Adaptive classifiers for dicentric chromosomes

JIM PIPER AND JOANNE SPREY

MRC Human Genetics Unit, Edinburgh, UK

Classification of dicentric chromosomes in a practical automatic screening system comprises three stages. The first generates plausible centromere candidates from each chromosome in an automatically segmented metaphase, and uses contextual knowledge to generate distributions of "probably true" and "probably false" centromeres, thus adapting to the conditions within a particular metaphase. The second stage classifier uses these distributions to re-classify the candidates as centromeres or non-centromeres. From this classification, likely dicentrics are found by counting centromeres; a third classifier attempts to reject false positives among the likely dicentric chromosomes, by comparing the feature values of the proposed centromeres of a chromosome and rejecting chromosomes for which these values do not satisfy certain similarity criteria.

The second stage classifier may be a simple box classifier, or may use a variety of parametric Bayesian methods. The performance of these alternatives has been tested both on reference data sets comprising about 600 metaphases, and on larger data sets when embedded in a practical fully automatic dicentric pre-screening system. When operating parameters were such that a similar number of true positives were found by both classifiers, the Bayesian classifier produced about half as many false positive errors as the box classifier, with the final false positive rate being in the region of one candidate dicentric chromosome in every four cells.

INTRODUCTION

The dicentric chromosome is a common cellular product of ionising radiation. Being characterised by two centromeres, it is relatively easily distinguishable from undamaged chromosomes which have just one centromere (figure 1), and thus presents a suitable target for automated systems for radiation dosimetry. However, dicentrics are rare in unexposed individuals, or in those exposed to low doses, and in order to measure doses down to about 0.1 Gy of either X- or γ-irradiation, it is necessary to analyse at least 1000 metaphases, preferably more.

Unaided visual microscopic analysis of 1000 cells in the metaphase of cell division takes several days and is thus both tedious and extremely costly. On account of the large number of cells and the expected rarity of dicentric chromosomes, recent work on automated analysis has concentrated on developing a pre-screening system which automatically selects likely dicentric chromosomes for subsequent visual review by a cytogeneticist. Existing systems achieve an overall performance on routine material that can be expressed as a true positive detection rate of about 40% of all dicentrics, for a cost in false positives of about one in every two or three cells analysed, or about 1% of chromosomes analysed. While our group has
already shown that it is both feasible and potentially cost-effective to determine radiation dose by semi-automatic screening for dicentric chromosomes with this level of performance\textsuperscript{6)}, clearly a relative improvement in the ratio of true to false positives would improve both the economic prospects and the ergonomics of this sort of analysis.

In this paper we describe the classifier used in\textsuperscript{6)} and some recently developed alternatives which improve the performance. In common with previous work by our group and others\textsuperscript{4,5,8)}, we do not aim to find the approximately 25\% of dicentrics which have one or more terminal or

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{A digitised human peripheral blood metaphase homogeneously stained with Giemsa, showing a dicentric chromosome \(d\) with two metacentric (non-terminal) centromeres, a dicentric \(a\) with one metacentric and one acrocentric (terminal) centromere, and two acentric fragments \(f\).}
\end{figure}
“acrocentric” centromeres (figure 1) since finding such centromeres appears to be significantly more error prone and is thus not cost-effective.

MATERIALS AND METHODS

Adaptive classifiers

In order to find dicentric chromosomes one must determine the number of centromeres per chromosome, which differentiates this task from the better known problem of classification of chromosome constitution (or “karyotyping”)9,10, where it is invariably assumed a priori that there is exactly one centromere per chromosome. Since we ignore the “acrocentric” centromeres, which in normal cells comprise about 25% of the total, it cannot even be assumed that each chromosome has “at least one” centromere. Also, centromere finding in a karyotyping context achieves an accuracy of, at best, less than 95%9,10,11.

Consider figure 2, which shows 10 chromosomes from each of four cells, together with their longitudinal “integrated density” profiles, obtained by summing pixel values along lines perpendicular to the chromosome axis of symmetry at equally spaced points along the axis9. Since

Fig. 2. Ten digitised metacentric chromosomes from each of four different metaphases from the same slide, showing the visual similarity of centromeres within each cell in contrast with the difference between the cells, together with longitudinal integrated density profiles.
chromosomes are usually paler at centromeres as well as being narrower, the integrated density profile tends to have a pronounced minimum at a centromere. It can be seen in figure 2 that there is indeed a profile minimum at each centromere, but there are other minima at non-centromere locations, and the “significance” of the minima varies substantially between the cells. Thus we cannot with any reliability count centromeres simply by counting profile minima, but the profile minima are a good guide to likely centromere locations, and usually (though by no means always) the most pronounced minimum is indeed at the centromere. What in fact is required is a classification of profile minima into true centromeres ($T$) and artefacts ($F$).

Figure 2 clearly shows that although there is wide variation among the cells, within each cell both the chromosomes themselves, and particularly their centromere regions, in some respects show considerable visual similarity. It is our aim to make use of this similarity, and to “adapt” the centromere classifier to the particular nature of each cell.

The variability in the appearance of metaphase cells is reflected in the variation in the values of numerical features likely to assist in the recognition of centromeres. Table 1 gives, for each of the sets of chromosomes in figure 2, the mean and standard deviation of chromosome width (defined as $\text{area/length}$), of the width at the centromere, and of the relative integrated density at the centromere. These figures confirm that the variation within a cell is much less than between cells, and other features behave in the same way. It follows that a centromere classifier must be adapted to the cell characteristics. One way to achieve adaptation is by appropriate within-chromosome or within-cell normalisation of feature values. Lörrch et al.\(^5\) have proposed an effective way of normalising measurements within each chromosome based on iterative transformation of profiles. We have taken an alternative approach that explicitly utilises the assumption of within-cell similarity of chromosome centromeres to generate different classifier parameters for each cell.

Table 1. The means and standard deviations of chromosome mean width (defined in 0.125 $\mu$m pixel units as $\text{area/length}$), centromere mean width (expressed as a proportion of chromosome mean width), and profile relative density at the centromere, for the four sets of ten chromosomes shown in figure 2.

| cell | chromosome width | centromere relative width | centromere relative density |
|------|------------------|--------------------------|----------------------------|
| 1    | 11.6±1.4         | .62±.09                  | .35±.07                    |
| 2    | 16.3±1.7         | .65±.13                  | .23±.12                    |
| 3    | 19.6±1.3         | .73±.09                  | .53±.14                    |
| 4    | 18.6±1.7         | .76±.05                  | .46±.08                    |

A “bootstrap” classifier scheme

By choosing “the most pronounced” minimum from the density profile of every chromosome in the cell we obtain a set $A$ of chromosome locations which will predominantly be centromeres. Occasionally an artefactual minimum will be more pronounced than that at the centromere, resulting in a non-centromere member of $A$; and in the case of a dicentric, obviously only one of the centromeres can be a member of $A$. However, the distribution of
values of a feature on set $A$ can be expected to be similar to those obtained from a set $T$ consisting entirely of centromeres.

We similarly expect a set $B$ of "not most pronounced" minima to consist largely of non-centromere locations, though $B$ will obviously contain those true centromeres which do not have the most pronounced corresponding profile minimum. The distributions of $A$ and $B$ may be used to construct a rule discriminating between "mostly centromeres" ($A$) and "mostly not centromeres" ($B$). It may then not be unreasonable to expect that this rule will actually classify the locations of all minima, i.e. $A \cup B$, into the two classes "true centromere" ($T$) or "artefact" ($F$) though with some error. This classifier can perform well even though the sets $A$ and $B$ used in its design do themselves contain some erroneous allocations in terms of the desired classification into $T$ or $F^{12,13}$. Following such a classification, simple counting leads directly to candidate dicentric chromosomes.

**Candidate centromeres**

All local minima of the density profile are regarded as "candidate" centromeres, and features are measured for each. We obtain one member of set $A$ from most of the approximately 46 chromosomes in a cell (profiles of acrocentric chromosomes typically have no minimum). However, in well-stained metaphase cells there are often few if any density profile minima except at true centromeres. In order to obtain an adequate number of "mostly not centromeres", the initial set $B$ consisting of the "not most pronounced minimum" locations is augmented by one additional location in each chromosome chosen at random.

**Simple box classifier**

The box classifier has already been described$^{4,8}$, and was used in a successful recent trial of the ability of the dicentric finding system to estimate unknown radiation dose$^6$. In outline, all features of candidate centromeres are transformed as necessary so as to have positive values which tend to be less at centromeres than at other locations$^8$. For example, the "width" feature satisfies these criteria without any transformation. The distribution of each feature in set $A$ is used independently to establish an upper allowable limit for that feature. Feature values of every candidate centromere in the cell are then compared with the appropriate limit; only a candidate for which every feature is less than its corresponding limit are classified as $T$ (true centromere).

**Linear and quadratic discriminants**

The other classifiers are based on Bayesian theory, which says that an object (a candidate centromere) is assigned to the class $T$ or $F$ for which it has the higher conditional probability. In practice, probability distributions have to be estimated from a "design" set of data, and here we use $A$ and $B$, making the assumption that they are in fact well representative of $T$ and $F$. In this way, we adapt the classifier by estimating different probability distributions for each metaphase cell.

There are relatively few design samples available in $A$ and $B$, and so in order to constrain the problem we assume that classes $T$ and $F$ both have multivariate Gaussian distribution.
Then the decision is based on the (log) posterior probability densities

\[ d_k(\hat{x}) \propto -\log(|\Sigma_k|) - D_k^2(\hat{x}, \mu_k) + 2 \log(P_k), \]

where for classes \( k = T, F, \mu_k \) is the class mean feature vector, \( \Sigma_k \) is the distribution covariance, and \( D_k^2(\hat{x}, \mu_k) = (\hat{x} - \mu_k)'\Sigma_k^{-1}(\hat{x} - \mu_k) \) is the Mahalanobis distance of feature vector \( \hat{x} \) to \( \mu_k \). A candidate \( \hat{x} \) is classified as \( T \) (true centromere) if \( d_T(\hat{x}) > d_F(\hat{x}) \), and as \( F \) otherwise.

The term \( P_k \) represents the prior probability of class \( k \). We use it also to represent the different notional misclassification costs for the two classes; modifying the \( P_k \) computed from \( A \) and \( B \) provides a simple way in a practical system to balance the costs of false negatives (reduced sensitivity) against false positive (more operator review).

In general, where the covariance matrices \( \Sigma_T, \Sigma_F \) differ, the decision surface is quadratic. However, \( \Sigma_T \) and \( \Sigma_F \) tend to be similar, and making the assumption that they are identical results in linear discrimination, where now.

\[ d_k(\hat{x}) \propto -D_k^2(\hat{x}, \mu_k) + 2 \log(P_k) \]

From experience, it is known that when the design set for a classifier is, as here, rather small, estimating complete covariance matrices does not always lead to the best possible classifier performance. Frequently in such cases, making some simplifying assumption to reduce the number of independent covariance terms will improve classification accuracy\(^1\). In particular, we have investigated setting all the off-diagonal elements of the matrix, i.e. the covariances, to zero. Substitution of the covariance matrix by this “variance matrix” means that the classifier no longer takes account of the correlation of the features.

**Feature measurement and selection**

A variety of features which discriminate centromeres from non-centromeres may be found from one of three alternative image transformations. Firstly, features may be derived from the longitudinal integrated density profile itself (figure 2, 3(b)), secondly from the shape of the chromosome’s boundary in the vicinity of the candidate centromere location (figure 3(c)), and thirdly from the “crossing-profile” of pixel density measured transversely across the chromosome.

**Fig. 3.** Chromosomes with centromere candidates showing (b) longitudinal density profile showing profile value at candidate, (c) boundary and transverse lines at candidates to obtain boundary curvature, (d) crossing profiles at candidates.
at the candidate location (figure 3(d)).

We have found by ad hoc testing that the most valuable features are the longitudinal profile value at the candidate relative to the profile maximum value, the sum of the boundary curvatures at either end of a transverse line crossing the boundary, and a moment statistic computed from the crossing profile as \( \sum m_i |d_i| / \sum m_i \), where \( m_i \) is the value of the crossing profile at point \( i \) distance \( d_i \) from the crossing profile centroid (figure 4).

The initial partition of the set of candidates into sets \( A \) and \( B \) is based on a single feature chosen from several suitable alternatives. We have found that either width at the candidate, or the moment feature defined above, lead to appropriate partitions into sets \( A \) and \( B \).

![Fig. 4. Computation of moment feature \( \sum m_i |d_i| / \sum m_i \) from crossing profile.](image)

"Post-processing" classifier

After selection of centromeres, simple counting determines which chromosomes are potential dicentrics. A post-processing stage may then be applied, aimed at reducing the number of false positives. This checks the relative value of certain features of the selected centromeres, notably the chromosome width, and the total density across each centromere. The reasoning here is that true dicentrics should have relatively similar centromeres. On the other hand, a false positive, which might for example result if a cluster of two small chromosomes was undetected by the automatic segmentation, and in which one "centromere" is in fact the join between two chromosomes, should not usually show such similar feature values.

The classifier consists of the following rules:

1. At a centromere candidate, the chromosome must be at least 6 pixels (0.75 \( \mu \text{m} \)) wide.
2. The difference between the widths at two candidate centromeres must be less than 33% of the mean of the two widths.
3. The difference between the density profile values at two candidate centromeres must be less than 20% of the profile peak value.
4. The crossing profiles (figure 3) must show the same modality (i.e. either both should have modality of one, or both two).

The assumption of particularly close similarity of the centromeres on one dicentric chromosome is in practice violated by a small proportion of true dicentrics, and the penalty of incorporating post-processing is that it contributes an additional false negative rate to the complete system.
Comparison of classifier performance

Comparison of classifier performance is complicated by the fact that it is possible to alter parameters that change the ratio of false positives to false negatives. In a practical system this is an important facility, since the two types of error have different costs; false negatives reduce the sensitivity of a system and in the present application result in a requirement to analyse more cells, while false positives must be rejected by operator interaction and thus cost operator time.

If a reference data base exists that has the “true” classification encoded by full visual analysis, then it can be re-classified automatically many times, with different parameter settings, in order to obtain the “receiver operating characteristic” (ROC). This may be plotted as the proportion of false negatives versus the corresponding proportion of false positives, resulting in a first quadrant curve that usually has characteristic, approximately hyperbolic, shape (figure 5). As a general interpretation rule, the closer such a curve lies to the axes and origin, the better the performance.

Fig. 5. Receiver operating characteristic curves for the two data bases of manually segmented irradiated metaphases, comparing the performance of the various classifiers, each of which was executed with a range of parameter settings. The false negative rate includes a figure of about 0.25 on account of acrocentric and tiny dicentrics, which the system deliberately does not attempt to find. The false positive rate is expressed in units of false dicentrics per cell analysed.

However, it is a common experience that the performance on such a data base may not reflect the “real world” behaviour of a system, because the reference data base has been specially treated in some way, for example in the selection of “good quality” metaphases, or in the use of manual interaction to obtain fully correct segmentation. We have therefore also compared the classifiers when embedded in a complete, fully automatic screening system.
Performance comparison on reference data bases

Two reference data bases of digitised metaphase cells have been established, in each of which the chromosomes have been segmented, with manual correction of the segmentation as necessary, and with every dicentric and acrocentric dicentric coded. One contains 457 orcein stained cells from material exposed in vitro to 250 kV X-rays, cultured and prepared for analysis in our laboratory; there are 230 dicentric chromosomes in this set, which was used in earlier reported experiments. The other data base comprises 162 metaphases homogeneously stained with Giemsa from material exposed to 60Co γ-rays, and contains 191 dicentrics. This set was collected by selecting cells containing a dicentric as determined by visual scoring, and is a small proportion of the set of metaphases used for a large dose-response trial. The ROCs in figure 5 summarise the performance of several classifiers on each data base; the post-processing stage was omitted.

Performance comparison in a practical system

Since on the reference data, it appeared that the parametric Bayesian classifiers were substantially superior to the box classifier, although there was little difference between the various parametric classifiers (figure 5), a further experiment was carried out using just the quadratic classifier with variance matrix, to see whether the improvement compared with the box classifier was maintained in the context of a complete system for automatic radiation dosimetry applied to routine data. The system incorporates fully automatic metaphase cell location, digitisation, and segmentation, and has rules for rejecting cells expected to be unsuitable for analysis. After these preliminary stages, the box classifier and the quadratic Bayesian classifier using the variance matrix were used in turn to find possible dicentrics. These were coded to indicate by which classifier they were selected, or whether by both. Of these, chromosomes which would be rejected by the post-processing stage were also appropriately coded. All objects selected by either classifier were subsequently reviewed by a cytogeneticist in the usual fashion. Nothing indicated to the reviewer which classifier was responsible for the positive object, or whether it would have been rejected by post-processing.

Three sets of data were used, (i) about 1400 cells from a range of in vitro γ-ray exposures, drawn from a calibration set used in a previous collaborative trial with Pittsburgh University, (ii) about 550 cells digitised at the UK National Radiological Protection Board (NRPB) on equipment similar to our own, obtained from a blood sample from an industrial radiologist accidentally exposed in vivo to γ-rays (for this data the slide was unavailable for reviewing, and the review was from the screen image only), (iii) about 600 cells from a microscope slide of a peripheral blood sample, exposed in vitro to 4 Gy of X-rays and cultured at the National Institute of Radiological Sciences, Japan (NIRS). In each case, the number of cells refers to the initial set, of which a proportion were rejected by the system.

The scale of this experiment precluded the use of many different operating points in order to generate ROCs, and so in a preliminary experiment, the Bayesian classifier costs \( P_k \) were set so as to generate a similar proportion of true positive (after post-processing) as had the box classifier in earlier experiments. The results of this experiment cannot include measurement of false negative rates since the experiment did not include full visual analysis of every cell;
Comparison of adaptive classifiers on reference data

Figure 5 shows that all of the parametric Bayesian classifiers perform substantially better than our box classifier on both reference data sets. Overall, we judged that the quadratic classifier using the variance matrix was best, and this was used for the comparison in the complete system.

Comparison of classifiers in a complete system

Table 2 shows numbers of true and false positives obtained both before and after the application of the post-processing classifier. Almost (but not exactly) all the Bayes true positives were included in the set found by the box classifier. In the NRPB set, the same set of cells was analysed by identical software at NRPB; the review using a microscope as well as the screen image of the positives produced by the Bayes classifier and post-processing resulted in 25 true positives scored (P. Finnon, personal communication), compared with our figure of 30 when reviewing screen images only. This discrepancy is comparable with those that we have previously observed between different reviewers\(^6\). The higher false positive rate and apparently low sensitivity obtained with the NIRS slide (by comparison with published dose-response data\(^15\)), we estimate the sensitivity to be about 30% may be attributed to the substantial level of damage caused by the high dose, and in particular to the large number of errors made by the automatic segmentation part of the system on this material. About 40% of the NIRS cells were rejected by the system, compared with about 20% in the other samples.

Table 2. True and false positive dicentrics found by the box and the Bayes classifier (quadratic, variance version), both with and without subsequent post-processing. Cells rejected by the system (between 20% and 40% of the initially digitised sets) are not included in the figures presented here.

| Material | No. of cells analysed | Post-processed? | Box | Bayes |
|----------|----------------------|----------------|-----|-------|
|          |                      |                | TP  | FP    | TP  | FP    |
| Pittsburgh | 1153                | no             | 96  | 1124  | .97 | 418  | .36 |
|          |                      | yes            | 77  | 485   | .42 | 74   | .19 |
| NRPB     | 459                  | no             | 39  | 645   | 1.41| 32   | .59 |
|          |                      | yes            | 32  | 250   | .54 | 30   | .24 |
| NIRS     | 362                  | no             | 199 | 673   | 1.86| 148  | .72 |
|          |                      | yes            | 136 | 231   | .64 | 112  | .32 |

however true positives correctly located and false positive rates are compared in table 2.

RESULTS
DISCUSSION

Figure 5 and table 2 show a convincing improvement in results obtained with a parametric Gaussian classifier compared to our earlier system based on a simple box classifier. This results in approximately a halving of the false positive rate, when classifier parameters are chosen so as to keep the true positive rate approximately constant.

A similar ratio of improvement was apparent both before and after the post-processing classifier. The post-processing classifier successfully rejected proportionally many more of the false positives in the practical system than had been predicted on the basis of earlier experiments using the 457-cell orcein reference data set; this was an unexpected and pleasant surprise, and we are now attempting to optimise this part of the system.

ACKNOWLEDGEMENTS

We thank Eric Thomson for his patience when reviewing and rejecting large numbers of false positives, and the following colleagues for generously preparing and supplying material: Neil Wald and Xia Chen for the “Pittsburgh” slides; Isamu Hayata for the “NIRS” slide; and Paul Finnon and David Lloyd for the “NRPB” digitised data.

REFERENCES

1. Bender, M. A., Awa, A. A., Brooks, A. L., Evans, H. J., Groer, P. G., Littlefield, L. G., Pereira, C., Preston, R. J. and Wachholz, B. W. (1988) Current status of cytogenetic procedures to detect and quantify exposures to radiation. Mutat. Res. 196: 103–159.
2. Finnon, P., Lloyd, D. C. and Edwards, A. (1986) An assessment of the metaphase finding capability of the Cytoscan 110. Mutat. Res. 164: 101–108.
3. Lloyd, D. C. (1989) Automated aberration scoring: the requirements of an end-user. In: “Automation of Cytogenetics”, Eds: C. Lundsteen, J. Piper, pp. 9–17, Springer-Verlag, Heidelberg.
4. Piper, J. (1990) Automated Cytogenetics in the Study of Mutagenesis and Cancer. In: “Advances in Mutagenesis 2”, Ed: G. Obe, pp. 127–153, Springer-Verlag, Berlin.
5. Lorch, T., Wittler, C., Stephan, G. and Bille, J. (1989) An automated chromosome aberration scoring system. In: “Automation of Cytogenetics”, Eds: C. Lundsteen, J. Piper, pp. 19–30, Springer-Verlag, Heidelberg.
6. Bayley, R., Carothers, A., Chen, X., Farrow, S., Gordon, J. Ji. L., Piper, J. Rutherford, D., Stark, M. and Wald, N. (1991) Radiation Dosimetry by Automatic Image Analysis of Dicentric Chromosomes. Mutat. Res. 253: 223–235.
7. T. Lorch (1990) Automated dicentric scoring; a PC-based implementation. Proc. XII European Workshop on Automated Cytogenetics, 19–23, September 1990, Segovia, Spain, p. 41.
8. Piper, J., Towers, S., Gordon, J., Ireland, J. and McDougall, D. (1988) Hypothesis combination and context sensitive classification for chromosome aberration scoring. In “Pattern Recognition and Artificial Intelligence”, Eds: E. S. Gelsema, L. N. Kanal, pp. 449–460, Elsevier, Amsterdam.
9. Piper, J. and Granum, E. (1989) On fully automatic feature measurement for banded chromosome
10. van Vliet, L., Young, I. T. and Mayal, B. H. (1990) The Athena semi-automated karyotyping system. Cytometry, 11: 51–58.
11. Gregor, J. and Granum, E. (1991) Finding chromosome centromeres using band pattern information. Comput. Biol. Med. (in press).
12. Chhikara, R. S. and McKeon, J. (1984) Linear discriminant analysis with misallocation in training samples. J. Am. Statistical Assoc. 79: 899–906.
13. Krishnan, T. (1988) Efficiency of learning with imperfect supervision. Patt. Recog. 21: 183–188.
14. Kirby, S. P. J., Theobald, C. M., Piper, J. and Carothers, A. D. (1991) Some methods of combining class information in multivariate normal discrimination for the classification of human chromosomes. Statist. Med. 10: 141–149.
15. Lloyd, D. C. (1984) An overview of radiation dosimetry by conventional cytogenetic methods. In: “Biological Dosimetry”, Eds: W. G. Eisert, M. L. Mendelsohn, pp. 3–14, Springer-Verlag, Berlin.