma	90-18593	3	131.27	114.17	178.32
Adenoma	90-18593	6	194.85	193.20	223.47
Adenoma	90-792	2	158.95	141.93	174.30
Adenoma	90-792	3	115.91	107.18	165.87
Adenoma	90-792	4	79.67	78.45	143.50
Carcinoma	90-10885	1	162.93	151.97	232.05
Carcinoma	90-10885	2	104.35	102.33	176.77
Carcinoma	90-10885	4	146.36	148.61	206.72
Carcinoma	91-22010	1	206.27	204.90	309.53
Carcinoma	91-22010	3	167.95	167.85	271.46
Carcinoma	91-22010	5	191.47	194.76	265.89
		Mean	**126.72**	**124.72**	**205.94**

* Normal Tissue in Patient with Parathyroid Adenoma

Figure 2

Overview of Results of Neural Network Diagnosis Based Upon Average Adjusted Nuclear Diffuseness and Average Nuclear Area

Case	Pathologist's Diagnosis	Avg. Diffuseness (rmax = 4)	Avg. Area (μm^2)	Normal Weight	Adenoma Weight	Carcinoma Weight	Neural Net's Diagnosis	Result
90-16719B	Normal *	129.84	18.10	0.999	0.001	0.003	Normal	Correct
JA-91-25	Normal	91.62	19.09	0.999	0.003	0.021	Normal	Correct
A91-199	Normal	110.6	16.61	0.999	0.002	0.004	Normal	Correct
A91-196	Normal	96.77	18.36	0.999	0.002	0.004	Normal	Correct
A91-132	Normal	173.21	15.49	0.999	0.001	0.001	Normal	Correct
A91-233	Normal	65.71	21.37	0.363	0.203	0.039	Normal ?	Correct
	Mean	**111.29**	**18.17**					
	SD	**37.02**	**2.04**					
90-17588	Adenoma	35.59	9.42	0.717	0.999	0.000	Adenoma	Correct
90-15699	Adenoma	27.25	21.76	0.016	0.873	0.032	Adenoma	Correct
90-1454	Adenoma	31.73	18.77	0.082	0.977	0.004	Adenoma	Correct
90-18593	Adenoma	15.94	15.38	0.005	0.999	0.000	Adenoma	Correct
90-792	Adenoma	20.61	18.40	0.007	0.999	0.002	Adenoma	Correct
	Mean	**26.22**	**16.75**					
	SD	**8.01**	**4.68**					
90-10885	Carcinoma	25.64	38.42	0.000	0.012	0.999	Carcinoma	Correct
91-22010	Carcinoma	46.73	43.50	0.000	0.001	0.999	Carcinoma	Correct
91-10223	Carcinoma	81.86	35.98	0.250	0.000	0.959	Carcinoma	Correct
87-27676	Carcinoma	92.7	42.75	0.009	0.000	0.999	Carcinoma	Correct
78-10898	Carcinoma	29.22	33.27	0.000	0.063	0.947	Carcinoma	Correct
	Mean	**55.23**	**38.78**					
	SD	**30.57**	**4.37**					

* Normal Tissue in Patient with Parathyroid Adenoma

Figure 1

Camera Characteristics: Relative Luminance *vs.* Relative Output Voltage

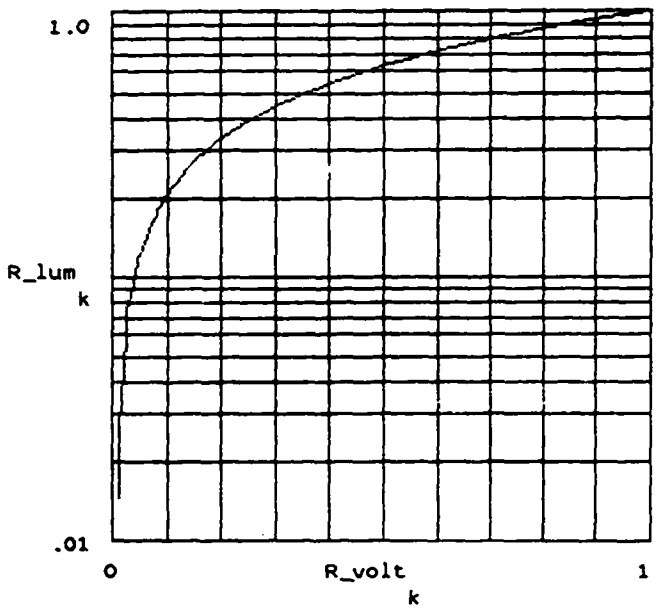


Figure 2

Digitized Images of Selected Parathyroid Chief Cell Nuclei

The nuclear profiles in the first row are from normal cases, those in the second row are from adenoma cases, and those in the third row are from carcinoma cases.

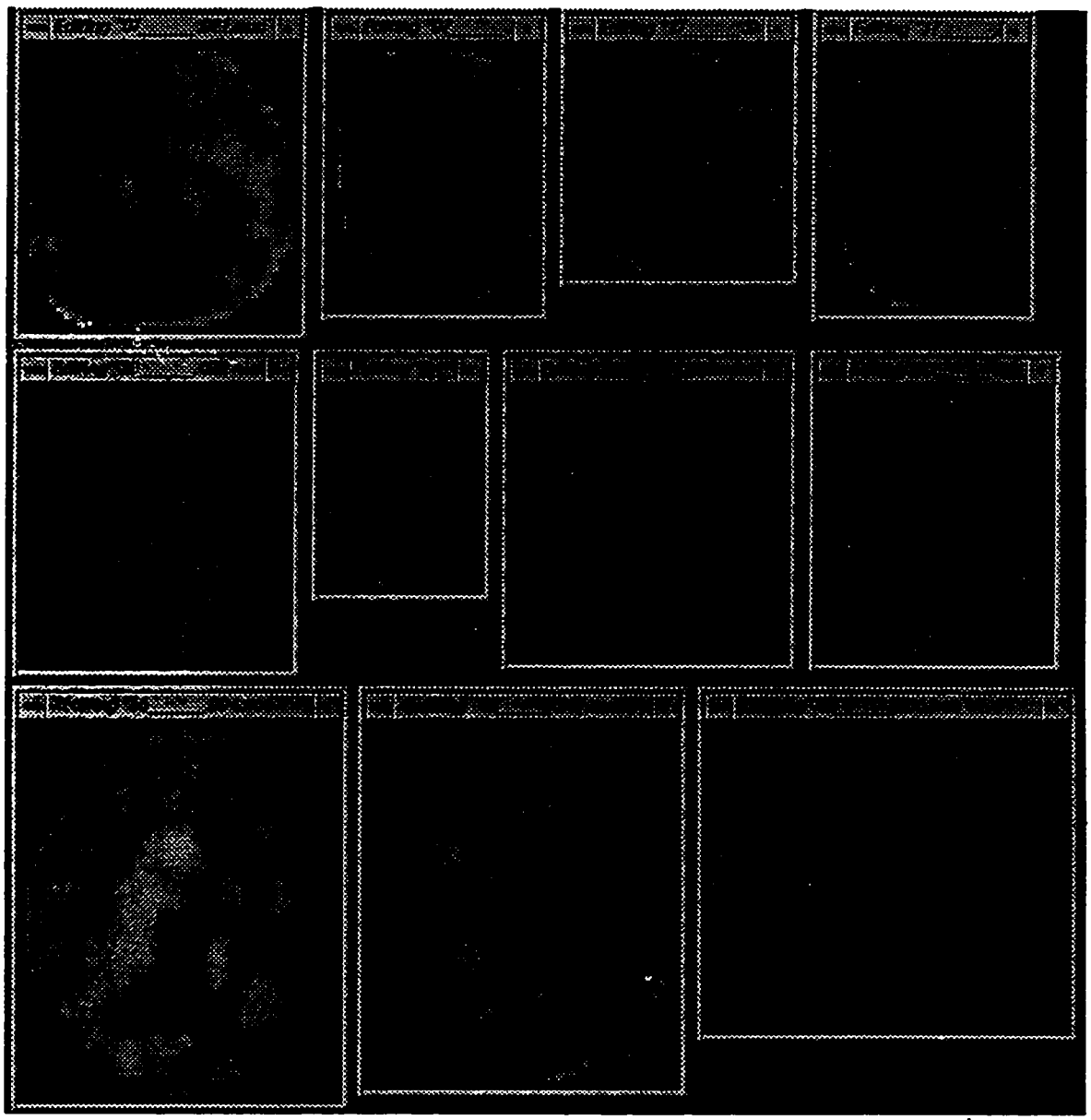
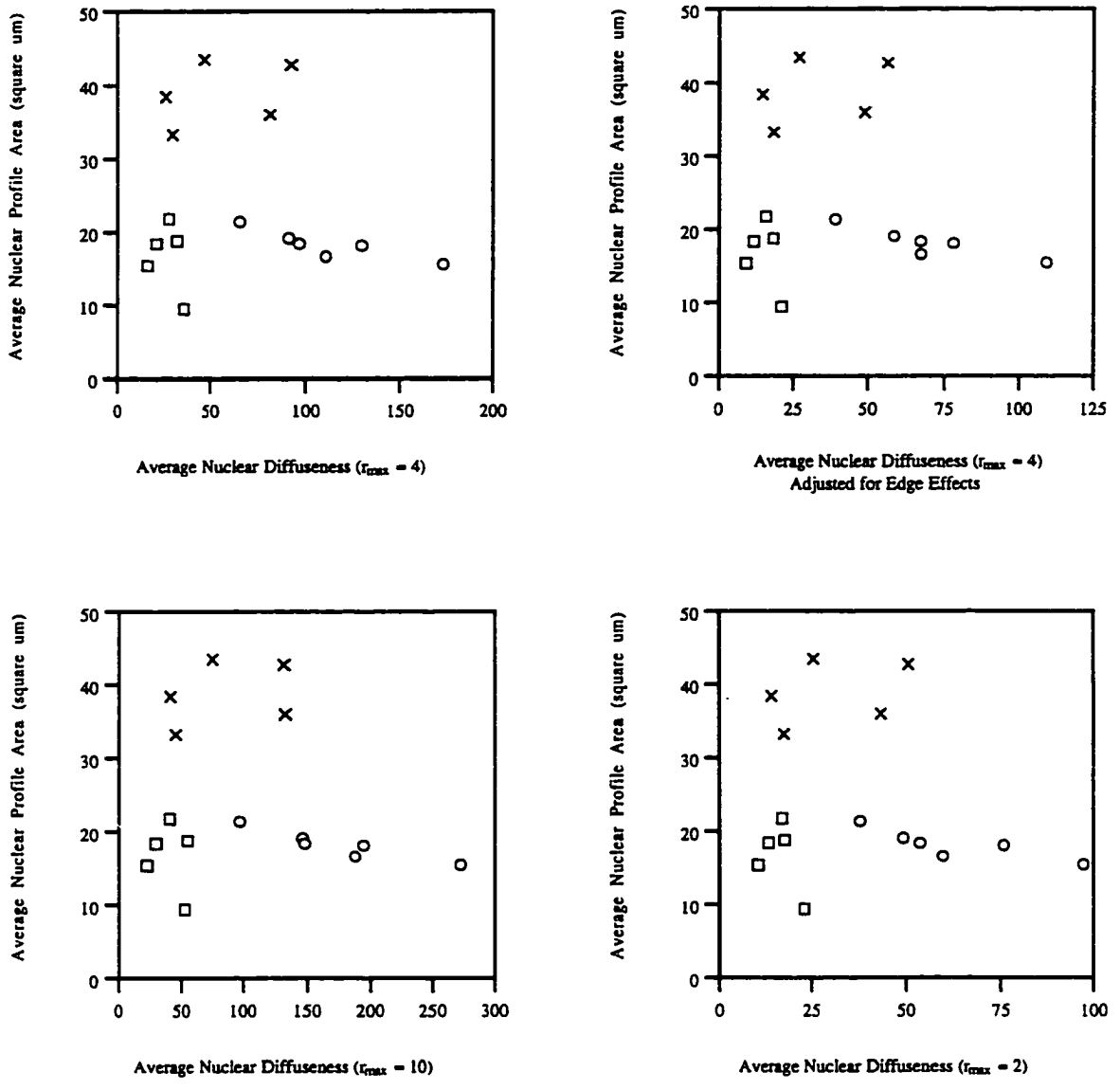


Figure 3

Characterization of Parathyroid Biopsies by Morphometry:

Scatter Plots of Average Nuclear Profile Area *vs.* Average Nuclear Diffuseness



Circles represent normal cases, squares represent adenomas, and ×'s represent carcinomas. Note the higher average areas of carcinomas as compared to the other two groups, and the lower diffusenesses of adenomas as compared to normal cases.

CHAPTER 5

Self-Affinity and Lacunarity of Chromatin Texture in Benign and Malignant Breast Epithelial Cell Nuclei

Andrew J. Einstein

Hai-Shan Wu

Joan Gil

Physical Review Letters; accepted for publication.

Preface to Chapter 5

This chapter and the two following it address the fractal characterization of chromatin appearance, and its application to the diagnosis of breast lesions. In these chapters we provide the first report of the fractal structure of chromatin appearance, introduce two new methods, an iterative approach to estimating spectral dimensions and a weighted measure of lacunarity, and demonstrate the use of these methods for providing a highly accurate cytologic diagnosis of breast cancer.

While their applications are virtually as diverse as is nature, fractals have found something of a home in the field of statistical physics. Since the cytology and statistical physics communities, with their considerably different vocabularies and emphases, have virtually no members in common, we decided to prepare separate accounts of this work for the two groups. Chapter 5 is written as a rapid communications Letter for physicists, while Chapters 6 and 7 are written for an audience of pathologists. Thus, this chapter addresses more theoretical issues relating to fractals not discussed later, such as concavity of log-log plots, connections between fractals and random processes, and the spectral reserve hypothesis which posits the loss of complexity in disease states, as well as offering more mathematical detail than Chapters 6 and 7. The subsequent chapters are more application oriented.

The iterative spectral dimension method was conceived of and developed primarily by Dr. Hai-Shan Wu in response to my questions about taking the two dimensional discrete Fourier transform of nonrectangular images. I wrote the description of it found below, and applied it to the breast cytology data set. Joe Silberfarb and Cindy Sheppard performed the nuclear segmentation.

Methods are presented for characterizing the self-affinity and lacunarity of arbitrarily shaped images. Chromatin appearance in breast epithelial cell nuclei is shown to be statistically self-affine. Spectral and Minkowski dimensions are lesser in nuclei of malignant cases than in nuclei of benign cases, and lacunarity further quantifies morphologic differences such as chromatin clumping and nucleoli. Fractal texture features are used as the basis for an accurate cytologic diagnosis of breast cancer.

PACS numbers: 87.10.+e, 87.56.Fc

Contemporary standards in oncology typically require a pathologic diagnosis prior to instituting therapy, and thus the definitive role in cancer diagnosis is borne by the pathologist. Despite the introduction of promising molecular biomarkers, diagnosis is still made by the pathologist largely on the basis of a subjective morphologic assessment of physical properties of cells and nuclei such as size, shape, adhesion, regularity, and chromatin appearance. Factors considered in evaluating chromatin include whether it takes a fine or coarse appearance, if it is clumped, whether marginated chromatin and/or nuclear voids are present, and the size and number of nucleoli. Considering the significant morbidity and mortality associated with cancer treatment, and the equally devastating consequences of withholding treatment when it is necessary, it is essential for a diagnosis to be correct. While in most cases, diagnosis is straightforward for an experienced pathologist, in some cases, it is not. The diagnosis of breast cancer poses a particular challenge. Fine-needle aspiration (FNA) cytology, which involves inserting a needle into the breast and aspirating individual cells from suspicious tissue, has become an increasingly important diagnostic procedure in the management of breast masses. An accurate diagnosis using FNA is highly dependent on how experienced the aspirator and cytopathologist are, and high false positive rates have been reported.[1]

Scale-invariant behavior is characteristic of many biological systems, ranging from microvascular network formation[2] to branching of the bronchial tree,[3] from heart interbeat intervals[4] to bacterial colony formation.[5] Recent studies have shown that aspects of morphology considered by pathologists, such as the shapes of cell membranes,[6] nuclear membranes,[7] and tissues,[8,9] exhibit fractal properties. Here we investigate scale-invariant properties of chromatin appearance in microscopic images of breast epithelial cell nuclei obtained by cytology, and show how they can be used as the basis for a more objective diagnosis between benign and malignant breast epithelial cell lesions. Chromatin appearance in nuclear images is manifested in terms of texture, the spatial distribution of grey values. Several studies have attempted to quantify texture in cytologic specimens, for both diagnostic[10,11] and prognostic[12-14] purposes.

As illustrated in Fig. 1, a nuclear image can be viewed as a surface for which the x- and y- coordinates represent position and the z- coordinate represents grey level. We evaluate the self-affinity of these surfaces, and characterize them by fractal dimensions and lacunarity. If nuclear images can be shown to be (statistically) self-affine, then these fractal parameters are an appealing approach to texture characterization. Pentland[15] has shown a high correlation between human visual roughness perception and fractal dimension, and thus fractal dimension is a good candidate for numerically characterizing the nuclear irregularity or pleomorphism subjectively observed by pathologists. Moreover, lacunarity[16,17] measures the largeness of gaps or holes. Since nucleoli form large "holes" in cytologic images, and increases in nucleolar size and irregularity tend to be indicative of malignancy,[18,19] lacunarity should also be of diagnostic utility.

We obtained cytology specimens from 41 patients. The specimens were stained using the ultrafast Papanicolaou protocol.[20,21] 22 patients were diagnosed with invasive ductal carcinoma and 19 as benign; each diagnosis was made by a cytologist and independently confirmed by at least one pathologist on the basis of surgical biopsy findings. Specimens were viewed on an image analysis system under oil-immersion light

microscopy, and an average of 64 randomly selected nuclear images were segmented for each patient by an individual blinded as to the cytologic and histopathologic diagnoses. Since in cytologic specimens, cells lie compressed between the slide and the cover slip, these images provide planar representations of complete nuclei. Images were normalized[22] to compensate for possible differences in staining and lighting conditions.

Several fractal dimensions can be used to characterize self-affine surfaces, each one contributing to an overall description.[23] We characterized the nuclear images in the spatial domain with a Minkowski-Bouligand dimension, and in the frequency domain with a spectral exponent. Minkowski dimensions were determined using a modification of the variation method of Dubuc *et al.*[24-26] The Minkowski sausage $M_\varepsilon(S)$ of a bounded set S in a metric space is the set of all points less than a specified distance ε from the given set; in terms of mathematical morphology,[27] it can be viewed as the dilation of the set by an isotropic structuring element.[28] The Minkowski-Bouligand dimension D_{MB} relates ε to the size of the Minkowski sausage. In $\Re^3$ (with $|\cdot|$ denoting volume), the Minkowski dimension of a surface S is defined[24-26] as

$$D_{MB}(S) = \lim_{\varepsilon \to 0} \left(3 - \frac{\log \, | \, (M_\varepsilon(S)) \, |}{\log \, \varepsilon}\right) \, .$$

In practice, other structuring elements can be used for dilation. While for a self-similar object the structuring element used in determining Minkowski dimension will typically be isotropic in all directions, for a self-affine object it is often appropriate to choose a structuring element isotropic in the directions of the independent variables. This approach is taken by the variation method, which is invariant with respect to affine transforms and has been shown to accurately determine the Minkowski dimensions of mathematical objects with known dimension, such as fractal brownian[29] and Takagi[24-26] surfaces. For a surface $z = f(x, y)$ at a given point (x', y'), the ε-oscillation is defined as the difference between the extreme values of f in a neighborhood of (x', y'), *i.e.*,

$$V_\varepsilon(x', y') = \max f(x, y) - \min f(x, y) \, .$$

We determine the extrema over all (x, y) on the surface such that $(x - x')^2 + (y - y')^2 \leq \varepsilon^2$. The ε-variation $V_f(\varepsilon)$ of f is the sum of $V_\varepsilon(x', y')$ over the entire surface. The variation method estimates the Minkowski dimension by three minus the slope of the least squares line fitting the plot $(\log \varepsilon, \log V_f(\varepsilon))$ over a suitable range of ε's. For the nuclear images, we determined $V_f(\varepsilon)$ at 145 values of ε, ranging between 0.134 μm (1 pixel) and 2.69 μm (20 pixels).

The plots $(\log \varepsilon, \log V_f(\varepsilon))$, an example of which is shown in Fig. 1, can be used for two purposes: to assess the validity of the self-affine model, and to estimate D_{MB}. Values of r^2, a standard measure of goodness of fit equal to the square of the Pearson correlation coefficient, ranged from 0.836 to 0.996 for the 2621 nuclei, with a median of 0.962. The good linear fit implies that chromatin appearance is indeed self-affine over the range of resolutions considered. That nuclear appearance is fractal is highly suggestive that the three dimensional organization of nuclear chromatin is also fractal, as has been hypothesized by Orlando and Paro.[30] Supportive of this, Pentland[15] has shown that under certain assumptions, a three dimensional surface is fractal iff its image intensity surface is fractal. Thus, these results extend the scope over which genetic material may have fractal organization, from single chromosomes* to nuclear organization. The mean Minkowski dimension per patient was 2.527 for the benign cases and 2.510 for the malignancies, a difference which was of borderline significance (significance level $p = 0.067$ in two-way ANOVA). Minkowski dimensions estimated from a smaller range of resolutions (21 values of ε between 0.134 μm and 0.940 μm) exhibited a statistically significant difference between benign and malignant cases ($p = 0.044$).

Despite the high coefficients of correlation between $\log \varepsilon$ and $\log V_f(\varepsilon)$, a slight concavity can be observed in the $(\log \varepsilon, \log V_f(\varepsilon))$ plots for most nuclei. This concavity

---

* Takahashi[31] proposes a fractal model of metaphase chromosomes while Grosberg *et al.*[32] suggest that long range correlations observed in DNA can be accounted for by a crumpled globule model in which three-dimensional DNA structure is self-similar.

may be due to edge effects, or to a slight intrinsic deviation of chromatin texture from mathematical fractality. Consistent with this, we observed that the slopes of the plots, and consequently their associated Minkowski dimensions, depend on the range of εs included in the regression, notwithstanding the high correlation coefficients common to various ranges. Minkowski dimensions were lesser over the smaller range of resolutions (21 εs), with mean dimensions per patient of 2.333 for the benign cases and 2.319 for the malignancies, than over the longer range (145 εs). Rigaut claims[33] that this concavity is present in almost all naturally occurring fractals, and cites[28] several empirical models which have been proposed to account for it, including his asymptotic fractal model[34] in which fractal dimension varies continuously as a function of the range of resolution. Alternatively, some (*e.g.*, Mandelbrot, as cited by Rigaut[34]) prefer to account for concavity with successive values of fractal dimension, as we have done here.

Recent studies[35,36] have shown that random processes can generate fractal structures. Interestingly, the fractal structures generated by these mechanisms exhibit log-log plots with curvature, and the associated fractal dimensions are not universal but rather depend on parameters of the random processes such as densities. While it is unclear what role, if any, random processes play in determining chromatin morphology, and how these processes differ between benign and malignant nuclei, these results suggest a feasible mechanism to account for the curved log-log plots seen in our data.

Fractal images are typified by a power spectrum in which there is a $1/f^{\beta}$ dependence on frequency. Voss[37] has shown that for statistically self-affine fractal Brownian motion, the spectral exponent β is related to a fractal similarity dimension D_S by

$$D_S = (7 - \beta) / 2 \ . \qquad (1)$$

Since power spectra are determined from an image's two-dimensional Fourier transform, this approach is limited to rectangular images. Because nuclei have irregular shapes, we developed a new method to determine the fractal dimension of images of arbitrary shape.[38] A nuclear image R_n is embedded in a rectangular array R. The region of R surrounding the

actual nuclear image is referred to as R_b; the goal of the iterative procedure is to fill in this background region in such a way that it is similar to the nucleus in the Fourier domain. Since convergence may be sensitive to initial conditions, it is desirable to begin as close as possible to the final image. Thus, R_b is initialized such that the statistical properties of the image in R_n are extended to the background. If we assume a first order Markov process for the image in both R_n and R_b, the initial image in the background region can be generated with a recursive equation whose coefficients are estimated based on the autocorrelation and variance of the image in the nuclear region. To avoid the discontinuities at the boundaries between the nuclear and background regions, the image outside the nuclear region is iteratively generated starting at the nuclear region and gradually dilated to the whole rectangular region.

Following initialization of the background region, the two dimensional Fourier transform of R is taken, from which its power spectrum is determined. β is estimated by fitting this computed power spectrum to the assumed model in which power spectrum equals $Cf^{-\beta}$, where C is a proportionality constant. The fitted power spectrum is combined with the actual phase of the Fourier transform of R to yield a new frequency domain signal. The inverse Fourier tranform of this signal is taken to produce another spatial domain image R'. The background area in R' has been determined both from the previous background and from spectral properties of the nucleus, and in this sense should be an improvement over the previous background. Thus, the image used in the next iteration contains the original nuclear image R_n padded with the improved background signal R_b'. This process is iterated until the change in the coefficients of the $1/f^{\beta}$ model is sufficiently small, and the final β is used to determine the fractal dimension from (1). In this study, iteration was stopped when the change in β was less than 0.01, or after 200 iterations. We have tested this method with synthetic "nuclei" of known fractal dimension generated using Voss' inverse Fourier filtering method,[37,39] and have found that β is estimated very accurately, *e.g.*, within 0.2% for a nucleus with $D_S = 2.75$.

The mean spectral dimension per patient was 2.853 for the benign cases and 2.796 for the malignancies, a difference which was determined to be very highly significant (ANOVA significance level $p = 0.000011$). These means exclude 239 of the 2621 nuclei for which the iterative method was considered to have converged incorrectly ($D_S < 2$, *i.e.*, spectral density approaching white noise, or $C > 2000$). Spectral and Minkowski dimensions for individual nuclei were weakly correlated ($r^2 = 0.027$ for the 2382 nonexcluded nuclei), as were mean spectral and Minkowski dimensions ($r^2 = 0.018$), indicating that they reflect different facets of the self-affine nature of chromatin in breast epithelial cell nuclei. The lower fractal dimensions exhibited in malignancy—in both the frequency and spatial domains—are consistent with the hypothesis advanced by Goldberger and colleagues[40] that disease is associated with a loss of biologic complexity.§

Two objects can exhibit different texture but still have the same fractal dimension. Often this difference can be characterized in terms of lacunarity. Figure 2 demonstrates three Sierpinski carpets and two nuclei; each set has the same fractal dimensions, but different texture. In particular, while the nuclei have similar chromatin appearance, the nucleus on the left has a prominent nucleolus, rendering the texture less translationally invariant and more lacunar. We measure weighted lacunarity on thresholded nuclear images using an extension of the gliding box method[44,45] to binary images of arbitrary shape.

In the gliding box method, an $s \times s$ pixel "gliding box" is initially placed at the upper left corner of the image, and the number of light pixels in the image contained in the gliding box is denoted n_1. The box "glides" over the entire image, moving to all N positions at which it covers at least one pixel of the image, at each location recording the number of light pixels n_i in the image. The sequence of values $\{n_i\}$ for $i \in \{1,2,\ldots,N\}$

---

§ See also Babloyantz and Destexhe,[41] and Watt and Hameroff,[42] where sleep and anaesthetic states are associated with lower dimensionality than consciousness, and Kaplan *et al.*,[43] where aging is associated with decreased biologic complexity.

defines a probability distribution $Q_n(M, s)$ which represents the probability that a gliding box of side r contains M light pixels. Lacunarity is defined in terms of the moments $Z_{Q_n}^{(q)}(s) = \sum_{M=1}^{s^2} M^q Q_n(M, s)$ of the distribution $Q_n(M, s)$, by

$$\Lambda(s) = Z_{Q_n}^{(2)}(s) / [Z_{Q_n}^{(1)}(s)]^2 \ .$$

When working with irregularly shaped images such as nuclei, at many positions—in particular near the periphery of an image—many of the pixels in the gliding box will not be part of the image, but "background" instead. The values of n_i at these positions will be low since there are few pixels of the image within the gliding box at these locations. The net result can be to skew the distribution $Q_n(M, s)$. To correct for this problem, we incorporate a weighting factor w_i, equal to the number of pixels in the gliding box that are part of the image. We define x_i, the weighted number of light pixels in the image contained in the gliding box, by

$$x_i = n_i \, (s^2 / w_i) \ . \tag{2}$$

Weighted lacunarity is defined in terms of weighted moments of the distribution $Q_x(M, s)$ of $\{x_i\}$. The rationale for weighting the moments as well as the number of pixels is that we do not wish a position for which the gliding box contains only a few pixels in the image to contribute as much to the lacunarity as a position for which it contains many pixels. Instead of summing over the number of light pixels, moments of $Q_x(M, s)$ can be calculated by summing over the N positions taken by the gliding box, *i.e.*,

$$Z_{Q_x}^{(q)}(s) = \sum_{M=1}^{s^2} M^q Q_x(M, s) = \sum_{i=1}^{N} x_i^q / N = \sum_{i=1}^{N} x_i^q \Big/ \sum_{i=1}^{N} 1 \ .$$

Weighted moments $Z'_{Q_x}{}^{(q)}(s)$ are given by

$$Z'_{Q_x}{}^{(q)}(s) = \sum_{i=1}^{N} w_i x_i^q \Big/ \sum_{i=1}^{N} w_i \ , \tag{3}$$

and weighted lacunarity is defined as

$$\Lambda'(s) = Z'_{Q_x}{}^{(2)}(s) / [Z'_{Q_x}{}^{(1)}(s)]^2 \ . \tag{4}$$

Equivalently, substituting (2) and (3) into (4) yields

$$\Lambda'(s) = \sum_{i=1}^{N} (n_i^2 / w_i) \sum_{i=1}^{N} w_i \Big/ \left(\sum_{i=1}^{N} n_i \right)^2 .$$

Normalized weighted lacunarity curves for the Sierpinski carpets and nuclei in Fig. 2 are shown below the images. While these fractals have the same dimensions, they can be distinguished on the basis of the lacunarity curves.

Weighted lacunarity functions were determined for each nucleus thresholded at the third quartile of its intensity histogram, with box side lengths ranging from $s = 2$ pixels (0.269 μm) to $s = 35$ pixels (4.70 μm). Patient means of weighted lacunarity were greater in malignant cases than in benign cases. As is shown in Fig. 3, this difference was significant at virtually all box sizes (maximum p of 0.0505 at $s = 6$), but especially at the two ends of the curve. These ends correspond, respectively, to fine-scale texture, such as chromatin clumping, and to large-scale structures such as nucleoli and voids. Thus, lacunarity complements fractal dimensions in characterizing chromatin texture.

Differences in dimensionality and lacunarity can be used as a basis for diagnosing unknown cases. We used logistic regression to classify cases by jackknife (leave-one-out) analysis, considering a number of models incorporating various combinations of fractal textural features as well as nuclear area. The best performance was achieved by a model predicting diagnosis for a patient on the basis of mean D_S, mean normalized $\Lambda'(35)$, mean area, standard deviation of C, and standard deviation of normalized $\log \Lambda'(2)$. This yielded a correct diagnosis for all 19 benign cases and 20 of the 22 malignant cases, an accuracy of 95.1%. These findings suggest that measures of self-affinity can be used as clinical parameters to assist cytologists in rendering an accurate diagnosis in FNA of the breast, and potentially in automated instruments for quality assurance in pathology.

The authors thank C. Benham, J. Malinsky, S. Wallenstein, and C. Bodian for helpful advice and discussion, M. Sanchez, G. Yang, and S. Hoda for materials, D. Burstein for cytologic diagnoses, I. Bleiweiss for histopathologic diagnoses, A. Berger for

French translation, and J. Silberfarb and C. Sheppard for technical assistance. A.J.E. was supported by a traineeship on NIH MSTP Training Grant GM07280.

References

1. Layfield LJ, Glasgow BJ, Cramer H. Fine-needle aspiration in the management of breast masses. *Pathol Annu* 1989; **24:** 23-62.

2. Gazit Y, Berk DA, Leunig M, Baxter LT, Jain RK. Scale-invariant behavior and vascular network formation in normal and tumor tissue. *Phys Rev Lett* 1995; **75:** 2428-2431.

3. West BJ, Bhargava V, Goldberger AL. Beyond the principle of similitude: renormalization in the bronchial tree. *J Appl Physiol* 1986; **60:** 1089-1097.

4. Peng C-K, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL. Long-range anticorrelations and non-Gaussian behavior of the heartbeat. *Phys Rev Lett* 1993; **70:** 1343-1346.

5. Vicsek T, Cserzö M, Horváth V. Self-affine growth of bacterial colonies. *Physica A* 1990; **167:** 315-321.

6. Losa GA, Baumann G, Nonnenmacher TF. The fractal dimension of pericellular membrane from lymphocytes and lymphoblastic leukemic cells. *Acta Stereol* 1992; **11/I:** 335-341.

7. Landini G, Rippin JW. An "asymptotic fractal" approach to the morphology of malignant cell nuclei. *Fractals* 1993; **1:** 326-335.

8. Cross SS, Bury JP, Silcocks PB, Stephenson TJ, Cotton DW. Fractal geometric analysis of colorectal polyps. *J Pathol* 1994; **172:** 317-323.

9. Landini G, Rippin JW. Fractal dimensions of the epithelial-connective tissue interfaces in premalignant and malignant epithelial lesions of the floor of the mouth. *Anal Quant Cytol Histol* 1993; **15:** 144-149.

10. Moragas A, Garcia-Bonafe M, de Torres I, Sans M. Textural analysis of lymphoid cells in serous effusions. A mathematical morphologic approach. *Anal Quant Cytol Histol* 1993; **15:** 165-170.

11. Banda-Gamoba H, Ricketts I, Cairns A, Hussein K. Spectral analysis of cervical cells using the discrete Fourier transform. *Anal Cell Pathol* 1993; **5:** 85-102.

12. Jørgensen T, Yogesan K, Tveter KJ, Skjorten F, Danielsen HE. Nuclear texture analysis: a new prognostic tool in metastatic prostate cancer. *Cytometry* 1996; **24:** 277-283.

13. Pauwels O, Kiss R. Monitoring of chemotherapy-induced morphonuclear modifications by means of digital cell-image analysis. *J Cancer Res Clin Oncol* 1993; **119:** 533-540.

14. Gimenez-Mas JA, Sanz-Moncasi MP, Remon L, Gambo P, Gallego-Calvo MP. Automated textural analysis of nuclear chromatin: a mathematical morphology approach. *Anal Quant Cytol Histol* 1995; **17:** 39-47.

15. Pentland AP. Fractal-based description of natural scenes. *IEEE Trans Pattern Anal Machine Intell* 1984; **PAMI-6:** 661-674.

16. Mandelbrot B. Corrélations et texture dans un nouveau modèle d'Univers hiérarchisé, basé sur les ensembles trémas. *C R Acad Sci Ser A* 1979; **288:** 81-83.

17. Gefen Y, Mandelbrot BB, Aharony A. Critical phenomena on fractal lattices. *Phys Rev Lett* 1980; **45:** 855-858.

18. Oertel YC. Fine Needle Aspiration of the Breast. Boston: Butterworth Publishers, 1987.

19. Frost JK. The cell in health and disease: an evaluation of cellular morphologic expression of biologic behavior. Basel: Karger, 1969.

20. Yang GC, Alvarez II. Ultrafast Papanicolaou stain. An alternative preparation for fine needle aspiration cytology. *Acta Cytol* 1995; **39:** 55-60.

21. Papanicolaou GN. A new procedure for staining vaginal smears. *Science* 1942; **95:** 438-439.

22. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

23. Mandelbrot BB. Self-affine fractals and fractal dimension. *Physica Scripta* 1985; **32:** 257-260.

24. Dubuc B, Zucker SW, Tricot C, Quiniou JF, Wehbi D. Evaluating the fractal dimension of surfaces. *Proc R Soc Lond A* 1989; **425:** 113-127.

25. Dubuc B, Roques-Carmes C, Tricot C, Zucker SW. The variation method: a technique to estimate the fractal dimension of surfaces. *SPIE Visual Communications and Image Processing II* 1987; **845:** 241-248.

26. Dubuc B, Quiniou JF, Roques-Carmes C, Tricot C, Zucker SW. Evaluating the fractal dimension of profiles. *Phys Rev A* 1989; **39:** 1500-1512.

27. Serra J. Image Analysis and Mathematical Morphology. London: Academic Press, 1982.

28. Rigaut JP. Fractal models in biological image analysis and vision. *Acta Stereol* 1990; **9:** 37-52.

29. Miller GSP. The definition and rendering of terrain maps. *Computer Graphics* 1986; **20:** 39-48.

30. Orlando V, Paro R. Chromatin multiprotein complexes involved in the maintenance of transcription patterns. *Curr Opin Genet Dev* 1995; **5:** 174-179.

31. Takahashi M. A fractal model of chromosomes and chromosomal DNA replication. *J theor Biol* 1989; **141:** 117-136.

32. Grosberg A, Rabin Y, Havlin S, Neer A. Crumpled globule model of the three-dimensional structure of DNA. *Europhys Lett* 1993; **23:** 373-378.

33. Rigaut JP, Schoëvaërt-Brossault D, Lu H: Asymptotic fractals *2nd International Symposium on Fractals in Biology and Medicine* 1996.

34. Rigaut JP. An empirical formulation relating boundary lengths to resolution in specimens showing 'non-ideally fractal' dimensions. *J Microsc* 1984; **133:** 41-54.

35. Hamburger DA, Biham O, Avnir D. Apparent fractality emerging from models of random distributions. *Phys Rev E* 1996; **53:** 3342-3359.

36. Avnir D, Biham O, Lidar (Hamburger) D, Malcai O. On the abundance of fractals. In: Novak MM, Dewey TG, eds. Fractal Frontiers. Singapore: World Scientific, 1997: 199-234.

37. Voss RF. Random fractal forgeries. In: Earnshaw RA, ed. Fundamental Algorithms for Computer Graphics. Berlin: Springer-Verlag, 1985: 805-835.

38. Wu H-S, Einstein AJ, Gil J. Fractal characterization of irregularly shaped images by frequency analysis. *Submitted* .

39. Anguiano E, Pancorbo M, Aguilar M. Fractal characterization by frequency analysis. I. Surfaces. *J Microsc* 1993; **172:** 223-232.

40. Goldberger AL, Findley LJ, Blackburn MR, Mandell AJ. Nonlinear dynamics in heart failure: implications of long-wavelength cardiopulmonary oscillations. *Am Heart J* 1984; **107:** 612-615.

41. Babloyantz A, Destexhe A. Low-dimensional chaos in an instance of epilepsy. *Proc Natl Acad Sci USA* 1986; **83:** 3513-3517.

42. Watt RC, Hameroff SR. Phase space analysis of human EEG during general anesthesia. *Ann N Y Acad Sci* 1987; **504:** 286-288.

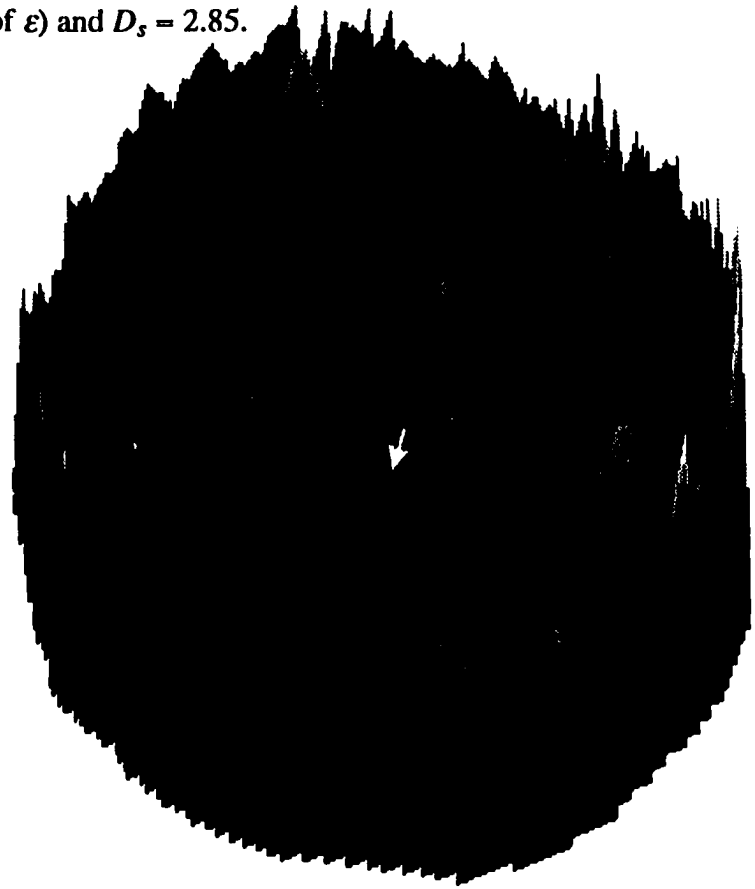
43. Kaplan DT, Furman MI, Pincus SM, Ryan SM, Lipsitz LA, Goldberger AL. Aging and the complexity of cardiovascular dynamics. *Biophys J* 1991; **59:** 945-949.

44. Allain C, Cloitre M. Characterizing the lacunarity of random and deterministic fractal sets. *Phys Rev A* 1991; **44:** 3552-3558.

45. Plotnick RE, Gardner RH, Hargrove WW, Prestegaard K, Perlmutter M. Lacunarity analysis: a general technique for the analysis of spatial patterns. *Phys Rev E* 1996; **53:** 5461-5468.

Figure 1

(a) Surface plot of a malignant breast epithelial cell nucleus; note the nucleolar "pit," marked by an arrow. Fractal dimensions, defined below in the text, are D_{mb} = 2.53 (as determined by least squares fit of all 145 values of ε) and D_s = 2.85.



(b) The associated plot of $log\ V_f(\varepsilon)$ versus $log\ \varepsilon$. r^2 = 0.97, indicating that chromatin texture is self-affine. Logs are base e.

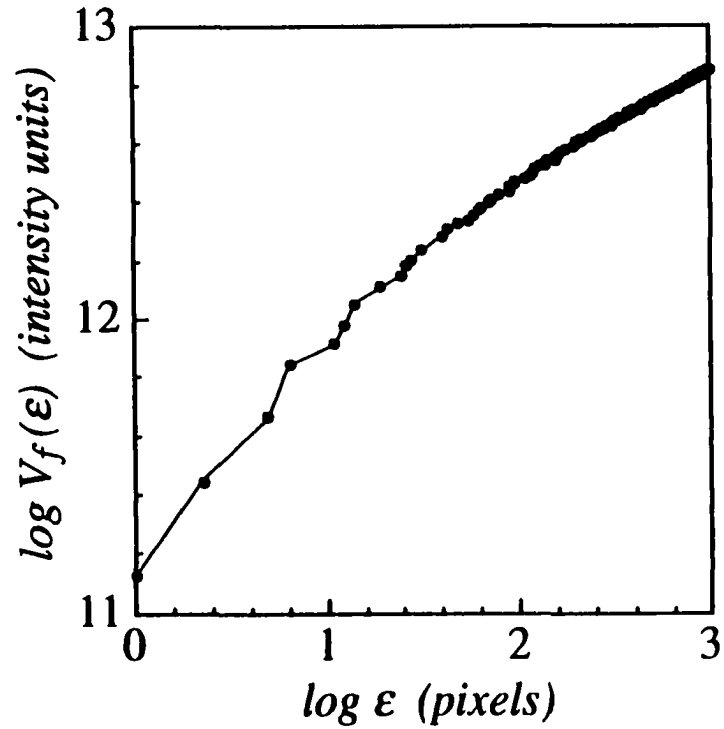
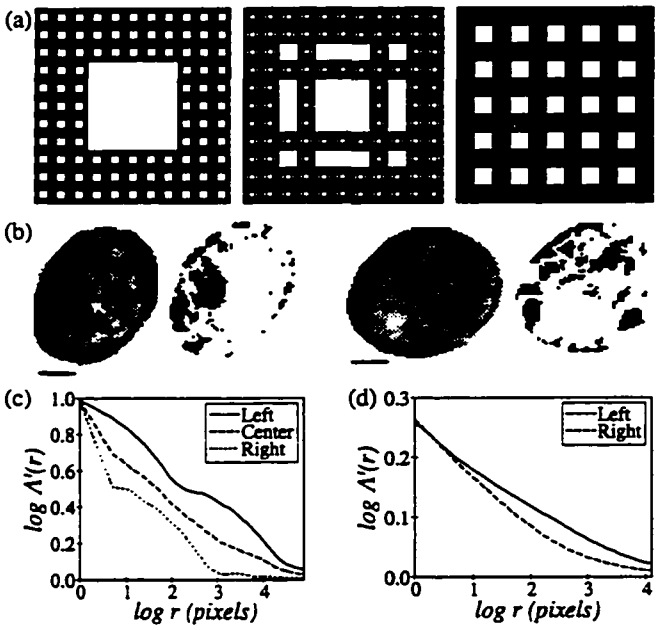


Figure 2

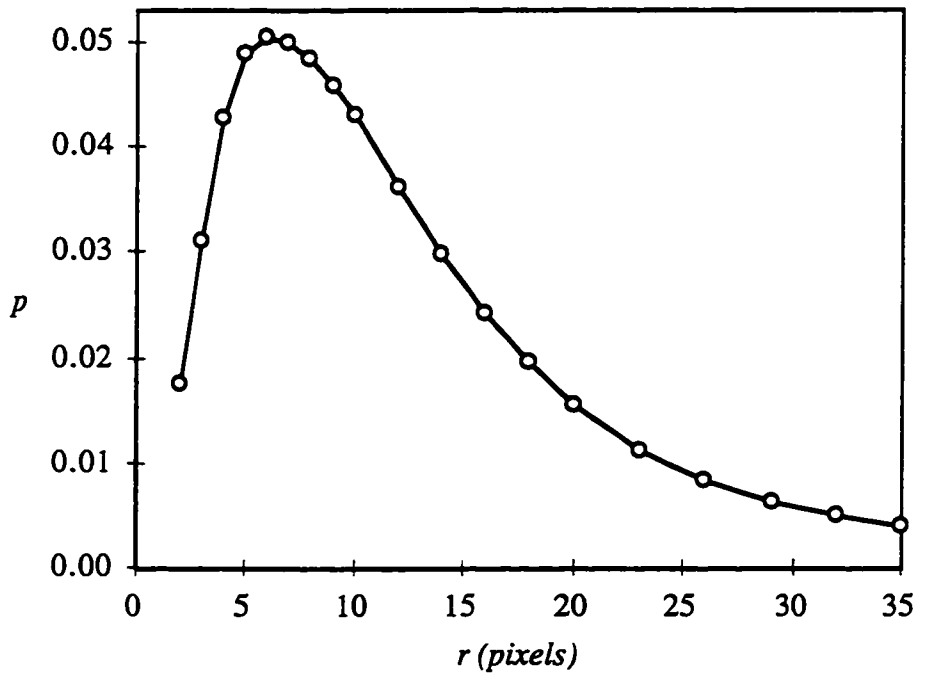


(a) Two stages of Sierpinski carpets (inverse images) with Hausdorf dimensions of log 96 / log 11 ≈ 1.90 and (c) their lacunarity curves. The carpets are denoted in (c) by their positions in (a) as left, right, and center. (b) Two breast epithelial cell nuclei with D_{mb} = 2.49 and D_s = 2.81, shown both as grayscale images and thresholded at the first quartile of the intensity histogram, and (d) the lacunarity curves for the two nuclei. The nucleus on the left has a prominent nucleolus, which is marked by an arrow, and marginated chromatin. The nucleus on the right, despite a nuclear void, has a more uniform distribution of chromatin and is less lacunar. Scale bars are 2μm. Logs are base *e*.

Figure 3

p Value as a Function of Box Size

p is the significance level of the Student's *t*-test comparing mean lacunarities of benign and malignant patients.



CHAPTER 6

Fractal Characterization of Chromatin Appearance for Diagnosis in Breast Cytology: I. Fractal Dimensions

Andrew J. Einstein

Hai-Shan Wu

Miguel Sanchez

Joan Gil

Submitted

Preface to Chapter 6

Chapter 6 discusses the fractal dimensions of chromatin texture. The focus here is on fractals as a tool for image analysis in pathology, rather than on the physics of fractals. Materials and methods are discussed in greater detail than in Chapter 5, and more data is presented, but there is less mathematical detail.

Summary

This study explores the possibility that fractal analysis may be used to characterize chromatin texture in breast cytology. Images of nuclei from fine needle aspiration biopsies of the breast were characterized in terms of their Minkowski and spectral fractal dimensions, for 19 patients with benign epithelial cell lesions and 22 with invasive ductal carcinomas. Chromatin appearance in breast epithelial cell nuclear images was demonstrated to be fractal. There was a very highly significant difference ($p = 0.000011$) between the mean spectral dimensions of the benign (2.853) and malignant (2.796) cases. The two fractal dimensions were very weakly correlated ($r = -0.13$). The fractal nature of breast epithelial cell nuclear images suggests that three dimensional chromatin structure in these cells has fractal properties.

Key Words: fractals, fractal dimension, morphometry, image analysis, breast cytology

Introduction

Numerous studies have aimed at developing image analysis procedures for the resolution of difficult differential diagnoses in cytology and histopathology. The general approach used involves characterizing nuclei with numerical measures of factors considered subjectively by pathologists. A diagnosis is assigned on the basis of these features or descriptors, in accordance with a prescribed classificatory approach determined and validated on the basis of a representative set of cases. Recent work has focused both on various aspects of this diagnostic paradigm—ranging from developing[1,2] and analyzing the reproducibility of[3,4] segmentation methods used to separate nuclear images from their background, to formulating effective descriptors,[5] to comparing[6,7] and optimizing[8] classificatory methods—and also on diverse applications, such as the diagnoses of ovarian

dysplasia,[9] prostate cancer,[10] parathyroid lesions,[11] and melanoma.[12] Among the most useful features for cytologic applications have been measures of nuclear size, pleomorphism[13] (the distribution of nuclear sizes), and chromatin appearance.[14] The numerical characterization of chromatin appearance is especially propitious for cytologic applications, where there are no sectioning effects as in histology, and the Papanicolaou stain commonly used is an excellent stain of chromatin. This study evaluates the use of fractal dimensions as descriptors of chromatin appearance in breast cytology.

In image analysis, chromatin appearance is expressed in terms of the texture of a digitized nuclear image, *i.e.*, the spatial distribution of grey values. A nuclear image can be viewed as a surface for which the x- and y- coordinates represent position and the z-coordinate represents grey level. Figures 1 and 2 illustrate the texture of representative benign and malignant breast epithelial cell nuclei. While qualitatively they look different, it has proven exceedingly difficult to quantitatively assess such subjective differences in texture in a way suitable for classificatory analysis. A number of approaches have been suggested, based on Markovian analysis,[15,16] run length statistics,[17] textons,[18] Fourier analysis,[19] mathematical morphology,[20] and local gray level variation.[11] The most popular approach, based on Markovian analysis, may yield over 60 highly correlated[21] features which are difficult to interpret and do not correspond with the visual impressions of pathologists.[22] While work by Dawson *et al.*[23] using nuclear models and malignant breast epithelial cell nuclei has identified a subset of Markovian descriptors that are optimal for discriminating among texture patterns, the correspondence between Markovian descriptors and factors such as nucleoli, granularity, and margination remains unclear. Fractal geometry offers an alternative approach to chromatin texture description.

Fractal geometry provides a framework to mathematically describe objects exhibiting structure over a range of scales.[24] Such objects, which have no characteristic size but rather exhibit similar detail on many scales, are called fractals and may be described by their fractal dimensions. In biology, fractals have recently been used to characterize a

wide range of complex systems including long-range correlations in DNA sequences,[25] myocardial blood flow,[26] and the branching of airways[27] and retinal vasculature.[28] Applications of fractal geometry to pathology[29,30] have concentrated on characterizing the structural complexity or irregularity of cell[31,32] and nuclear[33] membranes, of tissue shape,[34-36] and of cell growth *in vitro*.[37,38] In this work, we use methods of fractal analysis to describe the structure of chromatin in light microscopic images of breast epithelial cell nuclei. We will not discuss here the possible influence of staining techniques on fractal dimensions; rather, we will confine our study to ultrafast Papanicolaou stained cytologic specimens, normalized for contrast, and investigate differences in fractal properties between benign and malignant cases.

Materials and Methods

Cytologic and Histologic Materials

Cytology specimens were obtained from 41 patients, of whom 22 were diagnosed with invasive ductal carcinoma and 19 as benign with no atypia. Each diagnosis was made by a cytologist and independently confirmed by a pathologist on the basis of histopathologic findings. With the exception of two cases, all histopathologic diagnoses were reconfirmed by a second pathologist. Cytologic specimens were processed using the ultrafast Papanicolaou protocol.[39]

Image Analysis

Image acquisition was performed on a self-assembled system based on a Gateway 2000 486DX2/50E microcomputer augmented with Sprynt i860 image processing boards and the Semper 6 Plus graphics program, connected to a Nikon Optiphot microscope equipped with a 100× Nikon Plan objective through a Sony DXC-M2 camera. For each specimen, randomly selected epithelial cell nuclei were segmented, excluding overlapping

and damaged nuclei as well as those with insufficient contrast. Segmentation was performed using an arc-forming method,[4] involving the selection of three points on the margin of a nucleus, the computation and display of the arc connecting these points, and the extension of this contour with additional arcs until the whole profile is outlined. We have found this approach to be the most reproducible method for segmenting cytologic images. Subsequent image processing and computation of fractal dimensions and nuclear areas was performed using software written in C and run on a Silicon Graphics Indigo^2 workstation and Silicon Graphics Power Challenge XL supercomputer. Nuclear images were normalized[11] to compensate for possible differences in staining and lighting conditions and then screened to assure that they were complete, properly segmented, and of sufficient contrast for texture analysis. A total of 2621 images was collected, with a minimum of 55 and a mean of 64 nuclei per patient.

Fractal Dimension Calculations

Several related mathematical formulations of fractal dimension have been suggested, such as Hausdorf-Besicovitch dimension, Minkowski-Bouligand dimension, Kolmogorov box-counting dimension, spectral dimension, and Korcak dimension[40]. Each method for determining fractal dimension may characterize a particular aspect of the fractal nature of a nuclear image, and not all fractal dimensions are applicable in every context. In particular, it is important to make a distinction between self-similar and self-affine fractals, since the methods used to calculate dimensions differ between these two classes of fractals. Mandelbrot[41] illustrates this difference by comparing measurement on the earth's surface, where the choice of north-south and east-west as the coordinate axes is in a sense arbitrary and distances are meaningful, to measurement on a graph of volume versus pressure, where the choice of coordinate axes is canonical but distances $\sqrt{(\Delta V)^2 + (\Delta p)^2}$ are meaningless and therefore cannot be used in determining the fractal dimension. In the first case, linear log-log scaling would demonstrate that an object is self-similar, while in the

second case, linear log-log scaling would show that it is self-affine. Since for our surface plots the x- and y- coordinates represent position and the z- coordinate represents grey level, fractal characterization of the images should use methods appropriate for self-affine objects. In this study, we characterize breast epithelial cell nuclei using the Minkowski-Bouligand fractal dimension (D_{MB}) and the spectral fractal dimension (D_S).

Minkowski-Bouligand Dimensions—Minkowski-Bouligand dimensions were determined using a modification of the variation method of Dubuc *et al.*[42,43] This method is based on the concept of ε-variation. In the digitized gray scale image of a nucleus, an intensity value between 0 and 255 is associated with each pixel. The ε-oscillation for a given pixel is defined as the difference between the maximum and minimum intensity values of all pixels within a distance ε of the given pixel, *i.e.*, all pixels in a disc of radius ε, centered at the given pixel. In terms of the surface plot, the distance ε is measured in the x-y plane, thus avoiding the meaningless distances discussed in the previous paragraph. The ε-variation for the whole nucleus, denoted as $V_f(\varepsilon)$, is the sum of the ε-oscillations for all pixels in the image. Minkowski dimension is estimated by three minus the slope of the least squares line fitting the plot of $log\ \varepsilon$ versus $log\ V_f(\varepsilon)$ over a suitable range of ε s. In this study, $V_f(\varepsilon)$ was computed for each nucleus at 145 values of ε, ranging between 1 and 20 pixels (0.135 and 2.69μm), and also over a shorter range of resolutions, considering 21 values of ε between 1 and 7 pixels (0.135 and 0.942μm). This method is illustrated in Figure 3.

Spectral Dimensions—Spectral dimensions of the nuclei were determined using the iterative spectral dimension method. Spectral dimensions are based on Fourier analysis. In Fourier analysis, a signal such as a digitized image is expressed as the sum of sine and cosine functions of different frequencies. The two-dimensional discrete Fourier transform[19] of an M × N pixel image $z(x, y)$ is written as

$$Z(u,v) = \frac{1}{MN}\sum_{y=0}^{N-1}\sum_{x=0}^{M-1} z(x, y)\ \{\cos[2\pi(ux/M + vy/N)] - i\ \sin[2\pi(ux/M + vy/N)]\}\ .$$

The power spectrum $|Z(u,v)|^2$ of an image is the square of the magnitude of its Fourier tranform. Fractal images are typified by a power spectrum in which there is a $1/f^{\beta}$ dependence on frequency,[44] *i.e.*, $|Z|^2 = Cf^{-\beta}$. In particular, Voss[45] has shown that for statistically self-affine fractal Brownian motion, the spectral exponent β is related to a fractal similarity dimension D_s by the equation

$$D_s = (7 - \beta) / 2\ .$$

However, since power spectra are determined from an image's two-dimensional Fourier transform, conventional methods for determining the spectral dimension of surfaces[44] require rectangular images, which poses a problem since nuclei have irregular shapes. The requirement of rectangularity can be overcome using an iterative spectral dimension method. This approach embeds a nuclear image in a rectangular image and iteratively fills in the remainder of the image so that it shares similar properties to the nucleus in the Fourier domain. Mathematical details of the iterative procedure are provided elsewhere[46].

Statistical Analysis

Statistical analysis was performed on an Apple Power Macintosh 8100/110 using JMP and Microsoft EXCEL. Pearson correlation coefficients were determined for plots of $\log \varepsilon$ versus $\log V_f(\varepsilon)$ to assess the appropriateness of the fractal model. Mean areas and Minkowski dimensions were determined for each patient. Mean spectral dimensions were determined only considering nuclei with fractal dimensions D_S between two and three and multiplicative prefactors C less than 2000. The 239 nuclei not meeting these criteria were considered to have converged incorrectly to nonsensical values and therefore excluded from determinations of mean spectral dimensions. In addition, the area of each nucleus was determined. Fractal dimensions were compared between benign and malignant cases using two way analysis of variance (ANOVA), treating diagnosis as a fixed effect and patient as a

nested random effect. Correlation coefficients between Minkowski and spectral dimensions were determined.

Results

The log-log plots of $V_f(\varepsilon)$ vs. ε were linear, with correlation coefficients r greater than 0.95 in 98% of the nuclei (median $r = 0.981$, minimum $r = 0.914$), as illustrated in Figure 3, and fractal Minkowski dimensions of the nuclei were strictly greater than their topological dimensions. Thus, we conclude that the fractal model is an appropriate one; chromatin appearance in breast cytology specimens is fractal over the range of resolutions considered. As such, it is appropriate to use fractal dimensions as descriptors of chromatin appearance in image analysis. Over the smaller range of resolutions, the fit was even better, with a median r of 0.993 and a minimum of 0.985.

Minkowski dimensions for the 41 cases are shown in Figure 4. The patient means of Minkowski dimension averaged 2.527 for the benign cases and 2.510 for the malignancies. This difference approached statistical significance ($p = 0.067$). Over the smaller range, average patient means of Minkowski dimension were 2.333 for the benign cases and 2.319 for the malignant cases, a statistically significant difference ($p = 0.044$). Thus, Minkowski dimensions were smaller over this range. The dependence of Minkowski dimension on the range of resolutions considered may suggest that fractal dimension changes depending on the range of resolutions, or may simply be a consequence of edge effects.

Spectral dimensions for the 41 patients are shown in Figure 5. The average mean spectral dimension was 2.853 for the benign cases and 2.796 for the malignancies, a difference which was determined to be very highly significant ($p = 0.000011$). Spectral and Minkowski dimensions for individual nuclei were weakly correlated ($r = -0.166$ for the

2382 nonexcluded nuclei), as were patient means of spectral and Minkowski dimensions (r = -0.134). These correlations are illustrated in Figure 6.

Discussion

Although biological structures have certain characteristic size scales, some aspects of morphology are better described in terms of fractal geometry than in terms of Euclidean geometry. Our study shows that chromatin appearance in breast epithelial cell nuclei has fractal properties, and can be described by fractal dimensions which differ between patients with benign and malignant breast lesions. These dimensions can be used, in conjunction with other features, in the diagnosis of cytologic specimens. That nuclear appearance is fractal is highly suggestive that three dimensional chromatin structure is also fractal, as has been hypothesized by Orlando and Paro.[47] Supportive of this, Pentland[48] has mathematically proven that under certain assumptions, a three dimensional surface is fractal if and only if its image intensity surface is fractal.

An item that needs to be emphasized is that in order to be reproducible, texture descriptors such as fractal dimensions require, as a precondition, an excellent nuclear segmentation. The arc-forming segmentation method used in this study is fast and far more reproducible than the more automatic threshold-based methods preferred by many laboratories.[4]

While for some types of mathematical objects, different fractal dimensions may necessarily be equivalent, the two approaches to characterize fractal dimension that we considered here were remarkably different. Spectral dimensions were considerably greater than Minkowski dimensions, and there was only weak correlation between these two indices of fractal structure. Thus, the two approaches seem to measure different aspects of the fractal nature of nuclear chromatin, underscoring the fact that no single parameter can completely describe the fractal nature of biologic structure. This observation is consistent

with results obtained in geology by Cox and Wang.[49] They compared seven approaches for determining fractal dimensions of geologic surfaces such as rock fractures and fault traces of the San Andreas Fault, and found that fractal dimension may vary systematically, depending on the measurement method. Fractal structure in nature is apparently more heterogeneous than had been initially appreciated. As Mandelbrot[41] has recently commented, while initially a preeminent position was given to Hausdorf-Besicovitch dimension, it is clear now that fractal dimension is a multifaceted concept. The different computational methods need not even theoretically yield the same value for fractal dimension, so a surface may have several distinct fractal dimensions.[41]

In fact, aspects of fractal structure cannot be described by a fractal dimension. In the companion paper following,[50] we discuss another approach which further characterizes the fractal properties of chromatin texture, called lacunarity analysis, and apply measures of fractal dimension and lacunarity to the classification of breast epithelial lesions.

Acknowledgments

The authors thank Drs. Craig Benham, Joseph Malinsky, Sylvan Wallenstein, and Carol Bodian of the Department of Biomathematical Sciences, Mount Sinai School of Medicine, for their helpful advice and discussion. We also thank our pathology colleagues: Dr. Ira Bleiweiss for histopathologic diagnoses, Dr. David Burstein for cytologic diagnoses, Dr. Grace Yang (New York University Medical Center) and Dr. Syed Hoda (New York Hospital-Cornell Medical Center) for providing materials, and Joseph Silberfarb and Cynthia Sheppard for technical assistance. This material was presented in part at the Second International Symposium on Fractals in Biology and Medicine, Ascona, Switzerland, March 6-9, 1996. Andrew Einstein was supported by a traineeship on NIH MSTP Training Grant GM07280.

References

1. Arman F, Pearce JA. Unsupervised classification of cell images using pyramid node linking. *IEEE Trans Biomed Eng* 1990; **37:** 647-650.

2. Wu H-S, Barba J, Gil J. An iterative algorithm for cell segmentation using short-time Fourier transform. *J Microsc* 1996; **184:** 127-132.

3. MacAulay C, Palcic B. A comparison of some quick and simple threshold selection methods for stained cells. *Anal Quant Cytol Histol* 1988; **10:** 134-138.

4. Einstein AJ, Gil J, Wallenstein S, *et al.* Reproducibility and accuracy of interactive segmentation procedures for image analysis in cytology. *J Microsc* in press.

5. Gimenez-Mas JA, Sanz-Moncasi MP, Remon L, Gambo P, Gallego-Calvo MP. Automated textural analysis of nuclear chromatin: a mathematical morphology approach. *Anal Quant Cytol Histol* 1995; **17:** 39-47.

6. Erler BS, Hsu L, Truong HM, *et al.* Image analysis and diagnostic classification of hepatocellular carcinoma using neural networks and multivariate discriminant functions. *Lab Invest* 1994; **71:** 446-451.

7. Einstein AJ, Gil J. Classification procedures for diagnosis based on multiple morphometric parameters. *Acta Stereol* 1996; **15:** 15-24.

8. Truong H, Morimoto R, Walts AE, Erler B, Marchevsky A. Neural networks as an aid in the diagnosis of lymphocyte-rich effusions. *Anal Quant Cytol Histol* 1995; **17:** 48-54.

9. Deligdisch L, Miranda C, Barba J, Gil J. Ovarian dysplasia: nuclear texture analysis. *Cancer* 1993; **72:** 3253-3257.

10. Hutchinson ML, Schultz DS, Stephenson RA, Wong KL, Harry T, Zahniser DJ. Computerized microscopic analysis of prostatic fine needle aspirates: comparison with breast aspirates. *Anal Quant Cytol Histol* 1989; **11:** 105-110.

11. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

12. Fleming MG, Wied GL, Dytch HE. Image analysis cytometry of dysplastic nevi. *J Invest Dermatol* 1990; **95:** 287-291.

13. Unger PD, Watson CW, Liu Z, Gil J. Morphometric analysis of neoplastic renal aspirates and benign renal tissue. *Anal Quant Cytol Histol* 1993; **15:** 61-66.

14. Doudkine A, Macaulay C, Poulin N, Palcic B. Nuclear texture measurements in image cytometry. *Pathologica* 1995; **87:** 286-299.

15. Haralick RM, Shanumgam K, Dinstein I. Textural features for image classification. *IEEE Trans Systems Man Cybernet* 1973; **3:** 610-621.

16. Pressman NJ. Markovian analysis of cervical cell images. *J Histochem Cytochem* 1976; **24:** 138-144.

17. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975; **4:** 172-179.

18. Julesz B. Textons, the elements of texture perception, and their interactions. *Nature* 1981; **290:** 91-97.

19. Brigham E. Fast Fourier Transform and Its Applications. Englewood Cliffs, NJ: Prentice-Hall, 1988.

20. Serra J. Image Analysis and Mathematical Morphology. London: Academic Press, 1982.

21. Diegenbach PC, Baak JPA. Quantitative nuclear image analysis: differentiation between normal, hyperplastic, and malignant appearing uterine glands in a paraffin section. III. The use of texture features for differentiation. *Europ J Obstet Gynec Reprod Biol* 1978; **8:** 109-116.

22. Beil M, Irinopoulou T, Vassy J, Rigaut JP. Chromatin texture analysis in three-dimensional images from confocal scanning laser microscopy. *Anal Quant Cytol Histol* 1995; **17:** 323-331.

23. Dawson AE, Cibas ES, Bacus JW, Weinberg DS. Chromatin texture measurement by Markovian analysis. Use of nuclear models to define and select texture features. *Anal Quant Cytol Histol* 1993; **15:** 227-235.

24. Mandelbrot BB. The Fractal Geometry of Nature. New York: W. H. Freeman and Company, 1983.

25. Peng CK, Buldyrev SV, Goldberger AL, *et al.* Long-range correlations in nucleotide sequences. *Nature* 1992; **356:** 168-170.

26. Bassingthwaighte JB, King RB, Roger SA. Fractal nature of regional myocardial blood flow heterogeneity. *Circ Res* 1989; **65:** 578-590.

27. Weibel ER. Fractal geometry: a design principle for living organisms. *Am J Physiol* 1991; **261:** L361-369.

28. Landini G, Misson GP, Murray PI. Fractal analysis of the normal human retinal fluorescein angiogram. *Curr Eye Res* 1993; **12:** 23-27.

29. Cross SS. Fractals in pathology. *J Pathol* 1997; **182:** 1-8.

30. Losa GA, Nonnenmacher TF. Self-similarity and fractal irregularity in pathologic tissues. *Mod Pathol* 1996; **9:** 174-182.

31. Losa GA, Baumann G, Nonnenmacher TF. Fractal dimension of pericellular membranes in human lymphocytes and lymphoblastic leukemia cells. *Pathol Res Pract* 1992; **188:** 680-686.

32. Smith TG, Jr., Marks WB, Lange GD, Sheriff WH, Jr., Neale EA. A fractal analysis of cell images. *J Neurosci Methods* 1989; **27:** 173-180.

33. Landini G, Rippin JW. An "asymptotic fractal" approach to the morphology of malignant cell nuclei. *Fractals* 1993; **1:** 326-335.

34. Cross SS, Bury JP, Silcocks PB, Stephenson TJ, Cotton DW. Fractal geometric analysis of colorectal polyps. *J Pathol* 1994; **172:** 317-323.

35. Landini G, Rippin JW. How important is tumour shape? Quantification of the epithelial-connective tissue interface in oral lesions using local connected fractal dimension analysis. *J Pathol* 1996; **179:** 210-217.

36. Fazzalari NL, Parkinson IH. Fractal dimension and architecture of trabecular bone. *J Pathol* 1996; **178:** 100-105.

37. Vilela MJ, Martins ML, Boschetti SR. Fractal patterns for cells in culture. *J Pathol* 1995; **177:** 103-107.

38. Matsuyama T, Matsushita M. Fractal morphogenesis by a bacterial cell population. *Crit Rev Microbiol* 1993; **19:** 117-135.

39. Yang GC, Alvarez II. Ultrafast Papanicolaou stain. An alternative preparation for fine needle aspiration cytology. *Acta Cytol* 1995; **39:** 55-60.

40. Russ JC. Fractal Surfaces. New York: Plenum Press, 1994.

41. Mandelbrot BB. A fractal's lacunarity, and how it can be tuned and measured. In: Nonnenmacher TF, Losa GA, Weibel ER, eds. Fractals in Biology and Medicine. Basel: Birkhäuser Verlag, 1993: 8-21.

42. Dubuc B, Roques-Carmes C, Tricot C, Zucker SW. The variation method: a technique to estimate the fractal dimension of surfaces. *SPIE Visual Communications and Image Processing II* 1987; **845:** 241-248.

43. Dubuc B, Zucker SW, Tricot C, Quiniou JF, Wehbi D. Evaluating the fractal dimension of surfaces. *Proc R Soc Lond A* 1989; **425:** 113-127.

44. Anguiano E, Pancorbo M, Aguilar M. Fractal characterization by frequency analysis. I. Surfaces. *J Microsc* 1993; **172:** 223-232.

45. Voss RF. Random fractal forgeries. In: Earnshaw RA, ed. Fundamental Algorithms for Computer Graphics. Berlin: Springer-Verlag, 1985: 805-835.

46. Wu H-S, Einstein AJ, Gil J. Fractal characterization of irregularly shaped images by frequency analysis. *Submitted*.

47. Orlando V, Paro R. Chromatin multiprotein complexes involved in the maintenance of transcription patterns. *Curr Opin Genet Dev* 1995; **5:** 174-179.

48. Pentland AP. Fractal-based description of natural scenes. *IEEE Trans Pattern Anal Machine Intell* 1984; **PAMI-6:** 661-674.

49. Cox BL, Wang JSY. Fractal surfaces: Measurement and applications in the earth sciences. *Fractals* 1993; **1:** 87-115.

50. Einstein AJ, Wu H-S, Gil J. Fractal characterization of chromatin appearance for diagnosis in breast cytology: II. Lacunarity analysis and classification approaches. *Submitted* .

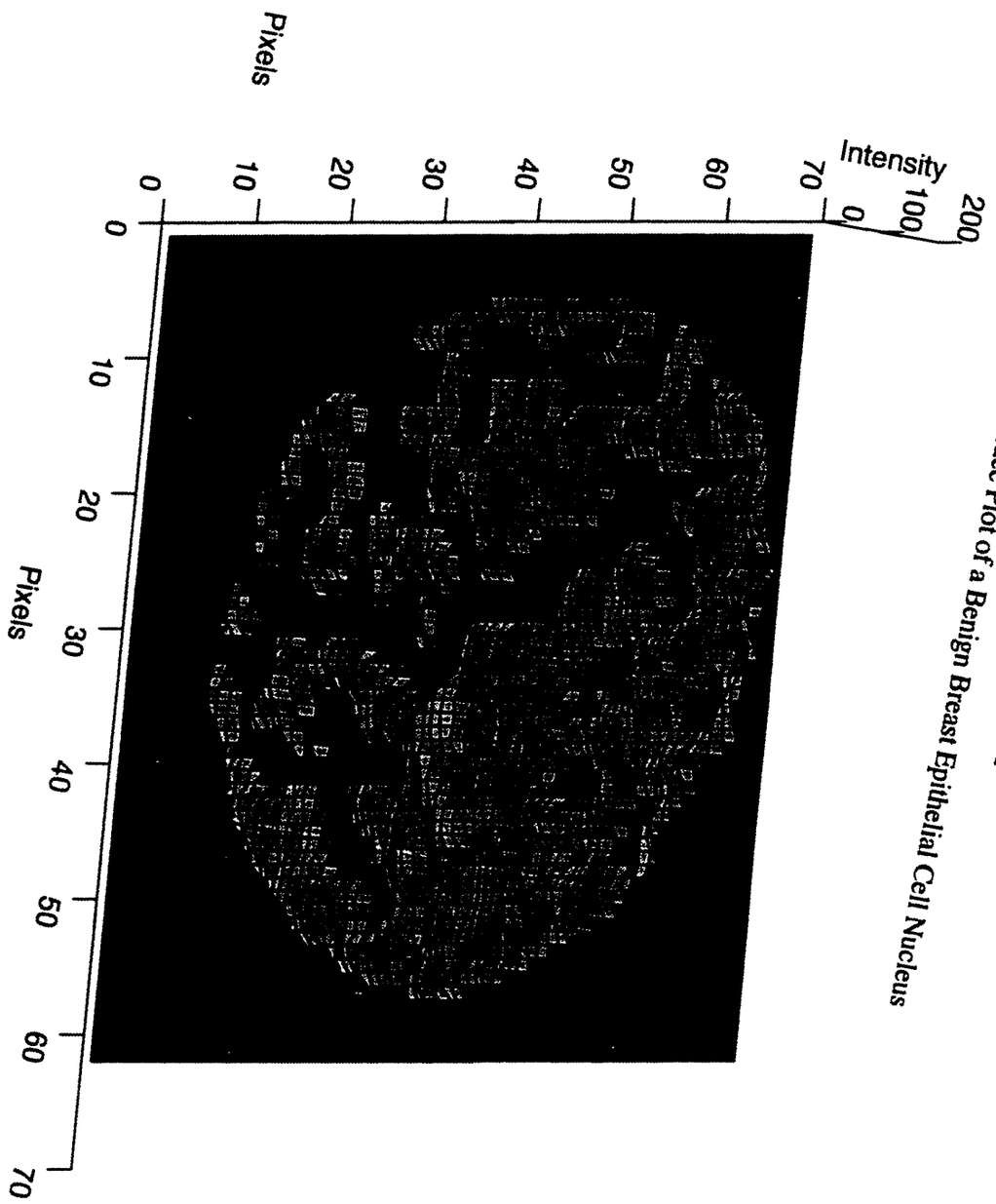


Figure 1

Surface Plot of a Benign Breast Epithelial Cell Nucleus

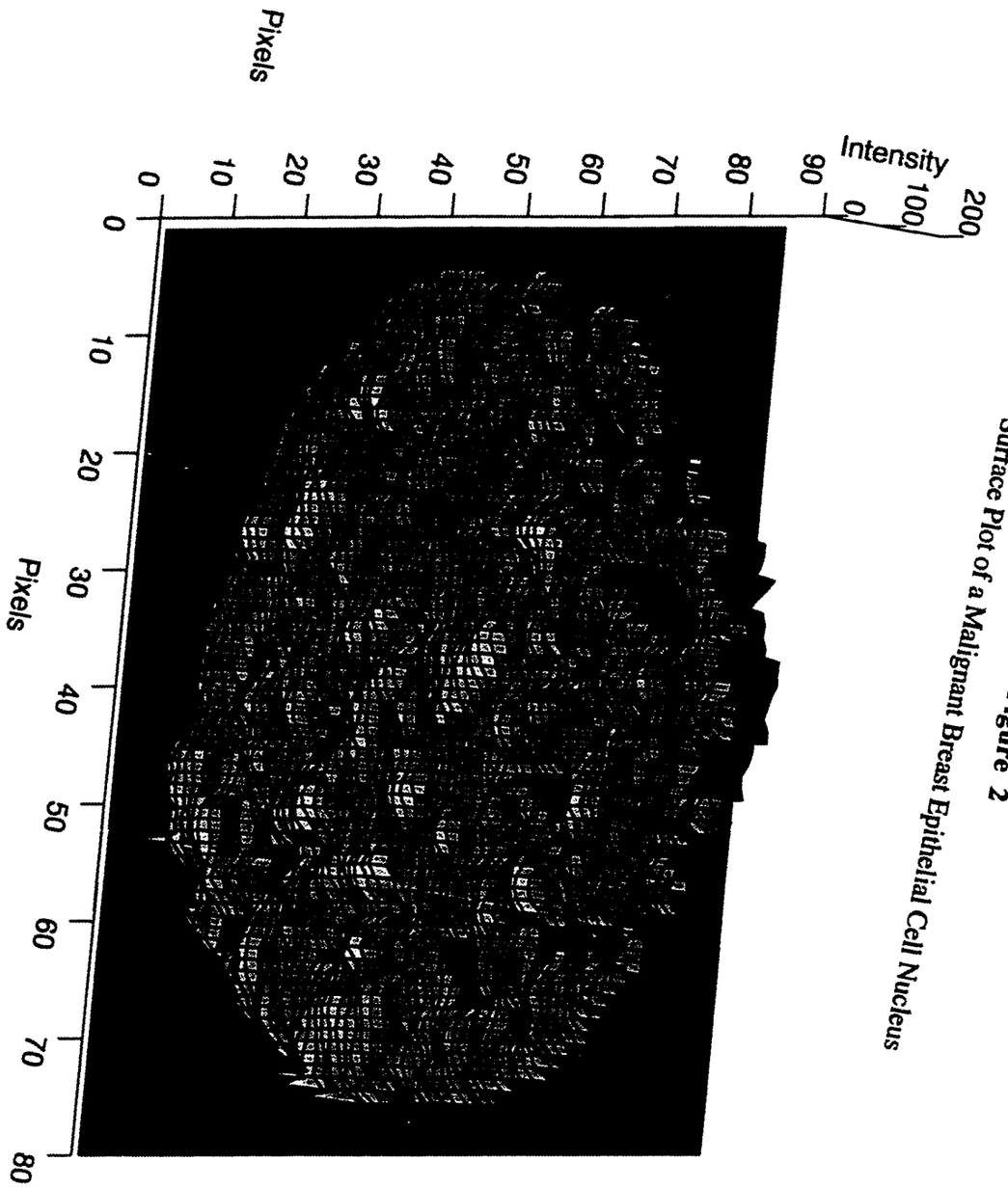


Figure 2

Surface Plot of a Malignant Breast Epithelial Cell Nucleus

Figure 3

Variation Method Plot for a Typical Malignant Breast Epithelial Cell Nucleus

$D_{MB} = 2.49$, $r = 0.991$

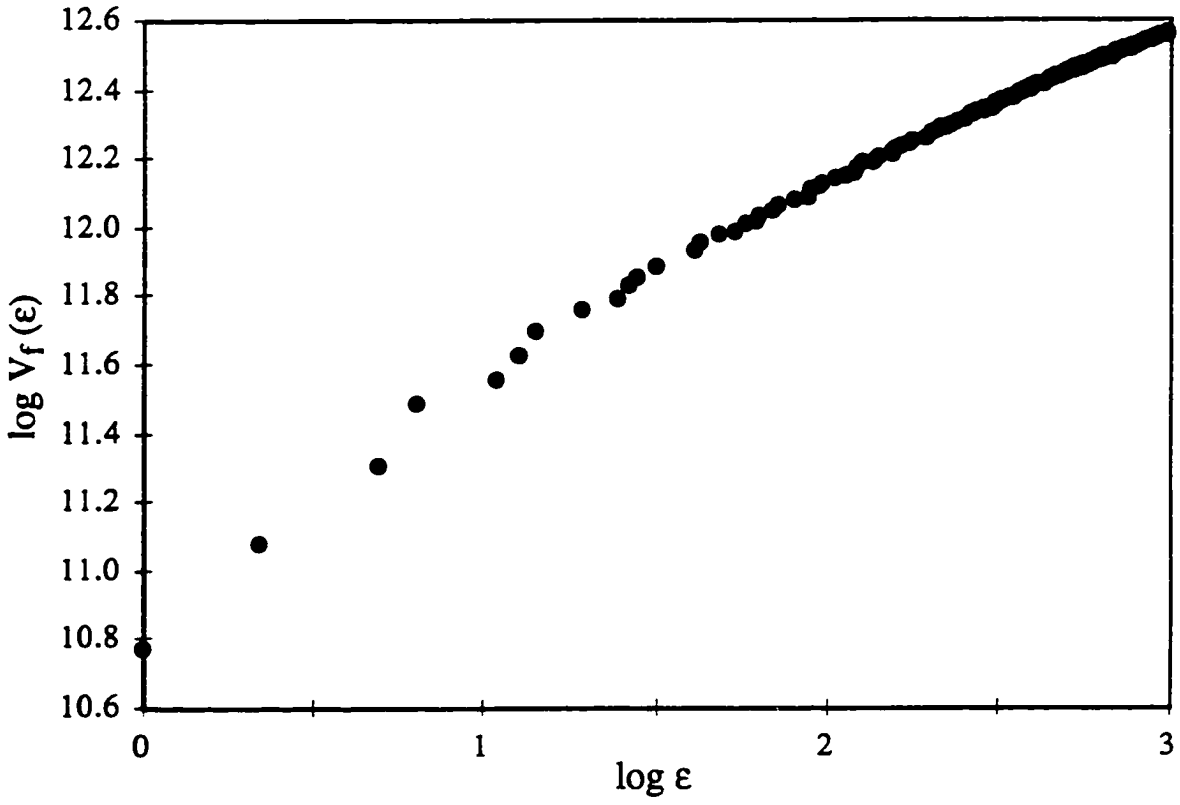


Figure 4

Mean Minkowski Dimensions for the 41 Cases

Horizontal lines represent mean values.

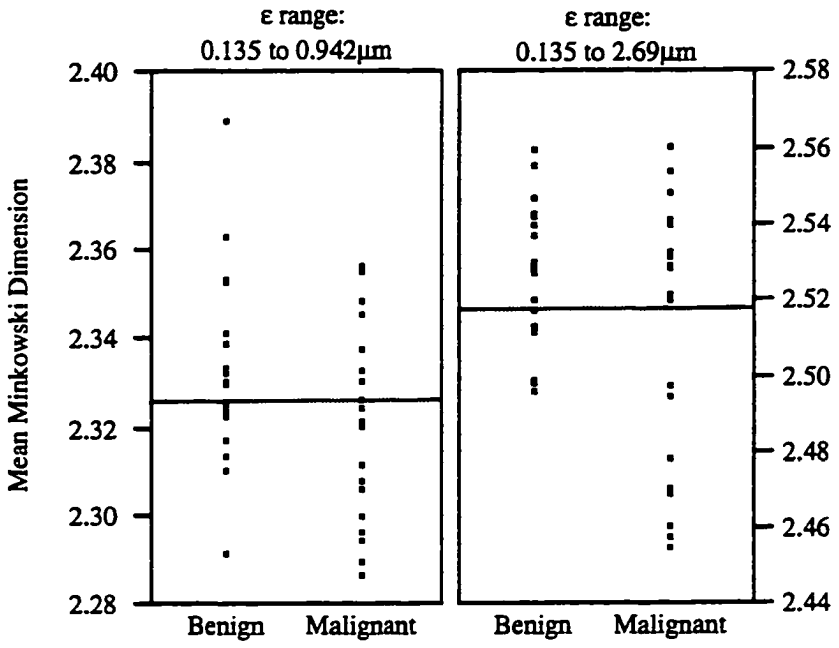


Figure 5

Mean Spectral Dimensions for the 41 Cases

Horizontal line represents mean value.

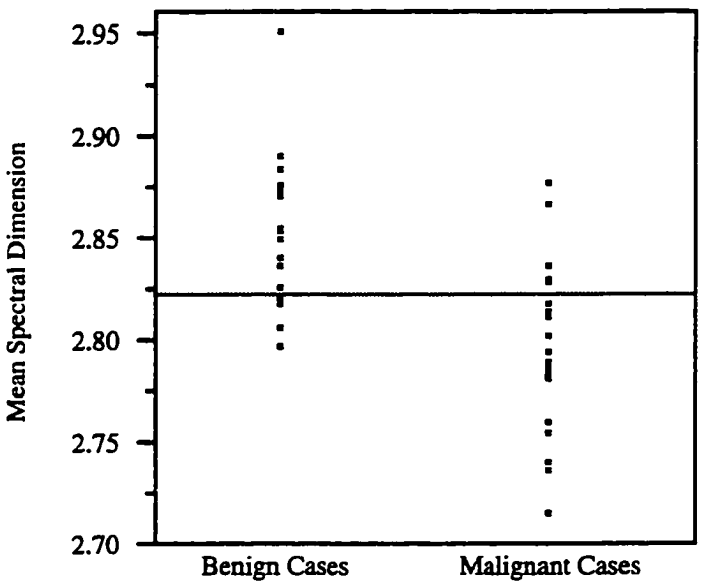
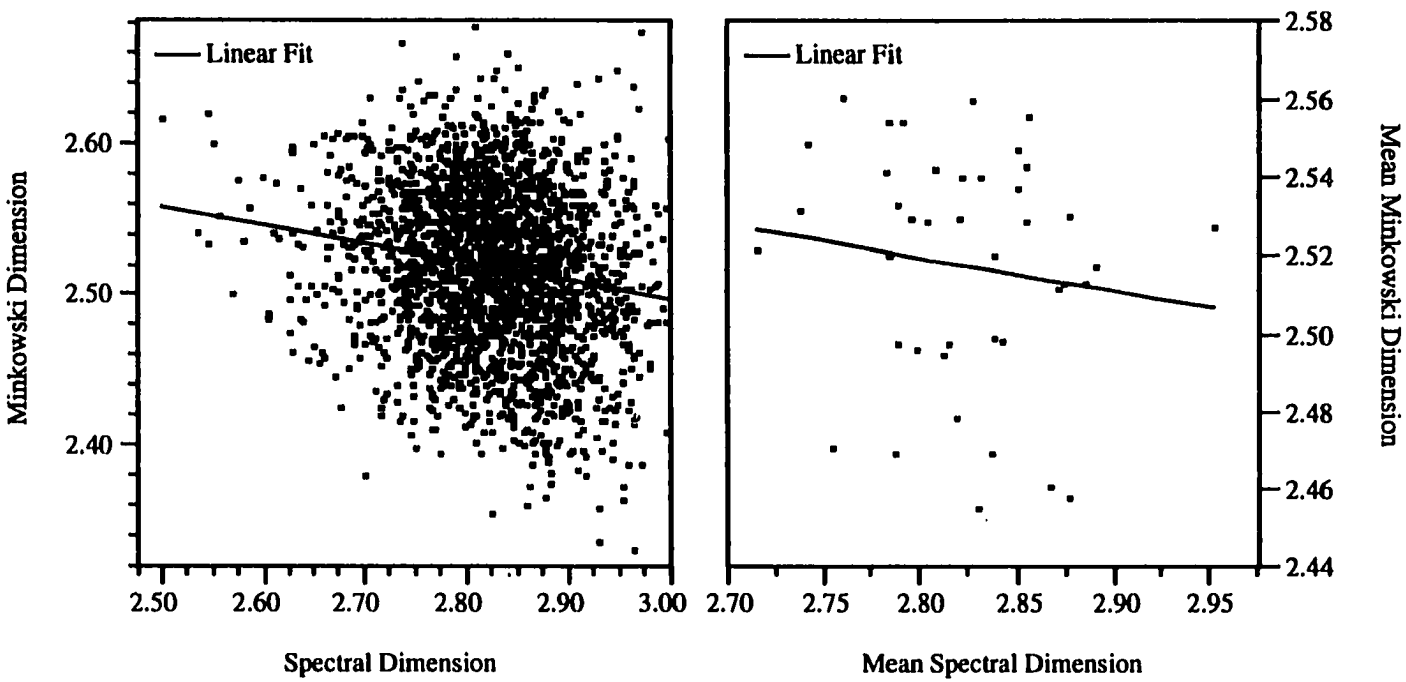


Figure 6

Correlations between Minkowski and Spectral Dimensions



CHAPTER 7

Fractal Characterization of Chromatin Appearance for Diagnosis in Breast Cytology: II. Lacunarity Analysis and Classification Approaches

Andrew J. Einstein

Hai-Shan Wu

Joan Gil

Submitted

Preface to Chapter 7

Chapter 7 is a continuation of Chapter 6. The first half of the chapter focuses on the concept of lacunarity, introducing a method for measuring lacunarity and applying it to characterize both a class of mathematical fractals called Sierpinski carpets, and benign and malignant breast epithelial cell nuclei. The second half of the chapter focuses on classification. Using the fractal texture measures presented in Chapters 5 through 7 and nuclear area, logistic regression and neural network classification models are developed for the diagnosis of breast epithelial cell lesions. Overtraining is addressed, and the chapter concludes with a discussion of issues involved in the development of image analysis devices for quality assurance in breast cytology.

Summary

Previously, we showed that chromatin texture in breast epithelial cells is fractal, and that fractal dimensions differ between benign and malignant cells. In this study we investigate lacunarity, a fractal property characterizing the size of holes or gaps in a texture. Pairs of mathematical fractals and nuclei with the same fractal dimensions are illustrated, and it is shown how the different textures within each pair can be distinguished on the basis of lacunarity. Lacunarity is compared between 19 benign cases and 22 cases with invasive ductal carcinomas. A statistically significant difference in lacunarity between the benign and malignant cases is found over a wide range of scales. These differences are particularly pronounced at the smallest and largest scales, corresponding respectively to fine-scale texture, indicating whether chromatin is clumped or fine, and to large-scale structures like nucleoli. Logistic regression and artificial neural network classification models are developed to classify unknown cases on the basis of fractal measures of chromatin texture. Using leave-one-out cross-validation, the best logistic regression classifier correctly diagnoses 95.1% of the cases. The best neural network model can correctly classify all of the cases, but it is unclear whether this is due to overtraining. Fractal dimensions and lacunarity are useful tools for the quantitative characterization of chromatin appearance, and can potentially be incorporated into image-analysis devices to assure the quality and reproducibility of diagnosis by breast fine-needle aspiration biopsy.

Key Words: fractals, lacunarity, logistic regression, artificial neural networks, morphometry, image analysis, breast cytology

Introduction

Fractal analysis provides a set of mathematical tools to concisely describe structures displaying similar detail over a range of scales. Recent work has demonstrated that many pathologies can be characterized in terms of their fractal properties. Fractal dimensions have been shown to differ from their normal values for the epithelial-connective tissue interface in squamous cell carcinomas of the floor of the mouth,[1] for pulmonary arteries in pulmonary hypertension,[2] and for adenomatous colorectal polyps.[3] We have demonstrated that fractal dimensions of chromatin texture differ between benign and malignant breast epithelial cell nuclei.[4] By viewing images of the nuclei as surfaces, fractal dimensions can characterize aspects of surface texture, which differs between the benign and malignant cases.

Fractal dimensions quantify textural complexity and irregularity and therefore can be used to discriminate between diagnostic categories for which these properties differ. However, they do not provide a unique description of an entire textured surface, for two fractals can have strikingly different appearances yet still possess the same fractal dimensions. In image analysis-based diagnostic approaches, fractal dimensions are useful as descriptors to be entered along with others in a multivariate classification algorithm. Another fractal measure that intuitively appears to be an attractive candidate to be added to the list of descriptors extracted from the nucleus for diagnostic purposes is lacunarity. Mandelbrot[5] introduced the notion of lacunarity to describe one particular aspect of the texture of a fractal: the largeness of its gaps or holes. As such, lacunarity would appear to capture features such as voids, chromatin clearings, and nucleoli, which pathologists regard as important in the diagnosis of malignancy. In this paper, we investigate the lacunarity of chromatin in benign and malignant breast epithelial cell nuclei, and show how fractal dimensions and lacunarity can be used in conjunction with logistic regression and artificial neural network classifiers to diagnose breast lesions.

Materials and Methods

Lacunarity Analysis

The notion of lacunarity can be illustrated using a class of mathematical objects called Sierpinksi carpets. In this section, we shall present examples of Sierpinski carpets which, though grossly different in appearance, have identical fractal dimensions. After introducing a method for measuring lacunarity, we will show how these Sierpinski carpets differ in terms of lacunarity. Finally, we will return to images from pathology and demonstrate the potential of lacunarity in distinguishing between nuclei of similar fractal dimensions.

Figure 1 shows two Sierpinski carpets. A Sierpinski carpet[6] is constructed[7,8] by beginning with a square, and subdividing it into b^2 subsquares, out of which l^2 subsquares are cut. This process can be iteratively repeated on each of the remaining subsquares, at each stage removing the l^2 subsquares from the same positions. Figure 1 shows three stages of Sierpinksi carpets. It can be mathematically shown that the fractal (Hausdorf) dimension of a Sierpinski carpet is given by the formula[7]

$$D = \log(b^2 - l^2) / \log(b) .$$

Since $b = 7$ and $l = 3$ for both the Sierpinski carpets in Figure 1, they have a common fractal dimension of log 40 / log 7 = 1.896. They differ in terms of which subsquares are removed. While fractal dimension is insufficient to discriminate between these carpets, the holes are larger in the carpet on the left, while the carpet on the right is more homogeneous. This difference can be quantified in terms of lacunarity.

Numerous expressions for lacunarity have been suggested in the physics literature. The approach we follow here is an based on the gliding box method of Allain and Cloitre.[9] This method has recently been applied successfully to the characterization of landscape textures[10,11] and ecological data,[12] and Allain and Cloitre claim that it overcomes difficulties of other lacunarity measures in describing a wide range of textures. The gliding

box method is designed for binary images, *i.e.*, images in which all pixels take the value 0 or 1, and while it can be modified for grayscale data this results in a loss of resolution.[12] Nuclear images can be binarized by thresholding: all pixels with gray values greater than a specified level take the value 1 (white) while the remaining pixels are assigned to 0 (black).

In this method, lacunarity is measured using an $s \times s$ pixel "gliding box." The box is initially placed at the upper left corner of the image. The number of white pixels in the image contained in the gliding box is counted. This value is denoted n_1 and can take any value from 0 to s^2. The gliding box is moved one pixel to the right, and the number of white pixels at this new location, denoted n_2, is counted. In a similar manner, the box glides over the entire image, moving to all positions at which it covers at least one pixel of the image, at each location recording the number of white pixels n_i in the image. The sequence of values $\{n_i\}$ (for i between 1 and N, the number of positions taken by the gliding box) defines a probability distribution $Q_n(M, s)$ which represents the probability that a gliding box of side s contains M white pixels. Lacunarity is defined in terms of the first (mean) and second statistical moments of the distribution $Q_n(M, s)$. The moments are given by the equations

$$Z_{Q_n}^{(1)}(s) = \sum_{M=1}^{s^2} M\, Q_n(M, s)$$

and

$$Z_{Q_n}^{(2)}(s) = \sum_{M=1}^{s^2} M^2\, Q_n(M, s) \ ,$$

and *lacunarity* is defined as

$$\Lambda(s) = Z_{Q_n}^{(2)}(s) \,/\, [Z_{Q_n}^{(1)}(s)]^2 \ .$$

Thus, lacunarity does not take a single value but is a function of the box size.

A difficulty may arise with the gliding box method when working with irregularly shaped images such as nuclei, and for this reason we introduce the weighted lacunarity. At many positions, in particular near the periphery of an image, many of the pixels in the

gliding box will be outside of the image. For example, for a nuclear image, when the gliding box is near the nuclear membrane, most of the pixels in the gliding box can fall in the cytoplasm, outside of the nuclear image. The values of n_i at these positions will be low, not because there are so many black pixels, but simply because there are few pixels of the image within the gliding box at these locations. The net result can be to skew the distribution $Q_n(M, s)$. To correct for this problem, we incorporate a weighting factor w_i, equal to the number of pixels in the gliding box that are part of the image. This weight can take any value between 0 and s^2. We define x_i, the *weighted* number of white pixels in the image contained in the gliding box, by the formula

$$x_i = n_i \, (s^2 / w_i) \, .$$

Weighted lacunarity can be defined in terms of the "weighted moments" of the distribution $Q_x(M, s)$ of $\{x_i\}$. The rationale for weighting the moments as well as the number of pixels is that we do not wish a position for which the gliding box contains only a few pixels in the image to contribute as much to the lacunarity as a position for which it contains a many pixels. Instead of summing over the number of white pixels, moments of $Q_x(M, s)$ can be calculated by summing over the N positions taken by the gliding box, *i.e.*,

$$Z_{Q_x}{}^{(q)}(s) = \sum_{M=1}^{s^2} M^q \, Q_x(M, s) = \sum_{i=1}^{N} x_i^q / N = \sum_{i=1}^{N} x_i^q \Big/ \sum_{i=1}^{N} 1 \, .$$

Weighted moments $Z'_{Q_x}{}^{(q)}(s)$ weight these terms:

$$Z'_{Q_x}{}^{(q)}(s) = \sum_{i=1}^{N} w_i \, x_i^q \Big/ \sum_{i=1}^{N} w_i \, .$$

Weighted lacunarity is defined as

$$\Lambda'(s) = Z'_{Q_x}{}^{(2)}(s) \, / \, [Z'_{Q_x}{}^{(1)}(s)]^2 \, ,$$

It can be shown mathematically that this is equivalent to

$$\Lambda'(s) = \sum_{i=1}^{N} w_i \sum_{i=1}^{N} (n_i^2 / w_i) \Big/ \left(\sum_{i=1}^{N} n_i \right)^2 ,$$

a form which is easy to implement on a computer.

Weighted lacunarity curves for the Sierpinski carpets in Figure 1 are shown in Figure 2. While these fractals have the same dimensions, they can be differentiated on the basis of the lacunarity curves. Thus, lacunarity complements fractal dimensions in characterizing texture. In fact, lacunarity is a useful method not just for Sierpinski carpets but for real-world data as well. Figure 3 illustrates a pair of nuclei with the same spectral fractal dimensions (2.36) and Minkowski dimensions (2.82) over the range 0.135 to 0.942μm, together with their normalized weighted lacunarity curves. The nucleus on the left has a prominent nucleolus while the nucleus on the right does not, so we would expect the nucleus on the left to be more lacunar. This is confirmed by the lacunarity curves.

Cytologic and Histologic Materials

Ultrafast Papanicolaou[13]-stained cytology specimens and corresponding histopathology materials were obtained from 41 patients (22 invasive ductal carcinomas, 19 benign with no atypia, based on consensus of cytology and histopathology).

Image Analysis

Image acquisition was performed on a self-assembled system based on a Gateway 2000 486DX2/50E microcomputer augmented with Sprynt i860 image processing boards and the Semper 6 Plus graphics program, connected to a Nikon Optiphot microscope equipped with a 100× Nikon Plan objective through a Sony DXC-M2 camera. For each specimen, randomly selected epithelial cell nuclei were segmented using an arc-forming method,[14] excluding overlapping and damaged nuclei as well as those with insufficient contrast. Subsequent image processing and determination of weighted lacunarities and nuclear areas was performed using software written in C and run on the Silicon Graphics Indigo^2 workstation and Power Challenger XL supercomputer. Nuclear images were normalized[15] to compensate for possible differences in staining and lighting conditions and then screened to assure that they were of adequate quality for texture analysis. A total of

2621 images was collected, with a minimum of 55 and a mean of 64 nuclei per patient. Lacunarity was determined for each nucleus at 20 box sizes, ranging from $s = 1$ to $s = 35$ pixels (0.134 to 4.71 μm), and at three threshold levels: the first, second, and third quartiles of the nucleus' gray level histogram.

Statistical Analysis

Statistical analysis was performed in Microsoft EXCEL on an Apple Power Macintosh 8100/110. Mean areas and lacunarities at each box size and threshold were determined for each patient. Patient means of lacunarity were compared between benign and malignant cases at each box size and threshold using Student's *t*-test.

Classificatory Approaches: Logistic Regression and Neural Networks

A variety of classificatory algorithms, such as logistic regression, discriminant analysis, and artificial neural networks, can be used to diagnose cases on the basis of morphometric descriptors (also referred to as features) such as fractal dimensions, lacunarities, and nuclear area.[16] Using a training set of cases with known features and diagnoses, a classification algorithm can be trained or optimized for the particular diagnostic setting of interest. This process involves determining an optimal set of features, (*e.g.*, mean spectral dimension and mean nuclear area) and parameters (*e.g.*, regression coefficients in logistic regression or neuron weights and number of hidden neurons in an artificial neural network) in the classificatory model. Subsequently, unknown cases can be diagnosed using the trained classifier. A useful approach to evaluate the efficacy of a classificatory algorithm for a data set is to perform leave-one-out cross-validation.[17] This is often referred to as jackknife analysis,[18] and here we use the terms interchangeably. Since here we consider 41 cases, a single run of a jackknife analysis involves training a classifier on 40 of the cases and testing the accuracy of its diagnosis on the remaining "masked case", repeating this process 41 times so that each case serves as the masked case.

We performed jackknife analyses using logistic regression models[19] with a variety of fractal features as well as measures of nuclear area. The jackknife analysis was repeated for each combination of features considered, except in a few instances where it was only necessary to consider a few patients to verify that a combination of features resulted in poor classificatory performance. In the terminology of logistic regression, the "covariates" were the features while the "independent variable" was the diagnosis. The fractal textural features included means and standard deviations of spectral and Minkowski-Bouligand fractal dimensions, of weighted lacunarity at a variety of box sizes, and of the multiplicative prefactor C, which can be viewed as a measure of lacunarity in the frequency domain. Weighted lacunarity measures considered included weighted lacunarity $\Lambda'(s)$, log weighted lacunarity $log[\Lambda'(s)]$ (also referred to as log $\Lambda'(s)$), normalized weighted lacunarity $\Lambda'(s)/\Lambda'(1)$ (also referred to as normalized $\Lambda'(s)$), and normalized log weighted lacunarity $log[\Lambda'(s)]/log[\Lambda'(1)]$ (also referred to as normalized log $\Lambda'(s)$). The weighted lacunarity measures considered were further restricted on the basis of the results of the statistical comparisons of these features between benign and malignant cases. Measures of nuclear area considered were the mean, minimum, and standard deviation per patient. Logistic regression was performed using JMP (SAS Institute, Cary, NC) on an Apple Power Macintosh 8100/110. The various classificatory models were evaluated in terms of accuracy, sensitivity, specificity, predictive values, and kappa (κ) statistics. In some models, the classifier was unable to predict the diagnosis for certain patients, resulting in three categories of "diagnoses": benign (negative test), malignant (positive test), and no prediction (uncertain test). For these models, sensitivity was determined as the ratio of the number of true positives to the total number of patients with a malignant lesion (true positives, false negatives, and malignant cases with uncertain results); specificity was defined as the ratio of the number of true negatives to the total number of patients with a benign lesion (true negatives, false negatives, and benign cases with uncertain results).[20]

Jackknife classification was also performed with backpropagation artificial neural networks.[21] General overviews of neural networks and their applications to pathology can be found in several recent reports.[18,22-24] Several neural network models were compared, differing from each other on the basis of architecture, the selection of input neurons, *i.e.*, area and fractal texture features, and the coding of diagnoses in output neurons. Neural networks were trained using the Levenberg-Marquardt learning rule,[25-27] which is markedly faster than standard conjugate gradient methods, thus enabling more neural network models to be considered. The models were implemented in the MATLAB Neural Networks Toolbox (The MathWorks, Inc., Natick, MA) on several Silicon Graphics computers, including a Power Challenger XL supercomputer.

Results

Lacunarity

Figure 4 illustrates weighted lacunarities of the 41 cases thresholded at the third quartile of the intensity histogram, as a function of box size. At each scale, all of the most lacunar cases are seen to be malignancies, and patient means of weighted lacunarities are greater in the malignant cases than in the benign ones. Figure 5 demonstrates the significance of these differences. For the images thresholded at the third quartile, the differences in patient means are significant at almost all box sizes, and the maximum p value of 0.0505 at $s = 6$ pixels (0.807 μm) is barely above the cutoff for significance. The difference is especially significant at the smallest and largest scales. These correspond respectively to fine-scale texture, such as whether chromatin is clumped or fine, and to large-scale structures like nucleoli. While increased nucleolar number or size is not pathognomonic for malignancy, nucleolar alterations provide diagnostically important information for Pap-stained breast cytology specimens[28]: eosinophilic macronucleoli (greater than 1.25 μm in diameter)[29] as well as excessive variability in size, shape, and

numbers of nucleoli[30] tend to be associated with malignancy. Differences in weighted lacunarity between benign and malignant cases are not as pronounced when the images are segmented at the first and second quartiles of their intensity histograms. While the plot of p as a function of s follows the same shape at the three thresholds, p values are higher at the second quartile than at the third quartile, and even higher at the first quartile.

Logistic Regression

The results of the jackknife analyses of logistic regression classification performance are summarized in Tables 1 through 3.

As lacunarity is not described by a single number as is, for example, spectral dimension, but rather, by a function, it was necessary to restrict the lacunarity measures considered in the logistic regression models to a few well-chosen ones. Since Figure 5 reveals that differences in lacunarity between benign and malignant cases are more pronounced when images are thresholded at the third quartile of the intensity histogram than at the first or second quartiles, the only lacunarity measures considered were those associated with third quartile thresholding. Further, since these differences are most pronounced at the smallest and largest scales considered (2 and 35 pixels), the only lacunarity measures incorporated into the classificatory models were for box sizes of 2 and 35 pixels.

Even having limited the number of lacunarity measures considered, it is still unfeasible to perform the jackknife analysis using all possible combinations of features to determine the group of features for which performance is best. Instead, it is necessary to adopt a strategy to achieve good performance while limiting the number of regressions performed. We began with a model including as covariates, members of the three classes of textural descriptors for which the difference between benign and malignant cases was most statistically significant: mean spectral dimension, standard deviation of the multiplicative prefactor C, and mean lacunarities at a box size of 35 pixels. As is shown in

Table 1, the best performance among the three-feature models was attained by a model including mean spectral dimension, standard deviation of C, and mean normalized log $\Lambda'(35)$. For this model there were three false positive diagnoses and one false negative, so 37 of the 41 cases were correctly diagnosed. Classificatory performance was poorer in models including just two of these textural features.

Our next attempt to improve classification performance, summarized in Table 2, was to add a fourth feature to the models, also chosen from those demonstrating a significant difference between benign and malignant cases. The added features included the standard deviations of the (four) weighted lacunarity functions at a box size of 2 pixels, and mean nuclear area. In one model there were only two false positives; however in this model there were also four cases for which maximum likelihood estimates of parameters in the model failed to converge, making it impossible to classify the cases as benign or malignant. Interestingly, while the three-feature model including mean normalized log $\Lambda'(35)$ performed better than the three-feature model including mean normalized $\Lambda'(35)$, the opposite was true for the four-feature models, where each model including mean normalized $\Lambda'(35)$ outperformed its counterpart with mean normalized log $\Lambda'(35)$. This finding suggests that a stepwise approach[19] to feature selection is likely to miss optimal sets of features in such classificatory scenarios. Thus, it is necessary to use a more flexible approach to feature selection, as we have done.

Continuing further, we considered five-feature models, adding both one of the four $\Lambda'(2)$ measures as well as mean area. They are summarized in Table 3. The best performance was attained by a model including mean normalized $\Lambda'(35)$ and the standard deviation of normalized log $\Lambda'(2)$. In this "best" model, all cases were classified correctly except for two false negatives. Additionally, other variations of five-feature models were considered, including features such as mean $\Lambda'(35)$, mean log $\Lambda'(35)$, Minkowski-Bouligand dimension, and minimum nuclear area, but none performed as well.

Finally, we considered some six-feature models, adding features such as Minkowski-Bouligand dimension and minimum nuclear area to the best five-feature model, but none improved on the performance of the best five-feature model. Thus, in this study, the highest accuracy attained using logistic regression to diagnose the breast cytology specimens was 95.1%.

Artificial Neural Networks

While the logistic regression models considered differed in terms of the features used as covariates, artificial neural network models can differ in terms of a host of parameters.[31] These include not only the features included as inputs, but also the number and arrangement of hidden layers, the coding of diagnoses, the transfer function(s) used, training and testing tolerances, the order in which training facts are presented, and the initial values chosen (typically at random) for the neuron weights. We began by considering neural networks with the five features in the "best" logistic regression model assigned to input neurons. Output diagnoses were coded both on a 0–1 scale, where 0 corresponded to a benign and 1 to a malignant diagnosis, and also on a -1–1 scale, where -1 corresponded to a benign and 1 to a malignant diagnosis. The number of hidden neurons was varied systematically from 2 to 20. Initially, a variety of transfer functions was considered, including linear, log-sigmoid, and tan-sigmoid (hyperbolic tangent) transfer functions. However, in these initial runs, convergence of the neural network weights was reliably attained only when transfer functions in both layers were tan-sigmoid, so subsequent runs were restricted to networks with this transfer function.

Jackknife analyses for neural networks with these five features are summarized in Table 4. Every entry in the table represents the number of incorrect diagnoses for a single jackknife analysis with the specified output coding and number of hidden neurons. It is evident from the table that better performance was attained using the -1–1 scale. Another feature of neural network training illustrated by the table is that repeating a jackknife

analysis need not yield identical performance. Due to the different initial values chosen at random for the neuron weights, repeat analyses resulted in differences of up to three misclassifications. Nevertheless, the two jackknife analyses with the highest classificatory accuracies were the two runs with 8 hidden neurons. These resulted in two and three of the 41 cases misclassified, performance comparable to that in the best logistic regression classification model.

Since the best features for neural network classification need not be the same as those for logistic regression classification, we next considered neural network models with a variety of features. Two hundred combinations of features, denoted "feature sets," were considered. Each of these feature sets included mean spectral dimension and mean nuclear area, and from 0 to 5 additional features:

1) mean Minkowski dimension (yes/no): 2 options
2) standard deviation (s.d.) of spectral dimension (yes/no): 2 options
3) mean multiplicative prefactor C (yes/no): 2 options
4) mean of (log/not log) (normalized/unnormalized) $\Lambda'(35)$ (yes/no): 5 options
5) s.d. of (log/not log) (normalized/unnormalized) $\Lambda'(2)$ (yes/no): 5 options.

Jackknife analyses were performed using neural networks with 8 hidden neurons, tan-sigmoid transfer functions, and each of the 200 feature sets. For two of the feature sets, only one of the 41 cases was misclassified, while for 11 feature sets there were two misclassifications, as in the "best" feature set for logistic regression. One feature set with only one misclassification included mean and s.d. of spectral dimension, mean area, mean normalized $\Lambda'(35)$, and s.d. of log $\Lambda'(2)$, while the other included mean and s.d. of spectral dimension, mean Minkowski dimension, mean area, mean log $\Lambda'(35)$, and s.d. of $\Lambda'(2)$.

Since Table 4 showed that there is some inter-run variability in neural network accuracy, owing to the dependence of the neuron weights converged upon on their randomly chosen initial values, we next repeated the jackknife analyses for the 13 features

sets with one or two misclassifications. For each feature set, the jackknife analysis was repeated 100 times. As observed previously, the number of misclassifications varied from run to run. Fig. 6 is a histogram of the number of misclassifications for 100 jackknife analyses using a neural network with mean spectral dimension, mean area, mean normalized $\Lambda'(35)$, and s.d. of normalized log $\Lambda'(2)$ as features. While in the original run using this feature set there were two misclassifications, over the 100 runs there were four misclassifications on average. However, for one run there were no misclassifications, and for two runs there was only a single misclassified case. This performance was typical for the 13 feature sets repeated.

Discussion

This study demonstrates that in our material, lacunarity differs between benign and malignant breast epithelial cell nuclei. Lacunarity provides useful diagnostic information, complementing fractal dimensions in characterizing fractal properties of chromatin texture in benign and malignant breast epithelial cell nuclei. Optimal feature sets for both logistic regression and neural network classifiers included lacunarity features, suggesting that fractal dimensions are insufficient to entirely characterize the fractality of nuclear texture. As mentioned above, the difference in weighted lacunarity between benign and malignant cases was especially significant on scales corresponding to fine-scale texture, such as chromatin coarseness, and to large-scale structures like nucleoli. The ubiquity of these qualitative features in pathologic diagnosis suggests that lacunarity, as a numerical measure of them, should have a broad range of applications in quantitative diagnostic pathology.

The logistic regression model with the optimal feature set resulted in classificatory accuracy exceeding 95%. This model shows particular promise and would be a good candidate for clinical trials to evaluate the diagnostic potential of fractal measures of chromatin appearance. The presence of jackknife runs with zero and one misclassifications

indicates that an optimized neural network can improve on logistic regression classification. However, this improved performance requires further validation since it could be attributable to overtraining.[16,32] Neural network overtraining is most commonly manifested in terms of a disparity in classifier performance between training and test sets. Two competing factors are at work in neural network training: 1) the network learns general characteristics of the classificatory task, and 2) the network learns particular characteristics of the members of the training set. As the criterion for ending neural network training becomes more stringent, and neural network training is allowed to proceed longer, the balance generally shifts towards the latter factor.[32-34] Consequently, the performance on new cases, such as the masked case in our jackknife analysis, for which the particular characteristics of the training set are not applicable, can be markedly less.

This "classical" form of overtraining is not present in our study. The optimized neural network achieved a 100% classification rate for test cases on which it had not been trained. Nevertheless, a more subtle form of overtraining may come into play. By considering many neural network models, the networks at each step in a single jackknife analysis may still focus on particular characteristics of the masked case. Although a network cannot learn particulars of the masked case through its training process, it may effectively "learn" these particulars through a process of selection. By considering for each case 100 nearly identical versions of the same network, differing only in the randomly selected initial weights, some of these may be predisposed to converging to neuron weights resulting in a correct classification for the masked case. A jackknife analysis in which such neural networks are trained for all of the most diagnostically challenging cases may exhibit perfect classificatory performance, although some of the networks trained in this jackknife run may not correctly diagnose further new cases. The scope of this potential problem remains unclear, and the issue of overtraining in repeated jackknife analyses requires further study. While a literature exists on neural network generalization, these studies

typically address the classical form of overtraining. For practical purposes, the generalizability of the "optimal" neural networks trained in the best jackknife analyses could be evaluated through prospective testing on new cases.

Thus, this study shows how fractal texture features can be used in the cytologic diagnosis of breast cancer to attain a diagnostic accuracy of 95.1% using logistic regression, and potentially approaching 100% using neural networks. Other investigators have also developed image analysis-based systems for the cytologic diagnosis of breast epithelial cell lesions. King *et al.*,[35] using texton[36]-based measures of chromatin texture[37] and stepwise discriminant analysis for classification, attained an accuracy of 92.3%, but a sensitivity of only 78.6%. Hutchinson *et al.*[38] used run-length[39] texture features as well as features from a low-resolution contextual analysis, combined with stepwise discriminant analysis for classification. Excluding fibroadenoma cases, this method correctly classified 93% of cases. Wolberg *et al.*[40,41] quantified texture using a measure of gray scale variation, and classified cases using MSM-Tree, a decision tree method. Tenfold cross-validation, which is comparable to the jackknife analysis but excludes one tenth of the cases from the training set at a time, rather than a single case at a time, estimated accuracy at 97.2%; a subsequent prospective study correctly classified all the benign and malignant cases. However, not all the benign masses in this study were histologically confirmed.

Recently there has been considerable interest within the cytology community in image analysis devices for quality assurance, particularly in the area of cervical cytology, where the United States Food and Drug Administration has approved two instruments for rescreening.[42,43] A review[44] of 83 published series on fine-needle aspiration (FNA) biopsy of the breast reported false-negative rates ranging from 0 to 35%, indicating that breast FNA diagnostic performance varies widely between institutions. Thus, ancillary image analysis techniques could serve an important role for quality assurance in this setting as well. Although diagnostic accuracy has been quite high in the pilot studies using image analysis for breast FNA diagnosis, it is commonly observed that test efficacy in practice is

lesser than that observed in pilot studies; this phenomenon may be attributable to a number of reasons, such as the broader spectrum of disease observed in practice and the differing conditions under which a test may be administered.[45-47] Ultimately, the most effective approach may well be some hybrid combining the best features of the various methods developed to date. In this study, we limited the number of features considered since our purpose was to test the usefulness of fractal texture descriptors. Nevertheless, the inclusion of other descriptors may contribute to higher diagnostic performance in practice. Observations in past studies suggest that, in many cases, features based on the size distribution of nuclear areas are extremely useful in cytologic diagnosis.[48] Other useful features may include contextual descriptors, such as those considered by Hutchinson *et al.*,[38] and other classes of texture descriptors. Similarly, the incorporation of fractal texture features could improve the performance of the classificatory models in other studies. A further area in which more sophisticated classificatory models may be needed is in the diagnosis of borderline epithelial lesions of the breast, where even the histopathologic diagnosis is fraught with high interobserver variability.[49] Ongoing efforts are aimed at developing image analysis techniques to assist cytologists in making this subtle distinction.

Acknowledgments

The authors thank Drs. Craig Benham, Joseph Malinsky, Sylvan Wallenstein, and Carol Bodian of the Department of Biomathematical Sciences, Mount Sinai School of Medicine, for their helpful advice and discussion. We also thank our pathology colleagues: Dr. Ira Bleiweiss for histopathologic diagnoses, Dr. David Burstein for cytologic diagnoses, Dr. Miguel Sanchez (Englewood Hospital and Medical Center), Dr. Grace Yang (New York University Medical Center), and Dr. Syed Hoda (New York Hospital-Cornell Medical Center) for providing materials, and Joseph Silberfarb and Cynthia Sheppard for

technical assistance. Andrew Einstein was supported by a traineeship on NIH MSTP Training Grant GM07280.

References

1. Landini G, Rippin JW. How important is tumour shape? Quantification of the epithelial-connective tissue interface in oral lesions using local connected fractal dimension analysis. *J Pathol* 1996; **179:** 210-217.

2. Boxt LM, Katz J, Liebovitch LS, Jones R, Esser PD, Reid L. Fractal analysis of pulmonary arteries: the fractal dimension is lower in pulmonary hypertension. *J Thorac Imaging* 1994; **9:** 8-13.

3. Cross SS, Bury JP, Silcocks PB, Stephenson TJ, Cotton DW. Fractal geometric analysis of colorectal polyps. *J Pathol* 1994; **172:** 317-323.

4. Einstein AJ, Wu H-S, Sanchez M, Gil J. Fractal characterization of chromatin appearance for diagnosis in breast cytology: I. Fractal dimensions. *Submitted*.

5. Mandelbrot B. Corrélations et texture dans un nouveau modèle d'Univers hiérarchisé, basé sur les ensembles trémas. *C R Acad Sci Ser A* 1979; **288:** 81-83.

6. Sierpinski W. Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée. *C R Acad Sci* 1916; **162:** 629-632.

7. Gefen Y, Aharony A, Mandelbrot BB. Phase transitions on fractals: III. Infinitely ramified lattices. *J Phys A: Math Gen* 1984; **17:** 1277-1289.

8. Lin B, Yang ZR. A suggested lacunarity expression for Sierpinski carpets. *J Phys A: Math Gen* 1986; **19:** L49-L52.

9. Allain C, Cloitre M. Characterizing the lacunarity of random and deterministic fractal sets. *Phys Rev A* 1991; **44:** 3552-3558.

10. Plotnick RE, Gardner RH, O'Neill RV. Lacunarity indices as measures of landscape texture. *Landscape Ecology* 1993; **8:** 201-211.

11. Henebry GM, Kux HJH. Lacunarity as a texture measure for SAR imagery. *Int J Remote Sensing* 1995; **16:** 565-571.

12. Plotnick RE, Gardner RH, Hargrove WW, Prestegaard K, Perlmutter M. Lacunarity analysis: a general technique for the analysis of spatial patterns. *Phys Rev E* 1996; **53:** 5461-5468.

13. Yang GC, Alvarez II. Ultrafast Papanicolaou stain. An alternative preparation for fine needle aspiration cytology. *Acta Cytol* 1995; **39:** 55-60.

14. Einstein AJ, Gil J, Wallenstein S, *et al.* Reproducibility and accuracy of interactive segmentation procedures for image analysis in cytology. *J Microsc* in press.

15. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

16. Einstein AJ, Gil J. Classification procedures for diagnosis based on multiple morphometric parameters. *Acta Stereol* 1996; **15:** 15-24.

17. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman & Hall, 1993.

18. Kattan MW, Beck JR. Artificial neural networks for medical classification decisions. *Arch Pathol Lab Med* 1995; **119:** 672-677.

19. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: John Wiley & Sons, 1989.

20. Garcia-Romero H, Garcia-Barrios C, Ramos-Gutierrez F. Effects of uncertain results on sensitivity and specificity of diagnostic tests. *Lancet* 1996; **348:** 1745-1746.

21. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986; **323:** 533-536.

22. Dytch HE, Wied GL. Artificial neural networks and their use in quantitative pathology. *Anal Quant Cytol Histol* 1990; **12:** 379-393.

23. Deligdisch L, Einstein AJ, Guera D, Gil J. Ovarian dysplasia in epithelial inclusion cysts. A morphometric approach using neural networks. *Cancer* 1995; **76:** 1027-1034.

24. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995; **346:** 1075-1079.

25. Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quart Appl Math* 1944; **2:** 164-168.

26. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Indust Appl Math* 1963; **11:** 431-441.

27. Hagan MT, Manhaj MB. Training feedforward networks with the Marquardt algorithm. *IEEE Trans Neural Net* 1994; **5:** 989-993.

28. Oertel YC. Fine Needle Aspiration of the Breast. Boston: Butterworth Publishers, 1987.

29. Kline TS, Kline IK. Guides to Clinical Aspiration Biopsy: Breast. New York: Igaku-Shoin, 1989.

30. Frost JK. The cell in health and disease: an evaluation of cellular morphologic expression of biologic behavior. Basel: Karger, 1969.

31. Truong H, Morimoto R, Walts AE, Erler B, Marchevsky A. Neural networks as an aid in the diagnosis of lymphocyte-rich effusions. *Anal Quant Cytol Histol* 1995; **17:** 48-54.

32. Astion ML, Wilding P. The application of backpropagation neural networks to problems in pathology and laboratory medicine. *Arch Pathol Lab Med* 1992; **116:** 995-1001.

33. Amari S, Murata N, Müller K-R, Finke M, Yang H. Statistical theory of overtraining — Is cross-validation asymptotically effective? In: Touretzky DS, Mozer MC, Masselmo ME, eds. Advances in Neural Information Processing Systems, vol. 8. Cambridge, MA: The MIT Press, 1996: 176-182.

34. Chauvin Y. Generalization dynamics in LMS trained linear networks. In: Lippmann RE, Moody JE, Touretzky DS, eds. Advances in Neural Information Processing Systems, vol. 3. San Mateo, CA: Morgan Kauffman Publishers, 1991: 890-896.

35. King EB, Chew KL, Duarte L, *et al.* Image cytometric classification of premalignant breast disease in fine needle aspirates. *Cancer* 1988; **62:** 114-124.

36. Julesz B. Textons, the elements of texture perception, and their interactions. *Nature* 1981; **290:** 91-97.

37. Young IT, Verbeek PW, Mayall BH. Characterization of chromatin distribution in cell nuclei. *Cytometry* 1986; **7:** 467-474.

38. Hutchinson ML, Isenstein LM, Zahniser DJ. High-resolution and contextual analysis for the diagnosis of fine needle aspirates of breast. *Anal Quant Cytol Histol* 1991; **13:** 351-355.

39. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975; **4:** 172-179.

40. Wolberg WH, Street WN, Mangasarian OL. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. *Anal Quant Cytol Histol* 1995; **17:** 77-87.

41. Wolberg WH, Street WN, Mangasarian OL. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Lett* 1994; **77:** 163-171.

42. Wilbur DC, Bonfiglio TA, Rutkowski MA, *et al.* Sensitivity of the AutoPap 300 QC System for cervical cytologic abnormalities. Biopsy data confirmation. *Acta Cytol* 1996; **40:** 127-132.

43. Koss LG, Lin E, Schreiber K, Elgert P, Mango L. Evaluation of the PAPNET cytologic screening system for quality control of cervical smears. *Am J Clin Pathol* 1994; **101:** 220-229.

44. Layfield LJ, Glasgow BJ, Cramer H. Fine-needle aspiration in the management of breast masses. *Pathol Annu* 1989; **24:** 23-62.

45. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978; **299:** 926-930.

46. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987; **6:** 411-423.

47. Einstein AJ, Bodian CA, Gil J. The relationships among performance measures in the selection of diagnostic tests. *Arch Pathol Lab Med* 1997; **121:** 110-117.

48. Unger PD, Watson CW, Liu Z, Gil J. Morphometric analysis of neoplastic renal aspirates and benign renal tissue. *Anal Quant Cytol Histol* 1993; **15:** 61-66.

49. Rosai J. Borderline epithelial lesions of the breast. *Am J Surg Pathol* 1991; **15:** 209-221.

Table 1

Two-feature and Three-feature Logistic Regression Models

Features Included			Misclassified Cases	SE	SP	PPV	NPV	ACC	κ
D_s (m)	Λ'(35) (m)	C (sd)	2, 10, 17, 18, 21, 23, 36, 39	0.82	0.79	0.82	0.79	0.80	0.61
D_s (m)	log Λ'(35) (m)	C (sd)	2, 10, 17, 18, 21, 23, 36, 39	0.82	0.79	0.82	0.79	0.80	0.61
D_s (m)	normalized Λ'(35) (m)	C (sd)	2, 10, 17, 18, 21, 23, 36, 39	0.82	0.79	0.82	0.79	0.80	0.61
D_s (m)	normalized log Λ'(35) (m)	C (sd)	10,17,18,39	0.95	0.84	0.88	0.94	0.90	0.80
D_s (m)	normalized log Λ'(35) (m)		2, 10, 17, 18, 23, 36, 39	0.86	0.79	0.83	0.83	0.83	0.66
D_s (m)		C (sd)	8, 10, 17, 18, 19, 32, 36, 37, 39, 41	0.77	0.74	0.77	0.74	0.76	0.51
	normalized log Λ'(35) (m)	C (sd)	1, 2, 10, 16, 20, 23, 27, 33, 38, 39, 40	0.68	0.79	0.79	0.68	0.73	0.47

SE = Sensitivity; SP = Specificity; PPV = Positive Predictive Value; NPV = Negative Predictive Value; ACC = Accuracy; D_s = spectral fractal dimension; (m) = mean; Λ'(s) = weighted lacunarity with box size of s; C = multiplicative prefactor; (sd) = standard deviation.

Cases 1-19 are benign, cases 20-41 are malignant.

Table 2

Four-feature Logistic Regression Models

Features Included				Misclassified Cases	No Convergence	SE	SP	PPV	NPV	ACC	κ
D_s (m)	normalized $\Lambda'(35)$ (m)	C (sd)	$\Lambda'(2)$ (sd)	10, 12, 17, 27, 39	22	0.86	0.84	0.86	0.89	0.85	0.71
D_s (m)	normalized $\Lambda'(35)$ (m)	C (sd)	log $\Lambda'(2)$ (sd)	10, 12, 17, 22, 27, 39		0.86	0.84	0.86	0.84	0.85	0.71
D_s (m)	normalized $\Lambda'(35)$ (m)	C (sd)	normalized $\Lambda'(2)$ (sd)	10, 23, 41	17	0.91	0.89	0.95	0.89	0.90	0.81
D_s (m)	normalized $\Lambda'(35)$ (m)	C (sd)	normalized log $\Lambda'(2)$ (sd)	10, 12	17, 23, 27, 41	0.86	0.84	0.90	1.00	0.85	0.73
D_s (m)	normalized $\Lambda'(35)$ (m)	C (sd)	Area (m)	17, 21, 26, 39	2	0.86	0.89	0.95	0.85	0.88	0.76
D_s (m)	normalized log $\Lambda'(35)$ (m)	C (sd)	$\Lambda'(2)$ (sd)	10, 12, 17, 22, 27, 39, 41		0.82	0.84	0.86	0.80	0.83	0.66
D_s (m)	normalized log $\Lambda'(35)$ (m)	C (sd)	log $\Lambda'(2)$ (sd)	10, 12, 17, 18, 22, 27, 39, 41		0.82	0.79	0.82	0.79	0.80	0.61
D_s (m)	normalized log $\Lambda'(35)$ (m)	C (sd)	normalized $\Lambda'(2)$ (sd)	10, 20, 23, 28, 39, 41	17	0.77	0.89	0.94	0.77	0.83	0.67
D_s (m)	normalized log $\Lambda'(35)$ (m)	C (sd)	normalized log $\Lambda'(2)$ (sd)	10, 12, 17, 23, 27, 28, 39, 41		0.77	0.84	0.85	0.76	0.80	0.61
D_s (m)	normalized log $\Lambda'(35)$ (m)	C (sd)	Area (m)	10, 17, 21, 26, 29, 39	2	0.82	0.84	0.90	0.80	0.83	0.67

SE = Sensitivity; SP = Specificity; PPV = Positive Predictive Value; NPV = Negative Predictive Value; ACC = Accuracy;
D_s = spectral fractal dimension; (m) = mean; $\Lambda'(s)$ = weighted lacunarity with box size of s; C = multiplicative prefactor; (sd) = standard deviation.

Cases 1-19 are benign, cases 20-41 are malignant.

For purposes of performance measures, benign nonconvergent cases are counted as false positives and malignant nonconvergent cases are counted as false negatives, following the approach of Garcia-Romero *et al.* [9]

Table 3

Five-feature Logistic Regression Models

Features Included					Misclassified Cases	SE	SP	PPV	NPV	ACC	κ
D_s (m)	normalized Λ'(35) (m)	C (sd)	Λ'(2) (sd)	Area (m)	12, 21, 26, 27, 39	0.82	0.95	0.95	0.82	0.88	0.76
D_s (m)	normalized Λ'(35) (m)	C (sd)	log Λ'(2) (sd)	Area (m)	12, 21, 27, 39, 40	0.82	0.95	0.95	0.82	0.88	0.76
D_s (m)	normalized Λ'(35) (m)	C (sd)	normalized Λ'(2) (sd)	Area (m)	20, 21, 23, 39	0.82	1.00	1.00	0.83	0.90	0.81
D_s (m)	normalized Λ'(35) (m)	C (sd)	normalized log Λ'(2) (sd)	Area (m)	21, 39	0.91	1.00	1.00	0.90	0.95	0.90
D_s (m)	normalized log Λ'(35) (m)	C (sd)	Λ'(2) (sd)	Area (m)	12, 21, 26, 27, 39, 40	0.77	0.95	0.94	0.78	0.85	0.71
D_s (m)	normalized log Λ'(35) (m)	C (sd)	log Λ'(2) (sd)	Area (m)	*	0.00	0.00	0.00	0.00	0.00	-0.04
D_s (m)	normalized log Λ'(35) (m)	C (sd)	normalized Λ'(2) (sd)	Area (m)	10, 12, 20, 21, 39	0.86	0.89	0.90	0.85	0.88	0.76
D_s (m)	normalized log Λ'(35) (m)	C (sd)	normalized log Λ'(2) (sd)	Area (m)	20, 21, 39	0.86	1.00	1.00	0.86	0.93	0.85

SE = Sensitivity; SP = Specificity; PPV = Positive Predictive Value; NPV = Negative Predictive Value; ACC = Accuracy;
D_s = spectral fractal dimension; (m) = mean; Λ'(s) = weighted lacunarity with box size of s; C = multiplicative prefactor; (sd) = standard deviation.

Cases 1-19 are benign, cases 20-41 are malignant.

* Cases 12, 39, and 40 were misclassified, while no convergence was attained in the remainder of cases.

Table 4

Numbers of Cases Incorrectly Diagnosed in Jackknife Analyses Using Various Artificial Neural Network Models with the Features Mean D_s, Mean Normalized $\Lambda'(35)$, Standard Deviation of C, Standard Deviation of Normalized Log $\Lambda'(2)$, and Mean Area

Output Coding	Number of Hidden Neurons																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0–1	7	10	8	6	7	10	12	8	11	11	11	12	8	8	10	8	9	7	11
-1–1	6	6	5	7	7	8	3	6	7	6	5	6	7	7	9	6	5	7	6
-1–1	6	4	7	5	8	6	2	6	6	8	6	4	5	6	6	8	7	5	6

Figure 1

Two Sierpinski Carpets with Fractal Dimensions of 1.896

Three stages shown.

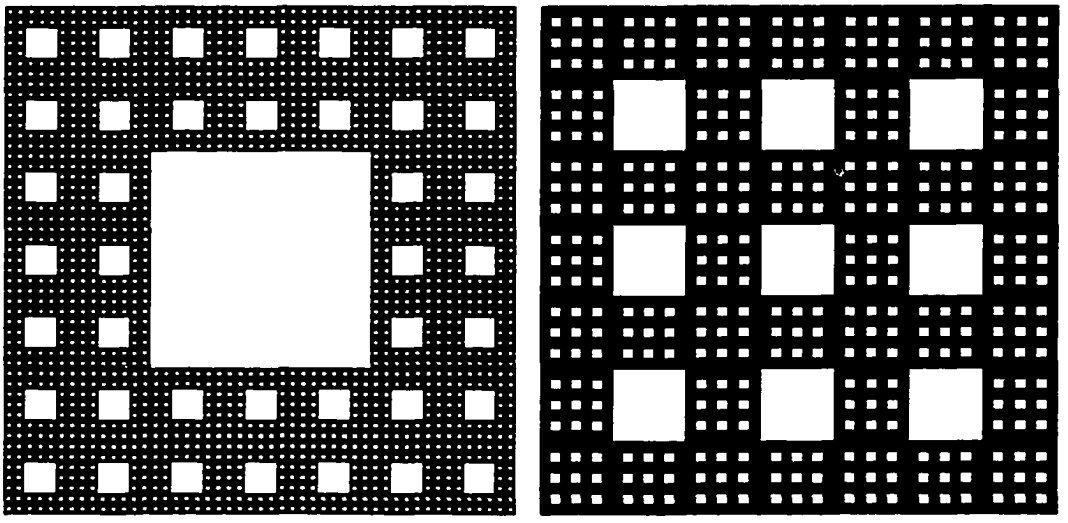


Figure 2

Weighted Lacunarity Curves of the Sierpinski Carpets in Figure 1

The carpet on the left is more lacunar.

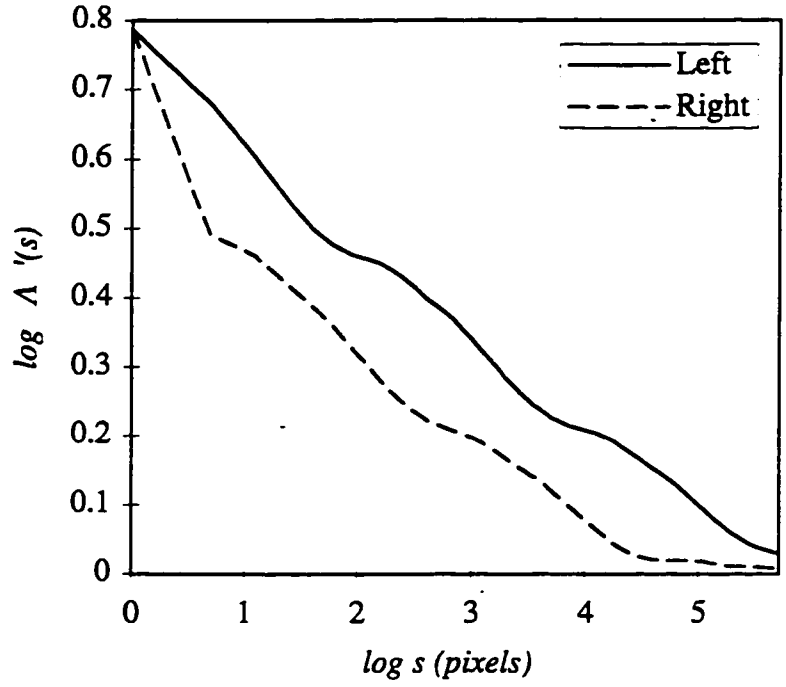


Figure 3

Two Nuclei with Equal Fractal Dimensions and Their Weighted Lacunarity Curves

The nucleus on the left is more lacunar.

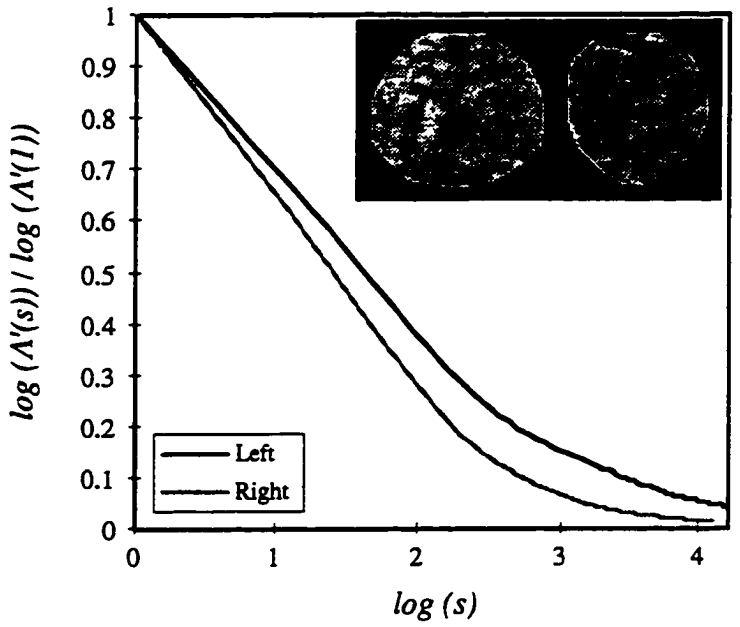


Figure 4

Weighted Lacunarities of the 41 Cases as a Function of Box Size

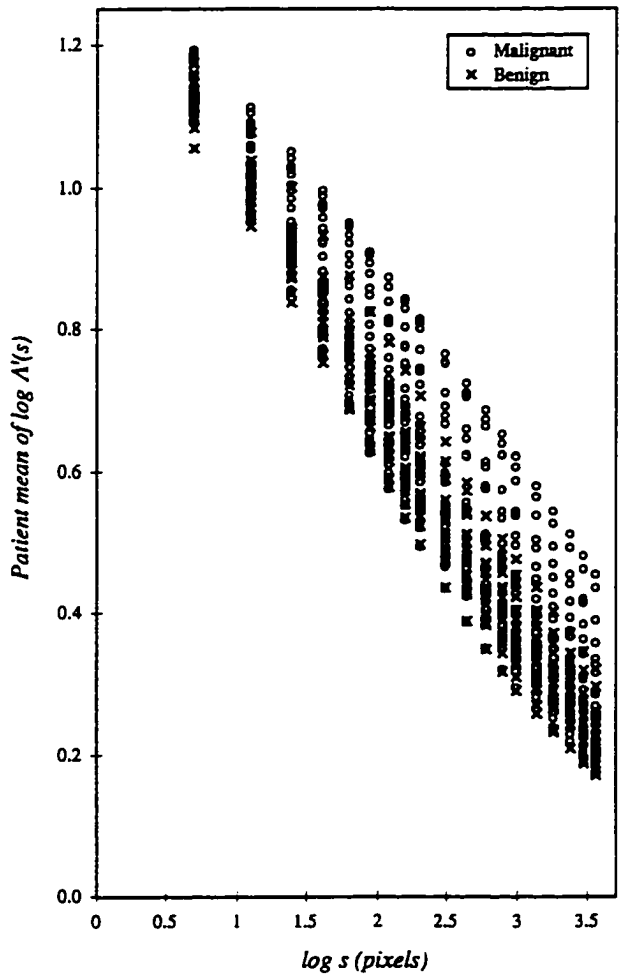


Figure 5

p as a Function of the Box Size *s* for Nuclear Images Thresholded at the First, Second, and Third Quartiles of the Intensity Histogram

p is the significance level of the *t*-test comparing the mean lacunarity per case of malignant and benign cases.

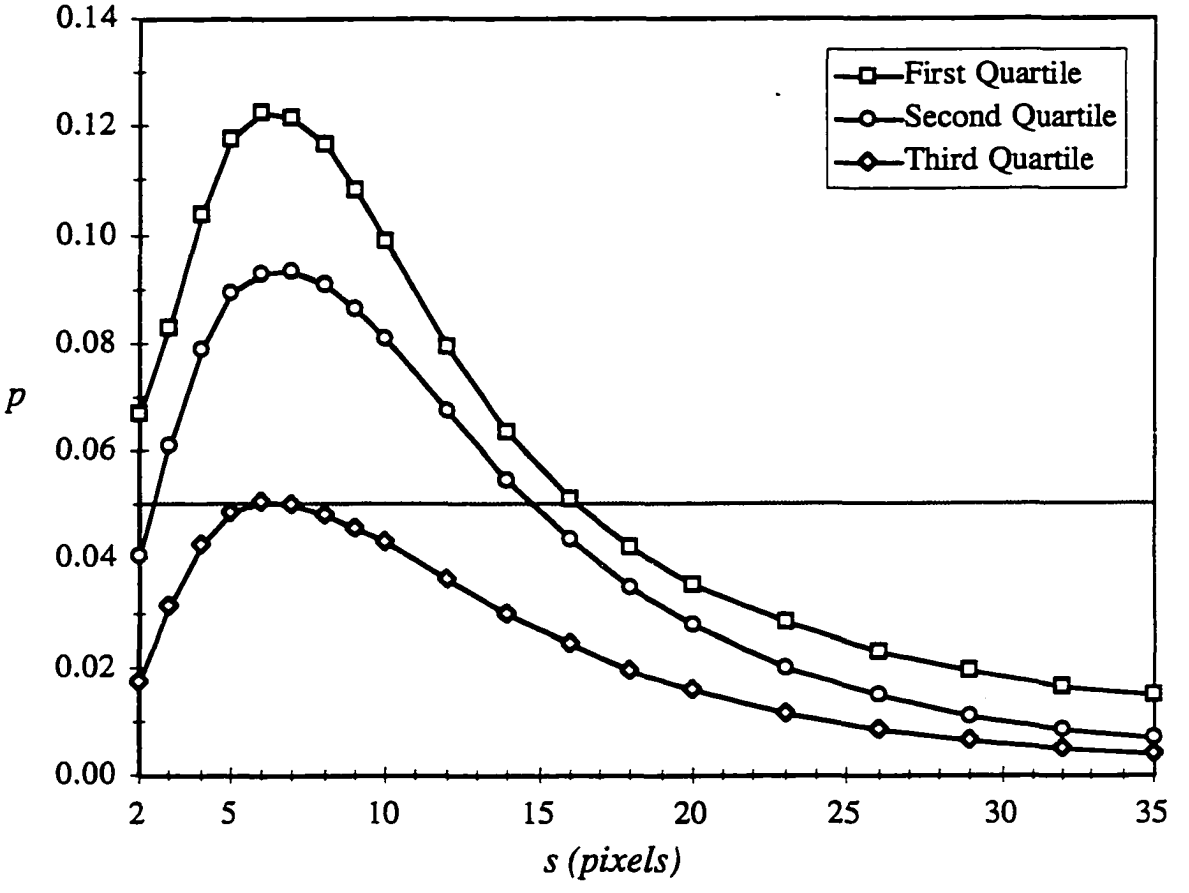
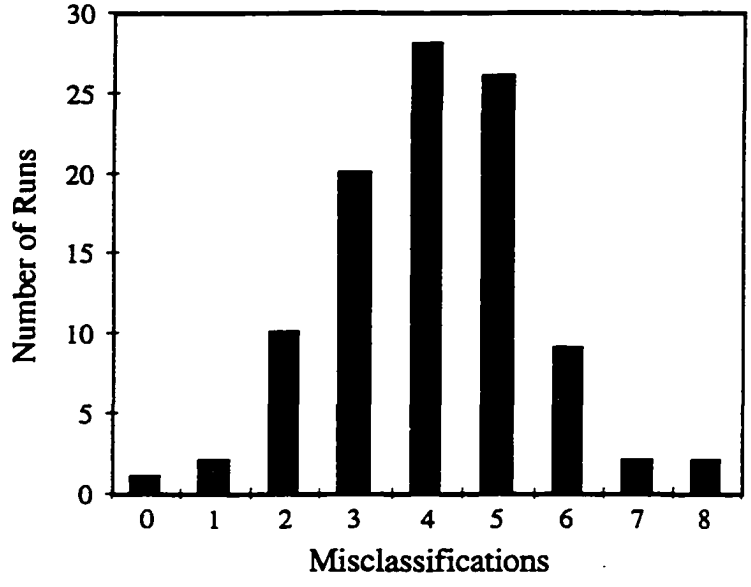


Figure 6

Histogram of the Number of Misclassifications out of 41 Cases for a Jackknifed Neural Network Model with 8 Hidden Neurons, Tan-sigmoid Transfer Functions, and Input Neurons Corresponding to Mean Spectral Dimension, Mean Area, Mean Normalized $\Lambda'(35)$, and Standard Deviation of Normalized Log $\Lambda'(2)$

100 observations, Mean = 4.08 misclassifications,

Standard deviation = 1.44 misclassifications.



Chapter 8

Classification Procedures for Diagnosis based on Multiple Morphometric Parameters

Andrew J. Einstein

Joan Gil

Acta Stereologica 1996; 15: 15-24.

Preface to Chapter 8

Chapter 7 compared and optimized logistic regression and neural network classifiers for the diagnosis of breast lesions. Chapter 8 continues on the theme of classification, in a more general context. Despite its importance, the problem of choice and optimization of classification methods has never been the subject of systematic study. In this chapter we discuss, in order, the role of computers in diagnosis, feature selection, types of classifiers, the details of some important classification algorithms, classifier overtraining/overfitting, and comparisons between classification methods.

The "morphometric parameters" referred to in the title of the chapter are synonymous with what we have referred to in previous chapters as features or descriptors. This paper was prepared for a stereology journal, and our brief perusal of the stereology literature indicated that the term "parameters" is used more commonly than these other terms.

Abstract

A diagnosis in pathology is a classification based upon multiple parameters. Therefore, an important challenge in morphometry is to determine what the most effective classification procedure is for a given setting; in practice, this applies to a narrow differential diagnosis between predetermined choices of similar morphology. To resolve this problem, a number of approaches have been recommended, both statistical and non-statistical. Among the statistical classificatory procedures are discriminant analysis, logistic regression, k-nearest neighbor analysis, and recursive partitioning. Among the non-statistical procedures, the community has expressed a preference for artificial neural networks. Experience in our laboratory using neural networks in the diagnosis of histology and cytology specimens suggests the tentative conclusion that in comparison with other classificatory algorithms, neural networks are the method of choice.

Key Words: classification, diagnosis, image analysis, morphometry

Introduction

Diagnosis is classification. In pathology, the clinical applicability of image analysis is dependent on prescribing a specific method of classifying new cases based on multiple parameters. Each case is represented by a vector of parameters; a sample of these vectors with known diagnoses, referred to as a training set, is used to determine a general method for assigning arbitrary parameter vectors to diagnostic groups. Although most morphometric classificatory studies to date have employed stepwise parameter selection and discriminant analysis to classify cases, there is no reason to assume that they are the most effective of the numerous methods which exist. The problem of choice and optimization of classification methods is a crucial one, yet a review of the literature shows that this issue

has rarely been addressed and never been the subject of systematic study. Mathematically, the problem can be framed in terms of how to best partition a multidimensional parameter space, where "best" can be understood in terms of various measures of diagnostic effectiveness, such as accuracy, sensitivity, specificity, and false negative rate.

It is important to note that the classifications we are dealing with are between very similar diagnostic groups, such as breast epithelial cell lesions, and not among broad categories such as cancer. The role of computers in diagnosis in the foreseeable future will be to assist with objective classification among narrow differentiations, not to replace physicians. While a cytopathologist has no trouble identifying breast tissue and hyperplastic breast epithelial cells within that tissue, he may find it difficult to reproducibly distinguish between mild and moderate hyperplasia. Moreover, as stressed by Rosai,[1] one cytologist's mild hyperplasia is the next one's moderate hyperplasia. Image analysis and classificatory methodology enable us to objectify this portion of the diagnostic process. This consultative role of computers in medicine is discussed with great insight and lucidity by Blois.[2] As illustrated in Figure 1, he describes the cognitive span required during diagnosis as a series of judgments, each of which narrows the possibilities for diagnosis. While at the beginning of this diagnostic process (Point A) the "totality of the world must be confronted," the role of computers is at the terminal stage represented by point B. Here, the "task domain has been structured through previous human effort, an abstraction is available, and little common-sense knowledge may be required."

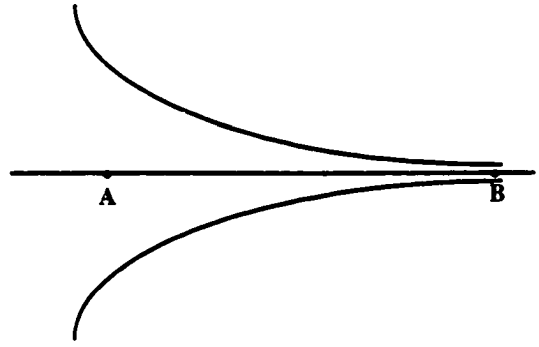


Fig. 1. The Cognitive Span Required during Diagnosis. Figure drawn after Blois.[2]

The classification problem is really a twofold one: 1) how to best select parameters, and 2) how to best select and optimize a classification algorithm.

Selection of Parameters

Parameters are known alternatively, depending on the context, as features, descriptors, inputs, variables, or covariates. They characterize properties of tissue deriving from morphometric or stereologic methods; in pathology these typically include measures of nuclear profile sizes and size distributions, chromatin texture, tissue architecture, etc. Besides from the parameter selection incorporated into some classification algorithms, numerous stand-alone methods are employed. These include both univariate approaches such as ANOVA and its nonparametric analog the Kruskal-Wallis test, Genchi-Mori ambiguity, area under an ROC curve,[3] and the Fisher ratio,[4] and multivariate methods, such as factor analysis and the popular stepwise procedure. While parameter selection merits treatment in its own right, the focus of the discussion here will be on classification algorithm selection and optimization. We shall treat, in order, types of classifiers, the details of some important algorithms, comparisons between classification methods, and overtraining/overfitting.

Types of Classifiers

Classifiers can be grouped as supervised or unsupervised, statistical or non-statistical, and parametric or nonparametric. In supervised learning, we know the "state of nature" (*e.g.*, normal, adenoma, carcinoma) for each sample, while in unsupervised learning we don't.[5] Unsupervised classifiers such as clustering methods, which are not generally used in morphometry, use the underlying structure of the data to assign group numbers to individual cases. The groups, or clusters, which are formed are not defined *a*

priori but rather are suggested by the data, although they can be compared to, and identified with, states of nature.[6] Numerous clustering methods exist. They vary in the nature of the clusters, which may be disjoint, hierarchical, overlapping, or fuzzy. Most classifiers are statistical, but some use alternative methods, *e.g.*, neural networks and rule-based expert systems. The distinction between statistical and non-statistical classifiers is an artificial one; it is easily seen that the statistical method CART effectively generates a rule-based expert system. Statistical classifiers assuming a normal distribution of covariates are called parametric. While they are generally robust, non-normality may require a transformation or rank ordering of covariates. Often, a simple logarithmic tranform will yield covariates with a Gaussian distribution. Properties of a number of popular classifiers are summarized in Table 1, and we discuss the rudiments of a few of these approaches below.

Table 1. Properties of Some Classification Methods.

Classifier	Supervised?	Statistical?	Parametric?
Artificial Neural Network (ANN)	Yes	No	No
Bayesian Belief Network	Yes	Yes	—
CART (Recursive Partitioning)	Yes	Yes	No
Cluster Analysis	No	Yes	No
Discriminant Analysis	Yes	Yes	Yes
k-Nearest Neighbor	Yes	Yes	No
Logistic Regression	Yes	Yes	No
Rule-based Expert System (LISP)	Yes	No	No

Discriminant Analysis

Discriminant analysis is a parametric classification method and as such is most appropriate when covariates have approximately normal within-class distributions. Nevertheless it is robust and may still serve as an effective classifier in the absence of normality (normality can be determined with the Shapiro-Wilk W test). The basic approach is to produce a mathematical rule, called a discriminant function, that determines the posterior probability that an unknown belongs to a particular group. Owing to the nature of

the resultant discriminant function, the method is called linear discriminant analysis if we assume equal within-group covariance matrices, and quadratic discriminant analysis if we assume unequal within-group covariance matrices. Diagnosis can be associated with the group of highest posterior probability. The procedure for computing posterior probabilities is given as follows:[6]

The generalized squared distance from an unknown parameter vector $\boldsymbol{x}$ to group i is given by

$$D_i^2(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{m}_i)' V_i^{-1} (\boldsymbol{x} - \boldsymbol{m}_i) + g_1(i) + g_2(i)$$

where

p_i	is the prior probability of membership in group i,
$\boldsymbol{m}_i$	is the vector containing the variable means in group i,
V_i	is the covariance matrix within group i if we assume unequal within-group covariance matrices, or the pooled covariance matrix if we assume equal within-group covariance matrices,
$g_1(i) = ln \mid V_i \mid$	if the within-group covariance matrices are used, or
$g_1(i) = 0$	if the pooled covariance matrix is used,
$g_2(i) = -2 \ln p_i$	if the prior probabilities are not all equal, or
$g_2(i) = 0$	if the prior probabilities are all equal.

Letting n denote the number of groups, the posterior probability of unknown $\boldsymbol{x}$ belonging to group i is defined as

$$p(i \mid \boldsymbol{x}) = exp\ (-0.5\ D_i^2(\boldsymbol{x})) \,/\, \textstyle\sum_{j=1,2,\ldots,n} exp\ (-0.5\ D_j^2(\boldsymbol{x}))\ .$$

Logistic Regression

Logistic regression[7] is a nonparametric statistical approach to classification. We want to discriminate among $n+1$ diagnostic categories numbered $0, 1, \ldots, n$ where category

0 is designated as the reference category. Let *y* be the outcome variable, taking the value of the diagnostic category, and let x be the $m+1$ element vector of covariates, where $x_0 = 1$ is constant and x_i represents the ith covariate for $i = 1, 2,..., m$. Define the n logit functions as

$$\begin{aligned} f_i(x) &= \ln [\, P(y = i \mid x) / P(y = 0 \mid x)] \\ &= \beta_{i0} + \beta_{i1} x_1 + \beta_{i2} x_2 + \ldots + \beta_{im} x_m \\ &= x^T \beta_i \end{aligned}$$

for $i = 1, 2,..., n$. Letting $f_0(x) = 0$, we then have that the conditional probability for an outcome given a covariate vector is

$$p(y = i \mid x) = exp\,(f_i(x)) \,/\, \textstyle\sum_{j=0,1,\ldots,n} exp\,(f_j(x)) \,. \qquad (1)$$

To compute maximum likelihood estimates of the β_{ij} we must set the partial derivatives of the log-likelihood function equal to zero and solve the resultant equations, typically by the Newton-Raphson method. Estimates of the information matrix and of its inverse, the covariance matrix, are similarly determined by computing the second partial derivatives of the log-likelihood function. These, in turn, are used to determine the standard errors of the estimates. To test the null hypothesis that a component of β is zero, and thereby determine if a variable contributes to the model, we can use the likelihood ratio test, Wald's test, or the score test. The score test is the least popular of these methods. Of the other two, the likelihood ratio test is regarded as the better method; both Hauck and Donner[8] and Jennings[9] have demonstrated that Wald's test behaves aberrantly. All of these calculations are performed by commercial packages such as SYSTAT. Having determined MLEs of the β_{ij} from the training set, we can now use them to assign posterior probabilities

to a new case by applying Eq. 1. Like in discriminant analysis, we can assign a diagnosis to a new case by choosing the outcome with the greatest posterior probability.

k-Nearest Neighbor Analysis

Another nonparametric classificatory algorithm is based on a case's k nearest neighbors in parameter space. Given an unknown case, we identify the k points (*e.g.*, 10) nearest to it in parameter space and determine to which diagnostic category they belong.[10] Nearness may be in terms of Euclidean or Mahalanobis distance. If all k nearest neighbors belong to the same group, then that is the classification for the unknown. If the neighbors belong to different groups, then posterior probabilities are computed for membership in each of the groups. Let n be the number of diagnostic groups, p_i $(i = 1, 2, \ldots, n)$ be the prior probabilities of membership in each group, and f_i $(i = 1, 2, \ldots, n)$ be the number of k-nearest neighbors of the unknown case belonging to each group, satisfying $\sum f_i = k$. Then the posterior probability $p(i \mid \boldsymbol{x})$ that the unknown belongs to group i is given by

$$p(i \mid \boldsymbol{x}) = f_i p_i / \sum_{j=1,2,\ldots,n} f_j p_j .$$

A diagnosis can be associated with the group of maximal posterior probability.

Recursive Partitioning with Classification and Regression Trees (CART)

A third nonparametric approach to classification is recursive partitioning, generally associated with the Classification and Regression Trees (CART) method of Breiman and Friedman.[11] Recursive partitioning is based on the process of binary stratification, done by analyzing each parameter to find its optimal cutpoint, which best separates patients into diagnostic groups.[12] Cutpoint determination may be achieved by a number of splitting criteria.[13] The test maximizing classification accuracy is chosen as the first partition, and

the binary stratification is repeated on each half of the partitioned parameter space. The recursive partitions can be described by a tree of decision rules.

Unconstrained partitioning will lead to a tree with as many terminal nodes as cases, decreasing the generalizability of the classifier. For this reason, the decision tree is "pruned," removing non-informative nodes. Here too there are a number of methods that can be used, such as the cross-validation method. Thus, a CART classifier is characterized by its splitting and pruning criteria. Alternative classifiers can be generated by varying the cost associated with an incorrect diagnosis, expressed in terms of the misclassification cost ratio.

CART's counterpart and competitor is the Fast Algorithm for Classification Trees (FACT) of Loh and Vanichsetakul.[14] FACT attempts to combine the advantages of CART with linear discriminant analysis; this parametric method places more emphasis on linear combination and non-binary splits, does not use cross-validation but rather employs a stop-splitting rule, and incorporates methods for dealing with missing values.[15] Despite the case made by FACT's proponents, CART remains the standard for recursive partitioning.

Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) are a type of artificial intelligence.[16-18] They can generalize from a training set, making few assumptions as to the underlying data structure. While several neural network designs have been studied and employed for classification,[19,20] backpropagation networks remain the standard, and the focus of this section. The fundamental unit of structure in artificial neural networks, like their biological counterparts, is the neuron. A backpropagation ANN has its neurons arranged in a multilayered architecture, with each neuron connected to the neurons in its adjacent layers. There is a layer of input neurons, one or more layers of hidden neurons, and a layer of output neurons. A value is associated with each neuron, and a weight with each connection. In addition, hidden and output neurons have biases associated with them;

these may be regarded as weights of special neurons having constant values of one. The value of a hidden or output neuron is computed by passing the biased weighted sum of values from the previous layer through a "transfer function," most typically a sigmoidal function. More formally, let n be the number of layers, n_k be the number of neurons in layer k ($k = 1, 2, \ldots, n$), w_{ijk} be the connection weight between the ith neuron in layer k and the jth neuron in layer k-1 ($i = 1, 2, \ldots, n_k$; $j = 1, 2, \ldots, n_{k-1}$; $k = 2, 3, \ldots, n$), and b_{ik} and v_{ik} be the bias and value of the ith neuron in layer k, respectively ($i = 1, 2, \ldots, n_k$; $k = 2, 3, \ldots, n$). Then the values v_{ik} are determined from the formula

$$v_{ik} = 1 / (1 + exp(-\Sigma_{j=1,2,\ldots,n_{k-1}} (w_{ijk} * v_{j(k-1)}) + b_{ik})) .$$

The architecture of a typical backpropagation network for use in morphometry is illustrated in Fig. 2. As is shown, input neurons represent morphometric parameters, while output neurons code for the diagnosis. The values taken by output variables are generally constrained to the interval [0, 1]; thus, *e.g.*, a value near one for the "Malignant" neuron is identified with a diagnosis of malignancy.

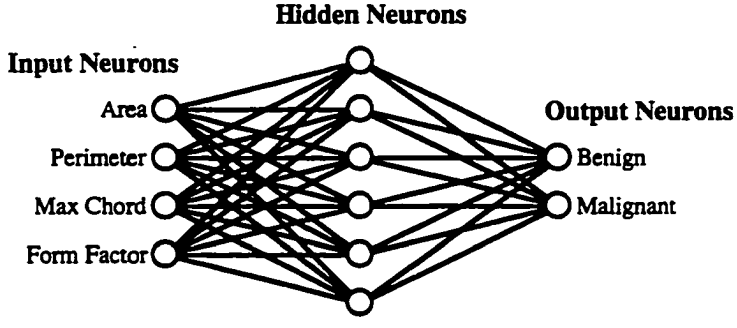


Figure 2. Neural Network Architecture. Example of a backpropagation neural network with four input neurons, six hidden neurons, and two output neurons.

The classificatory power of neural networks is contained in the weights and biases, known together as the connection matrix.[21] Determination of the optimal connection matrix

for a particular set of training facts is a computationally difficult problem; for even very simple networks, the problem has been mathematically demonstrated to be NP-complete (not computable in polynomial time).[22] The approach generally used in ANNs is an adaptive one, in which the weights and biases are "learned" over repeated iterations through the training set. The connection matrix is initially set at random, and the values of hidden and output neurons are computed for a parameter vector. Output neuron values are compared with their target values, *e.g.*, (0,1) for malignancy in the example. If the difference between the output value and its target value, known as the error factor, is less than a specified training tolerance, the case represented by the parameter value is regarded to be correctly classified, and the next case is considered. If the error factor is too large, then the connection matrix is modified using some learning rule, most typically the generalized delta rule or a modification thereof. These learning rules reflect factors such as the transfer function, neuron values, error factors, weights, and biases. Initially, output biases and connection weights between output neurons and the last layer of hidden neurons are adjusted. These changes are then backpropagated, one layer at a time, until the whole connection matrix has been modified. Cycling through the training set continues until all cases are correctly classified or, barring this, some other specified stopping point.

Using the final connection matrix, a new set of inputs can be used to predict output values. Numerous factors are incorporated into the design of a neural network, and these may affect its training and ability to generalize. Such factors include network topology, noise, order of training facts, choice of transfer function, initial weights, range specification, learning rate, training tolerance, testing tolerance, and the training set.

Other Methods

Besides the popular approaches discussed above, there exist numerous other classification methods, several of which are described in Duda and Hart's classic book.[5] Bartels and Thompson[23] have championed the use of Bayesian Belief Networks as a

classifier for quantitative pathology. Rule-based expert systems, often developed in the computer language LISP, have fallen out of favor for artificial intelligence applications. Other methods include Parzen Windows, the Furthest-Neighbor Algorithm, and linear programming.

Comparison of Classificatory Approaches

It is difficult to know which classifier will perform best for a given data set. Each method has its advantages and disadvantages. Neural networks probably make the least assumptions as to the surfaces that divide parameter space. They provide a number of options for fine-tuning. However, neural nets have been accused of being "black-box" classifiers, accurate but uninterpretable. While the effect of individual weights in a neural network is certainly less comprehensible than that of, for example, splits in CART, methods have been developed to determine how a neural network uses information. Ravdin and Clark[24] examined a trained network by selectively inactivating its input units, both individually and in sets. They claim that the effect of using only a single input neuron is analogous to evaluating its univariate importance, while inactivating the neuron may determine its multivariate importance. By including or omitting pairs of neurons, interactions between parameters can be analyzed. Wilding *et al.*[25] identify the most influential features in a jackknife analysis by individually varying inputs by ±10% of their range and determining the result on the output. These methods may also be used for parameter selection. The role of hidden neurons may be studied using a Hinton diagram,[26] which graphically represents connection weights. Apologetics aside, a neural network is undeniably more difficult to interpret than a linear discriminant function or CART tree. But this very complexity and nonlinearity is the source of neural networks' classificatory power.

Statistical approaches are more dependent on the validity of a model that is assumed. Linear discriminant analysis and CART should perform well if the "true" separating surfaces are roughly linear; whether this is a reasonable assumption for morphometric data is unclear. Quadratic discriminant analysis can model more complex separating surfaces than its linear counterpart, but it is more prone to overfitting of the training data.[27] As mentioned, both forms of discriminant analysis may suffer in the absence of normality. Ironically, logistic regression may fail if categories are too well separated in space. Using a data set in which each case had previously been correctly diagnosed using a jackknifed neural network approach,[28] we attempted to make a classification using logistic regression, but log-likelihood functions did not converge. Interestingly, we observed that convergence may in fact be dependent on the reference group chosen. While some classifiers interpolate diagnostic information for regions of parameter space where they have no training points, logistic regression effectively refuses to venture a classification.

Breiman *et al.*[11] suggest a number of advantages of their CART method. While a number of these could apply to most classifiers, some are specific to recursive partitioning. These include that it is invariant under monotone transformations of parameters, and that it is very robust with respect to misclassified points and outliers. One feature of CART is that the classification tree it yields bears a strong resemblance to clinical practice guidelines.[29] For applications where we want readily explainable classification rules, this is extremely beneficial. For applications where classification accuracy, regardless of complexity, is most important, this may be an unnecessary simplification.

Table 2. Selected Studies Comparing Classification Methods

Methods & Accuracies (%)			Reference Number
NN (93.3) >	LDA (84.4) >	QDA (82.2)	30
NN (47) >	Bayes (45)		31
NN (100) =	LDA (100)		32
*NN (77.7) >	DT (76.5) >	LDA (75.0)	33
LR (94.3) =	LDA (94.3)		34
NN (93.8, 82.2, 77.3) >	WDD (93.5, 81.7, 77.2)		35
CART (91.3) >	LR (87.4) >	QDA (84.5) > LDA (83.5)	12
CART (83.3) >	NN (73.8) =	LDA (73.8)	36
NN (80) >	QDA (75)		37
NN (94.1) >	LDA (85.7) >	QDA (78.0)	38
NN (80, 70) >	LDA (64, 65) >	QDA (69, 45)	39
LR (72) >	CART (68), other models		29
NN (89.9) >	LR (88.4)		40
NN (90.6) >	kNN (81) >	LDA (78.125)	20
Bayes (99.1) >	NN (95.4)		19
kNN (83) >	NN (79) >	MLM (73)	41
LR, LDA, CART (Not all accuracies given)			15

NN = Neural Network. LDA = Linear Discriminant Analysis. QDA = Quadratic Discriminant Analysis.
Bayes = Bayesian Classifier. DT = Decision Tree. LR = Logistic Regression.
WDD = Weighted Density Distribution Classifier, the best previously reported chromosome classifier.
kNN = k-Nearest Neighbors Method. MLM = Maximum Likelihood Method.
* But comparison differed for multi-stage classification: LDA (87.2) > NN (85.1) > DT (84.3).

Despite numerous papers using classificatory algorithms in a host of medical applications, there have been remarkably few comparing their accuracies. Table 2 summarizes some of this literature. We make no claims of comprehensiveness. The comparisons may not all be fair, in the sense that in a single study one method may be more optimized than the other(s). In some studies numerous classifications were compared and one representing each classifier had to be chosen. Perhaps the truth is as Hand[27] contends, that "careful and sensitive use of any method will probably yield very similar results." A few papers have been omitted in which classifier comparison using different parameter sets may have biased the results, accuracy was based primarily on the training set, or the emphasis was on survival models or different classifiers than those above.

In diagnostic anatomic pathology, we are only aware of three papers that have employed both neural networks and a statistical classificatory method;[30-32] in all cases neural networks were equal or better. Classifier performance should be expected to vary depending on the data structure, and therefore results from one study may not be applicable

to another. If it is possible to draw any conclusions, we would make the tentative claim that neural networks tend to perform better than statistical classifiers due to their minimal assumptions about data structure. Indeed, this flexibility has resulted in the application of neural networks to problems refractory to conventional approaches, such as speech pronunciation from text,[42] which can be viewed as a classification problem with strings of letters as parameter vectors and phonemes as output. Our preference for neural networks is shared by other reviewers[43] and lent support by the American Joint Committee on Cancer. To replace the TNM cancer staging system, the Committee has developed a new prognostic system using an artificial neural network to combine prognostic factors.[44] Nevertheless, the classification problem requires considerable further study.

Classifier Training and Overtraining

Regardless of the classifier used, training can be performed by three methods. The simplest approach is simply to divide the data set into two groups, a training set and a testing set. The training set is used to determine the classifier's parameters (such as weights in a neural network or coefficients of a discriminant function) while the testing set is left to evaluate the trained classifier. If it is difficult to obtain a sufficiently large data set, cross-validation can be used, allowing training on a larger number of patients. The data set is divided into k groups, of which k-1 are used for training and the remaining group is used to test the trained classifier. This is repeated with each of the k groups left as the testing set. A composite measure of accuracy can be used to evaluate the classificatory scheme. When n-way cross-validation is performed on n patients, this is called jackknife analysis or leave-one-out analysis. Any cross-validation scheme yields multiple trained classifiers, but if the accuracy is sufficiently high for each of them we might propose the same classifier trained on the entire data set as a classificatory approach. While these training methods

may seem obvious, the literature abounds with examples of poorly constructed studies in which training and testing sets overlap.

If we want to combine classifier training with optimization, the data can be divided into three parts: a training set, an optimization testing set, and an evaluation testing set.[21] If desired, this may be combined with a scheme for cross-validation. A variety of classifier models can be compared, *e.g.*, modifying tolerances, architecture, noise, and transfer function of a neural network. All the classifiers are tested against the optimization testing set, and the most accurate of the models is chosen. If we are cross-validating, the process is repeated for permutations of the three groups. The accuracy of this optimized classifier (or classifiers) may then be evaluated with the new testing facts left in the evaluation testing set (or sets). Accuracy determined by this method should generally be more realistic than accuracy estimated with a single testing set.

While at first glance it might seem reasonable to provide a classifier with as many parameters as are computationally tractable, leaving it to the algorithm to sort out which are important, this raises problems of overtraining, known in other contexts as overfitting. In any data classification scheme, two competing factors are at work as the classification rule is allowed to grow more complex: 1) identifying general features in a data set, those which are predictive of outcome, and 2) identifying the particular features of specific data points. If the ratio of the sample size to the number of weights or input parameters (the latter is called the S:I ratio) is low, a "clever" classifier may focus on the particulars and miss the big picture. Such a classifier will be able to classify its training cases with great accuracy, but not much more. While the problem of classifier complexity has been raised mostly as a criticism of overtraining neural networks,[45] it takes equally problematic forms in other classificatory approaches, *e.g.*, the number of splits allowed in CART and the type of discriminant function and number of covariates in discriminant analysis. In neural networks, overtraining may result from too many neurons or from too many cycles of training (a consequence of too small a training tolerance). Illustrating this was a study by

Lamb and Niederberger[39] using neural networks to predict fertility potential. Feeding a neural network their training set 10,000 times, it achieved a training classification rate of 92% and a testing rate of 81%. With 1.5 million cycles of training, the training rate increased to 98% but the testing rate decreased to 64%.

Complicating the matter even further is the fact that morphometric parameters, deriving from the same images, will in general be highly interdependent. Thus, formal treatments of sample size, based upon independent input variables, may not entirely apply, typically proposing more stringent sample size requirements than are necessary. Nevertheless, haphazard and nondiscriminatory parameter selection will still result in overtraining. The lack of a clear method of addressing this problem underscores the importance of using biologically relevant parameters, for a classificatory approach that simulates pathologist's decision-making processes and avoids multiple measures of biologic features should minimize some opportunities for the classifier to overtrain. Neural network overtraining may also be reduced by choosing a larger training tolerance and fewer hidden neurons.

In the literature, training problems in neural networks are often framed[21,25] in terms of shrinkage, defined as the difference in the classification rates for training and testing facts. This may just cloud the real issue of overtraining. Overtraining results in incorrect classifications for new testing facts. Shrinkage may be a second order measure of diagnostic accuracy, but ultimately what we care about is the accuracy itself, *i.e.*, how well "unknowns" will be classified. (Of course sensitivity, specificity, and predictive values are similarly of concern, and the argument applies equally to these measures.) Training classification rate, which certainly should be at least as high as testing rate, is not particularly important in its own right. In sum, a classifier should be chosen based upon its ability to classify unknowns.

Acknowledgments

Andrew Einstein gratefully acknowledges support from a Hans Elias Bursary and from United States Public Health Service Grant 5-T32-GM 7280-16.

References

1. Rosai J. Borderline epithelial lesions of the breast. *Am J Surg Pathol* 1991; **15:** 209-221.

2. Blois M. Clinical judgment and computers. *N Engl J Med* 1980; **303:** 192-197.

3. Bartels P. Numerical evaluation of cytologic data. III. Selection of features for discrimination. *Anal Quant Cytol* 1979; **1:** 153-159.

4. de Meester U, Young IT, Lindeman J, van der Linden HC. Towards a quantitative grading of bladder tumors. *Cytometry* 1991; **12:** 602-613.

5. Duda RO, Hart PE. Pattern Classification and Scene Analysis. New York: John Wiley, 1973.

6. SAS Institute Inc. SAS/STAT User's Guide, Version 6, vol. SAS Institute Inc. Cary, NC, 1989.

7. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: John Wiley & Sons, 1989.

8. Hauck WW, Donner A. Wald's test as applied to hypotheses in logit analysis. *J Am Stat Assoc* 1977; **72:** 851-853.

9. Jennings DE. Judging inference adequacy in logistic regression. *J Am Stat Assoc* 1986; **81:** 471-476.

10. Unger PD, Watson CW, Liu Z, Gil J. Morphometric analysis of neoplastic renal aspirates and benign renal tissue. *Anal Quant Cytol Histol* 1993; **15:** 61-66.

11. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. New York: Chapman & Hall, 1993.

12. Lacher DA. Comparison of nonparametric recursive partitioning to parametric discriminant analyses in laboratory differentiation of hypercalcemia. *Clinica Chimica Acta* 1991; **204:** 199-208.

13. Segal MR. Regression trees for censored data. *Biometrics* 1988; **44:** 35-47.

14. Loh W, Vanichsetakul N. Tree-structured classification via generalized discriminant analysis. *J Am Stat Assoc* 1988; **83:** 715-725.

15. Stewart PW, Stamm JW. Classification tree prediction models for dental caries from clinical, microbiological, and interview data. *J Dent Res* 1991; **70:** 1239-1251.

16. Lawrence J. Introduction to Neural Networks: Design, Theory, and Applications. Nevada City, CA: California Scientific Software, 1993.

17. Dytch HE, Wied GL. Artificial neural networks and their use in quantitative pathology. *Anal Quant Cytol Histol* 1990; **12:** 379-393.

18. Deligdisch L, Einstein AJ, Guera D, Gil J. Ovarian dysplasia in epithelial inclusion cysts. A morphometric approach using neural networks. *Cancer* 1995; **76:** 1027-1034.

19. Tourassi GD, Floyd CD. Lesion Size Quantification in SPECT using an artificial neural network classification approach. *Comp Biomed Res* 1995; **28:** 257-270.

20. DaPonte JS, Sherman P. Classification of ultrasonic image texture by statistical discriminant analysis and neural networks. *Comp Med Imaging Graph* 1991; **15:** 3-9.

21. Astion ML, Wilding P. The application of backpropagation neural networks to problems in pathology and laboratory medicine. *Arch Pathol Lab Med* 1992; **116:** 995-1001.

22. Blum A, Rivest RL. Training a 3-node neural network is NP-complete. In: Touretzky D, ed. Advances in Neural Information Processing Systems, vol. 1. San Mateo, CA: Morgan Kaufmann Publishers, 1989: 494-501.

23. Bartels PH, Thompson D, Bibbo M, Weber JE. Bayesian belief networks in quantitative histopathology. *Anal Quant Cytol Histol* 1992; **14:** 459-473.

24. Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Res Treat* 1992; **22:** 285-293.

25. Wilding P, Morgan MA, Grygotis AE, Shoffner MA, Rosato EF. Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. *Cancer Lett* 1994; **77:** 145-153.

26. Qian N, Sejnowski TJ. Predicting the secondary structure of globular proteins using neural network models. *J Mol Bio* 1988; **202:** 865-884.

27. Hand DJ. Statistical methods in diagnosis. *Stat Methods Med Res* 1992; **1:** 49-67.

28. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

29. Hadorn DC, Draper D, Rogers WH, Keeler EB, Brook RH. Cross-validation performance of mortality prediction models. *Stat Med* 1992; **11:** 475-489.

30. Erler BS, Hsu L, Truong HM, *et al.* Image analysis and diagnostic classification of hepatocellular carcinoma using neural networks and multivariate discriminant functions. *Lab Invest* 1994; **71:** 446-451.

31. Dawson AE, Austin RE, Weinberg DS. Nuclear grading of breast carcinoma by image analysis: classification by multivariate and neural network analysis. *Am J Clin Pathol* 1991; **95:** S29-S37.

32. Unger PD, Hoon V, Stone N, *et al.* Computerized interactive morphometry in the differential diagnosis of irradiated prostates. *Anal Quant Cytol Histol* 1995; **17:** 100-108.

33. Palcic B, MacAulay C, Shlien S, Treurniet W, Tezcan H, Anderson G. Comparison of three different methods for automated classification of cervical cells. *Anal Cell Pathol* 1992; **4:** 429-441.

34. Stenkvist B, Strande G. Analysis of machine-selected cells with an image analysis system in normal and abnormal cervical specimens. *Anal Cell Pathol* 1989; **2:** 1-13.

35. Errington PA, Graham J. Application of artificial neural networks to chromosome classification. *Cytometry* 1993; **14:** 627-639.

36. Reibnegger G, Weiss G, Werner-Felmayer G, Judmaier G, Wachter H. Neural networks as a tool for utilizing laboratory information: comparison with linear discriminant analysis and with classification and regression trees. *Proc Natl Acad Sci USA* 1991; **88:** 11426-11430.

37. Astion ML, Wilding P. Application of neural networks to the interpretation of laboratory data in cancer diagnosis. *Clin Chem* 1992; **38:** 34-38.

38. Erler BS, Vitagliano P, Lee S. Superiority of neural networks over discriminant functions for thalassemia minor screening of red blood cell microcytosis. *Arch Pathol Lab Med* 1995; **119:** 350-354.

39. Lamb DJ, Niederberger CS. Artificial intelligence in medicine and male infertility. *World J Urol* 1993; **11:** 129-136.

40. Doig GS, Inman KJ, Sibbald WJ, Martin CM, Robertson JM. Modeling mortality in the intensive care unit: comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression. *Proc Annu Symp Comput Appl Med Care* 1993 361-365.

41. Clarke LP, Velthuizen RP, Phuphanich S, Schellenberg JD, Arrington JA, Silbiger M. MRI: stability of three supervised segmentation techniques. *Magn Reson Imaging* 1993; **11:** 95-106.

42. Sejnowski TJ, Rosenberg CR. Parallel networks that learn to pronounce English text. *Compex Systems* 1987; **1:** 145-168.

43. Pun T, Gerig G, Ratib O. Image analysis and computer vision in medicine. *Comp Med Imaging Graph* 1994; **18:** 85-96.

44. Burke HB. Increasing the power of surrogate endpoint biomarkers: the aggregation of predictive factors. *J Cell Biochem Suppl* 1994; **19:** 278-282.

45. Clark GM, Hilsenbeck SG, Ravdin PM, De Laurentiis M, Osborne CK. Prognostic factors: rationale and methods of analysis and integration. *Breast Cancer Res Treat* 1994; **32:** 105-112.

SUMMARY

In this dissertation we have developed several image analysis and mathematical methods for diagnosis in pathology, and applied these methods to the diagnoses of neoplastic diseases of the breast, ovary, and parathyroid. In this Summary, we shall review these developments and suggest avenues for future investigation.

Chapters 1 and 2 emphasized quality control issues. Chapter 1 studied the effect of staining on nuclear morphology, focusing on the new ultrafast Papanicolaou procedure. We found no important differences in morphology between specimens processed with ultrafast Papanicolaou staining and those processed with conventional Papanicolaou staining. There were differences in nuclear size and shape and in chromatin appearance between each of these methods and Diff-Quik. These results thus provide the first objective data supporting claims that the ultrafast Papanicolaou procedure combines the superior speed of Diff-Quik with the detailed morphology of Papanicolaou staining. Chapter 2 focused on segmentation, a crucial step in the development of procedures using image analysis for the cytologic diagnosis of cancer, which can potentially serve as a source of great variability. We provided a framework for analyzing the variability of segmentation methods, using statistical variance component models. We compared several cytologic segmentation methods and found that an arc-fitting approach is most reproducible, while a thresholding-based method similar to the segmentation methods used in some commercial instruments is unacceptably variable. Accuracies of area measurements for the various methods, as determined with respect to point counting, paralleled their reproducibilities. Sample size requirements were observed to be more dependent on the biologic variability of the tissue sampled than on the particular segmentation method or on the number of individuals performing segmentation. The data suggest that in most instances, it should be sufficient to segment 50 or 60 nuclei per case.

Chapter 3 compared the morphology of normal ovaries, incidental findings of ovarian dysplasia, and ovarian dysplasia adjacent to frank carcinoma. By comparison of both the size, shape, and texture descriptors characterizing cases and the cases' neural network classifications, we found the two types of dysplasia to be morphologically indistinguishable. This suggests that the associated changes in tissue morphology reflect a single underlying pathologic process, in turn lending credence to the hypothesis that ovarian dysplasia seen in epithelial inclusion cysts may represent a preinvasive malignant change.

Chapter 4 introduced the nuclear diffuseness, a measure of local gray level variation, as an approach to quantitate chromatin appearance. A disc is moved over a nuclear image, and at each point the variance of the gray levels of all pixels in the nucleus that are within the disc is computed. Nuclear diffuseness is the mean value of these variances, taken over all pixels in the nuclear image. We studied the effects on diffuseness of changing the disc size, of adjusting for edge effects, and of changes in segmentation. Nuclear diffuseness was applied to the rare and challenging histopathologic diagnosis of parathyroid adenomas and carcinomas. Using just diffuseness and nuclear area, and an artificial neural network for classification, all 16 parathyroid biopsies considered were correctly diagnosed.

Chapters 5, 6, and 7 introduced methods for the fractal characterization of chromatin appearance, and applied these methods to the cytologic diagnosis of breast epithelial cell lesions. An iterative approach to estimating spectral dimensions and a weighted measure of lacunarity were presented. These new methods allow the fractal characterization of irregularly shaped (nonrectangular) images, such as those of nuclei. The ability of lacunarity to quantify morphologic differences between fractals with identical dimensions was demonstrated, using Sierpinski carpets as well as nuclear images. Several theoretical issues relating to fractals were discussed, including the concavity of log-log plots, correlations between Minkowski and spectral fractal dimensions, connections

between fractals and random processes, and the spectral reserve hypothesis which posits the loss of complexity in disease states. Logistic regression and neural network classifiers were compared and optimized for the diagnosis of breast lesions. The best logistic regression classifier correctly diagnosed 40 of 41 cases. The best neural networks correctly diagnosed all 41 cases, however it was unclear whether or not this performance could be attributed to overtraining.

Finally, Chapter 8 surveyed the problem of classification, treating the role of computers in diagnosis, feature selection, types of classifiers, the details of some important classification algorithms, classifier overtraining/overfitting, and comparisons between classification methods.

The work completed in this dissertation leads to numerous areas for future investigation. Current efforts focus on three areas: fractal characterization of chromatin appearance, breast cytology diagnosis, and the molecular genetics of ovarian dysplasia. Our ongoing research on fractals encompasses both methods for estimating the spectral dimension and methods for measuring lacunarity. In Chapters 5 and 6, we discussed the iterative spectral dimension method, which "reconstructs" a rectangular image embedding a nucleus by first modeling texture as a first order Markov process and then iterating background texture. The spectral exponent β can be readily obtained from this reconstructed image, and the spectral fractal dimension is derived from β by a simple linear relation. Our studies using synthetic nuclei generated using the inverse Fourier reconstruction method[1] show that the iterative spectral dimension method generally works well, although the estimation of β is less accurate when the nucleus is small relative to the embedding region and the true value of β is close to two.[2] We are presently collaborating with Dr. Hai-Shan Wu on the development of alternative methods for the estimation of β. Our plan is to develop these various methods and compare them to determine which is most accurate. For one dimensional (time series) data, a maximum likelihood estimator of the spectral exponent has proven to be superior to Fourier regression methods and Hurst

exponent methods.[3] While this may also be true for two dimensional data (images), maximum likelihood estimation may be too computationally demanding. We are also exploring methods using Hamming and Hanning filters, as well as averaging the exponents derived from one dimensional transects through images. In addition to these methods for estimating the spectral exponent, we are also considering alternative methods for the measurement of lacunarity, such as the correlation technique developed by Blumenfeld and Ball.[4,5] Beyond fractal dimensions and lacunarity is the notion of succolarity, introduced by Mandelbrot in his magnum opus[6] and thenceforth undeveloped and unused. Perhaps succolarity could be quantified and applied to chromatin texture description.

In Chapter 7, the best logistic regression model correctly classified 40 of 41 cases, while the best neural network correctly classified all 41 cases. This high performance attained suggests a potentially important quality assurance role in breast cytology for image analysis devices. However, it is commonly observed that test efficacy in practice is lesser than that observed in pilot studies. This phenomenon may be attributable to a number of reasons, such as the broader spectrum of disease observed in practice and the differing conditions under which a test may be administered. Thus, this diagnostic performance needs to be validated on a prospective basis prior to recommending either of these "best" classifiers for routine clinical use. We would hope to conduct such clinical trials. Moreover, Chapters 5 through 7 focused on distinguishing benign nonproliferative lesions from invasive carcinoma, and we would still like to evaluate the diagnostic performance of these classificatory approaches on intermediate lesions as well.[7] We had initially planned to include the spectrum of benign proliferative lesions, such as atypical hyperplasia, and carcinoma in situ in our study, but two years of data collection at the Mount Sinai, Englewood, Cornell, and NYU medical centers did not yield enough cases in this spectrum which met our inclusion criteria.

A third area of current investigation is the molecular genetics of ovarian dysplasia. Our group's previous studies of ovarian dysplasia have focused on its qualitative and

quantitative morphologic characterization. The cellular stratification, nuclear pleomorphism, and loss of polarity within the ovarian surface epithelium which characterize ovarian dysplasia must reflect some underlying genetic change. While preliminary data suggests a role for p53,[8] the mutations responsible for ovarian dysplasia are presently unknown. To begin to address this fundamental question, we have begun a study of BRCA1 and BRCA2 mutations in ovarian dysplasia.

References

1. Voss RF. Random fractal forgeries. In: Earnshaw RA, ed. Fundamental Algorithms for Computer Graphics. Berlin: Springer-Verlag, 1985: 805-835.

2. Wu H-S, Einstein AJ, Gil J. Fractal characterization of irregularly shaped images by frequency analysis. *Submitted* .

3. Pilgram B, Kaplan DT. A comparison of estimators for 1/f noise. *Physica D* submitted.

4. Blumenfeld R, Ball RC. Probe for morphology and hierarchical correlations in scale-invariant structures. *Phys Rev E* 1993; **47:** 2298-2302.

5. Blumenfeld R, Ball RC. Quantifying morphology of scale-invariant structures beyond the fractal dimension. *Fractals* 1993; **1:** 985-991.

6. Mandelbrot BB. The Fractal Geometry of Nature. New York: W. H. Freeman and Company, 1983.

7. Teague MW, Wolberg WH, Street WN, Mangasarian OL, Lambremont S, Page DL. Indeterminate fine-needle aspiration of the breast. Image analysis-assisted diagnosis. *Cancer (Cancer Cytopathol)* 1997; **81:** 129-135.

8. Deligdisch L, Berchuk A, Dottino PR, Cohen CJ. Ovarian dysplasia: a histologic, morphometric and immunohistochemical study. *Int J Gyn Cancer* 1995; **5 (Suppl. 1):** 37.

BIBLIOGRAPHY

1. Abbona GC, Papotti M, Gasparri G, Bussolati G. Proliferative activity in parathyroid tumors as detected by Ki-67 immunostaining. *Hum Pathol* 1995; **26:** 135-138.

2. Allain C, Cloitre M. Characterizing the lacunarity of random and deterministic fractal sets. *Phys Rev A* 1991; **44:** 3552-3558.

3. Amari S, Murata N, Müller K-R, Finke M, Yang H. Statistical theory of overtraining — Is cross-validation asymptotically effective? In: Touretzky DS, Mozer MC, Masselmo ME, eds. Advances in Neural Information Processing Systems, vol. 8. Cambridge, MA: The MIT Press, 1996: 176-182.

4. Anguiano E, Pancorbo M, Aguilar M. Fractal characterization by frequency analysis. I. Surfaces. *J Microsc* 1993; **172:** 223-232.

5. Arman F, Pearce JA. Unsupervised classification of cell images using pyramid node linking. *IEEE Trans Biomed Eng* 1990; **37:** 647-650.

6. Associated Press. Panel backs new tool to find cervical cancer New York Times. New York, August 8, 1995; C6.

7. Astion ML, Wilding P. The application of backpropagation neural networks to problems in pathology and laboratory medicine. *Arch Pathol Lab Med* 1992; **116:** 995-1001.

8. Astion ML, Wilding P. Application of neural networks to the interpretation of laboratory data in cancer diagnosis. *Clin Chem* 1992; **38:** 34-38.

9. Avnir D, Biham O, Lidar (Hamburger) D, Malcai O. On the abundance of fractals. In: Novak MM, Dewey TG, eds. Fractal Frontiers. Singapore: World Scientific, 1997: 199-234.

10. Baak JP, Noteboom E, Koevoets JJ. The influence of fixatives and other variations in tissue processing on nuclear morphometric features. *Anal Quant Cytol Histol* 1989; **11:** 219-224.

11. Babloyantz A, Destexhe A. Low-dimensional chaos in an instance of epilepsy. *Proc Natl Acad Sci USA* 1986; **83:** 3513-3517.

12. Bacus SS, Ruby SG. Application of image analysis to the evaluation of cellular prognostic factors in breast carcinoma. *Pathol Annu* 1993; **28:** 179-204.

13. Bak P, Chen K. The physics of fractals. *Physica D* 1989; **38:** 5-12.

14. Banda-Gamoba H, Ricketts I, Cairns A, Hussein K. Spectral analysis of cervical cells using the discrete Fourier transform. *Anal Cell Pathol* 1993; **5:** 85-102.

15. Barba J, Gil J: Image analysis in statistical texture analysis *Eighth International Congress for Stereology* 1991.

16. Barba J, Jeanty H, Fenster P, Gil J. The use of local entropy measures in edge detection for cytological image analysis. *J Microsc* 1989; **156:** 125-134.

17. Barba J, Yuan L, Gil J. Cell contour extraction on multi-threshold images. *Pathol Res Pract* 1992; **188:** 449-453.

18. Barry JD, Sharkey FE. Observer reproducibility during computer-assisted planimetric measurements of nuclear features. *Hum Pathol* 1985; **16:** 225-227.

19. Bartels P. Numerical evaluation of cytologic data. III. Selection of features for discrimination. *Anal Quant Cytol* 1979; **1:** 153-159.

20. Bartels PH, Thompson D. Scene segmentation. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994.

21. Bartels PH, Thompson D, Bibbo M, Weber JE. Bayesian belief networks in quantitative histopathology. *Anal Quant Cytol Histol* 1992; **14:** 459-473.

22. Bassingthwaighte JB, King RB, Roger SA. Fractal nature of regional myocardial blood flow heterogeneity. *Circ Res* 1989; **65:** 578-590.

23. Baum EB, Haussler D. What size net gives valid generalization? In: Touretzky DS, ed. Advances in Neural Information Processing Systems. San Mateo, CA: Morgan Kauffman Publishers, 1989: 81-90.

24. Baumann G, Barth A, Nonnenmacher TF. Measuring fractal dimensions of cell contours: practical approaches and their limitations. In: Nonnenmacher TF, Losa GA, Weibel ER, eds. Fractals in Biology and Medicine. Basel: Birkhäuser Verlag, 1994.

25. Beal SL. Sample size determination for confidence intervals on the population mean and on the difference between two population means. *Biometrics* 1989; **45:** 969-977.

26. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987; **6:** 411-423.

27. Beil M, Irinopoulou T, Vassy J, Rigaut JP. Chromatin texture analysis in three-dimensional images from confocal scanning laser microscopy. *Anal Quant Cytol Histol* 1995; **17:** 323-331.

28. Beltrami CA, Della Mea V, Finato N. Structure analysis of breast lesions using neighborhood graphs. *Anal Quant Cytol Histol* 1995; **17:** 143-150.

29. Berchuk A, Boente MP, Bast Jr. RC. The use of tumor markers in the management of patients with gynecologic carcinomas. *Clin Obstet Gynecol* 1992; **35:** 45-54.

30. Bibbo M, Michelassi F, Bartels PH, *et al.* Karyometric marker features in normal-appearing glands adjacent to human colonic adenocarcinoma. *Cancer Res* 1990; **50:** 147-151.

31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1(8476):** 307-310.

32. Blaustein A. Surface cells and inclusion cysts in fetal ovaries. *Gynecol Oncol* 1981; **12:** 222-233.

33. Blaustein A, Kantius M, Kaganowicz A, Pervez N, Wells J. Inclusions in ovaries of females aged day 1-30 years. *Int J Gynecol Pathol* 1982; **1:** 145-153.

34. Blichert-Toft M, Mollerup CL, Feldt-Rasmussen UF, Daugaard H, Engel UH. Primary hyperparathyroidism. An underdiagnosed disease in Denmark? *Ugeskr Læg* 1993; **155:** 765-769.

35. Blois M. Clinical judgment and computers. *N Engl J Med* 1980; **303:** 192-197.

36. Blum A, Rivest RL. Training a 3-node neural network is NP-complete. In: Touretzky D, ed. Advances in Neural Information Processing Systems, vol. 1. San Mateo, CA: Morgan Kaufmann Publishers, 1989: 494-501.

37. Blumenfeld R, Ball RC. Probe for morphology and hierarchical correlations in scale-invariant structures. *Phys Rev E* 1993; **47:** 2298-2302.

38. Blumenfeld R, Ball RC. Quantifying morphology of scale-invariant structures beyond the fractal dimension. *Fractals* 1993; **1:** 985-991.

39. Bondeson L, Sandelin K, Grimelius L. Histopathological variables and DNA cytometry in parathyroid carcinoma. *Am J Surg Pathol* 1993; **17:** 820-829.

40. Boon ME, van der Poel HG, Tan CJA, Kok LP. Effect of embedding methods versus fixative type on karyometric measures. *Anal Quant Cytol Histol* 1994; **16:** 131-136.

41. Bostwick DG, Burke HB, Wheeler TM, *et al.* Panel statement: the most promising surrogate endpoint biomarkers for screening candidate chemopreventive compounds for prostatic adenocarcinoma in short-term phase II clinical trials. *J Cell Biochem Suppl* 1994; **19:** 283-289.

42. Bowie JE, Young IT. An analysis technique for biological shape - II. *Acta Cytol* 1977; **21:** 455-464.

43. Boxt LM, Katz J, Liebovitch LS, Jones R, Esser PD, Reid L. Fractal analysis of pulmonary arteries: the fractal dimension is lower in pulmonary hypertension. *J Thorac Imaging* 1994; **9:** 8-13.

44. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. New York: Chapman & Hall, 1993.

45. Brigham E. Fast Fourier Transform and Its Applications. Englewood Cliffs, NJ: Prentice-Hall, 1988.

46. Burke HB. Increasing the power of surrogate endpoint biomarkers: the aggregation of predictive factors. *J Cell Biochem Suppl* 1994; **19:** 278-282.

47. Castleman KR. Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, 1979.

48. Cattoretti G, Becker MHG, Key G, *et al.* Monoclonal antibodies against recombinant parts of the Ki-67 antigen (MIB 1 and MIB 3) detect proliferating cells in microwave-processed formalin-fixed paraffin sections. *J Pathol* 1992; **168:** 357-363.

49. Chan JKC, Kung ITM. Rehydration of air-dried smears with normal saline: application in fine-needle aspiration cytologic examination. *Am J Clin Pathol* 1988; **89:** 30-34.

50. Chan KW, Chiu KY, Fu KH, Ling JML. Observer variability in microcomputer-assisted morphometric study of nuclear parameters. *Pathology* 1987; **19:** 407-409.

51. Charpin C, DeVictor B, Andrac L, *et al.* p53 quantitative immunocytochemical analysis in breast carcinomas. *Hum Pathol* 1995; **26:** 159-166.

52. Chauvin Y. Generalization dynamics in LMS trained linear networks. In: Lippmann RE, Moody JE, Touretzky DS, eds. Advances in Neural Information Processing Systems, vol. 3. San Mateo, CA: Morgan Kauffman Publishers, 1991: 890-896.

53. Christensson T, Hellström K, Wengle B, Alveryd A, Wikland B. Prevalence of hypercalcaemia in a health screening in Stockholm. *Acta Med Scand* 1976; **200:** 131-137.

54. Clark GM, Hilsenbeck SG, Ravdin PM, De Laurentiis M, Osborne CK. Prognostic factors: rationale and methods of analysis and integration. *Breast Cancer Res Treat* 1994; **32:** 105-112.

55. Clarke LP, Velthuizen RP, Phuphanich S, Schellenberg JD, Arrington JA, Silbiger M. MRI: stability of three supervised segmentation techniques. *Magn Reson Imaging* 1993; **11:** 95-106.

56. Coleman K, van Diest PJ, Baak JP, Mullaney J. Syntactic structure analysis in uveal melanomas. *Br J Ophthalmol* 1994; **78:** 871-874.

57. Cordon-Cardo C. Tumor suppressor genes. *Cancer* 1995; **75:** 2641.

58. Cornelisse JTWA, Berg JTP. Profile boundary length can be overestimated by as much as 41% when using a digitizer tablet. *J Microsc* 1984; **136:** 341-344.

59. Cox BL, Wang JSY. Fractal surfaces: Measurement and applications in the earth sciences. *Fractals* 1993; **1:** 87-115.

60. Cross SS. Fractals in pathology. *J Pathol* 1997; **182:** 1-8.

61. Cross SS, Bury JP, Silcocks PB, Stephenson TJ, Cotton DW. Fractal geometric analysis of colorectal polyps. *J Pathol* 1994; **172:** 317-323.

62. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995; **346:** 1075-1079.

63. Daniel W. Applied Nonparametric Statistics. Boston: PWS-KENT Publishing Company, 1990.

64. DaPonte JS, Sherman P. Classification of ultrasonic image texture by statistical discriminant analysis and neural networks. *Comp Med Imaging Graph* 1991; **15:** 3-9.

65. Dawson AE, Austin RE, Weinberg DS. Nuclear grading of breast carcinoma by image analysis: classification by multivariate and neural network analysis. *Am J Clin Pathol* 1991; **95:** S29-S37.

66. Dawson AE, Cibas ES, Bacus JW, Weinberg DS. Chromatin texture measurement by Markovian analysis. Use of nuclear models to define and select texture features. *Anal Quant Cytol Histol* 1993; **15:** 227-235.

67. de Meester U, Young IT, Lindeman J, van der Linden HC. Towards a quantitative grading of bladder tumors. *Cytometry* 1991; **12:** 602-613.

68. DeHoff RT. Sampling of material and statistical analysis in quantitative stereology. In: Elias H, ed. Stereology: Proceedings of the Second International Congress for Stereology. New York: Springer-Verlag, 1967: 119-130.

69. DeLellis R. Atlas of Tumor Pathology: Tumors of the Parathyroid Gland, 3rd series. Washington: Armed Forces Institute of Pathology, 1993.

70. Deligdisch L, Berchuk A, Dottino PR, Cohen CJ. Ovarian dysplasia: a histologic, morphometric and immunohistochemical study. *Int J Gyn Cancer* 1995; **5 (Suppl. 1):** 37.

71. Deligdisch L, Einstein AJ, Guera D, Gil J. Ovarian dysplasia in epithelial inclusion cysts. A morphometric approach using neural networks. *Cancer* 1995; **76:** 1027-1034.

72. Deligdisch L, Gil J. Characterization of ovarian dysplasia by interactive morphometry. *Cancer* 1989; **63:** 748-755.

73. Deligdisch L, Kerner H, Cohen CJ, Dargent D, Gil J. Morphometric differentiation between responsive tumor cells and mesothelial hyperplasia in second-look operations for ovarian cancer. *Hum Pathol* 1993; **24:** 143-147.

74. Deligdisch L, Miranda C, Barba J, Gil J. Ovarian dysplasia: nuclear texture analysis. *Cancer* 1993; **72:** 3253-3257.

75. Diegenbach PC, Baak JPA. Quantitative nuclear image analysis: differentiation between normal, hyperplastic, and malignant appearing uterine glands in a paraffin section. III. The use of texture features for differentiation. *Europ J Obstet Gynec Reprod Biol* 1978; **8:** 109-116.

76. Doig GS, Inman KJ, Sibbald WJ, Martin CM, Robertson JM. Modeling mortality in the intensive care unit: comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression. *Proc Annu Symp Comput Appl Med Care* 1993 361-365.

77. Doudkine A, Macaulay C, Poulin N, Palcic B. Nuclear texture measurements in image cytometry. *Pathologica* 1995; **87:** 286-299.

78. Drescher CW, Flint A, Hopkins MP, Roberts JA. Prognostic significance of DNA content and nuclear morphology in borderline ovarian tumors. *Gynecol Oncol* 1993; **48:** 242-246.

79. Dubuc B, Quiniou JF, Roques-Carmes C, Tricot C, Zucker SW. Evaluating the fractal dimension of profiles. *Phys Rev A* 1989; **39:** 1500-1512.

80. Dubuc B, Roques-Carmes C, Tricot C, Zucker SW. The variation method: a technique to estimate the fractal dimension of surfaces. *SPIE Visual Communications and Image Processing II* 1987; **845:** 241-248.

81. Dubuc B, Zucker SW, Tricot C, Quiniou JF, Wehbi D. Evaluating the fractal dimension of surfaces. *Proc R Soc Lond A* 1989; **425:** 113-127.

82. DuBuf JMH, Kardan M, Spann M. Texture feature performance for image segmentation. *Pattern Recognition* 1990; **23:** 291-309.

83. Duda RO, Hart PE. Pattern Classification and Scene Analysis. New York: John Wiley, 1973.

84. Dytch HE, Wied GL. Artificial neural networks and their use in quantitative pathology. *Anal Quant Cytol Histol* 1990; **12:** 379-393.

85. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman & Hall, 1993.

86. Einstein AJ, Barba J, Unger PD, Gil J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J Microsc* 1994; **176:** 158-166.

87. Einstein AJ, Bodian CA, Gil J. The relationships among performance measures in the selection of diagnostic tests. *Arch Pathol Lab Med* 1997; **121:** 110-117.

88. Einstein AJ, Gil J. Classification procedures for diagnosis based on multiple morphometric parameters. *Acta Stereol* 1996; **15:** 15-24.

89. Einstein AJ, Gil J, Wallenstein S, *et al.* Reproducibility and accuracy of interactive segmentation procedures for image analysis in cytology. *J Microsc* in press.

90. Einstein AJ, Wu H-S, Gil J. Fractal characterization of chromatin appearance for diagnosis in breast cytology: II. Lacunarity analysis and classification approaches. *Submitted* .

91. Einstein AJ, Wu H-S, Gil J. Self-affinity and lacunarity of chromatin texture in benign and malignant breast epithelial cell nuclei. *Submitted*.

92. Einstein AJ, Wu H-S, Sanchez M, Gil J. Fractal characterization of chromatin appearance for diagnosis in breast cytology: I. Fractal dimensions. *Submitted*.

93. Einstein AJ, Yang GCH, Silberfarb JB, Gil J. The effect of ultrafast Papanicolaou staining on nuclear and textural features in breast cancer cytology. *Anal Quant Cytol Histol* 1997; **19:** in press.

94. Erler BS, Hsu L, Truong HM, *et al.* Image analysis and diagnostic classification of hepatocellular carcinoma using neural networks and multivariate discriminant functions. *Lab Invest* 1994; **71:** 446-451.

95. Erler BS, Vitagliano P, Lee S. Superiority of neural networks over discriminant functions for thalassemia minor screening of red blood cell microcytosis. *Arch Pathol Lab Med* 1995; **119:** 350-354.

96. Errington PA, Graham J. Application of artificial neural networks to chromosome classification. *Cytometry* 1993; **14:** 627-639.

97. Esposito MJ, Fuchs A. Computerized image analysis and flow cytometric evaluation of ovarian borderline tumors: a study of 24 cases. *Cytometry* 1994; **18:** 218-222.

98. Esteban JM, Kandalaft PL, Mehta P, Odom-Maryon TL, Bacus S, Battifora H. Improvement of the quantification of estrogen and progesterone receptors in paraffin-embedded tumors by image analysis. *Am J Clin Pathol* 1993; **99:** 32-38.

99. Fazzalari NL, Parkinson IH. Fractal dimension and architecture of trabecular bone. *J Pathol* 1996; **178:** 100-105.

100. Ferrer-Roca O, Ballester-Guardia E, Martin J. Nuclear chromatin texture to differentiate follicular and papillary carcinoma of the thyroid. *Path Res Pract* 1989; **185:** 561-566.

101. Fisher B, Osborne CK, Margolese R, Bloomer W. Neoplasms of the breast. In: Holland JF, Frei E, Bast RC, Kufe DW, Morton DL, Weichselbaum RR, eds. Cancer Medicine. Philadelphia: Lea & Febiger, 1993.

102. Fleming MG, Friedman RJ. Multiparametric image cytometry of nevi and melanomas. *Am J Dermatopathol* 1993; **15:** 106-113.

103. Fleming MG, Wied GL, Dytch HE. Image analysis cytometry of dysplastic nevi. *J Invest Dermatol* 1990; **95:** 287-291.

104. Floyd CE, Lo JY, Yun AJ, Sullivan DC, Kornguth PJ. Prediction of breast cancer malignancy using an artificial neural network. *Cancer* 1994; **74:** 2944-2948.

105. Fontana X, Ferrari P, Namer M, Peysson R, Salanon C, Bussiere F. C-erb-B2 gene amplification and serum level of c-erb-B2 oncoprotein at primary breast cancer diagnosis. *Anticancer Res* 1994; **14:** 2099-2104.

106. Foran DJ, Berg RA. A method for quantitative image assessment based on redundant feature measurements and statistical reasoning. *Comput Methods Programs Biomed* 1994; **45:** 291-305.

107. Frost JK. The cell in health and disease: an evaluation of cellular morphologic expression of biologic behavior. Basel: Karger, 1969.

108. Fujimoto Y, Obara T. How to recognize and treat parathyroid carcinoma. *Surg Clin North Am* 1987; **67:** 343-357.

109. Futreal PA, Liu Q, Shattuck-Eidens D, *et al.* BRCA1 mutations in primary breast and ovarian carcinomas. *Science* 1994; **266:** 120-122.

110. Galera-Davidson H, González-Cámpora R, Mora-Marín JA, *et al.* Cytophotometric DNA measurements in medullary thyroid carcinoma. *Cancer* 1990; **65:** 2255-2260.

111. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975; **4:** 172-179.

112. Ganju V, Jenkins RB, O'Fallon JR, *et al.* Prognostic factors in gliomas: a multivariate analysis of clinical, pathologic, flow cytometric, cytogenetic, and molecular markers. *Cancer* 1994; **74:** 920-927.

113. Garcia-Romero H, Garcia-Barrios C, Ramos-Gutierrez F. Effects of uncertain results on sensitivity and specificity of diagnostic tests. *Lancet* 1996; **348:** 1745-1746.

114. Gazit Y, Berk DA, Leunig M, Baxter LT, Jain RK. Scale-invariant behavior and vascular network formation in normal and tumor tissue. *Phys Rev Lett* 1995; **75:** 2428-2431.

115. Gefen Y, Aharony A, Mandelbrot BB. Phase transitions on fractals: III. Infinitely ramified lattices. *J Phys A: Math Gen* 1984; **17:** 1277-1289.

116. Gefen Y, Mandelbrot BB, Aharony A. Critical phenomena on fractal lattices. *Phys Rev Lett* 1980; **45:** 855-858.

117. Ghadially FN. Ultrastructural Pathology of the Cell and Matrix. London: Butterworths, 1988.

118. Gil J, Barba J. Principles of stereology: computerized applications in anatomic pathology. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994: 79-124.

119. Gil J, Ciurea D, Lin Z, Barba J: Semi-automatic nuclear texture analysis based on three gray level zones *Second International Congress Histology Laboratory* 1992.

120. Gil J, Deligdisch L. Interactive morphometric procedures and statistical analysis in the diagnosis of ovarian dysplasia and carcinoma. *Pathol Res Pract* 1989; **185:** 680-685.

121. Gil J, Marchevsky AM, Silage DA. Applications of computerized interactive morphometry in pathology: I. Tracings and generation of graphic standards. *Lab Invest* 1986; **54:** 222-227.

122. Gil J, Silage DA. Touch sensitive screen as interactive peripheral in stereology and image analysis. *Acta Stereol* 1987; **6:** 282-288.

123. Gimenez-Mas JA, Sanz-Moncasi MP, Remon L, Gambo P, Gallego-Calvo MP. Automated textural analysis of nuclear chromatin: a mathematical morphology approach. *Anal Quant Cytol Histol* 1995; **17:** 39-47.

124. Goldberger AL, Findley LJ, Blackburn MR, Mandell AJ. Nonlinear dynamics in heart failure: implications of long-wavelength cardiopulmonary oscillations. *Am Heart J* 1984; **107:** 612-615.

125. Graham CH. Visual space perception Vision and Visual Perception. New York: John Wiley, 1965: 504-547.

126. Grob B. Basic Television Principles and Servicing. New York: McGraw-Hill, 1964.

127. Grosberg A, Rabin Y, Havlin S, Neer A. Crumpled globule model of the three-dimensional structure of DNA. *Europhys Lett* 1993; **23:** 373-378.

128. Gschwend JE, Vogel U, Bader C, Mattfeldt T, Hautmann RE. Predictive value of magnetic resonance imaging and computerized tomography for conservative renal surgery in an ex vivo tumor enucleation study followed by step-sectioning. *J Urol* 1996; **155:** 451-454.

129. Gschwind R, Umbrich CB, Torhorst J, Oberholzer M. Evaluation of shape descriptors for the morphometric analysis of cell nuclei. *Path Res Pract* 1986; **181:** 213-222.

130. Gundersen HJG, Jensen EB. The efficiency of systematic sampling in stereology and its prediction. *J Microsc* 1987; **147:** 229-263.

131. Gundersen HJG, Østerby R. Optimizing sampling efficiency of stereological studies in biology: or 'Do more less well!'. *J Microsc* 1981; **121:** 65-73.

132. Gusberg SD, Deligdisch L. Ovarian dysplasia: a study of identical twins. *Cancer* 1984; **54:** 1-4.

133. Hadorn DC, Draper D, Rogers WH, Keeler EB, Brook RH. Cross-validation performance of mortality prediction models. *Stat Med* 1992; **11:** 475-489.

134. Hagan MT, Manhaj MB. Training feedforward networks with the Marquardt algorithm. *IEEE Trans Neural Net* 1994; **5:** 989-993.

135. Hall EL. Computer Image Processing and Recognition. New York: Academic Press, 1979.

136. Hamburger DA, Biham O, Avnir D. Apparent fractality emerging from models of random distributions. *Phys Rev E* 1996; **53:** 3342-3359.

137. Hand DJ. Statistical methods in diagnosis. *Stat Methods Med Res* 1992; **1:** 49-67.

138. Hanselaar AGJM, Vooijs GP, Van't Hof-Grootenboer AE, Pahlplatz MMM. Cytophotometric analysis of cervical intraepithelial neoplasia grade III, with and without synchronous invasive squamous cell carcinoma. *Cytometry* 1990; **11:** 901-906.

139. Haralick RM, Shanumgam K, Dinstein I. Textural features for image classification. *IEEE Trans Systems Man Cybernet* 1973; **3:** 610-621.

140. Hauck WW, Donner A. Wald's test as applied to hypotheses in logit analysis. *J Am Stat Assoc* 1977; **72:** 851-853.

141. Henebry GM, Kux HJH. Lacunarity as a texture measure for SAR imagery. *Int J Remote Sensing* 1995; **16:** 565-571.

142. Hicks CR. Fundamental Concepts in the Design of Experiments. New York: Holt, Rinehart, and Winston, 1982.

143. Hoffman K, Gil J, Barba J, *et al.* Morphometric analysis of benign and malignant adrenal pheochromocytomas. *Arch Pathol Lab Med* 1993; **117:** 244-247.

144. Holmes EC, Morton DL, Ketcham AS. Parathyroid carcinoma: a collective review. *Ann Surg* 1969; **169:** 631-640.

145. Hölzel WGE, Beer R, Deschner W, Griesmacher A, Müller MM. Individual reference ranges of CA 15-3, MCA and CEA in recurrence of breast cancer. *Scand J Clin Lab Invest Suppl* 1995; **221:** 93-101.

146. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: John Wiley & Sons, 1989.

147. Hutchinson ML, Isenstein LM, Martin JJ, Zahniser DJ. Measurement of subvisual changes in cervical squamous metaplastic cells for detecting abnormality. *Anal Quant Cytol Histol* 1992; **14:** 330-334.

148. Hutchinson ML, Isenstein LM, Zahniser DJ. High-resolution and contextual analysis for the diagnosis of fine needle aspirates of breast. *Anal Quant Cytol Histol* 1991; **13:** 351-355.

149. Hutchinson ML, Schultz DS, Stephenson RA, Wong KL, Harry T, Zahniser DJ. Computerized microscopic analysis of prostatic fine needle aspirates: comparison with breast aspirates. *Anal Quant Cytol Histol* 1989; **11:** 105-110.

150. Hytiroglou P, Harpaz N, Heller DS, Liu ZY, Deligdisch L, Gil J. Differential diagnosis of borderline and invasive serous cystadenocarcinomas of the ovary by computerized interactive morphometric analysis of nuclear features. *Cancer* 1992; **69:** 988-992.

151. Irinopoulou T, Rigaut JP, Benson MC. Toward objective prognostic grading of prostatic carcinoma using image analysis. *Anal Quant Cytol Histol* 1993; **15:** 341-344.

152. Jacobi JM, Lloyd HM, Smith JF. Nuclear diameter in parathyroid carcinomas. *J Clin Pathol* 1986; **39:** 1353-1354.

153. Jagoe R, Sowter C, Slavin G. Shape and texture analysis of liver cell nuclei in hepatomas by computer aided microscopy. *J Clin Pathol* 1984; **37:** 755-762.

154. Jarman WT, Myers RT, Marshall RB. Carcinoma of the parathyroid. *Arch Surg* 1978; **113:** 123-125.

155. Jennings DE. Judging inference adequacy in logistic regression. *J Am Stat Assoc* 1986; **81:** 471-476.

156. Jørgensen T, Yogesan K, Tveter KJ, Skjorten F, Danielsen HE. Nuclear texture analysis: a new prognostic tool in metastatic prostate cancer. *Cytometry* 1996; **24:** 277-283.

157. Julesz B. Textons, the elements of texture perception, and their interactions. *Nature* 1981; **290:** 91-97.

158. Kalir T, Chan KS, Liu Z, Strauchen J, Gil J. Semi-automatic quantitation of nucleolar organizer regions in non-Hodgkin's lymphomas. *Pathol Res Pract* 1994; **190:** 124-128.

159. Kaplan DT, Furman MI, Pincus SM, Ryan SM, Lipsitz LA, Goldberger AL. Aging and the complexity of cardiovascular dynamics. *Biophys J* 1991; **59:** 945-949.

160. Kattan MW, Beck JR. Artificial neural networks for medical classification decisions. *Arch Pathol Lab Med* 1995; **119:** 672-677.

161. Keough KM, Hyam P, Pink DA, Quinn B. Cell surfaces and fractal dimensions. *J Microsc* 1991; **163:** 95-99.

162. Kim NW, Piatyszek MA, Prowse KR, *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* 1994; **266:** 2011-2014.

163. King EB, Chew KL, Duarte L, *et al.* Image cytometric classification of premalignant breast disease in fine needle aspirates. *Cancer* 1988; **62:** 114-124.

164. Kline TS, Kline IK. Guides to Clinical Aspiration Biopsy: Breast. New York: Igaku-Shoin, 1989.

165. Koch GG. Intraclass correlation coefficient. In: Kotz S, Johnson NL, eds. Encyclopedia of Statistical Sciences, vol. 4. New York: John Wiley & Sons, 1983: 212-217.

166. Koss LG, Lin E, Schreiber K, Elgert P, Mango L. Evaluation of the PAPNET cytologic screening system for quality control of cervical smears. *Am J Clin Pathol* 1994; **101:** 220-229.

167. Kraus MH, Issing W, Miki T, Popescu NC, Aaronson SA. Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: evidence for overexpression in a subset of human mammary tumors. *Proc Natl Acad Sci USA* 1989; **86:** 9193-9197.

168. Kronqvist P, Collan Y, Kuopio T, Kujari H. Nuclear morphometry in breast cancer: the influence of sampling rules and freezing of samples. *Mod Pathol* 1995; **8:** 187-192.

169. Kuwahara H, Shimazaki M, Morikita I, Chanoki Y, Sakurai M. Texture analysis of histological images of giant cell tumor of bone. *Path Res Pract* 1992; **188:** 565-569.

170. Lacher DA. Comparison of nonparametric recursive partitioning to parametric discriminant analyses in laboratory differentiation of hypercalcemia. *Clinica Chimica Acta* 1991; **204:** 199-208.

171. Ladekarl M. The influence of tissue processing on quantitative histopathology in breast cancer. *J Microsc* 1994; **174:** 93-100.

172. Lamb DJ, Niederberger CS. Artificial intelligence in medicine and male infertility. *World J Urol* 1993; **11:** 129-136.

173. Landini G, Misson GP, Murray PI. Fractal analysis of the normal human retinal fluorescein angiogram. *Curr Eye Res* 1993; **12:** 23-27.

174. Landini G, Rippin JW. An "asymptotic fractal" approach to the morphology of malignant cell nuclei. *Fractals* 1993; **1:** 326-335.

175. Landini G, Rippin JW. Fractal dimensions of the epithelial-connective tissue interfaces in premalignant and malignant epithelial lesions of the floor of the mouth. *Anal Quant Cytol Histol* 1993; **15:** 144-149.

176. Landini G, Rippin JW. How important is tumour shape? Quantification of the epithelial-connective tissue interface in oral lesions using local connected fractal dimension analysis. *J Pathol* 1996; **179:** 210-217.

177. Lawrence J. Introduction to Neural Networks: Design, Theory, and Applications. Nevada City, CA: California Scientific Software, 1993.

178. Layfield LJ, Chrischilles EA, Cohen MB, Bottles K. The palpable breast nodule: a cost-effectiveness analysis of alternate diagnostic approaches. *Cancer* 1993; **72:** 1642-1651.

179. Layfield LJ, Glasgow BJ, Cramer H. Fine-needle aspiration in the management of breast masses. *Pathol Annu* 1989; **24:** 23-62.

180. Lee S, Tolmachoff T, Marchevsky AM. DNA content analysis ("ploidy") by image analysis. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994.

181. Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quart Appl Math* 1944; **2:** 164-168.

182. Li FP, Schneider JA, Kantor AF. Cancer epidemiology. In: Holland JF, Frei E, Bast RC, Kufe DW, Morton DL, Weichselbaum RR, eds. Cancer Medicine. Philadelphia: Lea & Febiger, 1993.

183. Lin B, Yang ZR. A suggested lacunarity expression for Sierpinski carpets. *J Phys A: Math Gen* 1986; **19:** L49-L52.

184. Listrom MB, Fenoglio-Preiser CM. Cellular aspects of neoplasia. In: Fenoglio-Preiser CM, Weinstein RS, Kaufman N, eds. New Concepts in Neoplasia as Applied to Diagnostic Pathology. Baltimore: Williams & Wilkins, 1986.

185. Loh W, Vanichsetakul N. Tree-structured classification via generalized discriminant analysis. *J Am Stat Assoc* 1988; **83:** 715-725.

186. Losa GA, Baumann G, Nonnenmacher TF. The fractal dimension of pericellular membrane from lymphocytes and lymphoblastic leukemic cells. *Acta Stereol* 1992; **11/I:** 335-341.

187. Losa GA, Baumann G, Nonnenmacher TF. Fractal dimension of pericellular membranes in human lymphocytes and lymphoblastic leukemia cells. *Pathol Res Pract* 1992; **188:** 680-686.

188. Losa GA, Nonnenmacher TF. Self-similarity and fractal irregularity in pathologic tissues. *Mod Pathol* 1996; **9:** 174-182.

189. MacAulay C, Palcic B. A comparison of some quick and simple threshold selection methods for stained cells. *Anal Quant Cytol Histol* 1988; **10:** 134-138.

190. Makkus ACF, van't Hof-Grootenboer AE, Pahlplatz MMM, *et al.* Practical aspects of fixatives in high resolution nuclear image analysis. *Cytometry* 1994; **15:** 302-310.

191. Mandelbrot B. Corrélations et texture dans un nouveau modèle d'Univers hiérarchisé, basé sur les ensembles trémas. *C R Acad Sci Ser A* 1979; **288:** 81-83.

192. Mandelbrot BB. The Fractal Geometry of Nature. New York: W. H. Freeman and Company, 1983.

193. Mandelbrot BB. Self-affine fractals and fractal dimension. *Physica Scripta* 1985; **32:** 257-260.

194. Mandelbrot BB. A fractal's lacunarity, and how it can be tuned and measured. In: Nonnenmacher TF, Losa GA, Weibel ER, eds. Fractals in Biology and Medicine. Basel: Birkhäuser Verlag, 1993: 8-21.

195. Marchevsky AM, Erler BS. Morphometry in pathology. In: Marchevsky AM, Bartels PH, eds. Image Analysis: A Primer for Pathologists. New York: Raven Press, 1994: 125-179.

196. Marchevsky AM, Gil J, Jeanty H. Computerized interactive morphometry in pathology: current instrumentation and methods. *Hum Pathol* 1987; **18:** 320-331.

197. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Indust Appl Math* 1963; **11:** 431-441.

198. Matheron G. Random Sets and Integral Geometry. New York: John Wiley & Sons, 1975.

199. Mathieu O, Cruz-Orive LM, Hoppeler H, Weibel ER. Measuring error and sampling variation in stereology: comparison of the efficiency of various methods for planar image analysis. *J Microsc* 1981; **121:** 75-88.

200. Matsuyama T, Matsushita M. Fractal morphogenesis by a bacterial cell population. *Crit Rev Microbiol* 1993; **19:** 117-135.

201. Mattfeldt T. The accuracy of one-dimensional systematic sampling. *J Microsc* 1989; **153:** 301-313.

202. Mayall BH. Current capabilities in clinical applications in image cytometry. *Cytometry Suppl* 1988; **3:** 78-84.

203. McCance DR, Kenny BD, Sloan JM, Russell CFJ, Hadden DR. Parathyroid carcinoma: a review. *J R Soc Med* 1987; **80:** 505-509.

204. McGuire WL. Breast cancer prognostic factors: evaluation guidelines. *J Natl Cancer Inst* 1991; **83:** 154-155.

205. McKeown PP, McGarity WC, Sewell CW. Carcinoma of the parathyroid gland: is it overdiagnosed? A report of three cases. *Am J Surg* 1984; **147:** 292-298.

206. Merchant FA, Aggarwal SJ, Diller KR, Bovik AC. In-vivo analysis of angiogenesis and revascularization of transplanted pancreatic islets using confocal microscopy. *J Microsc* 1994; **176:** 262-275.

207. Miller GSP. The definition and rendering of terrain maps. *Computer Graphics* 1986; **20:** 39-48.

208. Miller R. Multiple comparisons. In: Kotz S, Johnson NL, eds. Encyclopedia of Statistical Sciences, vol. 5. New York: John Wiley & Sons, 1985: 679-689.

209. Mitmaker B, Begin LR, Gordon PH. Nuclear shape as a prognostic discriminant in colorectal carcinoma. *Dis Colon Rectum* 1991; **34:** 249-259.

210. Mitmaker B, Kyzer S, Begin LR, H GP. The value of nuclear morphometry in the management of patients with colorectal polyps that contain invasive adenocarcinoma. *J Surg Oncol* 1992; **51:** 42-46.

211. Mittal KR, Zeleniuch-Jacquotte A, Cooper JL, Demopoulos RI. Contralateral ovary in unilateral ovarian carcinoma: A search for preneoplastic lesions. *Int J Gynecol Pathol* 1993; **12:** 59-63.

212. Moragas A, Garcia-Bonafe M, de Torres I, Sans M. Textural analysis of lymphoid cells in serous effusions. A mathematical morphologic approach. *Anal Quant Cytol Histol* 1993; **15:** 165-170.

213. Morton DL. Principles of surgical oncology. In: Holland JF, Frei E, Bast RC, Kufe DW, Morton DL, Weichselbaum RR, eds. Cancer Medicine. Philadelphia: Lea & Febiger, 1993.

214. Nicholson S, Richard J, Sainsbury C, *et al.* Epidermal growth factor receptor (EGFr); results of a 6 year follow-up study in operable breast cancer with emphasis on the node negative subgroup. *Br J Cancer* 1991; **63:** 146-150.

215. Nilsson NJ. The Mathematical Foundations of Learning Machines. San Mateo, CA: Morgan Kaufmann Publishers, 1990.

216. Nouwen EJ, Hendrix PG, Eerdekens MW, DeBroe ME. Tumor markers in the human ovary and its neoplasms: a comparative immunohistochemical study. *Am J Pathol* 1987; **126:** 230-242.

217. Obara T, Fujimoto Y. Diagnosis and treatment of patients with parathyroid carcinoma: an update and review. *World J Surg* 1991; **15:** 738-744.

218. Oertel YC. Fine Needle Aspiration of the Breast. Boston: Butterworth Publishers, 1987.

219. Orlando V, Paro R. Chromatin multiprotein complexes involved in the maintenance of transcription patterns. *Curr Opin Genet Dev* 1995; **5:** 174-179.

220. Ostrowski JL, Sawan A, Henry L, *et al.* p53 expression in human breast cancer related to survival and prognostic factors: an immunohistochemical study. *J Pathol* 1991; **164:** 75-81.

221. Pal NR, Pal SK. A review on image segmentation techniques. *Pattern Recognition* 1993; **26:** 1277-1294.

222. Palcic B, MacAulay C, Shlien S, Treurniet W, Tezcan H, Anderson G. Comparison of three different methods for automated classification of cervical cells. *Anal Cell Pathol* 1992; **4:** 429-441.

223. Papanicolaou GN. A new procedure for staining vaginal smears. *Science* 1942; **95:** 438-439.

224. Parker SL, Tong T, Bolden S, Wingo PA. Cancer statistics, 1997. *CA Cancer J Clin* 1997; **47:** 5-27.

225. Pauwels O, Kiss R. Monitoring of chemotherapy-induced morphonuclear modifications by means of digital cell-image analysis. *J Cancer Res Clin Oncol* 1993; **119:** 533-540.

226. Peng C-K, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL. Long-range anticorrelations and non-Gaussian behavior of the heartbeat. *Phys Rev Lett* 1993; **70:** 1343-1346.

227. Peng CK, Buldyrev SV, Goldberger AL, *et al.* Long-range correlations in nucleotide sequences. *Nature* 1992; **356:** 168-170.

228. Pentland AP. Fractal-based description of natural scenes. *IEEE Trans Pattern Anal Machine Intell* 1984; **PAMI-6:** 661-674.

229. Pilgram B, Kaplan DT. A comparison of estimators for 1/f noise. *Physica D* submitted.

230. Plaxe SC, Deligdisch L, Dottino P, Cohen CJ. Ovarian intraepithelial neoplasia (OIN) demonstrated in patients with stage I ovarian carcinoma. *Gynecol Oncol* 1990; **38:** 367-372.

231. Plotnick RE, Gardner RH, Hargrove WW, Prestegaard K, Perlmutter M. Lacunarity analysis: a general technique for the analysis of spatial patterns. *Phys Rev E* 1996; **53:** 5461-5468.

232. Plotnick RE, Gardner RH, O'Neill RV. Lacunarity indices as measures of landscape texture. *Landscape Ecology* 1993; **8:** 201-211.

233. Plowman GD, Culouscou J-M, Whitney GS, *et al.* Ligand-specific activation of HER4/p180 erbB4, a fourth member of the epidermal growth factor receptor family. *Proc Natl Acad Sci USA* 1993; **90:** 1746-1750.

234. Porat M, Zeevi YY. Localized texture processing in vision: analysis and synthesis in the Gaborian space. *IEEE Trans Biomed Eng* 1989; **36:** 115-129.

235. Pressman NJ. Markovian analysis of cervical cell images. *J Histochem Cytochem* 1976; **24:** 138-144.

236. Pun T, Gerig G, Ratib O. Image analysis and computer vision in medicine. *Comp Med Imaging Graph* 1994; **18:** 85-96.

237. Qian N, Sejnowski TJ. Predicting the secondary structure of globular proteins using neural network models. *J Mol Bio* 1988; **202:** 865-884.

238. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978; **299:** 926-930.

239. Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Res Treat* 1992; **22:** 285-293.

240. Reibnegger G, Weiss G, Werner-Felmayer G, Judmaier G, Wachter H. Neural networks as a tool for utilizing laboratory information: comparison with linear discriminant analysis and with classification and regression trees. *Proc Natl Acad Sci USA* 1991; **88:** 11426-11430.

241. Reichl U, Treskatis S, Turner RFB. A backpropagation neural network for segmentation of optical microscopy images of cells. *Submitted* .

242. Remmelink M, Salmon I, Petein M, Henrion S, Pasteels J-L, Kiss R. Computer-assisted quantitative description of chromatin pattern in soft tissue tumors of the adult. *Am J Clin Pathol* 1994; **102:** 780-787.

243. Rickaert F, Gelin M, van Gansbeke D, *et al.* Computerized morphonuclear characteristics and DNA content of adenocarcinoma of the pancreas, chronic

pancreatitis, and normal tissues: relationship with histopathologic grading. *Hum Pathol* 1992; **23:** 1210-1215.

244. Rigaut JP. An empirical formulation relating boundary lengths to resolution in specimens showing 'non-ideally fractal' dimensions. *J Microsc* 1984; **133:** 41-54.

245. Rigaut JP. Fractal models in biological image analysis and vision. *Acta Stereol* 1990; **9:** 37-52.

246. Rigaut JP, Schoëvaërt-Brossault D, Lu H: Asymptotic fractals *2nd International Symposium on Fractals in Biology and Medicine* 1996.

247. Rimm DL, Sinard JH, Morrow JS. Reduced α-catenin and e-cadherin expression in breast cancer. *Lab Invest* 1995; **72:** 506-512.

248. Rosai J. Borderline epithelial lesions of the breast. *Am J Surg Pathol* 1991; **15:** 209-221.

249. Rosner B. Fundamentals of Biostatistics. Belmont, CA: Wadsworth Publishing Company, 1995.

250. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986; **323:** 533-536.

251. Russ JC. Fractal Surfaces. New York: Plenum Press, 1994.

252. Salmon I, Kiss R, Franc B, *et al.* Comparison of morphonuclear features in normal, benign and neoplastic thyroid tissue by digital cell image analysis. *Anal Quant Cytol Histol* 1992; **14:** 47-54.

253. Sanders H, Crocker J. A simple technique for the measurement of fractal dimensions in histopathological specimens. *J Pathol* 1993; **169:** 383-385.

254. SAS Institute Inc. SAS/STAT User's Guide, Version 6, vol. SAS Institute Inc. Cary, NC, 1989.

255. Schantz A, Castleman B. Parathyroid carcinoma: a study of 70 cases. *Cancer* 1973; **31:** 600-605.

256. Schulte E, Wittekind C. The influence of the wet-fixed Papanicolaou and the air-dried Giemsa techniques on nuclear parameters in breast cancer cytology: a cytomorphometric study. *Diagn Cytopathol* 1987; **3:** 256-261.

257. Schulte E, Wittekind C. The influence of Romanowsky-Giemsa type stains on nuclear and cytoplasmic features of cytological specimens. *Anal Cell Pathol* 1989; **1:** 83-86.

258. Scully RE. Ovary. In: Henson DE, Albores-Saavedra J, eds. Pathology of Incipient Neoplasia. Philadelphia: W.B. Saunders, 1993: 283-300.

259. Searle SR, Casella G, McCulloch CE. Variance Components. New York: John Wiley & Sons, 1992.

260. Segal MR. Regression trees for censored data. *Biometrics* 1988; **44:** 35-47.

261. Sejnowski TJ, Rosenberg CR. Parallel networks that learn to pronounce English text. *Compex Systems* 1987; **1:** 145-168.

262. Serra J. Image Analysis and Mathematical Morphology. London: Academic Press, 1982.

263. Shane E, Bilezikian JP. Parathyroid carcinoma: a review of 62 patients. *Endocrine Reviews* 1982; **3:** 218-226.

264. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979; **86:** 420-428.

265. Sierpinski W. Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée. *C R Acad Sci* 1916; **162:** 629-632.

266. Silage DA, Gil J. The use of a touch-sensitive screen in interactive morphometry. *J Microsc* 1984; **134:** 315-321.

267. Skates SJ, Xu F-J, Yu Y-H, *et al.* An optimal algorithm for screening for ovarian cancer with longitudinal tumor markers. *Cancer* 1995; **75:** 2637.

268. Slamon DJ, Godolphin W, Jones LA, *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989; **244:** 707-712.

269. Smeulders AWM, Dorst L. Measurement issues in morphometry. *Anal Quant Cytol Histol* 1985; **7:** 242-249.

270. Smith TG, Jr., Marks WB, Lange GD, Sheriff WH, Jr., Neale EA. A fractal analysis of cell images. *J Neurosci Methods* 1989; **27:** 173-180.

271. Snover DC, Foucar K. Mitotic activity in benign parathyroid disease. *Am J Clin Pathol* 1981; **75:** 345-347.

272. Spina D, Disanto A, Luzi P, *et al.* Novel, contrast gradient-oriented, automated chromatin texture analysis: I. Feasibility study on nuclei from benign and malignant breast epithelial cell lines in fine needle aspirates. *Virchows Archiv B Cell Pathol* 1992; **62:** 119-124.

273. Steeg PS, Bevilacqua G, Kopper L, *et al.* Evidence for a novel gene associated with low tumor metastatic potential. *J Natl Cancer Inst* 1988; **80:** 200-204.

274. Stenkvist B, Strande G. Analysis of machine-selected cells with an image analysis system in normal and abnormal cervical specimens. *Anal Cell Pathol* 1989; **2:** 1-13.

275. Stewart PW, Stamm JW. Classification tree prediction models for dental caries from clinical, microbiological, and interview data. *J Dent Res* 1991; **70:** 1239-1251.

276. Stock JL, Weintraub BD, Rosen SW, Aurbach GD, Spiegel AM, Marx SJ. Human chorionic gonadotropin subunit measurement in primary hyperparathyroidism. *J Clin Endocrinol Metab* 1982; **54:** 57-63.

277. Stoecker WV, Chiang C-S, Moss RH. Texture in skin images: comparison of three methods to determine smoothness. *Comput Med Imaging Graph* 1992; **16:** 179-190.

278. Stolz W, Abmayr W, Schmoeckel C, Landthaler M, Massoudy P, Braun-Falco O. Ultrastructural discrimination between malignant melanomas and benign nevocytic nevi using high-resolution image and multivariate analyses. *J Invest Dermatol* 1991; **97:** 903-910.

279. Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. *Comput Vision Graph Image Process* 1982; **23:** 341-352.

280. Takahashi M. A fractal model of chromosomes and chromosomal DNA replication. *J theor Biol* 1989; **141:** 117-136.

281. Tao L-C. Transabdominal Fine-Needle Aspiration Biopsy. New York: Igaku-Shoin, 1990.

282. Teague MW, Wolberg WH, Street WN, Mangasarian OL, Lambremont S, Page DL. Indeterminate fine-needle aspiration of the breast. Image analysis-assisted diagnosis. *Cancer (Cancer Cytopathol)* 1997; **81:** 129-135.

283. Teich JM, Young IT, Sher SE, Lee JS. Transformation of nuclear morphology during cellular maturation. *J Histochem Cytochem* 1979; **27:** 193-198.

284. Thunnissen FB, Diegenbach PC, van Hattum AH, *et al.* Further evaluation of quantitative nuclear image features for classification of lung carcinomas. *Pathol Res Pract* 1992; **188:** 531-535.

285. Tourassi GD, Floyd CD. Lesion Size Quantification in SPECT using an artificial neural network classification approach. *Comp Biomed Res* 1995; **28:** 257-270.

286. Truong H, Morimoto R, Walts AE, Erler B, Marchevsky A. Neural networks as an aid in the diagnosis of lymphocyte-rich effusions. *Anal Quant Cytol Histol* 1995; **17:** 48-54.

287. Unger PD, Hoon V, Stone N, *et al.* Computerized interactive morphometry in the differential diagnosis of irradiated prostates. *Anal Quant Cytol Histol* 1995; **17:** 100-108.

288. Unger PD, Watson CW, Liu Z, Gil J. Morphometric analysis of neoplastic renal aspirates and benign renal tissue. *Anal Quant Cytol Histol* 1993; **15:** 61-66.

289. van Diest PJ, Baak JPA. Morphometry. In: Bibbo M, ed. Comprehensive Cytopathology. Philadelphia: W. B. Saunders Company, 1991: 946-964.

290. van Diest PJ, Smeulders AW, Thunnissen FB, Baak JP. Cytomorphometry. A methodologic study of preparation techniques, selection methods and sample sizes. *Anal Quant Cytol Histol* 1989; **11:** 225-231.

291. Verhest A, Kiss R, d'Olne D, *et al.* Characterization of human colorectal mucosa, polyps, and cancers by means of computerized morphonuclear image analyses. *Cancer* 1990; **65:** 2047-2054.

292. Vicsek T, Cserzö M, Horváth V. Self-affine growth of bacterial colonies. *Physica A* 1990; **167:** 315-321.

293. Vilela MJ, Martins ML, Boschetti SR. Fractal patterns for cells in culture. *J Pathol* 1995; **177:** 103-107.

294. Voss RF. Random fractal forgeries. In: Earnshaw RA, ed. Fundamental Algorithms for Computer Graphics. Berlin: Springer-Verlag, 1985: 805-835.

295. Wali R, Colef M, Barba J. Best fit ellipse for cell shape analysis. *SPIE Visual Communications and Image Processing* 1991; **1606:** 665-674.

296. Watt RC, Hameroff SR. Phase space analysis of human EEG during general anesthesia. *Ann N Y Acad Sci* 1987; **504:** 286-288.

297. Weibel ER. Fractal geometry: a design principle for living organisms. *Am J Physiol* 1991; **261:** L361-369.

298. Weidner N, Folkman J, Pozza F, *et al.* Tumor angiogenesis: a new significant and independent prognostic indicator in early-stage breast carcinoma. *J Natl Cancer Inst* 1992; **84:** 1875-1887.

299. West BJ, Bhargava V, Goldberger AL. Beyond the principle of similitude: renormalization in the bronchial tree. *J Appl Physiol* 1986; **60:** 1089-1097.

300. Wheeles LL, Robinson RD, Lapets OP, *et al.* Classification of red blood cells as normal, sickle, or other abnormal, using a single image analysis feature. *Cytometry* 1994; **17:** 159-166.

301. Wied GL. Industrial developments in automated cytology as submitted by their developers. *Anal Quant Cytol Histol* 1993; **15:** 358-370.

302. Wilbur DC, Bonfiglio TA, Rutkowski MA, *et al.* Sensitivity of the AutoPap 300 QC System for cervical cytologic abnormalities. Biopsy data confirmation. *Acta Cytol* 1996; **40:** 127-132.

303. Wilding P, Morgan MA, Grygotis AE, Shoffner MA, Rosato EF. Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. *Cancer Lett* 1994; **77:** 145-153.

304. Winchester DP, Senen S, Immerman S, Blum M. A systematic approach to the evaluation and management of breast masses. *Cancer* 1983; **51:** 2535-2540.

305. Wolberg WH, Street WN, Mangasarian OL. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Lett* 1994; **77:** 163-171.

306. Wolberg WH, Street WN, Mangasarian OL. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. *Anal Quant Cytol Histol* 1995; **17:** 77-87.

307. Wu H-S, Barba J. An efficient semi-automatic algorithm for cell contour extraction. *J Microsc* 1995; **179:** 270-276.

308. Wu H-S, Barba J, Gil J. An iterative algorithm for cell segmentation using short-time Fourier transform. *J Microsc* 1996; **184:** 127-132.

309. Wu H-S, Barba J, Gil J. Morphological segmentation of textured cell images. *J Imaging Sci Tech* 1996; **40:** 265-270.

310. Wu H-S, Barba J, Gil J. Region growing segmentation of textured cell images. *Electr Lett* 1996; **32:** 1084-1085.

311. Wu H-S, Einstein AJ, Gil J. Fractal characterization of irregularly shaped images by frequency analysis. *Submitted* .

312. Yang GC. The mathematical basis for the increased sensitivity in cancer detection in air-dried cytopreparations. *Mod Pathol* 1994; **7:** 681-684.

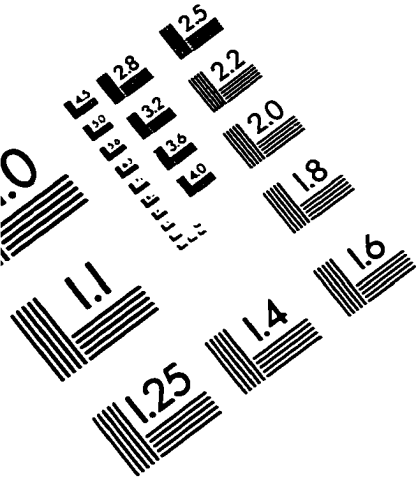
313. Yang GC, Alvarez II. Ultrafast Papanicolaou stain. An alternative preparation for fine needle aspiration cytology. *Acta Cytol* 1995; **39:** 55-60.

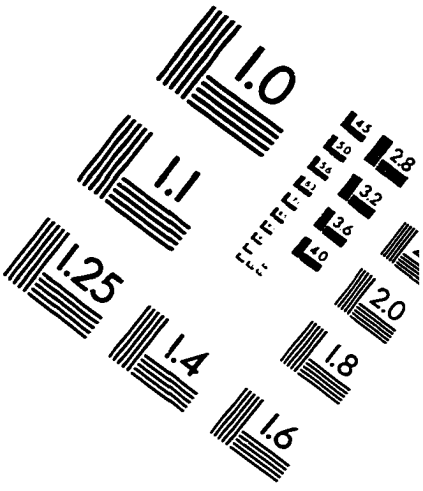
314. Yang GCH. Ultrafast Papanicolaou stain: a superior stain for fine-needle aspiration cytology applied in conjunction with the rehydration of air-dried smears by normal saline solution technique. *Adv Anatom Pathol* 1995; **2:** 208-211.

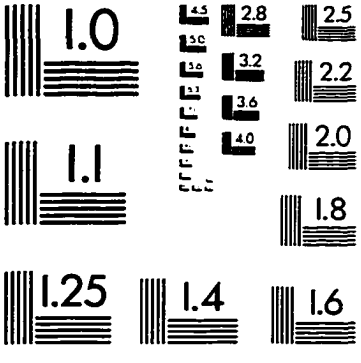
315. Yang GCH, Hoda SA. Combined use of "scratch and smear" technique and ultrafast Papanicolaou stain enhances intraoperative cytology. *Acta Cytol* in press.

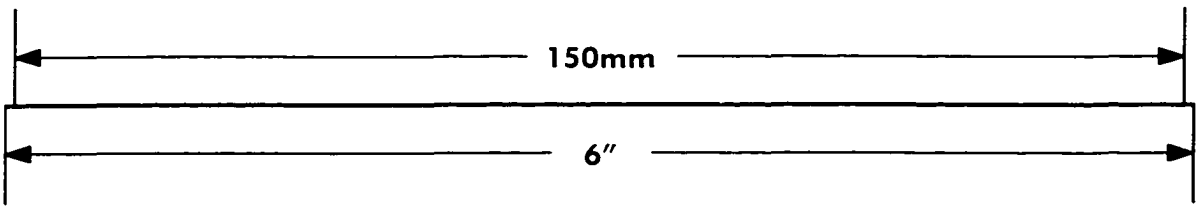
316. Young IT, Verbeek PW, Mayall BH. Characterization of chromatin distribution in cell nuclei. *Cytometry* 1986; **7:** 467-474.

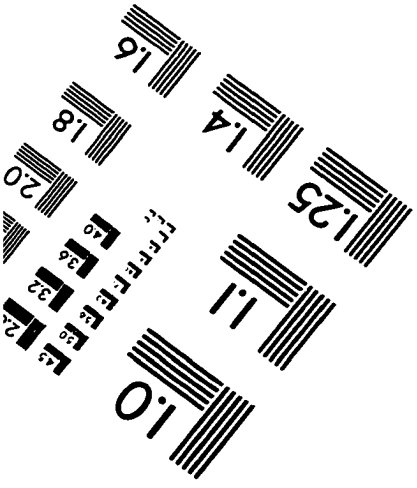
IMAGE EVALUATION TEST TARGET (QA-3)

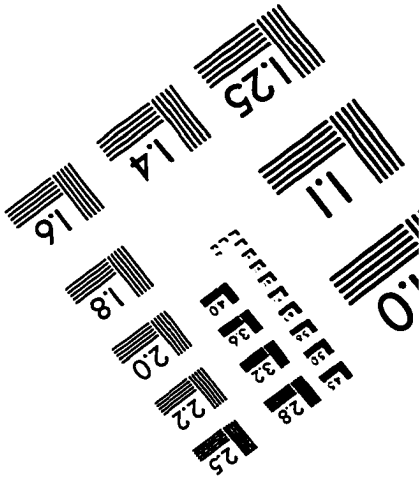












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989