

FILE COPY

CERN - DATA HANDLING DIVISION
DD/Internal Report/71/10
E.S. Gelsema
November 1971

Impressions of a Visit to some
Bio-Medical Computer Groups in the U.S.

In the period from October 26th to November 3rd I visited several groups doing Bio-Medical computer work, especially work on white blood cells, and talked to a number of people involved in such work. This note is a collection of the main impressions of these conversations.

DD-jh

1. Stanford Research Institute

Lack of time prevented me from visiting the group working here, but I had a conversation by telephone with Dr. Alfred Brain, who has been doing, and is hoping to be able to continue, work on white blood cell recognition. His work is supported by a commercial client (an insurance company). He had stopped the work for the moment, being short of money.

He uses a television camera digitizing the blood smear (stained with Wright's stain) through a microscope into 16 grey levels. He does not use filters, therefore discarding all colour information. He thought this is a severe limitation (both the relatively small number of grey levels and the colour rejection) but he is forced to do so because of considerations of economy. He went into great detail explaining the practical implications.

- i) A practical machine should not cost more than \$ 75 K.
- ii) It should replace at least 3 technicians.
- iii) It should analyse one cell in about one second on a small machine, like a PDP-8. This would amount to a ratio of 400 to 500 smears per day.

The advantage of such a machine would be that it would do the work more reliably and more reproducibly than a team of technicians.

He also stressed the point that the feasibility of such a machine would only be guaranteed, provided that great care would be taken in the preparation of the smears. The range of variations in staining and in overall quality of the slide from one hospital to another, or even within some hospitals, are much too wide for any machine to handle.

His procedure of analysis starts out with locating the nucleus of the cell. It then encloses the nucleus in a rectangular box, containing (apart from the nucleus) only cytoplasm. It then evaluates a nuclear shape factor, which does 80% of the classification work. Up to a maximum of 7 other parameters may be evaluated if ambiguities remain. Dr. Brain could not reveal the definition of this shape factor, nor of the other parameters.

With this procedure he classifies 7 cell types (he distinguishes bands from segmented neutrophils and small from large lymphocytes). When he resumes this work he will include more cell types in his scheme. He did not think this would be a big problem.

Finally he stated that he did not think that a continuous flow

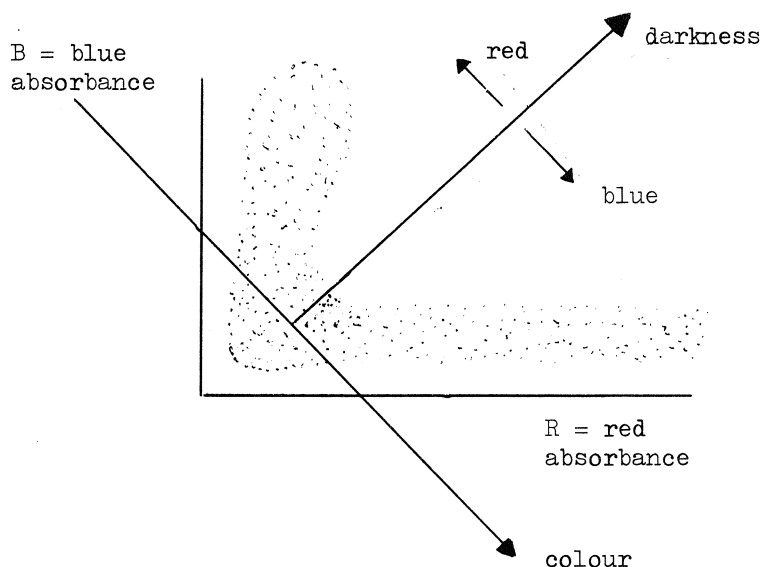
technique would be a very serious competitor. Hematologists would be reluctant to abandon their vast amount of knowledge on morphological cell features in favour of a particle counting technique.

2. Hospital of the University of Pennsylvania, Department of Radiology, Philadelphia

Here I discussed with Dr. B.H. Mayall some features of his classification procedure, especially the evaluation of the colour information. In fact, this classification scheme is based on information obtained from grey scale histograms only. He treats the whole cell, without separating nucleus from cytoplasm. There is a selection procedure ensuring that the field of view contains only the white cell and background. Touching and nearby red cells are thus avoided.

His equipment consists of a CRT scanner. The spot is focused onto the preparation through a microscope. The digitizings are recorded on tape. Up to a maximum of 256 grey levels can be distinguished, but for the white blood cell work the two least significant bits are discarded.

Based on the assumption that there are only two components in the Writht stain and that therefore a two-colour scan gives the maximum colour information, he scans the smears through two colour filters, which are matched with the two peaks in the absorption spectrum of the stain. He then constructs a scatter diagram of all the points, plotting optical density using one filter against optical density using the other. He then shifts the origin to the center of the background peak. The axes with an angle of 45° with respect to the original ones now indicate darkness and colour.



He then constructs 6 histograms from this plot. What I understood of it is that he makes histograms of R-B and R+B for the red points, for the blue points and for all points.

From these 6 histograms he extracts of the order of 60 parameters. (mean values, standard deviation, skewness etc. and also ratios of these). There will of course be strong correlations between the parameters. He then uses a stepwise linear discriminant technique to select the most powerful parameters for classification.

This procedure was used on a sample of 30 cells (5 white blood cells of each of the 5 normal types and 5 red blood cells). The classification result, using the three most powerful parameters, showed 2 errors.

Mayall has stopped this work for the time being. He is now more interested in chromosome work. His attitude towards this work is that it is an interesting intellectual exercise. He did not show much interest in further experimenting with this technique, e.g. testing it on a larger sample. He did not seem to think much in terms of a practical instrument. He did look at the continuous flow method as a serious competitor, this being one of his reasons to stop the work. He is going to publish his results shortly and put our group on his mailing list, at my request.

3. M.I.T., Boston

At M.I.T. I had a conversation with Professor Murray Eden, Dr. Young being absent.

Their hardware consists of a CRT scanner, which scans a 24×36 mm colour slide of the blood smear. The output signal goes through a set of dichroic mirrors, yielding three signals (red, blue and green). He admitted that Mayall's technique of two colours might give the maximum colour information, provided carefully selected filters are used. On the other hand, he did not feel the three-colour treatment much of a burden and moreover this avoided the problem of matching the filters.

He could not give me a plot of the points of one cell on the standard 3-colour diagram but promised to send it as soon as Young came back.

An interesting feature about their colour measurement is that they adjust the origin of the plot on the basis of the measured colour of the red cells. In fact, at regular intervals they do a special scan on

red cells only. On their subsequent scans of white cells they shift the colour coordinates in such a way that the red cell points always cluster in the same region of the colour diagram. This is a way of normalizing the wide range of staining variations which may appear from one slide to another.

In addition to the colour analysis they also evaluate nuclear shape factors and other geometrical quantities. In particular they have an algorithm which determines the number of lobes in a segmented neutrophil.

Discussing the practical implications of a machine doing the white cell differential count, Professor Eden stated essentially the same figures as Dr. Brain in Stanford. The cost should not exceed \$ 100 K. The machine should replace 3 technicians (assuming they work at a rate of 10 counts/hour). This means a processing speed of 1 sec/cell.

The answer to the question whether a machine should be able to recognize more than the 5 common cell types depended, according to Professor Eden, on the application of it. He saw two distinct possibilities.

- i) Population scanning. In this case it would be sufficient if the machine would recognize the 5 main types, provided it would also flag the abnormal types.
- ii) Hematological Department of a Hospital. Here would be a clear need for the machine to include the abnormal types of cells in the classification.

They have now an experimental set-up of a microscope scanner. They plan to eventually replace their current instrument with this one, thus avoiding the intermediate colour film.

4. New England Medical Centre, Boston

I talked for about an hour with Dr. P. Neurath and spent the rest of the day with Bill Selles and Richard Hertzberg.

According to Dr. Neurath, the photographic problem was now much better understood. They have done a lot of work on special exposures, photographing a wedge of known optical densities. They now know how to proceed to always work in the correct range of the density curve.

The PIQUANT units will now be defined as 0 at optical density 1.81 and 63 at optical density 0.05. The new data will have the background peak at grey level 7. Moreover, the logarithmic as well as the

linear amplifier will be used with the next set of data.

There were, however, severe problems with the display CRT, which is a vital part of their data-acquisition equipment. It was estimated that digitizing could be resumed towards the end of November. I left a tape to remind them of our wish to receive new data as soon as possible.

It was felt that the data we have so far (which is also their most recent lot) is of no use in trying to do any colour evaluation "à la Mayall", because of one of the filters (5900) being very badly matched with the Wright's stain absorption spectrum. I stated that we had made an attempt to do a colour analysis "à la Young" and would very much like to pursue this some more with new data. Thinking back on this conversation I feel that for some reason this point was not emphasized strongly enough; I did, however, also mention it to Selles. This topic should be taken up again in our next letter to Dr. Neurath.

In general it was made clear that they were very happy with this collaboration. They showed much interest in what we have been doing so far and asked me for a copy of the listing of our routines. Dr. Neurath expressed the wish that we should try to communicate more frequently (he suggested a two-weeks regular basis) and that we should try to divide the work between the two groups, thus avoiding doing everything twice.

I spent the rest of the day with Selles and Hertzberg and found that they have the same points of view about the collaboration. I was shown the results of their work on recognition. They have been working in three fields.

- i) Geometrical parameters: The procedures amounted to essentially the same as what we have been doing here.
- ii) Colour evaluation: This had failed because of the wrongly chosen filter.
- iii) Texture evaluation: This work was still in progress. The original idea of looking at the shape of one scan line (monodimensionally) had apparently been abandoned in an effort to combine the results of more adjoining scan lines (two-dimensionally).

It was hard to evaluate their achievements on the basis of what I was shown. The problem was at least far from being solved.

This visit has been useful, particularly in that it has established a direct contact between our group and NEMC.

5. BNL, Brookhaven

In Brookhaven I saw Dr. P. Hough, who has taken up a project in the bio-medical field. He showed me the draft of a report to the National Institute of Health (NIH) from which he hopes to obtain a funding for the automatic analysis of cervical smears. He collaborates with a specialist in pathology and with Dr. M.L. Ingram, who has worked with Dr. K. Preston (Perkin Elmer Corporation) on white blood cell analysis. This group will be backed up by the resources of BNL and the State University of New York at Stony Brook.

The proposal to the NIH states that though much time and effort has already been spent on this problem no satisfactory system has as yet been produced. It then emphasizes the fact that it is of prime importance to reconsider the preparation stage of the analysis in an attempt to yield smears which are more suitable for automatic examination. They will reconsider the smearing and flattening technique as well as the staining technique.

The design specifications of the cell-scanning instrument will be tailored to the result of these investigations and to the result of an examination of algorithms to discriminate a cancer cell. A CRT generated flying spot scanner was favoured at the time of writing the proposal.

A response from NIH was expected at the end of November.

Dr. Hough stated that if we would be interested in this problem we were welcome to collaborate if a suitable way of doing so could be found.

6. Conclusions

As a summary of the foregoing the following conclusions emerge:

1. The visit to NEMC has been fruitful in establishing a contact which we will need now that Pino has left us. The group in NEMC expressed their interest in a close collaboration with us.
2. In white blood cell work and presumably in all cytological work, when an attempt at automatic analysis is made, the preparation technique (smearing and staining) must be standardised and well controlled.
3. In a practical instrument the intermediate photographic state should be avoided. The data should be taken directly from the smear through a microscope. All groups concerned about practicability of the

instrument are working or experimenting with or at least thinking of this type of solution.

4. The practical instrument should replace at least 3 technicians. It should find, digitize and analyze a cell in 1 second. The cost of the instrument should not exceed \$ 100 K.
5. It is advisable to consider already in an early stage the possibility of recognizing more than 5 normal types of cells. This implies that sufficient data on abnormal cells should become available as soon as possible.
5. The continuous flow method is rated very differently by different people. A crucial question seems to be what it can do with abnormal white blood cells. In their proposal to the NIH the group of Hough enumerates four arguments in favour of the morphological approach to the problem of cancer cell detection. These hold equally well for the problem of white blood cell differential work.

