DETECTION OF MALIGNANCY ASSOCIATED CHANGES IN THIONIN STAINED CERVICAL CELLS

Jennifer Hallinan and Paul Jackway

Cooperative Research Centre for Sensor Signal and Information Processing
Department of Electrical and Computer Engineering,
University of Queensland, Brisbane 4072, Australia.
email: hallinan@elec.uq.oz.au
Tel: +61 7 3653284; Fax: +61 7 3653684

September 25, 1995

ABSTRACT

This paper presents the results of a feasibility study into the detection of Malignancy Associated Changes in visually normal intermediate cells from the uterine cervix. Forty-two features drawn from the literature were measured and the resulting data subjected to statistical analysis. The results indicate that discrimination between populations of cells from normal and abnormal slides is feasible.

1. INTRODUCTION

Most previous research in the area of automated Pap smear screening has concentrated on the detection of frankly abnormal cells in microscopic images of a slide [1][2][3]. Since a single slide may contain hundreds of thousands of cells, most or all of which are non-malignant, many of which are in overlapping clumps and obscured or only partly visible, as well as pathogens and other artifacts, and since every cell in up to a thousand images must be measured and analyzed in a short time and at very high accuracy, this is a complex task for a machine [4]. Three decades of research have led to only two products so far being trialled for limited use.

A complementary approach which we believe offers considerable promise is the detection of so-called Malignancy Associated Changes (MACs) in visually normal cells from patients with cancerous or precancerous lesions [5]. It has been suggested since at least the 1950's that malignant cells affect the arrangement of DNA in the surrounding tissue [6], but such changes are subtle and difficult to detect, and little progress was made in this area for several decades. The advent of powerful microscopes, CCD cameras, fast desktop computers, and the growth of the field of image analysis has led to a recent resurgence of interest in this area, with several groups reporting success in the detection of MACs [7][8]. The major advantage is that the changes occur in the population of "normal" appearing cells and can be detected by examining a sample from this population. We now have the luxury of selecting only those cells for analysis which can be clearly imaged and segmented prior to measurement. This makes the design of the machine much more tractable, but requires much more sensitive image analytical techniques.

This paper presents the results of a preliminary feasibility study into the detection of MACs within the Pap Smear Group of Cooperative Research Centre for Sensor Signal and Information Processing at the University of Queensland.

2. SLIDE COLLECTION AND STAINING

Slides were prepared by the Queensland Cytology Service. For several months a 'Thin Prep[1]' slide was made from the swab accompanying each conventional Pap smear which passed thought the laboratory. This type of slide carries a monolayer of cells rather than the irregular clumps which occur on a conventional Pap smear, and so is ideal for computerized imaging. Both the conventional Pap smear and a Pap stained Thin-Prep slide were examined by cytologists and diagnoses made. A second, unstained Thin-Prep slide from each swab was provided to the CSSIP Pap smear project.

Slides were stained with a thionin-SO_2 stain, using a protocol proprietary to Xillix[2]. This stain, unlike the conventional Pap stain, is stoichiometric for DNA, which means that the intensity of dye in a cell image is directly representative of the amount of DNA in the nucleus. Since malignancy associated changes in the nuclear texture are postulated to be due to alterations in the amount and arrangement of DNA, this stain should make detection of these changes more sensitive.

The staining protocol we used did not include a counterstain for cytoplasm. This meant that cytoplasmic features often used for the detection of abnormality, notably the nucleus:cytoplasm ratio, were not available to this analysis. Since our primary interest is in the changes occurring in the nucleus, this was not regarded as a significant loss, although cytoplasmic changes as part of the MAC spectrum have been reported in the literature.

3. IMAGE COLLECTION

Image capture was performed by the use of an imaging system consisting of a Nikon Labophot 2A microscope and JVC TK1070E colour video camera. A x100 oil-immersion lend with numerical aperture of 1.25 provided images with resolution near the optical limit (0.25μm). The digitized images were of 7-bit photometric resolution with a per-pixel spatial resolution of 0.12μm. Two hundred cell images were collected from each slide. The images were saved by the image capture program, Video Pro 32 Ver 2.2[3] by Leading Edge, in Windows .bmp format. They were cropped to eliminate unwanted background, and converted to .gif format using Paint Shop Pro[4]. The images were then transferred to a DEC workstation under UNIX, and converted to .pgm format using XV Ver 2.21 by John Bradley.

4. FEATURE EXTRACTION

4.1. Software

All image processing from this point was carried out using a Motif-based, custom-written cytological image analysis application known as XCyte, running under XWindows on the DEC workstation. For each image the nucleus was segmented from the background using a protocol based on the morphological top-hat transform, which takes advantage of the smoothly rounded shape of the normal intermediate cell. Each image was inspected after segmentation, and those which had failed to segment properly were discarded. Failure to segment was usually due to the presence of cellular or other debris in the background of the image. Discarding such images ensures that the background grey values, used

---

[1]Cytotec Ltd.
[2]Xillix Technologies Corp, Vancouver B.C.
[3]Leading Edge Pty. Ltd., Adelaide SA
[4]JASC, Inc, Minnetonka MN

to compute optical density and other features, were valid, and that all images used in the analysis were segmented using the same protocol. Since many of the nuclear features have been shown to be highly sensitive to differences in segmentation [9], this should help to minimize procedure-related variation in the extracted features. This freedom to select clear, individual cells for analysis is one of the strengths of the MAC approach, in contrast with conventional 'rare event' detection, where every object encountered on a slide must be analyzed to avoid the possibility of missing a malignant cell.

The results of the nuclear segmentation procedure are shown for a typical normal intermediate cell in Fig. 1.

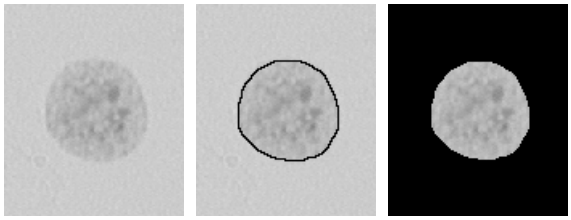

Fig. 1. Nuclear Segmentation

4.2. Features

A total of forty-two features were extracted from each nucleus. These are features which have been described in the literature as being more or less useful for MAC analysis (Table 1.). Of these, only the features which were found to be useful in discriminating between cells from normal slides and those from abnormal slides are described here.

Table 1. Nuclear Features Measured from Cervical Cells

Morphological Features:			
Area	Area of nucleus	Perim	Perimeter of nucleus
FormFac	Form factor	Circ	Circularity
Round	Roundness	Phi1	1st invariant moment
Phi2	2nd invariant moment	Phi3	3rd invariant moment
Phi4	4th invariant moment	Dist	Displacement of geometric centre
Grey Value Features:			
Gmean	Mean grey value*	GreySD	SD of grey value*
GRange	Range of grey values in nucleus	PGMean	Grey mean of 3 pixel perimeter
PGSD	Perimeter grey standard deviation*	Neighb	Mean no of identical neighbours*
IGMean	Grey mean of nuclear interior	IGSD	SD of nuclear interior
Optical Density Features:			
IOD	Integrated optical density	MaxOD	Maximum optical density
MinOD	Minimum optical density	MeanOD	Mean optical density
VarOD	Variance of optical density*	SDOD	SD of optical density
SkewOD	Skewness of OD distribution	KurtOD	Kurtosis of OD distribution
DenMax	Density of single pixel maxima	DenMin	Density of single pixel minima
RangeOD	Range of optical density	AveDiff	Averaged range of optical density
ARLow	Area ratio of low density chromatin	ARMed	Area ratio med density chromatin
ARHigh	Area ratio of high density chromatin	ERLow	Extinction ratio low density
ERMed	Extinction ratio medium density	ERHigh	Extinction ratio high density
LowSing	No low density single pixel groups	MedSing	No med density single pixel groups
HighSing	No high density single pixel groups	ERMA	Avg extinction ratio med density
ERHA	Avg extinction ratio high density	ERMHA	Avg extinction ratio med & high

** Features used in discriminant function*

5. DATA ANALYSIS

Data from four “normal” and four “abnormal” (CIN 2 or 3) slides, (553 and 575 cells respectively) was subjected to analysis using the statistical package SPSS-X. The aim of the analysis was to reduce the forty-two features measured for each nucleus to a smaller subset which could be used for discriminating between cells from normal slides and those from abnormal slides, without significant loss of sensitivity.

Sixteen of the features measured showed statistically significant differences in mean value between the two groups of cells, using a two-tailed independent samples Student’s t-test.. Of these, fourteen had a P value of less than 0.025. Despite these differences, none of the features individually was adequate to reliably distinguish between the two groups. Fig. 2. illustrates this with the histogram for the feature ‘Mean Grey Value’. While the means of the two distributions are significantly different, there is considerable overlap between the two.

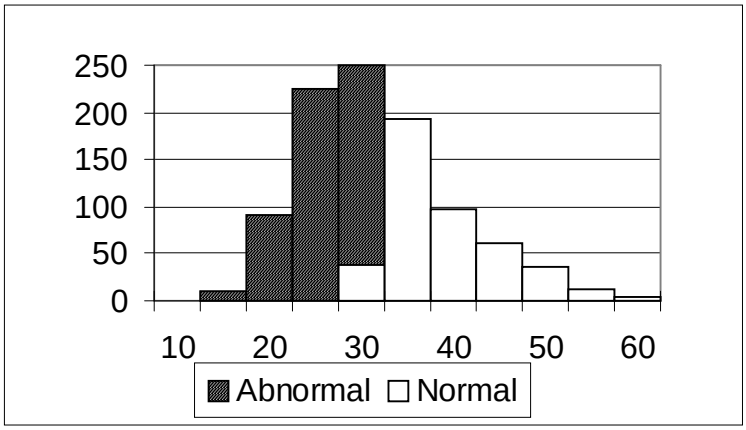

Fig. 2. Histogram of Scores for Mean Grey Value

A linear stepwise discriminant analysis was then performed, to examine the discriminating power of groups of variables. Discriminant analysis aims to form linear combinations of independent variables to compute a discriminant score as

$$\hat{F}_j = \sum_{i=1}^{p} W_{ij} X_{ij}$$

where W_{ij} is the calculated coefficient for variable *i* for case *j*, and X_{ij} if the observed value for variable *i* for case j, and there are *p* variables. This score can then be used as a basis for assigning the training set data to groups, and for classifying new individuals.

For the procedure to yield valid results, a number of assumptions regarding the data must be met. Each group should be a sample from a multivariate normal population [10]. In the case of this data set, a Kolmogorov-Smirnov test of identity with the normal distribution suggests that 31 of the variables are individually normally distributed. Individual normal distribution does not necessarily imply a multivariate normal distribution, and it is apparent that in this case, this assumption is probably not fully met. However, as long as the covariance matrices for all groups are reasonably equal, the discriminant function still performs adequately [10]. In addition, a large data set, with 3 - 10 times as many individuals as features is required [9]. This data set contains 1128 individuals and 42 original

features, of which only 6 are used in the final discriminant function. The number of individuals in each group should also be approximately the same, which it is for this data set.

The method used for the discriminant analysis was minimization of the Mahalanobis distance, D^2, between the means of the two groups, using an F-to-enter of 9 and an F-to-remove of 9. This reduced the set of 42 features to a discriminant subset of 6 features, as described in Table 2.

The means of the discriminant function scores are significantly different for the two groups of cells (chi square of 231.025 with 6 degrees of freedom, significance level of <0.025) (Fig. 3). Despite this, the classification of individual cells is relatively poor, with only 26.8% of normal and 90.1% of abnormal cells classified correctly when using prior probabilities of group membership of .7 for abnormals and .3 for normals, in order to minimize false negatives. This produces a false positive rate of 73.2% and false negative rate of 9.9%. This result is not unexpected, since the changes in nuclear texture in which we are interested are too subtle to be reliably detected by eye, and several workers have made the point that discrimination is only possible on the basis of populations of cells rather than individuals [7].

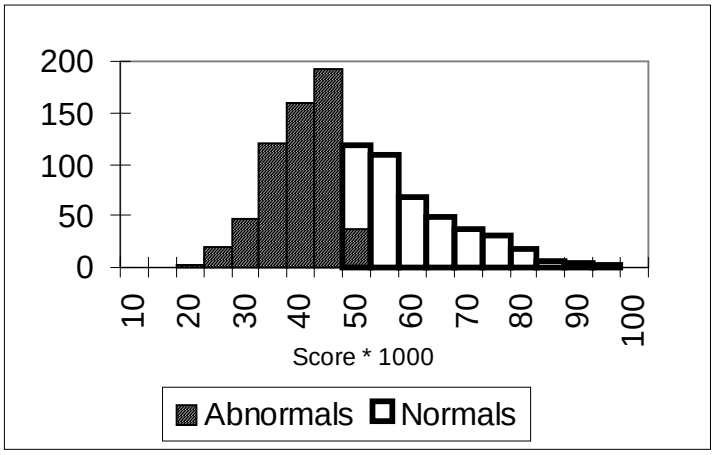

Fig. 3. Histogram of Discriminant Scores

6. DISCUSSION

The cells used in this analysis were selected as visually normal intermediate cells. It would not, therefore, be expected that features such as size, shape, or amount of DNA would be significantly different between the two groups, since such relatively gross features should be easy to pick by eye. In fact, of these features, size alone differs between normals and abnormals, with the abnormal cells being slightly smaller. The integrated optical density, a measure of the amount of DNA in the nucleus, was not significantly different, and inspection of the histogram for this feature shows no large outliers. A Kolmogorov-Smirnov test indicates that IOD is approximately uniformly distributed. This indicates that the amount of DNA is approximately constant between cells, and thus that most or all of the cells contain a normal diploid complement of DNA. Malignant cells are frequently polyploid.

Our results indicate that while there appear to be differences between normal cells from normal slides and those from abnormal slides, such differences are very subtle, and not practical to detect on a cell-by-cell basis. The histogram of discriminant scores (Fig. 3), however, shows reasonable separation on a population basis. Previous workers have found that not all cells from abnormal slides contain MACs, and of those which do, apparently only about 80% of cells from any individual show the changes [11][12]. This variability complicates the detection of MACs, and means that any grouping of cells must be done on a slide-by slide, rather than a cell-by-cell basis. The calculated discriminant score may be a valid basis for such separation.

This study was intended to establish the feasibility of detecting MACs in slides of cervical cells. The features measured from this set of cells were taken from the literature, and in general are fairly crude, obvious measurements. Our results indicate that even under these circumstances, it is possible to discriminate, on a slide-by-slide basis, between cells from normal and abnormal slides, and this provides a basis for future work. Such work will include refinement of the statistical techniques used to analyze the data, enlarging the current database of just over one thousand cell images, and the investigation of hopefully more sensitive nuclear texture features, such as fractal features [13] and grey-level coocurrence matrix features, already under investigation in this laboratory [14].

7. REFERENCES

[1] H. Banda-Gamboa, I. Ricketts, A. Cairns, K. Hussein, J. Tucker & N. Husain, “Automation in Cervical Cytology: An Overview”. Anal. Cell. Path. 4: 25-48, 1992

2] P. H. Bartels & G. L. Weid, “Automated Image Analysis in Clinical Pathology”. American J. of Clin. Path. 75:489-493, 1981.

[3] O. A. Nasseem Husain, “The History of Automated Cell Scanners” in ‘Automated Cervical Cancer Screening’ H. K. Grohs & O. A. Nasseem Hussain (eds) Igaku-Shoin: New York, 1994.

[4] T. L. Anderson, “Automatic Screening of Conventional Papnicolaou Smears”. in ‘Compendium on the Computerized Cytology and Histology Laboratory’ G. L. Weid , P. H. Bartels, D. L. Rosenthal & U. Schenk (eds.) Tutorial of Cytology: Chicago, Ill., 1994.

[5] P. H. Bartels, M Bibbo, H. E. Dytch, F. T. Pishotta, K. Yamauchi & G. L. Weid, “Diagnostic Marker Displays for Intermediate Cells from the Uterine Cervix”. Acta Cytologica 26:29-34, 1982.

[6] H. E.Nieburgs, R.G. Zak, D.C. Allen, H. Reisman, and T. Clardy, “Systemic Cellular Changes in Material from Human and Animal Tissues in the Presence of Tumours”. In “Trans. of the 7th Annual Meeting of the Int. Society for Cytology”, 137-144, 1959.

[7] B. Palcic & C. MacAulay, “Malignancy associated changes: Can they be employed clinically”. in 'Compendium on the Computerized Cytology anf Histology Laboratory' G. L. Wied, P. H. Bartels, D. L. Rosenthal, & U. Schenk (eds), Tutorials of Cytology: Chicago, 1994.

[8] D. M. Garner, A. Harrison, C. MacAulay & B. Palcic, “Cyto-Savant(TM) - It’s Use in Automated Screening of Cervical Smears” in 'Compendium on the Computerized Cytology anf Histology Laboratory' G. L. Wied, P. H. Bartels, D. L. Rosenthal, & U. Schenk (eds), Tutorials of Cytology: Chicago, 1994.

[9] C. E. MacAulay, “Development, Implementation and Evaluation of Segmentation Algorithms for the Automatic Classification of Cervical Cells”. PhD Thesis, University of British Columbia, 1989.

[10] M. J. Norusis, “SPSS-X Advanced Statistics Guide”. SPSS Inc: Chicago, Ill.

[11] G. Burger, U. Jetting and K. Rodenacker, “Changes in Benign Cell Populations in Cases of Cervical Cancer and its Precursors”. Anal. and Quant. Cytology and Histology 3:261-271, 1981.

[12] M. Bibbo,, P. H. Bartels, J. J. Sychra, and G. L. Weid, “Chromatin Appearance in Intermediate Cells from Patients with Uterine Cancer”. Acta Cytologica 25:23-28, 1981.

[13] T. J. Dennis and N. G. Dessipris, “Fractal Modelling in Image Texture Analysis”.IEEE Proc. in Radar and Signal Processing 136(5):227, 1989.

[14] R. F. Walker, P. Jackway, B. Lovell & I. D. Longstaff, "Classification of cervical cell nuclei using morphological segmentation and textural feature extraction". in 'Proc. of the 2nd Australiqn and New Zealand Conference on Intelligent Information Systems (ANZIIS)', Brisbane, Australia, 1994.