Automatic scanning for cervical smears

A. I. SPRIGGS

From the United Oxford Hospitals

In approaching the question of automatic recognition of malignant cells, it is necessary first to make one point about neoplasia in general. It would be convenient to have a detectable or measurable character which is specific for malignant cells, for instance a staining reaction. The fact that none has ever been found is itself an important datum which tells us something about cancer. There are plenty of specific characters for various normal cells—things like secretion of insulin or keratin, or the possession of basophil granules or cilia: all these differentiated characters are specific because the cells are working to a program. In malignant transformation the various programs are lost or debased, and probably no common feature is to be expected. Instead, we find the utmost diversity.

The above statements do not by any means imply that an experienced observer cannot ever identify malignant cells. In some situations, and this includes the cervix uteri, the cells shed from a carcinoma are completely different in appearance from the normal epithelial cells, though they differ greatly among themselves. In discussions about the possibility of automatic screening an engineer is apt to be baffled because he cannot get a straight answer to his question, "What do you notice when you identify a cell as malignant?"

There are therefore two sides to the problem of automatic screening for malignant cells. The first is the matter of deciding on one or more marker characters or parameters, and how to use or interpret them so as to screen specimens. The second is the technical one, how to serve up the cells to the sensing device and how to measure the parameters chosen at a speed high enough to satisfy population screening requirements. Both these aspects involve peculiar difficulties and these will be discussed below. But at first it is worth examining the standard against which our hypothetical machine will be measured.

MEANING OF 'POSITIVE' AND 'NEGATIVE'

In most diagnostic applications, for instance in the examination of sputum for malignant cells, the cytologist can achieve quite a high degree of accuracy, because he finds out in due course whether his opinion was right or wrong. If it ever becomes possible to identify malignant cells in sputum automatically, the 'diagnosis' of the machine will be checked against the results of follow up, so that the accuracy in terms of 'false positive' and 'false negative' will be known and due adjustment made.

The only application of major practical importance in population screening at present is the search for precancerous lesions of the cervix uteri. Nearly everybody conversant with this subject is convinced by the evidence that the so-called 'carcinoma in situ' is a stage in the development of cancer of the cervix. Although much remains to be learnt about this process, there is no doubt that a substantial proportion of women known to have carcinoma in situ in a biopsy sample will later develop invasive carcinoma if followed without excision of the lesion. What this proportion would be if there was no interference remains unknown. But there is a double difficulty, highly relevant to the automatic scanning problem.

1 Carcinoma in situ is not a clear-cut entity, but instead there is a whole spectrum of changes from normal right up to micro-invasion. Nobody knows where to draw the line and when to call the lesion definitely precancerous. We therefore do not really know which cases we wish to find. Some lesions would demand excision in every competent opinion; some classifiable as carcinoma in situ by one pathologist would not be called so by others. Some 'dysplasias', not classifiable as carcinoma in situ according to the accepted definitions, are almost certainly precancerous and we would like to find these too, but the dividing line is even more uncertain.

2 The cytological opinions of different observers also vary, and there is no accepted classification. The five classes introduced by Papanicolaou (which I do not personally use) were defined by him as follows (Papanicolaou, 1954):

CLASS I, absence of atypical or abnormal cells
CLASS II, atypical cytology but no evidence of malignancy
CLASS III, cytology suggestive of, but not conclusive for, malignancy

J. clin. Path. 22, suppl. (Coll. Path.), 3, 1-6
CLASS IV, cytology strongly suggestive of malignancy
CLASS V, cytology conclusive for malignancy.

It will be seen at once that these are not five different kinds of smear, still less five different kinds of cell. They are five degrees of confidence felt by the observer. It is therefore nonsensical to specify for a machine that it should identify these classes. However, it remains inescapable that the performance of a machine has to be measured against the opinion of a clinical cytologist, because a check against the histological findings is no more decisive, and a check against the clinical outcome is of course irrelevant as the whole lesion will normally be removed by the biopsy whenever a positive report is given.

SORTING OR MARKING

To reduce the need for trained screeners, it might therefore appear necessary to devise a machine which will classify as 'negative' a large proportion of smears which all cytologists would accept as negative, and as 'positive' all those which any trained person might reasonably think deserving of examination. If this system is adopted, and if cellular criteria are chosen which are thought to be indicators of malignancy, a further question arises. Are we to classify as positive a smear which contains even one cell signalled according to our criteria? If we do this, an absolute avoidance of artefacts is mandatory, and this is almost inconceivable. Are we, then, to count the signals and call a smear positive if it has more than a given number? If this system is used, a smear with many artefacts will be positive, and one with very few malignant cells will be negative. It must be remembered that in some 'positive' smears the malignant cells amount to less than 1% of all the cells present. It also needs to be said that the opinion given about a smear is the result of examining large numbers of cells, and one does not classify an individual cell as malignant or benign.

We therefore have an unusual situation that we wish to search for objects present in low concentration, the quantity of which is irrelevant, and the final classification of which is not agreed by different experts. At this point the engineer may well begin to feel despair. There is, however, a neat solution proposed to me by Mr F. H. Smith (Spriggs, Diamond, and Meyer, 1968). It is not necessary to attempt to classify whole specimens into 'positive' and 'negative'. An equal saving of manpower could be achieved if there was simply an automatic selection of abnormal fields. If the human pathologist only had to look at fields containing cells picked out according to the chosen criterion, he would be saved the labour of scanning large amounts of normal material, and would be able to make his own decision on the basis of the marked fields. Even if he has to examine some part of every smear, he must be spared a proportion, perhaps a large proportion, of the repetitive work. Moreover the occurrence of artefacts, even if they outnumbered the genuine abnormal cells, would not be serious as the human eye would reject them.

CHOICE OF PARAMETERS

Malignant cells are very variable. To program a computer to recognize all the abnormal appearances in perhaps hundreds of cells, which together convince a cytologist that the smear is 'positive', would certainly be difficult and expensive; and since there may be up to 100,000 cells in a smear of the usual size, the process of characterizing them automatically might even be relatively slow. It seems inevitable that a few simple criteria must be used, and that less than 100% of the malignant cells present in the smear will measure up to the criteria chosen. If the system described above is to be used, that is, marking fields rather than sorting specimens, then only a small minority of the malignant cells need to be found at all: in principle, only one per smear, so that the observer is alerted by seeing it and goes back to study the whole smear manually.

Criteria must therefore be found which will be met by some, but not necessarily many, of the malignant cells in something near to 100% of positive samples. This is a much simpler requirement than to distinguish all or most of the malignant cells from the normal cells by which they are surrounded.

The choice of possible parameters is quite limited and all are well known to pathologists. Malignant cells frequently have nuclei which are hyperchromatic, representing a high concentration of DNA per unit area, but of course the most intensely stained of all are the nuclei of mature superficial squamous cells in a normal smear. The content of DNA per nucleus is perhaps the favourite parameter. This varies in most normal tissues from the 2n to the 4n value, but in malignant cells there is wider variation, often with abnormal modal peaks, and usually extends up to higher values (Atkin, Richards, and Ross, 1959; Sandritter and Fischer, 1962). The right-hand tail of the distribution of DNA values could be used as a distinctive character, and this is valid for carcinoma in situ as well as for invasive cancer (Wilbanks, Richart, and Terner, 1967). Unfortunately staining methods based on the Feulgen reaction as used with the Cytoanalyzer (see p. 4) are not easy to automate as they involve...
Automatic scanning for cervical smears

3

acid hydrolysis and rather long staining times. A rough estimate can be obtained from nuclear basophilia using haematoxylin; or fluorochromes, such as berberine sulphate as used in the first automatic scanning machine built nearly 20 years ago (Mellors and Silver, 1951); or gallocyanin chrome alum as used by Sandritter, Lobel, and Kiefer (1964). Ultraviolet absorption measurements are similarly likely to be confused by the presence of cytoplasmic RNA, but Koenig, Brown, Kamentsky, Sedlis, and Melamed (1968) use a criterion based on this.

The nucleo-cytoplasmic ratio is frequently higher in malignant cells than in the tissue of origin (Table I), and might be a useful measurement if it could be obtained accurately (Reagan, Hamonic, and Wentz, 1957; Tolles, Horvath, and Bostrom 1961). But it would be most difficult to stain the cells in such a way that the borders of nucleus and cytoplasm were clearly recognizable in every cell; moreover there are frequently many bare nuclei, and sometimes endocervical cells are numerous in tightly packed formations where the cytoplasm appears scanty and the cell borders are almost invisible. Cytoplasmic features, I am convinced, are going to be of very little use in automated systems.

The identification of a second cell population with a larger mean volume was proposed as a screening test by Ladinsky, Sarto, and Peckham (1964). This test, using a Coulter counter, seems to have been empirical, and its reliability has not been confirmed (Vaillant and Richart, 1966). Measurements of cell diameter or area (Table I) have shown that the cells of invasive carcinoma are much smaller, on average, than the normal squamous cells, and that carcinoma in situ and dysplasia give intermediate values (Reagan et al, 1957). From this it might be suggested that a population of exceptionally small cells might identify a positive smear; but this is clearly absurd because of the presence in smears of other even smaller cells such as leucocytes, not to mention endocervical cells and the small basal cells of atrophic epithelium.

Finally, there is the nuclear size. Taking means of nuclear measurements on identifiable cells, Reagan et al (1957) found that the mean diameter and area were highest in dysplasia, less in carcinoma in situ, less still in invasive carcinoma, and smallest of all in normal squamous cells (Table I). (In this series no account was taken of the large variations in nuclear size of normal cells in the presence or absence of hormonal stimulation, e.g., after the menopause.)

There is also a wide variation of nuclear size in any one malignant or premalignant lesion, with an extensive scatter to the right of the distribution. It is characteristic that a few extremely large nuclei are nearly always present, and these certainly might identify positive smears. (More will be said about this on page 4.) If nuclear diameter or area were coupled with optical density measurements this would be even better.

Tolles et al (1961) made numerous measurements of diameter and optical density of stained nuclei in cervical smears, and showed in contour diagrams that positive smears bulked together display a different distribution from normal, with more cells showing higher optical density and larger diameter. Shifts in the modal values are obscured by the presence of large numbers of normal cells, so that the extreme values are more important for our purpose than the maxima in the distribution.

**THE CYTOANALYZER AND ITS DESCENDENTS**

The above data provided the basis for the criteria used in the Cytoanalyzer, a cervical smear screening

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>MEASUREMENTS OF CELL AREA¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>No. of cases</td>
<td>50</td>
</tr>
<tr>
<td>Total cells measured</td>
<td>2,500</td>
</tr>
<tr>
<td>Cell Diameter (µ)</td>
<td>44.9 ± 4.28</td>
</tr>
<tr>
<td>Area (sq. µ)</td>
<td>1604.15 ± 312.35</td>
</tr>
<tr>
<td>Nucleus Diameter (µ)</td>
<td>6.75 ± 1.20</td>
</tr>
<tr>
<td>Area (sq. µ)</td>
<td>36.51 ± 13.31</td>
</tr>
<tr>
<td>Relative nuclear area</td>
<td>2.32% ± 0.84</td>
</tr>
</tbody>
</table>

¹From Reagan et al (1957). Only abnormal cells were measured, not all cells present on the smear.
apparatus built in the United States with the support of the National Cancer Institute and US Public Health Service. The principle was mechanical scanning of Feulgen-stained preparations measuring the area and absorption of nuclei only. Classification into positive and negative depended on the ratio of abnormal to normal nuclei, taking a dividing line based on the relationship between the numbers of false positive and false negative decisions at different levels using manual measurements. For instance, using the criterion of $> 10 \mu$ diameter and $e_N > 0.5$, a 10% false negative rate should be associated with 28% false positive. Using $> 11 \mu$ diameter, at 10% false negative there would be 20% false positive (Tolles et al., 1961). In field trials using the instrument, there was a 63% false positive rate with 10% false negatives, and it transpired that positive decisions depended to a large extent on overlapping leucocytes and debris (Spencer and Bostrom, 1962).

The Cytoanalyzer was succeeded by other instruments in which optical density readings obtained by electronic or mechanical scanning are converted to digital form, and the information (which can be printed out to reproduce a picture of a cell) is analysed for certain characters or patterns (Nadel, 1965; Prewitt and Mendelsohn, 1966; Wied, Bartels, Bahr, and Oldfield, 1968; McMaster, 1968). So far as I know, no instrument of this type has been adapted so as to scan cervical smears. Machines have been built for serving up cells on moving plastic tape (Boddington and Diamond, 1967; Dawson, Heanley, Heber-Percy, and Tylko, 1967) and these could be adapted in the future for use with a pattern recognition computer.

A quite different type of device has been made by IBM in which cells in liquid suspension are classified by absorption at 2537 Å (for nucleic acids) and scattering of light (for cell volume). Here, the occurrence of small numbers of abnormal cells is detected, and the decision depends on the number of cells present exceeding the chosen criteria. This instrument was reported not to give satisfactory results in practice, partly because of a high proportion of 'unreadable' samples (Koenig et al., 1968).

Yet another system has been suggested by Roth (1966); this is to treat a constant number of cells with a solution of acriflavine, which mainly stains the nucleic acids, and to measure the amount of stain remaining after absorption. Confirmation of the efficacy of this has not, to my knowledge, been published.

**THE ONE-PARAMETER APPROACH USING NUCLEAR DIAMETER**

It is by no means certain that malignant cell populations from cervical carcinoma always have a raised mean nuclear diameter as compared, for instance, with cervical parabasal cells. It is, however, a matter of common observation that some of the malignant nuclei are usually of large size. We therefore thought that this feature alone, applied to the system of automatic marking of abnormal fields described on page 3, might be a sufficiently reliable signal. In order to explore this possibility Dr R. A. Diamond in my laboratory made a series of nuclear measurements using a micrometer eyepiece. This work was not the same as that performed by Tolles et al. (1961); instead of spending large amounts of time and effort on measuring the whole population of cells, most of which are usually normal even in a positive smear, the effort was concentrated on searching for the largest nuclei and measuring these. In this way the right-hand tail of each distribution was studied, and it was confirmed that every 'positive' smear examined did in fact contain some very large nuclei. One of his tables (Diamond, 1967) is reproduced here (Table II), based on the examination of 30 positive and 30 negative smears. These figures suggest that an automatic device marking nuclei of diameter 12 $\mu$ or 13 $\mu$ would not fail to mark a few abnormal cells in every positive smear, and would not make too many 'false' marks on negative ones.

A striking demonstration of the extent of scatter to the right in the nuclear size distribution is given by Fig. 1 (Diamond, 1967). This is constructed by determining the largest nucleus found in the whole of a smear, and recording only these 'maximal

<table>
<thead>
<tr>
<th>Nuclear Diameters</th>
<th>No. of Positive Slides with Five or More Nuclei above this Diameter</th>
<th>No. of Negative Slides with Five or More Nuclei above this Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt; 11 \times 11 \mu$</td>
<td>30 (100%)</td>
<td>3 (10%)</td>
</tr>
<tr>
<td>$&gt; 12 \times 12 \mu$</td>
<td>30 (100%)</td>
<td>1 (3.3%)</td>
</tr>
<tr>
<td>$&gt; 13 \times 13 \mu$</td>
<td>30 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>$&gt; 14 \times 14 \mu$</td>
<td>26 (86.7%)</td>
<td>0</td>
</tr>
<tr>
<td>$&gt; 15 \times 15 \mu$</td>
<td>21 (70%)</td>
<td>0</td>
</tr>
</tbody>
</table>

1From Diamond (1967).
nuclear diameters' in different cell populations. It will be seen that the lowest value found in carcinoma exceeded the highest found in the controls, except for one postnatal smear.

Similar measurements have been made by Diane Rimmer (1968) in Birmingham, examining the specific proposition that positive smears have at least 12 cells on the slide with minimum diameters of 12 \( \mu \) and over. She has found three out of 57 slides, classified as class IIIA, IIIB, and IV, to have fewer than 12 nuclei with the smallest diameters above 12 \( \mu \); moreover six out of 10 classified as class V failed to meet this particular dividing line. A class V smear as reported in that laboratory comes from clinically apparent invasive cancer (M. E. Attwood, personal communication), so that failure to detect it would not be serious in populations of symptomless women.

The best cut-off point at which a nucleus should produce a signal cannot be determined from measurements made manually like the above, but will have to await the development of effective automatic instruments. The Vickers instrument, which presents stained cells to an automatic photometric device, is being developed with this principle in view. I believe that the nuclear diameter with or without a separate estimation of optical density, \textit{i.e.}, above the level necessary to record nuclei at all, offers the best prospect at present for automated prescreening, but only if these measurements are used for marking places of particular interest for human viewing. There does not seem to be a workable criterion in sight for sorting whole smears into positive and negative.

Even the simplest system, the automatic marking of nuclei with a diameter exceeding say 12 or 13 \( \mu \), is at present beyond the capability of existing machines. One difficulty is a mathematical one. Counts and size distributions can be obtained by line scans with a slit or spot, where the correct values are calculated from the frequency of occurrence of intercepts of different lengths. To identify an individual particle as being within certain limits of diameter, even assuming a spherical shape, is much less simple, and the best way to achieve this with speed and economy is not yet clear. Another intractable problem is that the cells do not lie isolated on a clean background, but in groups, with overlapping cytoplasm, together with leucocytes, and with extra confusion caused by specks of dust or accidental marks. In this respect, where the Cytoanalyzer failed our present machines fail too.

**ALTERNATIVE TO AUTOMATION**

The whole objective of attempting to automate cervical smear screening is to save the time and expense of the current manual method, with all its problems of staff recruitment and training. Strangely, there has been very little investigation into the possibility of streamlining and modernizing the scanning of smears by the human eye. A technician generally examines about 5,000 cervical smears per annum. This comes to about 20 per day. Why so few?

Cervical smears are normally spread out on 1 in. \( \times \) 3 in. slides, the size being traditional. The amount examined is that quantity which can conveniently be spread on this size of slide. Technicians are taught that they must examine the whole smear, field by field, and this is a sensible rule for maintaining quality. To scan the whole of one of these smears takes about five minutes on average, and in one hour it is easy to examine about 10, but this concentration cannot be kept up for long at a time with comfort.

Another system used in some centres is to put more than one smear on each slide, for instance one cervical scrape, one vaginal aspirate, and one endometrial aspirate (Wied and Bahr, 1959). Each of these has to be much smaller than the traditional smear, but according to Wied there is no loss of accuracy. Again, in the British Columbia project each technician is estimated to spend three minutes per smear (Bryans, Boyes, and Fidler, 1964) and this is hardly long enough for a complete examination of every field under a 22 \( \times \) 40 mm coverslip. It is therefore worth finding out what is the smallest size of smear to produce an acceptable accuracy.

Boddington (1968) has examined this question by estimating, in 67 positive smears, the number of cells present and the proportion which were ab-
normal; also how many low-power fields1 needed to be examined for finding the first abnormal cell. (In this context ‘abnormal’ means that the cells had the visible changes on which the original positive report had been based; and as only histologically proved cases are included in the 67, the ‘abnormal’ cells can be assumed to come from carcinoma, carcinoma in situ, or dysplasia.)

In 53 cases (79%) there were abnormal cells in practically every field. Eight more required less than 20 fields to find abnormal cells, and none required as many as 70 fields, which can be scanned in a minute. There is therefore a good case for scanning only a marked area of a size which can be covered in one minute. If this were adopted, each technician could do five times as many examinations without working any harder, that is, up to 150 per day. When any abnormality was detected, naturally the remainder of the slide would be examined in order to get the maximum information. Any automated system should perhaps be measured against this rapid manual method, rather than against the traditional way of examining all of the material submitted.

SUMMARY

Difficulties in the automation of cervical smear screening are not only of a technical kind. Options differ about (1) which kinds of cell we wish to detect; (2) which criteria distinguish putatively malignant cells from benign ones; (3) whether the relative number of abnormal cells should be used to divide smears into positive and negative.

In the author’s opinion, the possibility of programming a computer to imitate the ‘recognition’ of positive smears, as a human observer does, is rather remote. Much closer to fulfilment would be a marking system, by which a few of the most aberrant cells are automatically marked on a slide (or tape); the whole smear would not be recorded as positive or negative by the instrument, but instead of this the human observer would be led to examine a limited number of fields, from which to decide whether the rest of the sample deserved examination.

Of several possible single parameters, nuclear diameter is probably the most useful one, as practically every ‘positive’ smear contains at least a few very large nuclei. A machine capable of marking nuclei above a predetermined size is still not satisfactorily developed.

---

1 Objective × 6·3, Optovar × 1·6, eye pieces × 8.

REFERENCES


